Durham University

Centre for Medical Education

Research Report

The Validity of the Professional and Linguistic Assessments Board (PLAB) Exam: Research Report

A Report for the General Medical Council

December 2013

Dr Paul A. Tiffin
Prof Jan Illing
Dr Lisa Webster
Prof John C. McLachlan

Centre for Medical Education Research
Durham University

Research Report: The Validity of the Professional and Linguistic Assessments Board (PLAB) Exam

Executive Summary

Background

This programme of analysis was commissioned by the General Medical Council (GMC) as part of the internal review of the Professional and Linguistic Assessments Board (PLAB) assessments. The aims of the study were;

a) *To evaluate to what extent the knowledge and skills of PLAB International Medical Graduates (IMGs) could be considered equivalent to those doctors who had completed their Foundation Year 1 (FY1) of their training in the UK.*

b) *To evaluate whether the likelihoods of referral or censure in relation to fitness to practice (FtP) were equivalent between PLAB IMGs and FY1s.*

A secondary set of aims were to explore the data supplied in order to attempt to uncover any underlying reasons for any inter-group differences observed. These analyses also included evaluating what impact shifting the pass-marks for various components of the PLAB was likely to have on the outcomes observed for the groups of doctors studied.

Summary of Principal Findings

Part 1: Equivalence in terms of Clinical Competence: Comparing ARCP Outcomes between Groups

In order to compare the extent to which PLAB international medical graduates (i.e. those doctors graduating from non-EU Countries) could be considered equivalent to UK graduates, the outcomes at post-graduate assessment were explored. These postgraduate assessment outcomes were available in the form of Annual Review of Competence Progression (ARCP) outcome categories and equivalent outcomes for the former Record of In-Training Assessments (RITAs). A table of these outcomes is provided in the Technical Appendix. For the purposes of our evaluation the outcomes were generally collapsed into two categories. These two 'analytic' categories were: a) *satisfactory progress/eligibility to apply for Certification of Completion of Training-CCT*, and b) another, less satisfactory category, such as a requirement that *training time be extended.* At times, a modified form of four outcome categories could be used as 'ordered' categorical (ordinal) data in the analysis, but often this was inadvisable as the mathematical assumptions underpinning such 'ordinal logistic regression' analyses were not supported on testing. As each doctor had undergone varying numbers of ARCPs we used the lowest (poorest) outcome received as the outcome under investigation (i.e. we examined the probability that a doctor received a less than satisfactory ARCP outcome on at least one occasion). Selection of poorest rating at ARCP had the advantage of providing a reasonable amount of variation in the outcome variable, as roughly a third of doctors in the dataset had been recorded as having at least one ARCP that resulted in an outcome other than '*satisfactory progress/eligibility to apply for Certification of Completion of Training-CCT'.*

Our analyses strongly suggested that PLAB IMGs were not equivalent to UK graduates and were much more likely to experience at least one poor outcome at ARCP. There are a number of possible confounding variables between the populations. These include the number of ARCPs taken, age, sex, ethnicity, first language and educational background. Of these, we had relatively complete data only on the number of ARCPs taken, age and sex. Ethnic status was not included in the multivariable models as it was missing in approximately 22% of cases and was almost perfectly correlated with world region of qualification. The odds for PLAB IMGs to have obtained satisfactory outcomes at all ARCPs taken were roughly 40% lower compared to UK graduates, even after adjusting for the number of ARCPs taken, age and sex. Of course, from a patient perspective, the presence of these confounders may not be relevant: they are interested in their quality of care, not why it may be deficient. This is why, where appropriate, we also present the 'raw', or 'unadjusted' odds ratios (ORs) for the outcomes of interest between groups.

Interestingly, performance at ARCP between PLAB IMGs and EEA doctors who had not taken the PLAB was not greatly different; the odds of a PLAB IMG receiving at least one poor ARCP outcome was only about 20% less than for an EEA graduate. This difference disappeared once age and sex were adjusted for.

We looked at the potential effect of increasing the pass-mark of the various components of the PLAB (i.e. the PLAB exam parts 1 and 2). In addition we explored the potential impact on ARCP performance of raising the IELTS threshold score for entry to the PLAB assessments. The IELTS is the main (though not only) route by which IMGs are able to qualify for entry to the PLAB. Our findings suggest that raising the IELTS pass score to 8.5 or 9.0 is likely to reduce the difference in performance at ARCP between IMGs and UK graduates. As might be expected, increasing the IELTS score at which the PLAB could be taken would be unlikely to eradicate such an inter-group difference, even allowing for the number of ARCPs taken, sex, age and years of UK-based experience. Increasing the PLAB Part I pass mark by around 27 points is likely to result in IMG and UK graduate performance at ARCP being comparable (adjusting for the effects of the potential confounders listed earlier), though only around one sixth of candidates achieve such high marks currently. Likewise, raising the PLAB part 2 pass score by around 12 points will result in IMGs that perform roughly as well as UK graduates at ARCP, adjusting for the effects of potential confounding factors. However, it should be noted that currently only around one twelfth of PLAB candidates achieve that score level.

A number of components of the IELTS and the PLAB examination process were predictive of later performance at ARCP amongst PLAB IMGs:

1. Increased IELTS overall score independently predicted a higher probability of obtaining satisfactory outcomes at all ARCPs taken.
2. Fewer attempts at both PLAB part 1 and part 2 were predictive of an increased likelihood of obtaining satisfactory outcomes at all ARCPs taken.
3. Higher scores, relative to the pass mark, at both the first attempt at PLAB part 1 and part 2 were predictive of an increased likelihood of obtaining satisfactory outcomes at all ARCPs taken.

Analyses were also conducted for subscale scores for the components of the IELTS and PLAB. Few conclusions can be drawn from these results, as with the exception of the IELTS, subscales scores are not standardised in any way and therefore are not

comparable between candidates in differing diets. This could be considered a major weakness in the design of the PLAB at present. Regarding the IELTS, *reading* was the subtest score most predictive of good ARCP performance and *speaking* the least. Caution should be exercised in interpreting other findings relating to PLAB subtest scores, but the PLAB part 1 subtest (at first attempt) most predictive of later ARCP performance was *diagnosis* and the least was *context:* the PLAB part 2 subtest (at first attempt) most predictive of later ARCP performance was *communication* and the least was *practical skills*.

Analyses were also carried out separately according to the speciality group in which the ARCP was conducted. These sub-analyses were performed as it was postulated that ARCP stringency and leniency might vary across medical specialities. We observed that PLAB scores were more predictive of ARCP performance in some specialities compared to others. However, our ability to draw conclusions was limited at times by the low number of PLAB graduates in some of the smaller specialities (e.g. *public health* and *ophthalmology*), which would have led to the analysis being under-powered to highlight any differences, should they have existed. There were some noteworthy findings. For example PLAB part 1 scores, relative to pass mark at first attempt, were most predictive of ARCP progress in the speciality groups of *anaesthetics and intensive care* and *radiology*, with every point above pass increasing the odds of ARCP success by about 3%. In contrast PLAB part 2 scores, relative to pass mark, at first attempt were most predictive of ARCP progress in the speciality groups of *general practice*, *paediatrics*, *surgery*, *medicine* and *psychiatry*, with every point above pass increasing the odds of ARCP success by about 6 to 10%.

Part 2: Equivalence in terms of Professionalism: Comparing Fitness to Practice (FtP) Referrals and Outcomes between Groups

Regarding FtP, PLAB IMGs were more likely than UK graduates to be referred or censured in relation to FtP. The differences in rates of PLAB IMGs referred for FtP were still significant, though greatly reduced in magnitude, once the effects of sex, age and UK-based experience were controlled for. However, after adjusting for these potential confounding factors there was no difference in the odds of a PLAB IMG being actually censured in relation to FtP. In this case ‘censure’ indicates that a referral for FtP resulted in a warning or sanction being issued (e.g. conditions being placed on registration etc. Some aspects of PLAB performance were predictive of the probability of a FtP referral or censure) in IMGs:

1. Multiple attempts at both PLAB part 1 and part 2 were associated with an increased probability of eventual censure by the GMC. Those who took either part of the PLAB three times or more were independently more likely to receive censure than those who passed at first sitting.
2. PLAB part 1 scores, both at first attempt and at pass statistically significantly predicted whether an IMG would receive censure from the GMC. For part 2 of the PLAB only the score at first sitting was statistically significantly predictive of the likelihood of GMC censure.
3. For those IMGs who were referred in relation to FtP concerns their PLAB part 2 score at first sitting had some ability to predict the likelihood of eventual censure.

Summary

Even after controlling for the effects of the available potential confounding factors, PLAB IMGs who passed the PLAB system demonstrate, on average, poorer performance on ARCP compared to UK graduates. Raising the pass mark for the IELTS or PLAB may reduce the magnitude of this difference but is unlikely to eradicate it completely. In addition, changing the structure or scoring for PLAB is may have some effect on FtP referral rates for PLAB IMGs though is much less likely to impact on the proportions receiving censure from the GMC.

Summary of Considerations

Considerations regarding the mode of use of PLAB test scores

1. The primary focus of this report was on the PLAB exam and exploring evidence for the validity of the assessment system. However, as the IELTS is the main route by which IMGs qualify to take the PLAB we felt it was vital to explore this stage of (linguistic) evaluation. In the light of our findings the GMC PLAB review panel may wish to consider raising the threshold scores for IELTS. We note that, according to our results, candidates obtaining the highest score (9.0) approached equivalence with 'English as first language' candidates in terms of ARCP performance. This would seem to highlight the important role that language and communication ability potentially plays in clinical performance.

2. The GMC PLAB review panel may wish to consider raising the threshold scores for both parts of the PLAB; such an action will reduce the differences observed in ARCP performance between PLAB IMGs and UK. However, unless these new thresholds are very high (i.e. such that only one sixth to one twelfth of candidates are likely to pass at each sitting) then it is unlikely that equivalence between PLAB IMGs and UK graduates will be achieved.

3. As we have previously proposed after reviewing the performance of PLAB against the international literature, the number of resits should be limited. We proposed a limit of three re-sit attempts, followed by a personal development refractory period of at least two years. The data in this study is concordant with that recommendation.

Considerations regarding further development of the PLAB

4. It would be greatly desirable that the subscale scores for the PLAB, parts 1 and 2 are implemented in a manner that renders them comparable both within and between diets. This would involve both some form of standardisation and 'test equating'.

5. It would be valuable to obtain more detailed information on the psychometric properties of the PLAB (e.g. dimensionality, inter-rater reliability etc).

6. Overall our findings suggest that the PLAB may be too easy to pass, given the aims of the assessment (i.e. to help ensure equivalence between PLAB IMGs and UK graduates at the end of FY1). Nevertheless, it should be borne in mind that the outcomes examined in this study (ARCP performance and FtP processes) occurred, in general, after FY1 (or equivalent). Therefore, we cannot rule out other causes of divergence between the performance of the two groups following completion of the Foundation Years programme. However, on the balance of probabilities, baseline differences in ability are still likely to be the main reason for the differences observed.

Full Report on PLAB Validity

Background and context

The National Health Service (NHS) has traditionally relied on overseas-qualified staff to ensure it can effectively deliver healthcare. Indeed, the healthcare workforce is becoming increasingly internationalised.[1] In 2010, 37% of the doctors registered with the UK medical regulator qualified in other countries.[2] There has been a recent trend for this proportion to be decreasing, following increased output from UK medical schools and restrictions on visa requirements.[3] However, International Medical Graduates (IMGs) are still entering the UK workplace in significant numbers. At the same time there are concerns that overseas qualified doctors are over-represented in cases where a doctor's clinical performance or professional behaviour come to the attention of the GMC and/or the National Clinical Assessment Service (NCAS).[4] Indeed, there is evidence to suggest that overseas doctors are over represented in later stages of the GMC Fitness to Practice (FtP) processes.[5] Possible reasons for this over-representation may include: (i) pre-existing deficiencies in doctors' clinical performance before they come to the UK, (ii) biased reporting of doctors by other staff or patients; or (iii) difficulties in performance which arise as a consequence of the move to the UK [6] as well as cultural differences.[7]

In line with these concerns, the failure rate for IMGs (for the purposes of this report defined as those doctors whose primary medical qualification (PMQ) was awarded outside the European Economic Area (EEA)) taking the Clinical Skills Assessment component of the Royal College of General Practitioners (RCGP) Membership examination is 63% compared to 9% of UK graduates.[8] These findings have been much debated and there have been some suggestions that the possibility of assessor bias as the cause of these observations cannot be ruled out.[8]

IMGs face many issues when entering the UK workplace such as social and cultural isolation, communication and cultural issues, financial problems and discrimination. A study carried out by Illing et al in 2009 found that there were differences in the training culture of IMGs particularly with regard to communication, teams and hierarchy.[6] The patient-centred approach adopted in the UK was an area that many IMGs felt was different from the medical culture in their home country. This had implications relating to communication involving patients in decision making, taking consent at all stages of a clinical pathway, and the need to inform patients of clinical details and decisions at all stages of their care.[6] Healthcare systems and regulatory frameworks in different countries may have different values and expectations, causing difficulties when practising. Moreover it appears that IMGs require support in three main areas: before

coming to the UK, at the point at which they start work and on-going support when they are in their post.[9]

In order for IMGs to be able to register with the GMC they must satisfy the following criteria:

- hold an acceptable overseas qualification or have passed the examinations needed for such a qualification.
- have the knowledge, skills and experience needed to practise as a fully registered doctor in the UK.
- unless they are an exempt person, have the necessary knowledge of English
- their fitness to practise is not impaired.

The Medical Act does not specify how IMGs should satisfy these criteria but the usual route is completion of the Professional and Linguistic Assessments Board (PLAB) assessments, an offer of sponsorship with an organisation approved by the GMC for that purpose or having an acceptable postgraduate qualification. A majority of IMGs sit the PLAB tests. In 2010 the GMC granted 2,709 IMGs registration in practice setting and out of those 1,490 (55%) were granted registration through the PLAB test.[10]

The PLAB test is in two parts and is designed to ensure equivalence of competencies with those achieved by UK graduates by the end of their Foundation Year One (FY1) placements. Part 1 of the PLAB test can be taken in either the UK or in certain countries abroad. The test is a written exam consisting of 200 questions where the candidate has to provide a single best answer and lasts three hours. It is designed to test the skills of the candidate in the following areas: applying knowledge and experience to clinical practice; good clinical care; assessment; and good clinical care management. The questions cover illnesses which are important or common in the UK context.

The second part of the PLAB exam must be undertaken at the Clinical Assessment Centre in the UK and must be passed within three years of sitting Part 1. The Objective Structured Clinical Examination (OSCE) consists of 14 stations. Each station consists of a five minute clinical scenario where candidates are observed and scored on whether the candidate can use their skills and knowledge appropriately. The skills assessed are: clinical examinations; practical skills; communication skills; and history taking. As with PLAB Part 1 the skills are set out in the blueprint. The standard setting method used for scoring the OSCE stations is the Borderline Group Scoring method. In addition to fulfilling all of the above requirements all doctors must be familiar with the GMC's 'Good Medical Practice' which lays out the requirements of being a good doctor in the UK.

It is important that the PLAB exam is fit for purpose and that it is rigorous so that all stakeholders involved see it as a valid and reliable test of knowledge and skills to enable IMGs to enter into the UK workplace. However, as mentioned above IMGs are over represented in FtP cases. There is no limit to the number of times a candidate can re-sit the exam. Candidates who only just fail are likely to pass on re-sit. However there is moderately strong evidence to indicate that there is no further benefit after four exam attempts (first exam plus three re-sits). The GMC's own data on PLAB Part 1 is consistent with this effect. [11] In addition there has been almost no evidence relating the number of re-sits directly to later performance in practice. However, there is moderate to

strong evidence to support a correlation between examination performance in general and later performance in practice. [11]

In addition to passing the PLAB test many IMGs have passed an International English Language Test System (IELTS) assessment with a score of at least band seven (good user) as an overall score and in all subtests prior to applying for PLAB Part 1.[12] Although the majority of PLAB IMG candidates enter the PLAB exam via the IELTS assessment other forms of evidence of linguistic competency in English are accepted by the GMC. However, as communication skills are known to influence clinical competency, the IELTS as well as the PLAB Parts 1 and 2 can be seen as potential predictor variables for doctor performance.

This report outlines the results of a programme of analyses performed on data supplied by the GMC in order to explore the validity of the PLAB as assessment of a doctor's clinical competence as well as professional behaviour:

Aims

To explore and report on evidence to support or refute the validity of the PLAB assessment and its components with regard to subsequent clinical practice.

Principal Objectives

- To compare the post-graduate performance of PLAB IMGs and UK graduates (the latter having completed their Foundation year 1 (FY1) programme as assessed by the Annual Review Competency Panel (ARCP). Thus, evidence will be collated to support or refute the PLAB as ensuring the equivalence of clinical competence in PLAB IMGs as compared to UK medical graduates.

- To compare the likelihood of referral to fitness to practise procedures for each group of doctors in order to assess the extent to which the knowledge and application of *Good Medical Practice* demonstrated by a pass in both parts of the PLAB test is equivalent to those of UK doctors who have successfully completed FY1.

In addition, the results of other sub-analyses will be reported. For example, whether PLAB performance is more predictive of ARCP performance in some medical specialities compared to others.

For this programme of work several hundred statistical analyses were conducted; the results of many of these are contained in the Technical Appendix to this report (still under construction at the time of writing). In order to maximise the accessibility and usefulness of this report only the key findings considered to be of critical interest to the GMC PLAB Review Panel are included in the main body of this report. Where results are presented these are accompanied by interpretations. Moreover, for clarity the main report is structured as a series of questions and answers, concluding with considerations. These concern the future use of the test, its design and others issues, such as data collection.

The report is divided into two sections dealing with the two principle objectives; the first section is devoted to the relationship between IELTS/PLAB performance and ARCP outcomes. The second section is concerned with the ability of the IELTS and the PLAB tests to predict the likelihood of FtP events and outcomes.

Summary of Methods

Available data and preparation

In order to compare the extent to which PLAB IMGs (i.e. those doctors graduating from non-EEA Countries via the PLAB system) could be considered equivalent to UK graduates the outcomes at post-graduate assessment were explored. Postgraduate assessment outcomes were available in the form of Annual Review of Competence Progression (ARCP) outcome categories and equivalent outcomes for the former Record of In-training Assessments (RITAs). A table of these outcomes is provided in the Technical Appendix. The GMC provided data on 148,143 ARCP outcomes relating to 60,654 individual doctors. Of these only 112 outcomes (0.08%) were missing or could not be coded. There were data relating to a relatively small number of IMGs who were not reported as having completed the PLAB (n=3,449) and these were excluded from the analysis. There were also data relating to the Record of In-Training Assessment (RITA) outcomes, for those doctors being evaluated under the previous system. These RITA outcomes were converted to equivalent categories which matched the ARCP outcomes, where applicable. In addition, we were only concerned with outcomes that reflected competence, as rated by the ARCP/RITA panels. Therefore certain categories, such as 'out of programme experience' (RITA 'F'/ ARCP '8') were not used in the data analysis. Therefore data was potentially available on 56, 417 doctors with ARCP (or equivalent data) available. However, for the multivariable analyses where predictor variables, such as age, sex or year of registration were missing such cases had to be excluded. This left data relating to 56,410 doctors with complete information.

For the purposes of analysis by speciality, medical specialisms were grouped via theme. For example, surgical specialities were grouped together, psychiatric sub-specialities were also collapsed, as were those concerned with physician training. This categorisation of specialities was validated by examining the number of doctors classified as straddling more than one of the 'super-specialities'. In the event only 2.5% of doctors with a lowest ARCP outcome reported were 'cross classified' in this way. On further exploration these two specialities often made sense in relation to each other; for example, a paediatric surgical trainee may have undergone ARCPs in both paediatrics and a surgical speciality.

An initial exploratory analysis indicated, that in relation to the PLAB part 1 or Part 2 exam results as predictors, the different category groups (e.g. '*satisfactory progress/completion'*, '*targeted training required*, '*inadequate evidence presented*' '*extended training required*' etc.) could not always be mathematically differentiated in a multinomial logistic regression, in particular, the two categories '*inadequate evidence presented*' and '*additional training [though not time] required*' could not generally be reliably discriminated between. For this reason these outcomes were collapsed into the same category for the purposes of analysis. Indeed, for the purposes of our evaluation the outcomes were generally further collapsed into just two outcome categories. These

two ‘analytic’ categories were: (a) *satisfactory progress/eligibility to apply for Certification of Completion of Training-CCT*, or (b) another, less satisfactory category*, such as a requirement that training time be extended.* Occasionally, a modified form of four outcome categories could be used as ‘ordered’ categorical (ordinal) data in the analysis, but often this was inadvisable as the mathematical assumptions underpinning such ‘ordinal logistic regression’ analyses were not supported on testing. As each doctor had undergone varying numbers of ARCPs we used the lowest (poorest) outcome received as the outcome under investigation (i.e. we examined the probability that a doctor received a less than satisfactory ARCP outcome on at least one occasion).

1 IELTS and PLAB performance as predictors of ARCP outcome in International Medical Graduates

IELTS performance as a predictor of ARCP outcome in PLAB IMGs

PLAB IMGs have usually passed an International English Language Test System (IELTS) assessment with overall and sub-test scores of band seven (‘good user’) or above prior to applying for PLAB Part 1, although other forms of evidence of English competency are accepted by the GMC. The IELTS test is made up of four sub-tests; *listening*, *speaking*, *reading* and *writing*. Each part is graded between band one (non-user) to band nine (expert user). The IELTS test can be taken as many times as the candidate wishes in order to obtain the desired outcome.

1.1.1 *Question:* ***How do PLAB IMGs, UK and EEA graduates compare in terms of ARCP performance?***

Summary Answer: PLAB IMGs are much more likely than UK graduates to experience a ‘suboptimal’ outcome (e.g. *category 3*; **‘**inadequate progress by trainee**’**) following at least one ARCP. In addition**,** they tend to perform more poorly than EEA (non-UK) medical graduates. However, as can be seen in Figure 1, both IMGs (who are PLAB graduates) and, to a lesser extent, EEA doctors (note: a small number of EEA PLAB graduates were excluded from the analysis) are over represented in the less desirable categories of ARCP outcome, especially those representing ‘*targeted training required*’; ‘*extended training required*’ and ‘*release from programme with/without necessary competencies*’ ($p<.001$ for intergroup difference on chi-squared testing).

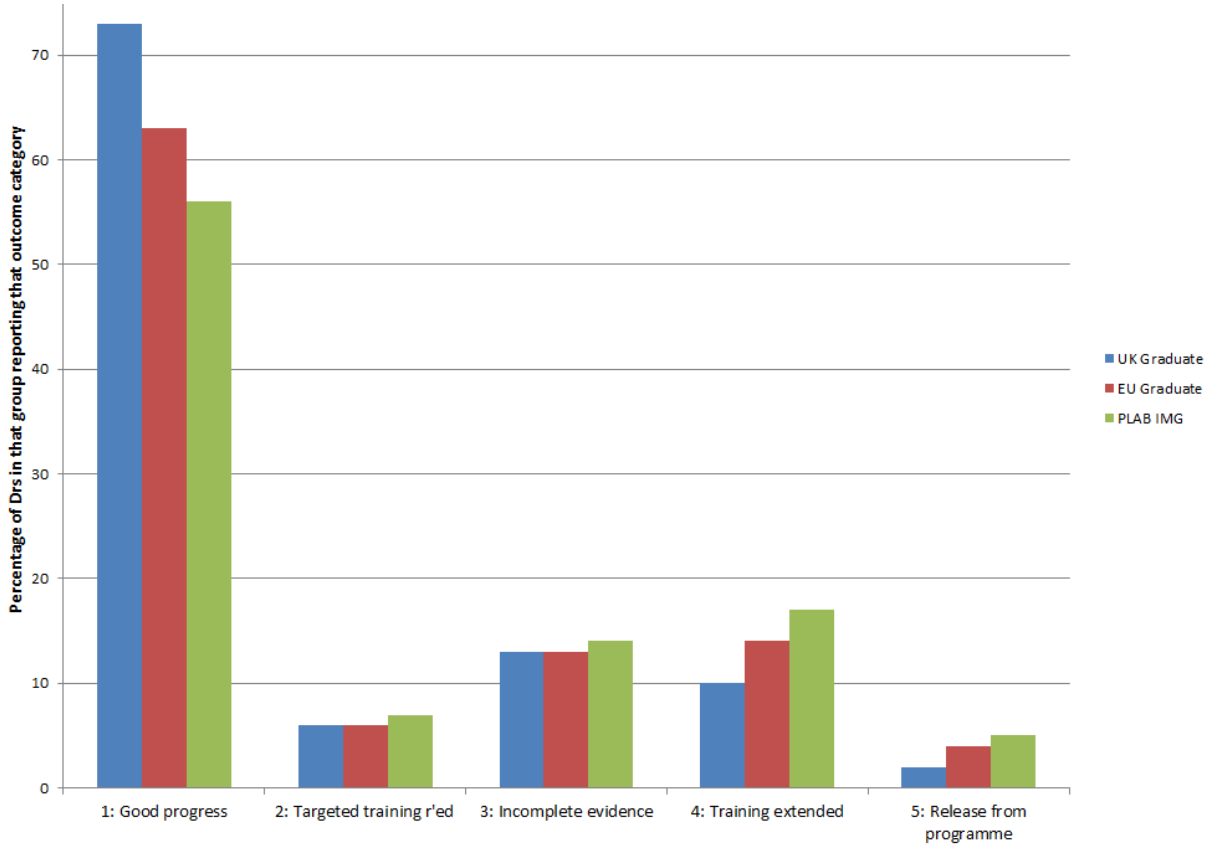

Figure 1. The lowest ARCP outcome recorded for post-graduate training doctors: percentages shown according to the world region where the Primary Medical Qualification (PMQ) was obtained (EEA PLAB graduates not included for this analysis).

Expansion: We used multivariable logistic regression in order to develop models for the prediction of a doctor receiving a 'sub-optimal' outcome on at least one ARCP taken. These models were developed using a forward stepwise procedure and adjusted for the potential effects of age, experience, sex and the number of ARCPs taken. There is some evidence that women may outperform men in post-graduate medical examinations. [12b] The decision was taken to also provide results adjusted for age and years of UK-based experience was taken as it was apparent from univariable analyses that both these factors had a significant non-linear relationship with the probability of successfully passing all ARCPs taken. Age was calculated based on the year of birth and the time the ARCP taken (where the poorest outcome was obtained). Duration of UK-based experience was calculated based on the year of provisional registration and time of ARCP. From our models we produced predicted probabilities that an individual doctor would obtain satisfactory outcomes at all ARCPs undergone. The predicted probabilities are depicted in graphical form in the 'box plots' in Figure 2 and Figure 3.

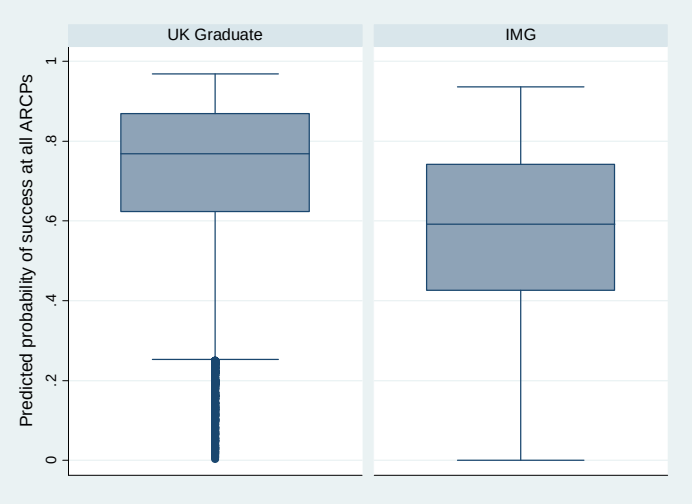

Figure 2. Box plots depicting the predicted probabilities of a UK graduate and PLAB IMG, respectively, receiving satisfactory outcomes at all ARCPs taken. The predicted probabilities are derived from multivariable logistic regression models.

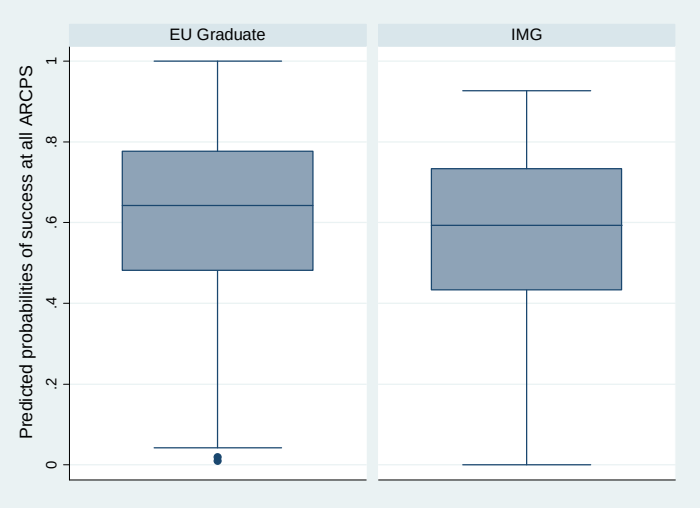

Figure 3. Box plots depicting the predicted probabilities of a UK and EEA graduate, respectively, receiving satisfactory outcomes at all ARCPs taken. The predicted probabilities are derived from multivariable logistic regression models.

As can be seen in Figure 2 UK graduates were generally predicted to experience a higher probability of passing all ARCPs with no negative outcomes compared to PLAB IMGs (all of whom had undergone the PLAB)($p<.0001$ for inter-group difference on Analysis of Variance [ANOVA]) . In Figure 3 it can be seen that EEA graduates had a tendency to be predicted a higher probability of passing all ARCPs with no negative outcomes compared to PLAB IMGs but these differences are slight, though statistically significant ($p<.0001$ for inter-group difference on Analysis of Variance [ANOVA]).

1.1.2 *Question:* ***does the IELTS overall score predict later success at the ARCP in PLAB IMGs?***

Summary Answer: Yes- IELTS score statistically significantly predicts the likelihood that a PLAB IMG will pass all ARCPs taken satisfactorily, even after controlling for the possible effects of other factors, such as the number of ARCPs taken, years of UK practice experience, sex and age.

Expansion: According to our models, the odds of a PLAB IMG having all satisfactory outcomes at ARCP increases by approximately 50% for every point scored on the IELTS. This holds true even after controlling for the effect for potential confounders, such as age, experience, sex and the numbers of ARCPs taken ($p<.001$).

1.1.3 *Question:* ***Which sub-scores of the IELTS (i.e. Writing, Speaking etc) are most predictive of ARCP performance in PLAB IMGs?***

Summary Answer: The reading score is most strongly predictive of ARCP success in PLAB IMGs.

Expansion: According to our models the odds of a PLAB IMG having all satisfactory outcomes at ARCP increases by approximately 30% for every point on the IELTS *reading* scale scored ($p<.001$). This holds true even after controlling for the effect of potential confounders, such as age, experience, sex and the numbers of ARCPs taken. The less predictive subscales of the IELTS were *comprehension* and *writing* (every point increasing the odds of success by around 12%, $p<.001$ in both cases).

1.1.4 *Question:* ***Would changing the IELTS mark at which IMGs could take the PLAB exam be likely to influence the difference observed between PLAB IMGs and UK graduates in terms of ARCP performance?***

Summary Answer: Yes.

Explanation: Although there are likely to be many factors contributing towards ARCP performance, English language, as indicated by IELTS score, is likely to be a key factor. Therefore we simulated changing the 'PLAB entry' mark of the IELTS by looking at those PLAB IMGs who had overall scores of at least 7.5, 8.0 etc on the assessment. Those PLAB IMGs who scored 8.5 or more on the IELTs were, on average, not statistically significantly poorer on ARCP performance compared to UK graduates, even after controlling for the potential effect of confounders (see Figure 4 for point estimates and error bars, indicating the magnitude and certainty regarding this inter-group difference). However, this is partly an artefact of the low numbers of doctors achieving such high IELTS scores, leading to lack of study power, as can be seen by the wide error bars in Figure 4. Certainly ARCP performance flattens out at an IELTS total score of 8.5 so there is little rationale for raising the pass mark beyond this, especially given the very small numbers of doctors achieving this mark. We would highlight that this does not necessarily imply that better language ability would not lead to further improvements in the postgraduate performance of PLAB IMGs; rather this may reflect the ability of the IELTS to differentiate candidates at the higher end of the communication ability range.

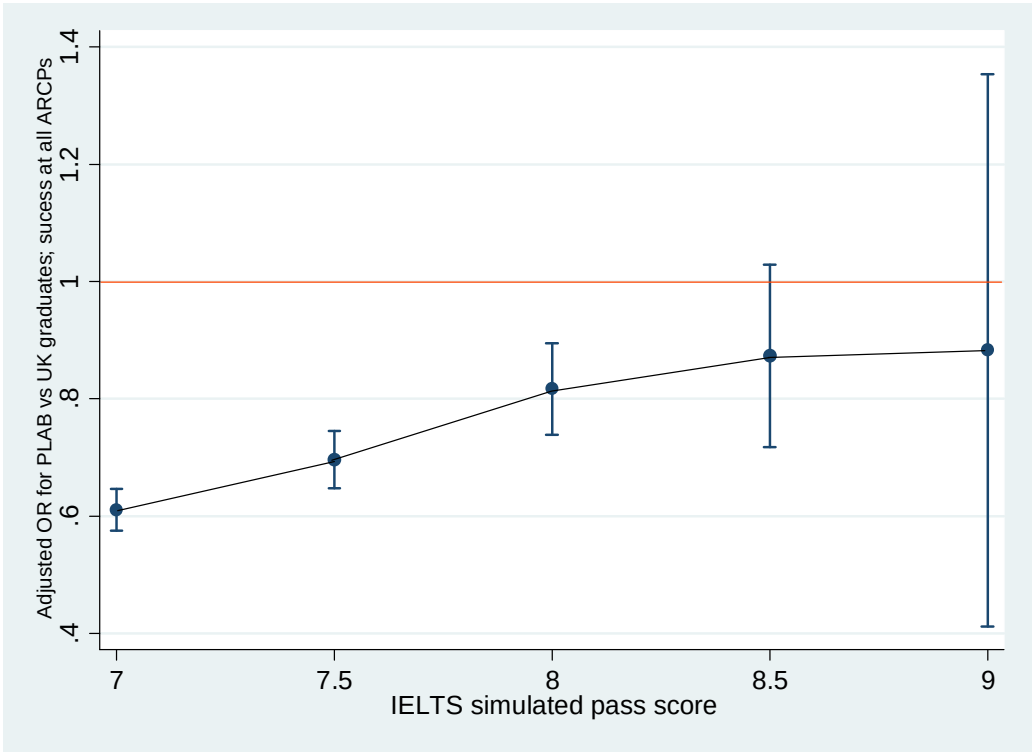

Figure 4**.** The adjusted odds ratios (ORs) for the risk that a doctor does not experience at least one poor ARCP outcome (PLAB vs UK graduates) for different simulated IELTS pass score thresholds. Note that the error bars cross an OR of 1.0 at an IELTS score of 8.5, which would indicate no significant difference between groups at the p<.05 level.

1.1.5. *Question:* ***Is there an IELTS threshold mark for which IMGs could take the PLAB exam that may lead to PLAB IMGs being equivalent, in terms of ARCP outcomes, to EEA graduates?***

Summary Answer: Yes. PLAB IMGs with an IELTS score of 8.0 or more, on average, perform better on the ARCPs compared to EEA graduates, controlling for the effects of potential confounders. This is depicted in Figure 5. Note that for three sets of analyses (those relating to paragraphs 1.1.5, 1.2.7 and 1.3.6), following advice from the GMC, it was decided to retain the small number of EEA graduates who had taken the PLAB (N=863) in the analysis. This is because we were advised that the most probable reason that these few EEA graduates were recorded as having sat the PLAB was that they had originated from European countries that had recently joined the European Union. We therefore felt, at least for some analyses, it was important to compare IMGs with doctors from the EEA, as it stands currently. As might be expected given the small numbers of EEA PLAB graduates, our sensitivity analyses suggested very little difference in results whether they were included or not. However, we noted that, on average, PLAB EEA graduates performed more poorly at ARCP than EEA doctors who had not taken the PLAB. Therefore the lines depicted in Figures 5, 9, 10 , 13 and 14 would be expected to shift very slightly in favour of EEA graduates if these European doctors who sat the PLAB were excluded from the analyses.

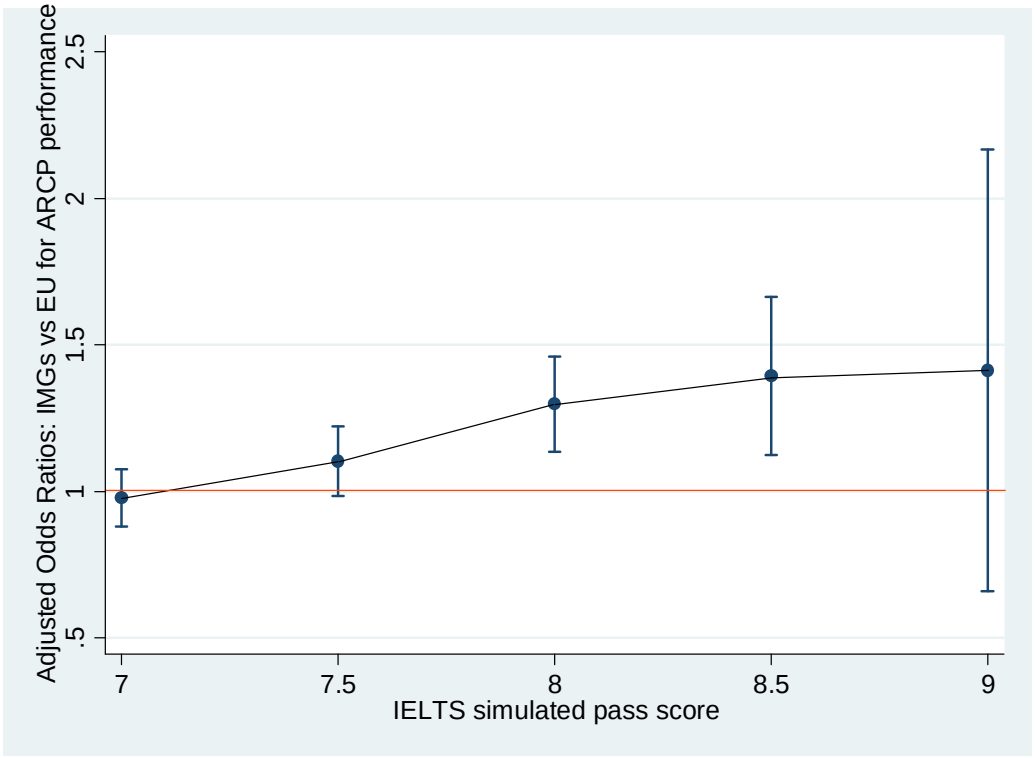

Figure 5. The adjusted odds ratios (ORs) for the risk that a doctor does not experience at least one poor ARCP outcome (PLAB IMGs vs EEA graduates) for different simulated IELTS pass score thresholds. Note that the error bars cross an OR of 1.0 at an IELTS score of 8.0, which would indicate a significant difference between groups.

Expansion: We simulated changing the pass mark of the IELTS by looking at those PLAB IMGs who had overall scores of at least 7.5, 8.0 etc on the assessment. Those PLAB IMGs who scored 8.0 or more on the IELTs, on average, performed statistically significantly *better* at ARCPs compared to EEA graduates, even after controlling for the potential effect of confounders (the magnitude and certainty around these estimates can be seen from the error bars in Figure 5).

1.2 Results relating to PLAB Part 1 Scores

1.2.1 *Question:* ***Are the number of attempts at PLAB part 1 related to later ARCP performance in PLAB IMGs?***

Summary Answer: Yes. The number of attempts at PLAB part 1 predict poorer performance at ARCP in PLAB IMGs.

Expansion: In this dataset the odds of candidates passing the PLAB part 1 fall steeply but relatively steadily after the first attempt, as depicted in Figure 6. As one would expect, given the pass rates, the number of candidates taking each resit falls exponentially with each attempt. For example, in the present dataset roughly 30,000 candidates took the PLAB part 1 at least once, almost 9,000 had at least one resit and

only around 3,000 took a third attempt. Thus, most (around 90%) of candidates in our dataset passed part 1in the first three attempts. Note, that the observed pass rates for both parts of the PLAB in the candidates for which data were supplied to us are significantly higher than those generally reported (usually cited as around 40-50% for part 1 and around 70% for part 2). This is because the GMC mainly only provided data on doctors who had passed the exam within the study timeframe, although a small number of candidates who failed to do so did appear in the data extract (around 1,500) and were excluded from the analyses.

In terms of the predictive properties of resitting; according to our multivariable model the odds of passing all later ARCPs satisfactorily reduces by around 20% for every attempt at PLAB part 1, even after controlling for other factors such as sex, age, UK experience and the number of ARCPs taken (p<.001). If we break IMG candidates down into those who took the PLAB part 1 once (n=19,610), twice (n=5,095), three times (n=1,766) or four times or more (n=1,297) we can explore this effect in more detail. Using this approach we noted that, on univariable analysis, within the sample of PLAB IMGs, passing the PLAB part 1 at first sitting (as opposed to needing at least one resit) was associated with a 53% higher odds of having satisfactory ARCP outcomes in all cases (p<.001). This value rose to around 70% when the effects of the number of ARCPs taken, sex, age and UK-based experience were controlled for in a multivariable model. Compared to passing first time, having one resit reduced the odds of consistently good ARCP outcomes by around 28% and two or more resits by roughly 45-40%. When these results were adjusted for potential confounding factors (as above) the values were 36% (one resit only) and around 50% respectively (two or more resits required)(p<.001 in all cases).

1.2.2 *Question:* ***Is the PLAB part 1 total score at initial sitting, relative to the pass mark for that sitting, predictive of ARCP performance in PLAB IMGs?***

Summary Answer: Yes. PLAB part 1 score at initial sitting (relative to pass) statistically significantly and independently predicts later ARCP performance.

Expansion: According to our multivariable logistic regression model the odds of being successful at every ARCP undergone increases by roughly 1-2% for every point scored above the pass mark (p<001). This effect is apparent even after controlling for potential confounding variables.

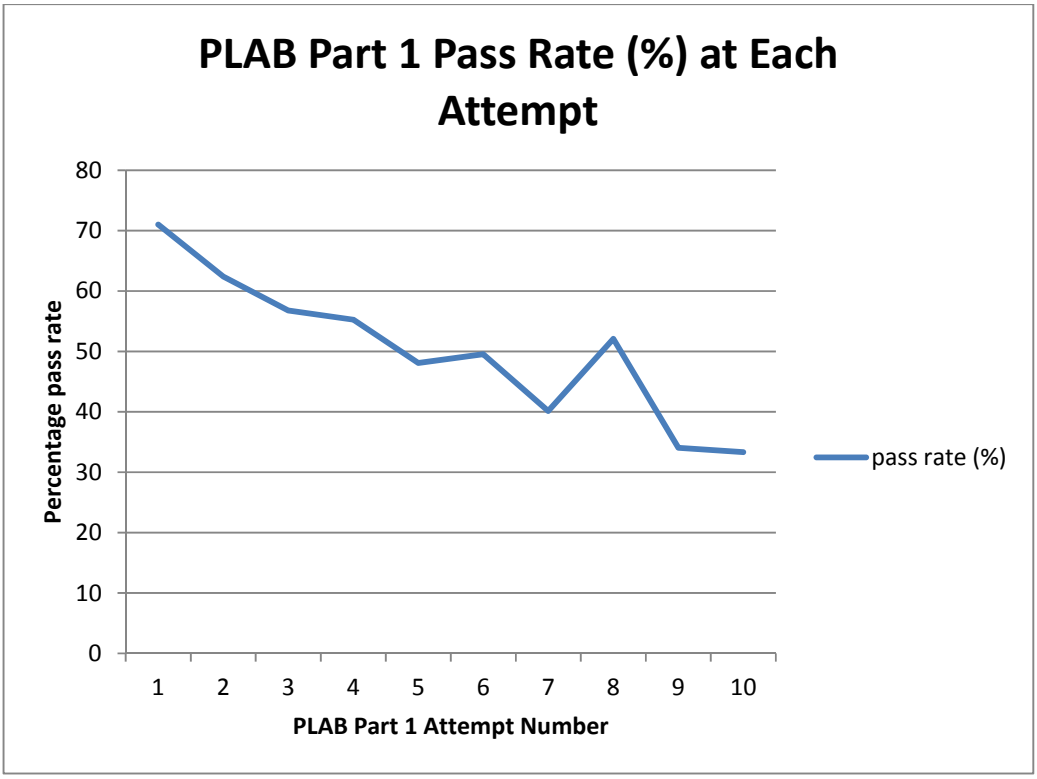

Figure 6. The proportion of candidates passing the PLAB 1 at each attempt in the dataset supplied to us (this included 27,768 doctors who had attempted the PLAB part 1 at least once and passed the overall PLAB during the study period).

1.2.3. *Question:* ***Is the PLAB part 1 total score at initial sitting, relative to the pass mark for that sitting, predictive of ARCP performance by PLAB IMGs in all medical specialities?***

Summary Answer: No. The findings suggest this effect may vary by medical speciality group.

Expansion: PLAB part 1 score is independently predictive (i.e. adjusted for confounders) of later ARCP performance in most specialities ($p<.05$ in all cases) but not *emergency medicine*, *ophthalmology* and *public health* ($p>.05$ in all three cases). However, for *public health* only data on 23 PLAB IMGs were available and so there was limited ability to demonstrate an effect due to lack of study power.

1.2.4. *Question:* ***Which sub-scores of the PLAB Part 1, at initial sitting, are most predictive of ARCP performance for PLAB IMGs in general?***

Summary Answer: Our results must be interpreted very cautiously as PLAB subscale scores are not standardised in any sense. However, overall the *diagnosis* section of the PLAB part 1 was most strongly predictive of ARCP performance with every percentage scored increasing the odds of success at all ARCPs taken by just under 3% ($p<.001$).

Expansion: When PLAB part 1 subscale scores (entered as percentages of correct answers) are entered into a multivariable model all have similar ability to independently predict later ARCP performance in a PLAB IMG. Overall each percentage point subscale increases the odds of having success at all ARCPs by roughly 2.5% (p<.001). This effect persists after controlling for the effects of sex, age, experience in the UK and number of ARCPs taken (p<.001 in all cases).

1.2.5. *Question:* ***Which sub-scores of the PLAB Part 1, at initial sitting, are most (and least) predictive of ARCP performance for PLAB IMGs in different medical speciality groups?***

Summary Answer: Our results must be interpreted very cautiously as PLAB subscale scores are not standardised in any sense. However, the *context* score for PLAB 1 is most predictive of ARCP performance for *anaesthetics*, *medicine*, *paediatrics*, *pathology and clinical sciences* and *radiology*. In these specialities each percentage point increases the odds of ARCP success by roughly 3-6% (p<.001 in all cases). The *context* subscale does not statistically significantly and independently predict later ARCP outcomes for *emergency medicine*, *ophthalmology* and *public health*.

The scores for the *investigations* subscale of the PLAB part 1 were most predictive of later ARCP performance in *anaesthetics*, *paediatrics* and *radiology*. In these specialities each percentage point increases the odds of ARCP success by roughly 3-4% (p<.001 in all cases). In contrast, the *investigations* subscale was not independently and statistically significantly related to ARCP performance in *emergency medicine*, *ophthalmology, pathology and clinical sciences* and *public health* (p>.05 in all cases).

The scores for the *management* subscale of the PLAB part 1 were most predictive of later ARCP performance in *anaesthetics, emergency medicine* and *pathology and clinical sciences*. In these specialities each percentage point increases the odds of ARCP success by roughly 4-5% (p<.001 in all cases). In contrast, the *management* subscale was not independently and statistically significantly related to ARCP performance in *surgery*, *ophthalmology, public health and radiology* (p>.05 in all cases).

The scores for the *diagnosis* subscale of the PLAB part 1 were most predictive of later ARCP performance in *anaesthetics*, *medicine* and *paediatrics*. In these specialities each percentage point increases the odds of ARCP success by roughly 4-6% (p<.001 in all cases). In contrast, the *diagnosis* subscale was not independently and statistically significantly related to ARCP performance in *emergency medicine*, *ophthalmology, public health and radiology (*p>.05 in all cases).

Expansion: When PLAB part 1 subscale scores (entered as percentages of correct answers) are entered into a multivariable model all have a variable ability to independently predict later ARCP performance in a PLAB IMG, depending on the medical speciality. However, in general, most of the subscales scores improve the odds of ARCP success by roughly 2% for every percentage point of items answered correctly(p<.05 in general).

1.2.6. *Question:* ***Is there a pass mark for the PLAB Part 1that, if set, is likely to lead to PLAB IMGs having equivalent performance at ARCP compared to UK graduates?***

Summary Answer: 'No', and 'Yes' depending whether you wish the PLAB part 1 to control for potential confounding factors or not.

Expansion: If you look at the 'raw odds' for PLAB IMGs obtaining satisfactory outcomes at all ARCPs there is no pass mark that can be set that results in equivalence to UK graduates. This can be clearly seen in Figure 7. In Figure 7 the raw odds ratios for ARCP success of PLAB IMGs vs UK graduates are plotted for groups of PLAB candidates divided into different duodeciles ('twelfths') according to their PLAB part 1 score at pass. It can be see that even those PLAB IMGs who were in the top duodecile (with a score at least 32 marks above the pass mark) have, on average, only 80% of the odds of having all satisfactory outcomes at ARCP compared to UK graduates.

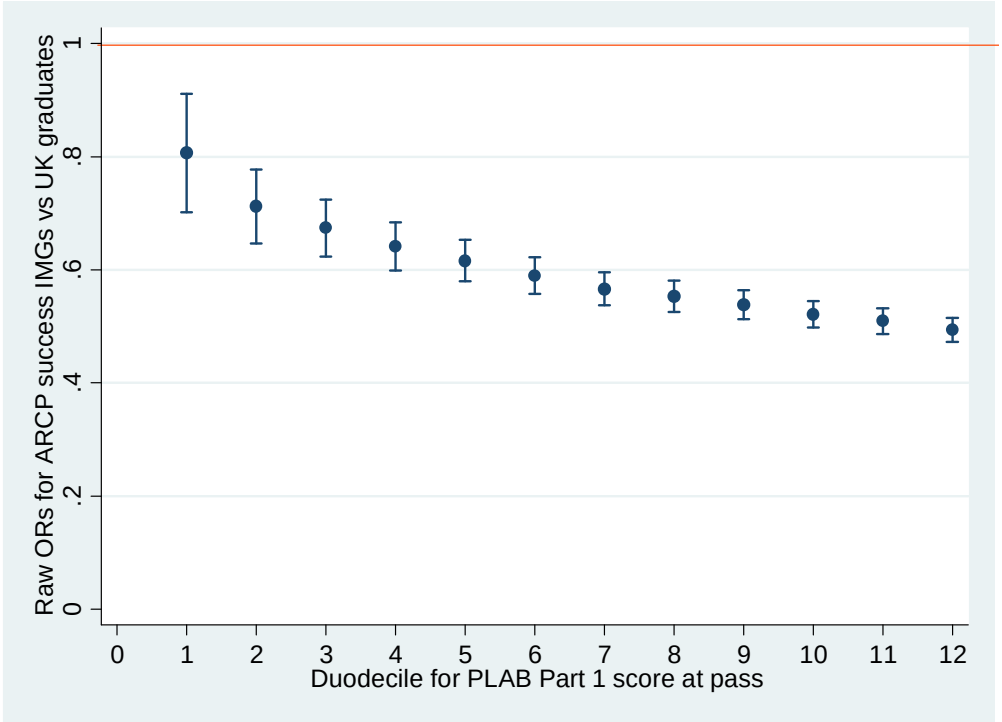

Figure 7. The estimated raw odds ratios (ORs) for a PLAB IMG vs a UK graduate experiencing satisfactory outcomes at all ARCPs according to PLAB part 1 pass score, relative to the pass mark for that sitting. The red line indicates where the OR=1 (i.e. no significant inter-group difference).

We now adjust the predicted odds for the effects of sex, age, years of UK-based experience and the number of ARCPs taken candidates. The results are presented graphically in Figure 8. We can now see, after adjusting for these potential confounders, that the candidates in the top two duodeciles perform as least as well as UK graduates at ARCP. Specifically, those with scores at pass of at least 27 points above the pass mark are equivalent to UK graduates.

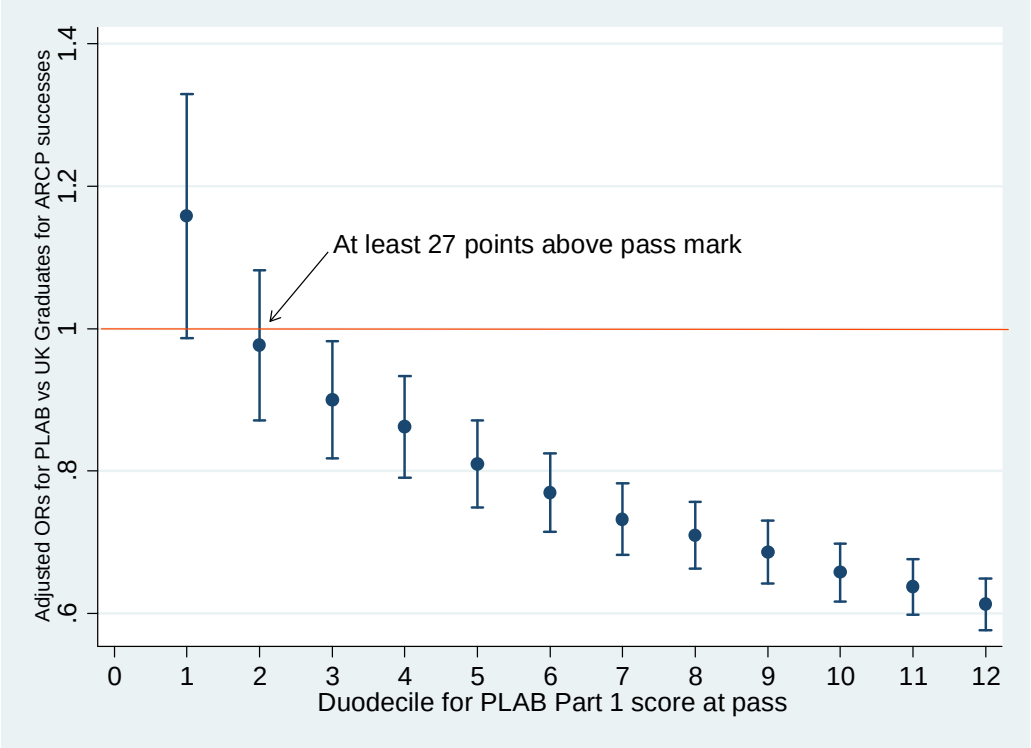

Figure 8. The odds ratios (ORs; adjusted for sex, age, number of ARCPs taken and years of UK experience) for a PLAB IMG vs a UK graduate experiencing satisfactory outcomes at all ARCPs according to PLAB part 1 pass score, relative to the pass mark for that sitting. The red line indicates where the OR=1 (i.e. no significant inter-group difference). It can be seen that candidated with a pass score of 27 above pass, placing them in the top sixth (two duodeciles) of testees perform at least as well as UK graduates.

1.2.7. *Question:* ***Is there a pass mark for the PLAB Part 1 that, if set, is likely to lead to PLAB IMGs having equivalent performance at ARCP compared to EEA graduates?***

Summary Answer: Yes. Even if UK based experience is not adjusted for then there are pass marks at PLAB part 1 that can potentially be used to achieve inter-group equivalence at ARCP performance.

Expansion: We repeat our approach to the comparison of UK vs PLAB IMGs according to different potential 'simulated pass scores' but for IMG PLAB candidates compared to EEA graduates. If you look at the 'raw odds' for PLAB IMGs obtaining satisfactory outcomes at all ARCPs the difference in performance is lost at around duodecile 7 or 8 (a score of at least 10 points above the pass mark)(see Figure 9). At duodecile 4 or above (a score of at least 20 above the pass mark), on average, ARCP performance is better amongst PLAB IMGs compared to EEA graduates. We can also see that if we adjust for all potential confounding factors (where data are available) PLAB IMGs and

EEA graduates do not, overall, differ in their later performance at ARCP. However, if we take only PLAB IMGs who scored 10 marks or higher above the pass level we can see they are, on average, statistically significantly *better* at ARCP compared to EEA graduates (Figure 10). The certainty of these inferences is communicated via the width of the error bars depicted in the relevant figures (wider bars indicating a greater degree of uncertainty regarding the magnitude of the inter-group differences).

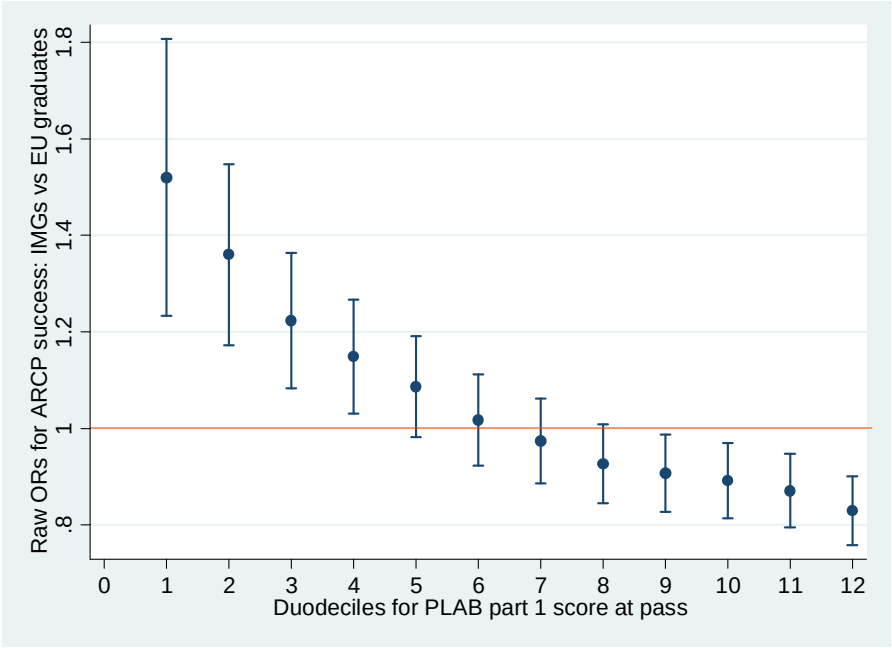

Figure 9. The estimated raw odds ratios (ORs) for a PLAB IMG vs an EEA graduate experiencing satisfactory outcomes at all ARCPs according to PLAB part 1 pass score, relative to the pass mark for that sitting. The red line indicates where the OR=1 (i.e. no significant inter-group difference).

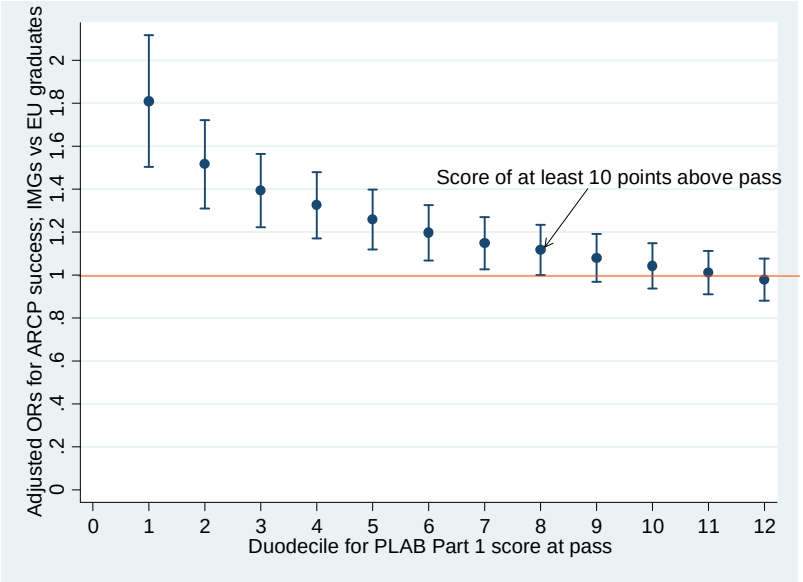

Figure 10. The odds ratios (ORs; adjusted for sex, age, number of ARCPs taken and years of UK experience) for a PLAB IMG vs an EEA graduate experiencing satisfactory outcomes at all ARCPs according to PLAB part 1 pass score, relative to the pass mark for that sitting. The red line indicates where the OR=1 (i.e. no significant inter-group difference). It can be seen that candidated with a pass score of 10 above pass, placing them in the eight duodeciles of testees, perform statistically significantly better at ARCP compared to EEA graduates.

1.3 Findings in relation to the PLAB Part 2

1.3.1. *Question:* ***Is the PLAB part 2 total score at initial sitting, relative to the pass mark for that sitting, predictive of ARCP performance in PLAB IMGs?***

Summary Answer: Yes.

Expansion: For PLAB IMGs the raw odds of obtaining a satisfactory outcome in all ARCPs increases by around 6% for every point above the pass mark for that diet ($p<.001$). When controlling for the effects of sex, age, number of ARCPs sat and duration of UK experience this effect remains, and indeed is slightly exaggerated, with the odds of ARCP success increasing by about 7% for every point scored above pass ($p<.001$).

1.3.2. *Question:* ***Is the PLAB part 2 total score at initial sitting, relative to the pass mark for that sitting, predictive of ARCP performance by PLAB IMGs in different groups of medical specialities?***

Summary Answer*:* No. This relationship is not observed for all groups of specialities.

Expansion: According to the multivariable models we developed PLAB part 2 score at first attempt statistically significantly predicts ARCP outcome in *general practice*, *paediatrics*, *surgery*, *medicine* and *psychiatry* ($p<.05$ in all cases). Specifically, for every point scored above the pass mark increases the odds of ARCP success at all assessments by roughly 6-10%. Some caution must be exercised in interpreting these findings as for the small specialities study power may be lacking and insufficient to demonstrate an effect at the $p<0.05$ level of significance.

1.3.3. Question: ***Which sub-scores of the PLAB Part 2, at initial sitting, are most predictive of ARCP performance for PLAB IMGs in general?***

Summary Answer: The performance on the *history taking* and *communication* subscales of the PLAB part 2 are most strongly predictive of later ARCP performance. *Examination* and *practical skills* are the least predictive subscales.

Expansion: According to the multivariable models the odds of success at all ARCPs taken increase by approximately 70% for every average point achieved on the *history taking* subscale, independent of potential confounders ($p<.001$). This value is 65% for *communication*, 47% *examination* and only 20% for *practical skills* ($p<.001$ in all cases). Such findings should be treated cautiously as sub-scores are not standardised between diets and therefore cannot be confidently interpreted across sittings.

1.3.4. Question: ***Which sub-scores of the PLAB Part 2, at initial sitting, are most predictive of ARCP performance for PLAB IMGs in different medical speciality groups?***

Summary Answer: The *communication* subscale scores are most predictive of ARCP success for *ophthalmology* (ORs 5.10) and *paediatrics* (OR 2.14). The *examination skills* sub-score is most predictive of ARCP performance in *medicine* (OR 1.77) and anaesthetics (OR 1.61). The *history taking* subscale score is most predictive of ARCP outcome in *pathology and clinical science* (OR 2.66) and *medicine* (OR 2.08). The *practical skills* sub-score is most predictive of ARCP in *general practice* (OR 1.38) and *ophthalmology* (OR 1.27). These subscale scores were all statistically significant predictors of ARCP performance in the respective specialities at the $p<.05$ level.

Expansion: According to the multivariable models the odds of success at all ARCPs taken independently increases by approximately 510% (i.e. increased by around five times) for every average point achieved on the *communication* subscale in *ophthalmology and roughly doubles for every point in paediatrics.* The odds of success at all ARCPs taken increases by approximately 77% for every average point achieved on the *examination* subscale in *medicine* by 61% *for* every point in *anaesthetics. T*he odds of success at all ARCPs taken increase by approximately 266% for every average point achieved on the *history taking* subscale in pathology by 208% *for* every point *in medicine.* The odds of success at all ARCPs taken increase by approximately 38% for every average point achieved on the *practical skills* subscale for ARCPs in *general practice* and by 27% for every point in *ophthalmology.*

1.3.5. *Question:* ***Is there a pass mark for the PLAB Part 2 that, if set, is likely to lead to PLAB IMGs having equivalent performance at ARCP compared to UK graduates?***

Summary Answer: 'No' and 'yes'.

Expansion: If you look at the 'raw odds' for PLAB IMGs obtaining satisfactory outcomes at all ARCPs there is no PLAB part 2 pass mark that can be set that results in equivalence to UK graduates. This can be clearly seen in Figure 11. In Figure 11 the odds ratios for ARCP success of PLAB IMGs vs UK graduates are plotted for groups of PLAB candidates divided into different duodeciles ('twelfths') according to their PLAB part 1 score at pass. It can be see that even those PLAB IMGs who were in the top duodecile (with a score at least 12.1 marks above the pass mark) have, on average, only 70% of the odds of having all satisfactory outcomes at ARCP compared to UK graduates.

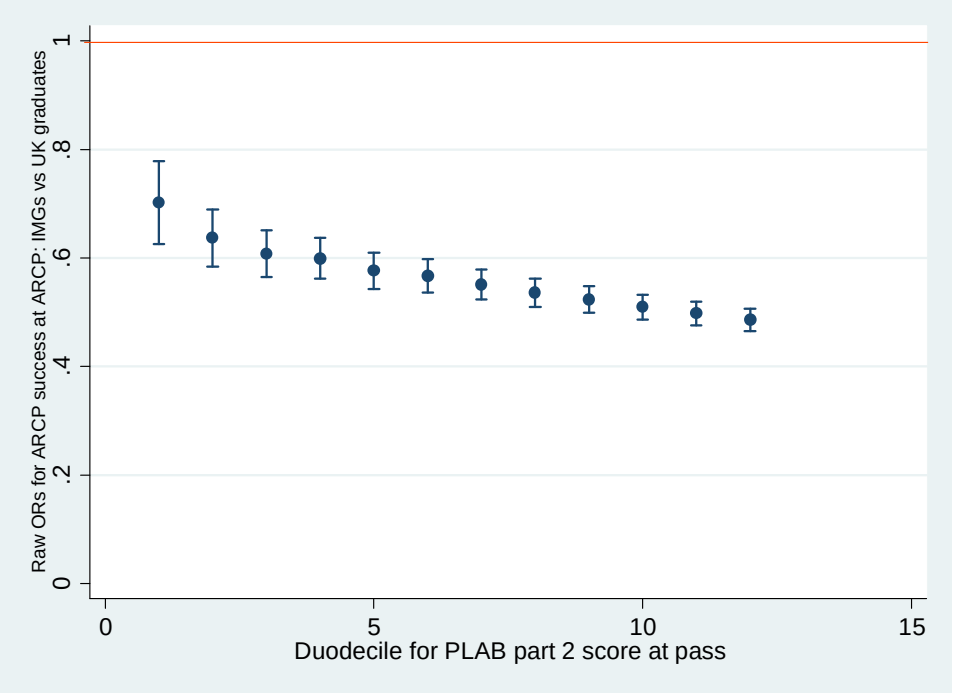

Figure 11 The estimated raw odds ratios (ORs) for a PLAB IMG vs a UK graduate experiencing satisfactory outcomes at all ARCPs according to PLAB part 2 pass score, relative to the pass mark for that sitting. The red line indicates where the OR=1 (i.e. no significant inter-group difference).

We now adjust the predicted odds for the effects of sex, age, years of UK-based experience and the number of ARCPs taken by candidates. The results are presented graphically in Figure 12. We can now see, after adjusting for these potential confounders that the candidates in the top duodecile for PLAB part 2 scores perform approximately as well as UK graduates at ARCP. Specifically, those with scores at pass of at least 12.1 points above the pass mark are roughly equivalent to UK graduates, or at least not statistically significantly worse. However, it should be noted that, in total, only 3.03% (842 of 27,763) PLAB IMGs obtained PLAB scores at pass at *both* parts 1 and 2 that

would render them roughly equivalent in ARCP performance to UK graduates. To recap this is equivalent to obtaining a score of at least 27 marks relative to pass for PLAB part 1 and at least 12.1 marks relative to pass for part 2.

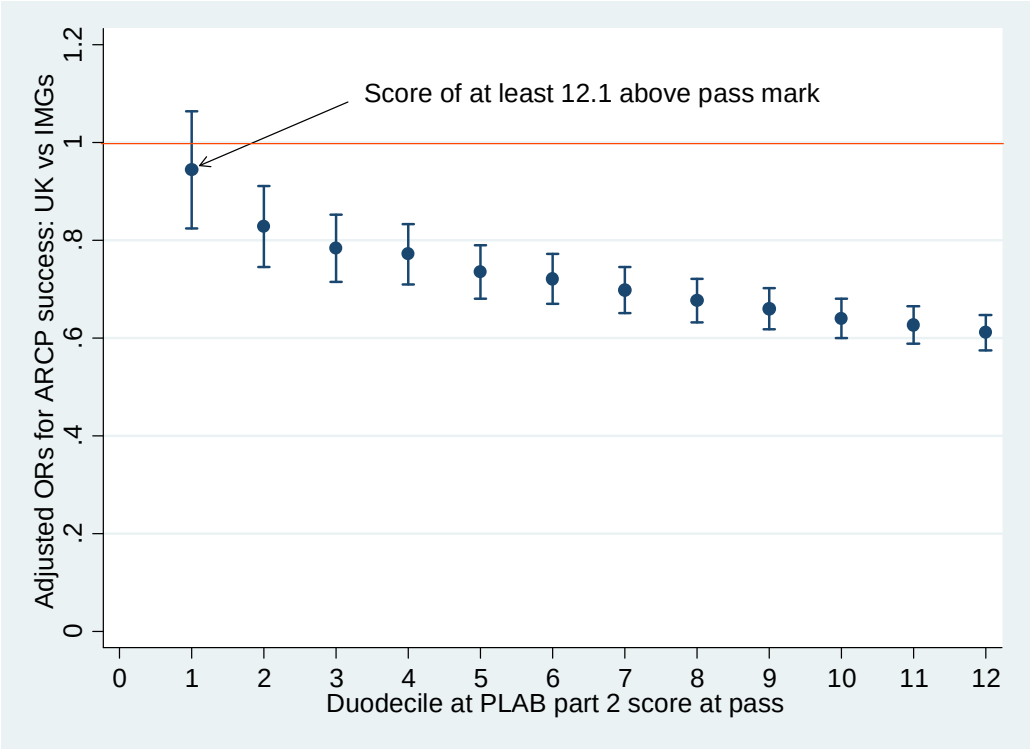

Figure 12. The odds ratios (ORs; adjusted for sex, age, number of ARCPs taken and years of UK experience) for a PLAB IMG vs a UK graduate experiencing satisfactory outcomes at all ARCPs according to PLAB part 2 pass score, relative to the pass mark for that sitting. The red line indicates where the OR=1 (i.e. no significant inter-group difference). It can be seen that candidated with a pass score of 12.1, or more, above pass, placing them in the top duodecile of testees, perform roughly similarly to UK graduates.

1.3.6. *Question:* ***Is there a pass mark for the PLAB Part 2 that, if set, is likely to lead to PLAB IMGs having equivalent performance at ARCP compared to EEA graduates?***

Summary Answer: Yes.

Expansion: If you look at the 'raw odds' for PLAB IMGs obtaining satisfactory outcomes at all ARCPs once PLAB part 2 pass marks fall into, or above, the fourth duodecile (i.e. 8.5 marks or more above pass) PLAB IMGs perform at least as well as EEA graduates. Indeed those with a pass mark of 6.05 or more above the pass level (duodecile 7 or more) are not statistically significantly worse at ARCP than EEA graduates.

This can be clearly seen in Figure 13. In Figure 13 raw odds ratios for ARCP success of PLAB IMGs vs UK graduates are plotted for groups of PLAB candidates divided into different duodeciles ('twelfths') according to their PLAB part 2 score at pass.

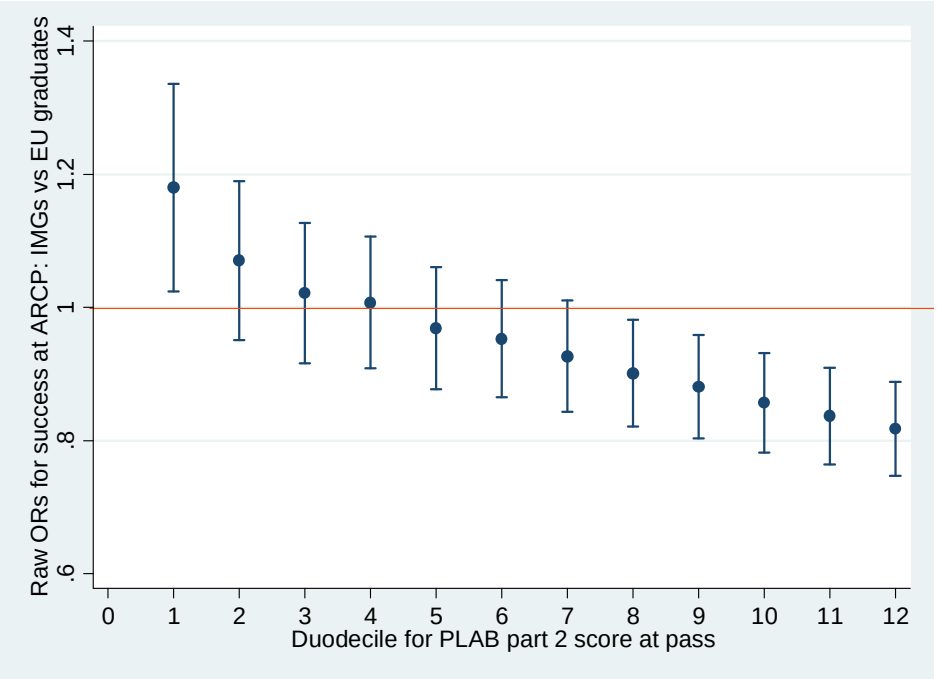

Figure 13. The estimated raw odds ratios (ORs) for a PLAB IMG vs an EEA graduate experiencing satisfactory outcomes at all ARCPs according to PLAB part 2 pass score, relative to the pass mark for that sitting. The red line indicates where the OR=1 (i.e. no significant inter-group difference).

We now adjust the predicted odds for the effects of sex, age, years of UK-based experience and the number of ARCPs taken by candidates. The results are presented graphically in Figure 14. We can now see, after adjusting for these potential confounders that, in general, PLAB IMGs perform approximately as well as EEA graduates at ARCP. Specifically, those PLAB IMGs with scores at least 6.85 points above the pass mark are, on average, superior to EEA graduates in this respect.

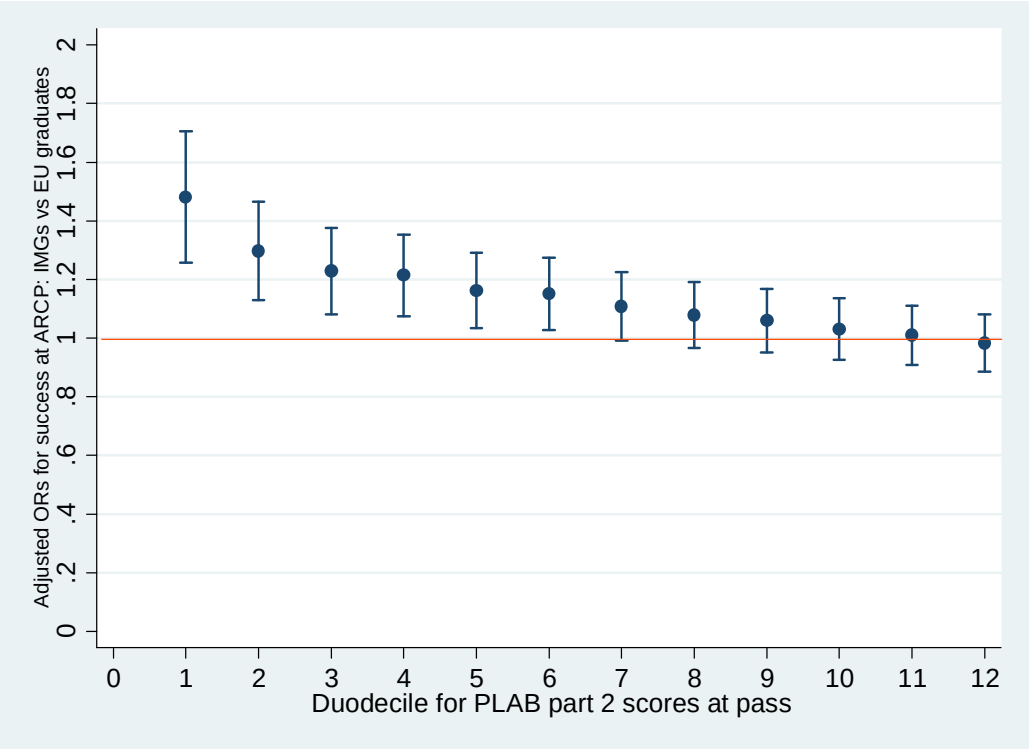

Figure 14. The odds ratios (ORs; adjusted for sex, age, number of ARCPs taken and years of UK experience) for a PLAB IMG vs an EEA graduate experiencing satisfactory outcomes at all ARCPs according to PLAB part 2 pass score, relative to the pass mark for that sitting. The red line indicates where the OR=1 (i.e. no significant inter-group difference).

1.4.1 *Question:* ***Are the number of attempts at PLAB part 2 predictive of later ARCP performance?***

Summary Answer: Yes.

Expansion: In this dataset the odds of passing the PLAB part 2 dramatically fall after the fourth attempt, as depicted in Figure 15. This is in contrast to the pattern we observed for PLAB 1 where pass rates showed a more steady decline from the first attempt. Analysis of pass rates was not possible after the fifth attempt for part 2 of the PLAB as, in this dataset, only very small numbers of doctors took more attempts. For example, whilst 157 IMG candidates took the PLAB part 2 for a fourth time, only 20 took it a fifth and only 8 were recorded as taking it a sixth time. Again, as in section 1.2.1, pass rates were higher than those generally reported (usually cited as around 70% for PLAB part 2) as our data extract was restricted to include only those doctors who had passed the PLAB exam within the study timeframe, although as noted earlier, a relatively small number of candidates were identified who had not (around 1,500 in number who were excluded from the analysis). Thus, it would be important to repeat this analysis in a more representative sample of PLAB candidates. It should be noted, that, as with PLAB

part 1, in the sample of candidates we studied, almost all doctors passed the PLAB part 2 within the first three attempts.

As with the PLAB part 1, according to our multivariable model, the odds of passing all ARCPs satisfactorily reduces by around 30% for every attempt at PLAB part 2, even after controlling for other factors such as sex, age, UK experience and the number of ARCPs taken (p<.01). Moreover, we can break down candidates into those IMGs that passed the PLAB part 2 first time (n=22,023), those who only took one resit (n=4,791) and those who took the exam three or more times (n=954).

Using this approach we noted that, on univariable analysis, within the sample of PLAB IMGs, passing the PLAB part 2 at first sitting (as opposed to needing at least one resit) was associated with a 55% higher odds of having satisfactory ARCP outcomes in all cases (p<.001, note; this is similar to the value we observed for those passing part 1 at first sitting). This value rose to around 61% (p<.001) when the effects of the number of ARCPs taken, sex, age and UK-based experience were controlled for in a multivariable model. Compared to passing first time, having one resit reduced the odds of consistently good ARCP outcomes by around 34% (p<.001) and two or more resits by roughly 47% (p<.001). When these results were adjusted for potential confounding factors (as above) the values were little changed, at 37% (one resit only) and around 46% respectively (two or more resits required) (p<.001 in both cases).

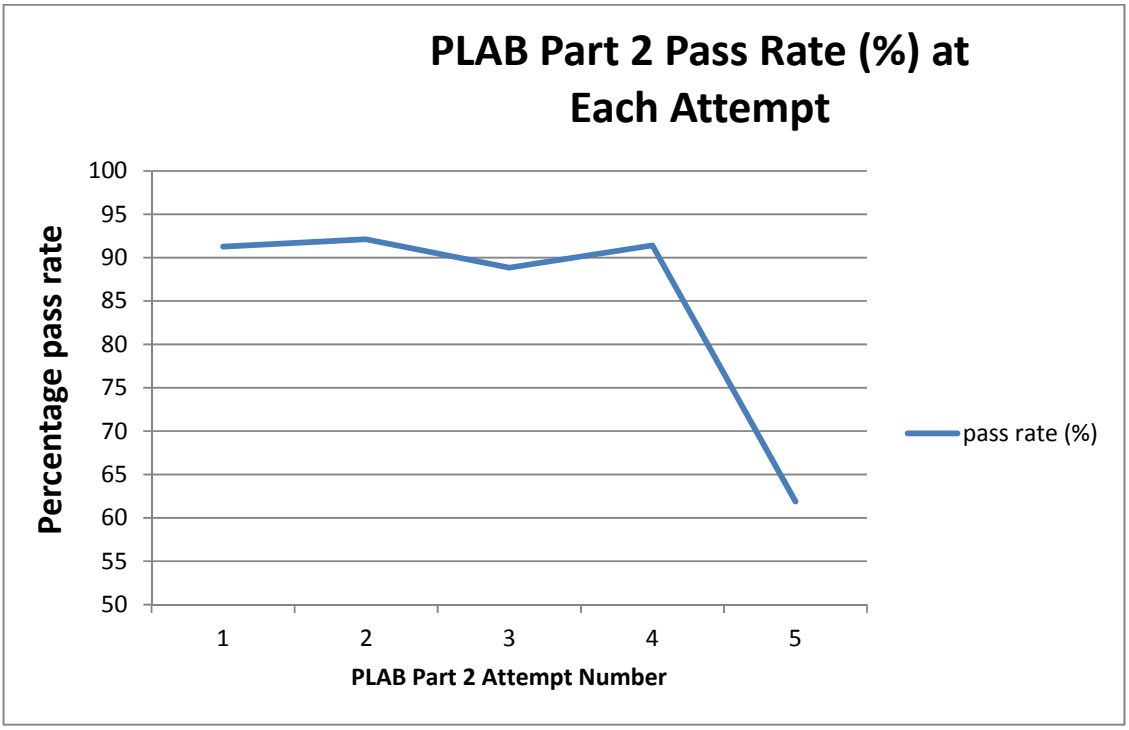

Figure 15. The proportion of IMG candidates passing the PLAB 2 at each attempt in the dataset supplied to us. Only IMGs who pass the PLAB during the study timeframe were included (N=27,768).

1.4.2. *Question:* ***Is the time elapsed between passing part I and taking part II predictive of success at the exam?***

Summary Answer: Yes and No.

Expansion: If you consider only the raw (unadjusted) odds of success at the PLAB part 2 there is a slight effect of the number of months that have elapsed between passing part 1 and taking part 2 of the PLAB on the success rate at first sitting of this latter exam. The odds of passing the PLAB part 2 at first sitting increase by 2% for every month elapsed ($p<.001$. However, if you adjust for the effect of potential confounders (i.e. sex, age, UK experience and the number of ARCPs taken) there is no significant effect of the time elapsed between the exams.

2. Fitness to Practice

Background

Powers are granted to the General Medical Council (GMC) under the Medical Act to investigate, and if appropriate, take action against doctors who are suspected of falling below the standards of competence or professionalism expected by the public and their peers. Reasons for referral to the GMC include the following: [13]

- misconduct
- poor performance
- a criminal conviction or caution in the UK or elsewhere for an offence that would be a criminal offence if committed in the UK
- physical or mental ill health
- a determination (decision) by a regulatory body either in the UK or overseas.

If a doctor is referred to the GMC regarding concerns about fitness to practise (FtP) then they may be referred to the Medical Practitioners Tribunal Service (MPTS) for a hearing. The MPTS is operationally separate to the GMC and directly accountable to parliament. The MPTS has the power to censure the doctor if they have been found to have significantly departed from the principles set out in *Good Medical Practice*. For the purposes of this report 'censure' is defined as the issuing of a formal warning or placing a sanction on the doctor, following the conclusion of a FtP process. Such sanctions range from the placing of restrictions on practice to erasure from the medical register.

The number of complaints made to the GMC about doctors has been steadily rising since 2007: in 2010 there were 7,153 complaints and in 2011 this rose by 23% to 8,781. However, the number of doctors annually censured after investigation remains relatively low at roughly 500 per year.[14]

2.1 Methods

Data on FtP processes relating to doctors provisionally or fully registered after 1998 were abstracted. It is important to note that only data on 5,572 cases that were *closed* (i.e. concluded) by the GMC between 2006 and 2012 (inclusive) were available. This

related to 5,472 doctors as some practitioners were referred to the GMC multiple times. In practice, due to the time it sometimes takes to process FtP referrals, the year of first FtP referral ranged from late 2001 to the end of 2011, though obviously most lay towards this latter period. However, it is important to note that the dataset may not have capture FtP cases on these doctors that were closed prior to 2006. This limitation must be borne in mind when interpreting the findings. In addition, some very early year of birth dates were noted in the overall dataset of registered practitioners- even for doctors who had registered after 1998 (one was recorded as 1894). On exploration of the data there were 11,601 doctors with registration recorded after 1998 but whose age was coded as more than 80 years. For those doctors with at least one FtP case closed in the study period the earliest year of birth recorded was 1935 (plausibly, only 27 doctors in this subset were recorded as being born before 1945). Therefore, for the purposes of the modelling only doctors born after 1934 were included. Also, it should be noted that both UK experience and age of the doctor was defined for these analyses as related to the end of the study period. This was to avoid confounding age and experience by defining it according to the time a doctor was first referred for a FtP concern. If this latter approach had been used to define age then the practitioners who had been recorded as being the subject of FtP referrals may have tended to automatically been recorded as (on average) to have been younger and less experienced than doctors in the wider 'control' dataset. Thus, we used the data available on all doctors born after 1934 and registered after 1998 in order to identify the predictors of both FtP referral and eventual censure (that is 124,986 'non-PLAB' and non-IMG graduates in total, which includes 89,463 UK and 35,523 EEA graduates). For these analyses the doctors were separated by region of PMQ as appropriate. The analytic approach used was predictive modelling via logistic regression. It should be noted that even following an introduction of this age and registration time cut-off for the final analytic dataset that PLAB IMGs were statistically significantly older (mean age; 38.1 years, SD; 5.25) compared to UK graduates (mean age; 32.9 years, SD; 4.75, $p<.001$ for difference on analysis of variance).

2.1.1. *Question:* ***Are PLAB graduates more likely to be referred for FtP concerns compared to UK and EEA graduates?***

Summary Answer: Yes.

Expansion: PLAB graduates are more frequently referred for FtP concerns compared to both UK and EEA graduates. Indeed, on univariable logistic regression analysis the odds that a PLAB IMG is referred for FtP is about 60% higher than for a UK graduate ($p<.001$) and 61% higher compared to an EEA graduate ($p<.001$). It should be noted, that in this dataset, EEA graduates were no more likely to be referred for FtP concerns compared to UK graduates. The proportions of doctors registered after 1998, born after 1934, referred for FtP concerns are displayed in Figure 16. The breakdown of numbers referred for FtP by region of PMQ are also depicted in Table 1.

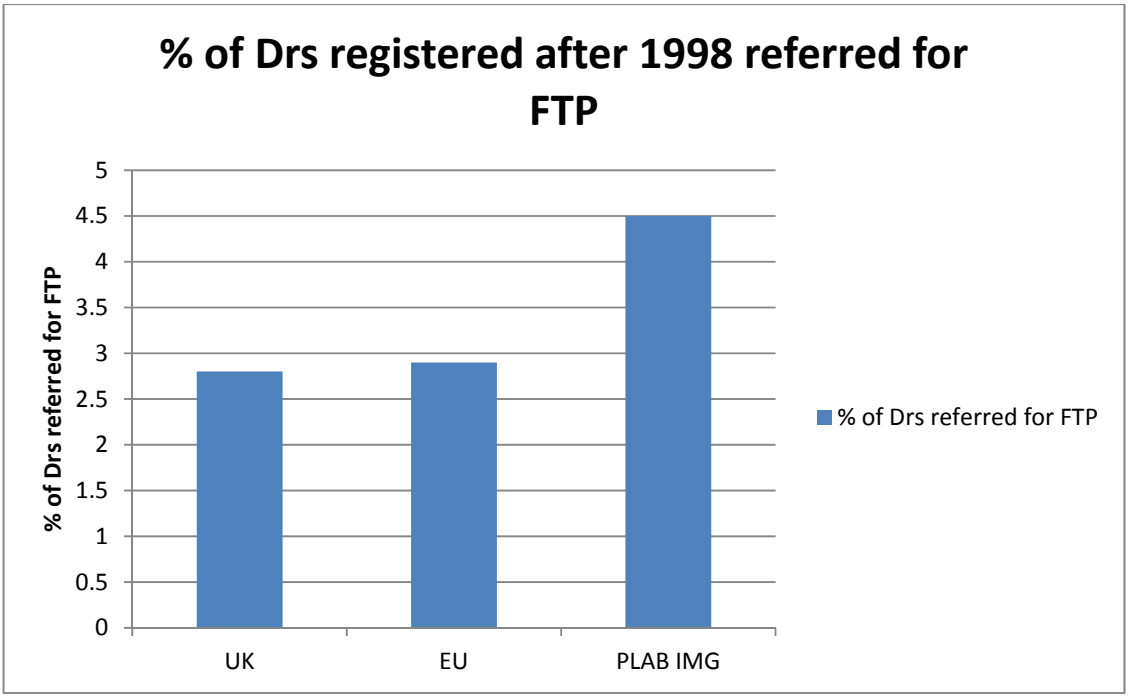

Figure 16. Proportion of doctors registered after 1998 (and born after 1934) referred to the GMC for Fitness to Practice (FtP) concerns according to the region of primary medical qualification.

	World Region of Primary Qualification		
Referred for FtP?	**UK**	**EU**	**PLAB IMG**
No	76,525	33,938	26, 573
Yes	2,135 (2.8%)	985 (2.9%)	1,187 (4.5%)

Table 1. The numbers and percentages of doctors registered after 1998 (and born after 1934) referred to the GMC for Fitness to Practice (FtP) concerns according to the region of primary medical qualification.** Note, that these figures do not include doctors who are IMGs but did not obtain registration via the PLAB process.

**Note where year of registration is present
Pearson Chi ^{2} = 177.46 $p < 0.001$

2.1.2. *Question:* ***Are PLAB graduates more likely than UK graduates to be referred for FtP concerns, even allowing for differences in age, UK-based experience and sex?***

Summary answer: Yes.

Expansion. It was important to control for the effects of age and UK-based experience (as obtained by the end of the study period). Previous research has shown that male sex is an independent risk factor for professional misconduct in UK graduates. [15] Age has a slightly more complex relationship with FtP; in the UK doctors under 30 years are less likely to be complained about than older practitioners but a higher proportion of concerns

directed at younger doctors are investigated by the GMC and result in censure.[16] Moreover, increasing age has not always been observed to be associated with an increased risk of being subject to censure; one study of 485 Australasian doctors who had been disciplined for professional misconduct found no significant difference in age between this group and the general doctor population. [17] Age may also correspond, to some degree with years in practice (in effect 'exposure time' for the risk of a FtP referral and censure) which further complicates the modelling process. Therefore in this case we used the results of exploratory statistical analysis as well as substantive reasons to guide the model building.

On univariable analysis the odds that a doctor was involved in a FtP process during the study period was around 4.5% higher per year of age (that is the age they were at the end of the study period in 2013). Similarly, on univariable analysis the odds that a doctor had been referred for FtP concerns were around 17% higher for every year of UK experience obtained by the start 2013. However, as with ARCP outcomes, experience showed a curvilinear relationship to the likelihood of FtP referral. Therefore polynomial terms were used to capture this. In the case experience terms up to the quartic (fourth power) were used to improve model fit. As the effect on risk of a FtP (and also censure) was more accurately predicted by years of UK practice than by age the former variable was entered into the model first. Once the relationship with UK experience was controlled for age then showed a linear relationship with risk. In our dataset, on univariable analysis, the odds of being referred for a FtP process were around 83% higher for males than for female doctors.

Once these factors were controlled for PLAB graduates had a very reduced but still higher probability of being referred for FtP; the odds were roughly 15% higher for PLAB IMGs compared to UK graduates (p=.005). However, by controlling for the effects of sex, age and UK experience the OR for a PLAB IMG being referred for FtP compared to an EEA graduate *increased* to 1.69 (i.e. roughly 69% higher odds, p<.001). This suggests that the mechanism by which EEA graduates end up being referred to FtP processes may be mediated by sex, age and/or UK experience to a greater degree compared to PLAB IMGs (see also section 2.1.4). To explain further; this observation is what you would expect to happen if the rates of FtP referrals in EEA graduates were mainly the result of the age and gender mix of this group of doctors i.e. once you compare younger and 'female' doctors in both groups the PLAB IMGs appear to have further relatively increased risks of referral for FtP. This suggests that it is age and 'maleness' that is, at least partly, mediating the odds of referral in EEAs, relative to PLAB IMGs.

2.1.3. Question: ***Are PLAB graduates more likely than UK graduates to be censured in any way following investigation for FtP concerns?***

Summary answer. Yes and No.

Expansion. Overall, PLAB IMGs have odds roughly 79% higher (p<.001) of receiving censure by the GMC in relation to FtP compared to UK graduates. The breakdown of the numbers and proportions by region of PMQ are depicted in Table 2 and in Figure 17. However, in contrast, within the pool of doctors *that have already been referred for FtP concerns* PLAB IMGs are not much more likely than UK (OR 1.16. p=.12) and only slightly more likely than EEA graduates (OR 1.26, p=.05) to have the FtP process conclude in censure. This suggests that the mechanism driving the over-representation

of PLAB IMGs in within the group of doctors eventually censured is operating at the *referral* stage of the process.

	World Region of Primary Qualification		
Censured for FtP?	**UK**	**EU**	**PLAB IMG**
No	78,318	34,115	27,545
Yes	342 (0.44%)	147 (0.41%)	215 (0.78%)

Table 2. The numbers and percentages of doctors registered after 1998 and born after 1934 censured following Fitness to Practice (FtP) processes according to the region of primary medical qualification.** Note, that these figures do not include doctors who are IMGs but did not obtain registration via the PLAB process.

**Note where year of registration is present
Pearson Chi ^{2} = 53.35 $p < 0.001$

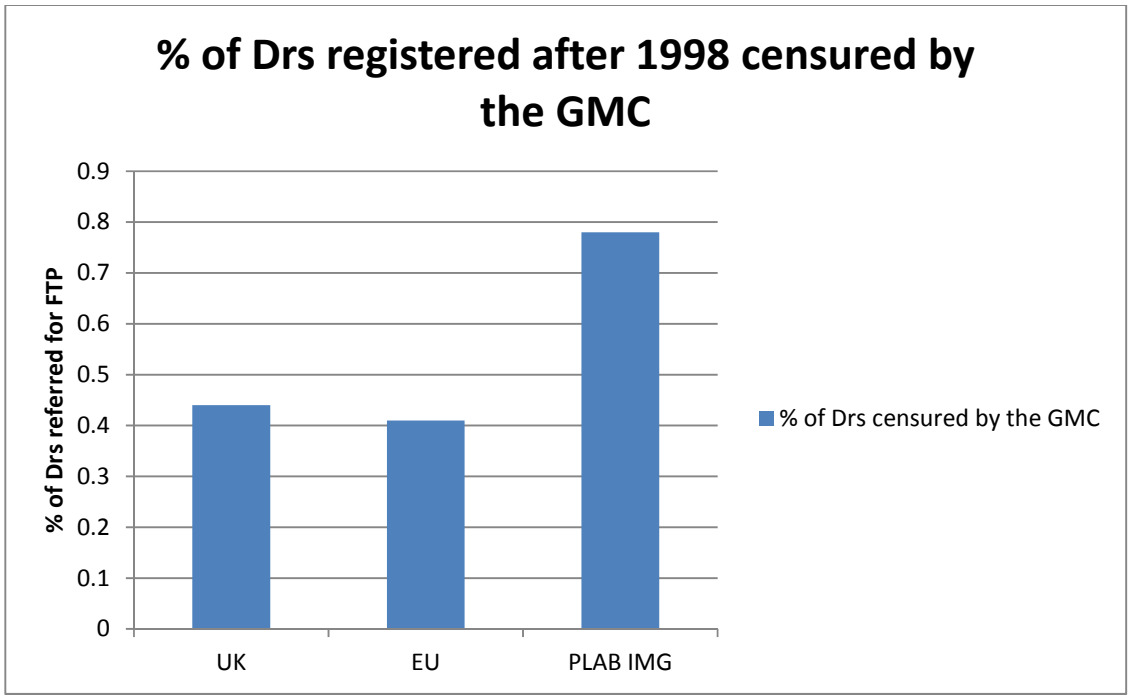

Figure 17. Proportion of doctors registered after 1998 (and born after 1934) censured by the GMC following investigation for Fitness to Practice (FtP) concerns. Proportions are broken down according to the region of primary medical qualification of the doctor.

2.1.4. *Question:* ***Are PLAB graduates more likely, overall, than UK graduates to be censured in any way following investigation for FtP concerns, even allowing for differences in age, UK practice experience and sex?***

Summary answer. No.

Expansion. As outlined in section 2.1.2, we used substantive and statistical findings to guide multivariable model building to relation to GMC censure as an outcome variable. The univariate relationships between the sociodemographic predictor variables and censure were similar to those observed for FtP referral. This is unsurprising given that referral is a pre-condition of censure. As with FtP referral as an outcome, UK-based experience showed a similar, curvilinear relationship with the likelihood of censure, modelled using terms up to the quartic. Again, once this effect had been adjusted for, age had a linear relationship with the outcome of interest, with older doctors having slightly higher odds of censure. Male sex was very strongly predictive of eventual censure, with the raw odds ratio being 3.60 (note: this figure was for all doctors in the analysed data set and suggested the odds of males being censured was over three and a half times greater than for females; for those referred for FtP concerns the OR was 2.35). After controlling for the influence of age, UK-based experience and sex the association between receiving censure for a FtP concern and being a PLAB IMG was no longer significant (OR .98, p=.4). Indeed, there was no statistically significantly increased risk of a PLAB IMG being censured following a FtP process, compared to UK graduates, even if only sex and age were controlled for (OR 1.09, p=.44). A similar picture emerged when comparing PLAB IMGS with EEA graduates who had been referred for FtP concerns. In this case once age, UK-based experience and sex were adjusted for there was no significant difference in the odds of a PLAB IMG receiving censure compared to an EEA graduate who had been referred for FtP concerns (ORs 1.24, p=.13).

2.1.5. *Question:* ***Does performance on any aspect of the IELTS or PLAB assessment, at pass, predict the likelihood of a fitness to practice referral, even allowing for differences in age, sex and UK-based experience?***

Summary answer. Yes.

Expansion. Although the IELTS is not part of the PLAB as such, it is the major gateway to the exam. Therefore we also explored whether performance on any component of the IELTS also predicted the probability of a FtP referral. There was a significant trend for higher IELTS *reading* scores (OR .88), p<.001) to independently predict reduced odds of a FtP referral. Conversely there was a trend of borderline significance for IELTS *speaking* scores to independently indicate slightly *higher* odds of a FtP referral (around for 4% every point, p=.05). Performance at PLAB part 1, both at first attempt and at pass statistically significantly predict a reduced likelihood of being referred for FtP concerns; the odds of this occurring roughly reduce by around 1% for every point scored above the pass mark. Likewise performance at PLAB part 2, both at first attempt and at pass statistically significantly predict a reduced likelihood of being referred for FtP concerns;

the odds of this occurring roughly independently reduce by around 2.5% for every point scored above the pass mark.

2.1.6. *Question:* ***Does performance on any aspect of the IELTS or PLAB assessment, at pass, predict the likelihood of censure in relation to a fitness to practice, even allowing for differences in age, sex and UK-based experience?***

Summary answer. Yes.

Expansion: None of the IELTS scores statistically significantly and independently predicted censure in PLAB IMGS, although there was a non-significant trend for IELTS *reading* scores (OR .84), p=.06) to predict reduced odds of censure. Performance at PLAB part 1 at first attempt (OR .99, p=.007) statistically significantly and independently predict a reduced likelihood of being censured for FtP concerns; the odds of this occurring roughly reduce by around 1% for every point scored above the pass mark. Likewise, performance at PLAB part 1 at pass (OR .98, p=.01) statistically significantly and independently predict a reduced likelihood of being censured for FtP concerns; the odds of this occurring roughly reduce by around 2% for every point scored above the pass mark. Performance at PLAB part 2 at first attempt (OR .95, p=.001), but not at pass, statistically significantly predicts a reduced likelihood of being censured for FtP concerns; the odds of this occurring roughly independently reduce by around 5% for every point scored above the pass mark at first attempt (p=.001).

In addition there was a trend for more attempts at both PLAB part 1 and part 2 to predict censure.

For PLAB part 1 we categorised IMGs into those with:

- 1 attempt only at PLAB part 1 (n=19,605)
- 2 attempts (n=5,094)
- 3 attempts (n=1,765)
- 4 or more attempts (n=1,296)

Note, these numbers are very slightly different to those cited for the section examining PLAB IMG performance in relation to ARCP outcome as eight PLAB IMGs with dates of birth recorded as prior to 1935 are excluded. Data on these doctors was not present in the dataset with ARCP outcomes recorded and so there was no need to exclude these, apparently anomalous, observations from the previous analyses concerned with ARCP. We used a multivariable logistic regression that treated PLAB part 1 attempts as a 'factor variable' with four levels to compare those who took the exam multiple times with those who passed at first sitting. We observed that, even after controlling for the effects of sex, age and UK-based experience, there was no significant difference between those who had taken the PLAB twice and those who passed first time (OR 1.23, p=.26); a trend of borderline significance between those who took the PLAB three times and those who passed at first sitting (OR 1.69, p=.04), and; a large difference between those who passed first time and those who had four or more attempts (OR 2.47, p=.001), in terms of the likelihood of GMC censure.

In addition there was a trend for more attempts at PLAB part 2 to predict censure. Exploring this using a multivariable analysis to control for the effects of sex, age and experience we divided IMGs into those with:

- 1 attempt only at PLAB part 2 (n=22,016)
- 2 attempts(n=4,790)
- 3 or more attempts (n=954)

Again, we observed that, even after controlling for the effects of sex, age and UK-based experience, there was no significant difference between those who had taken the PLAB part 2 twice and those who passed first time (OR 1.27, p=.17). However, those who took the PLAB three times or more were independently far more likely to receive censure than those who passed at first sitting (OR 2.75 p<.001).

2.1.7. *Question:* ***Of those candidates referred for FtP concerns, which sociodemographic variables predict eventual censure?***

Summary answer. Male sex and years of UK-based practice experience.

Expansion: Within the subset of FtP referred doctors for whom we had data, age did not predict eventual censure. However, male sex did statistically significantly predict eventual censure: male doctors referred for FtP had almost two and half times the odds of eventual censure compared to women (OR for male sex 2.35, p<.001). As with the broader set of doctors, UK-based experience had a curvilinear relationship with eventual censure. Those doctors who had practiced in the UK for five or six years were predicted to be most at risk and then a declining likelihood of censure following this period was modelled. It should be noted that this pattern was observed for the subset of doctors we had data for (and in effect were already partially matched for age and period of registration) and may not apply to the more general population of medical practitioners.

2.1.8. *Question:* ***Of those PLAB IMGs referred for FtP concerns, do scores on any elements of the PLAB independently predict eventual censure?***

Summary answer. Yes. The PLAB part 2 score on first attempt.

Adjusting for sex and UK-based experience in a multivariable model we noted that in the 1,187 IMG PLABs referred for FtP PLAB 2 score at first sitting (though not at pass) independently predicted a reduced odds of eventual censure by roughly 4% for every mark above the pass level. This was a trend of only borderline statistical significance (p=.04) though this may reflect the relatively small absolute numbers of PLAB IMGs eventually censured (n=215). It is worth noting that none of the IELTS scores were independently predictive of eventual censure following a FtP referral.

2. Overall Summary and Considerations

Even after controlling for effects of the available potential confounding factors, PLAB IMGs who passed the PLAB system demonstrate, on average, poorer performance on

ARCP compared to UK graduates. Raising the pass mark for the IELTS or PLAB may reduce the magnitude of this difference but is unlikely to eradicate it completely. In addition, changing the scoring for PLAB may impact on the numbers and proportions of PLAB IMGs referred for FtP concerns, though is less likely to affect proportions of overseas doctors eventually censured. It should also be remembered that this was a particular subset of IMGs we had access to data on and the results may not generalise to IMGs who do not obtain registration via the PLAB system, although such doctors will be included in published NCAS and more general FtP figures.

Considerations:

Considerations regarding the mode of use IELTS and PLAB test scores

1. Although not directly within the scope of the present review, the GMC PLAB review panel may wish to consider the effect on IMG performance of raising the threshold scores for IELTS, which is generally taken prior to the PLAB exams. We noted that at an IELTS score of 8.0 or more PLAB graduates outperform EEA graduates, as measured by ARCP outcomes. However, only at the highest IELTS scores do candidates approach equivalence with UK trained graduates, if at all, although clearly other factors, alongside English language ability, are likely to be important mediators of clinical competence. However, overall, this evidence suggests that IELTS performance is a significant predictor of later ARCP performance.

2. The GMC PLAB Review Panel may wish to consider raising the pass scores for both parts of PLAB. We understand that a separate independent report using alternative postgraduate educational outcomes recently included similar suggestions and we believe our present findings further support this position.

    a. At 27 or more points above pass at PLAB part 1 PLAB IMGs perform roughly as well at ARCP as UK graduates, adjusting for potential confounding effects.

    b. At 10 or more points above pass at PLAB part 1 PLAB IMGs outperform EEA graduates, as measured by ARCP outcomes, adjusting for potential confounding effects.

    c. At 12.1 or more points above pass at PLAB part 2 PLAB IMGs perform roughly as well at ARCP as UK graduates, adjusting for potential confounding effects

    d. At 6.85 points or more above pass at PLAB part 2, PLAB IMGs outperform EEA graduates, as measured by ARCP outcomes, adjusting for potential confounding effects

e. However, this will significantly reduce the number of IMG candidates passing these tests (to one sixth for PLAB part 1 and one twelfth for PLAB part 2)

3. As we have previously proposed after reviewing the performance of PLAB against the international literature, the number of resits should be limited. We proposed a limit of three resit attempts, followed by a personal development refractory period of at least two years. The data in this study is concordant with that recommendation.

Considerations regarding further development of the PLAB

4. It would be greatly desirable that the subscale scores for the PLAB, parts 1 and 2, are implemented in a manner that renders them comparable both with and between diets. A number of approaches are available to achieve such test-equating and standardisation.

5. More information on the dimensionality of the test would important (i.e. this may dictate how items are divided into subscales).

6. More information work on inter-rater reliability and the possible impact on candidate pass rates for the PLAB part 2 would be desirable. We are aware that previously OSCE station difficulty and examiner leniency or stringency has been explored using the Many Facets Rasch Model (MFRM)- a form of item response modelling. It would be also be interesting to evaluate how the results of such item response modelling may impact on examiner behaviour. For example, if examiners were aware that they tended to be 'hawkish' would their marking behaviour change?

7. Overall, the PLAB appears to be poorly calibrated. Specifically, the test appears to be too easy to pass given the aims of the assessment (i.e. to help ensure equivalence between PLAB IMGs and UK graduates at the end of FY1).

Other considerations

8. The GMC may wish to consider ways of strengthening the quality and completeness of data available on registered doctors, including PLAB candidates. This could be achieved via linking with other datasets, such as that held by the UKCAT consortium. During this analysis there were areas where the accuracy of the data could have been questioned; for example, the number of doctors on the medical register with very early dates of birth but with dates of registration after 1998. This is highly unlikely to have affected the validity of the present findings relating to ARCP outcomes. However, there is a slight, but significant chance that data quality could have potentially adversely affected the validity of the findings in relation to FtP, where a larger subset of data abstracted from the list of registered medical practitioners (LRMP) was used. It may therefore be worth considering repeating an analysis of FtP referrals and

outcomes in PLAB graduates once data quality is better assured and when a larger sample is available. Moreover, as more information becomes available relating to FtP it may be possible to conduct adequately powered studies where fitness concerns are divided by category (e.g. those relating to clinical competence, rather than conduct issues). More generally, there could also be opportunities to learn lessons by understanding more about the characteristics of those IMGs who demonstrate equivalent performance to UK graduates, as well as sub-analyses based on country of origin and place of graduation (these are not always the same).

9. The present report employed 'single level' analysis, in that each outcome (e.g. lowest ARCP grade) was linked uniquely to each doctor. However, it would be possible to explore the data within a multi-level modelling (MLM) framework given that multiple outcomes (i.e. more than one ARCP) could be conceptualised as nested within each individual doctor. Thus, the data could be considered hierarchical in nature. However, it is highly unlikely that the results of a more complex analysis would lead to significantly different conclusions than those drawn in the present report unless new data were made available that could enrich a MLM approach (e.g. the identity of the deanery in which each ARCP took place).

10. IMG PLAB graduates are, overall more likely to be referred to the GMC for FtP issues, and indeed, more likely to be censured. However, these differences can be mainly explained on the basis of differences in age and sex. Once these effects are controlled for there is relatively little difference in the risk of referral for FtP between PLAB IMGs and UK graduates. Moreover, after controlling for these potentially confounding factors there is no evidence that PLAB IMGs are more likely to be censured in relation to FtP concerns compared to UK graduates. Elements of the PLAB performance may be 'protective' against the risk of referral for FtP, if not eventual censure. By this we mean that high performance on some elements of the PLAB are *negatively* predictive of referral for FtP or censure. It is however, noteworthy, that there was a slight trend for higher scores on the *speaking* subtest of the IELTS to be modestly associated with a *greater* risk of referral for FtP in PLAB IMGs. This observation is somewhat anomalous in the context of the overall results of this study and may warrant further future exploration.

Acknowledgements

We are indebted to Thomas Jones and Andy Knapton at the General Medical Council for their consistent help and support in providing access to the data used in our analysis.

References

1. Clark PF et al, The globalisation of the labour market for health-care professionals. *Int Labour Rev*. 2006;145(1-2): 37-64.
2. General Medical Council 2011. The state of medical education and practice in the UK http://www.gmc-uk/publications/10586.asp [accessed 29.4.13]
3. BMA. 2013. Doctors new to the UK. http://bma.org.uk/practical-support-at-work/immigration/doctors-new-to-the-uk/immigration-regulations. [accessed 30.4.13]
4. NCAS 2009. National Clinical Assessment Service. NCAS Casework: The first eight years. http://www.ncas.npsa.nhs.uk/resources/publications/caseworkanalyses/; NCAS 2011. Concerns about professional practice and associations with age, gender, place of qualification and ethnicity – 2009/10 data. www.ncas.nhs.uk/publications/statistics.
5. Humphry C, Shaista Hickman, Gulliford MC; Place of medical qualification and outcomes of UK General Medical Council "fitness to practise" process: cohort study. *BMJ.* 2011;342:d1817.
6. Illing J, Kergon C(Rothwell), Morrow G, Burford B. The experiences of UK, EU and non-EU medical graduates making the transition to the UK workplace: Full Research Report, ESRC End of Award Report, RES-153-25-0097 [Internet]. Swindon: ESRC. 2009. http://www.esrc.ac.uk/my-esrc/grants/RES-153-25-0097/outputs/read/bb9d26ec-1a43-46eb-b8bd-b94d446cf984.
7. Morrow, G, Rothwell, C, Burford, B & Illing, J (2013). Cultural dimensions in the transition of overseas medical graduates to the UK workplace. *Medical Teacher* e1-e9, Early Online.
8. Esmail A and Roberts C. (2013). Academic performance of ethnic minority candidates and discrimination in the MRCGP examinations between 2010 and 2012: analysis of data. BMJ. 2013;347:f5662.
9. Rothwell C, Morrow G, Burford B, Illing J. 2013. Ways in which healthcare organisations can support overseas-qualified doctors in the UK. *Int J Med Ed*. 2013;4:75-82
10. General Medical Council. The Legal framework: Registration for International medical graduates.https://**gmc**.e-consultation.net/.../The%20**legal**%20**framework**%20-%20**re**... [accessed 28.4.13]
11. McLachlan J., Illing J, Rothwell C, Margetts JK, Archer J, Shrewsbury D. *Developing an evidence base for the Professional and Linguistic Assessments Board (PLAB) Test*. Report to the GMC, August 2012 (currently under embargo).
12. General Medical Council. Applying for registration as an International Medical Graduate. http://www.gmc-uk.org/doctors/before_you_apply/imgs.asp [accessed 15.11.2013].

12b. 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.

13. General Medical Council. A guide for doctors referred to the GMC. http://www.gmc-uk.org/concerns/doctors_under_investigation/a_guide_for_referred_doctors.asp [accessed 15.11.2013].
14. The General Medical Council. Record numbers of complaints against doctors-GMC report [press release, 18th September 2012](http://www.gmc-uk.org/news/13895.asp).[accessed 01.10.2012].
15. Yates J and James D. Risk factors at medical school for subsequent professional misconduct: multicentre retrospective case-control study. *BMJ*. 2010;340:c2040.
16. General Medical Council 2013. The state of medical education and practice in the UK http://www.gmc-uk.org/SOMEP_2013_web.pdf_53703867.pdf [accessed 14.11.13]
17. Elkin K, Spittal MJ, Elkin DJ, Studdert DM. Doctors disciplined for professional misconduct in Australia and New Zealand, 2000–2009. *Med J Aust*. 2011; 194 (9): 452-456.