Technical Appendix and Supplementary Results for:

Annual Review of Competence Progression (ARCP) performance of doctors who passed the Professional and Linguistic Assessments Board (PLAB) tests compared with UK medical graduates: national data linkage study

Paul A Tiffin MD, Jan Illing PhD, Adetayo S. Kasim PhD, John C McLachlan PhD

Foreword

In order to improve the readability of the linked article a number of technical details relating to the *data collection*, *modelling process* and *data imputation* are outlined in this appendix. In addition there are a number of supplementary results tables included, accompanied by an outline of interpretations.

Data collection processes

Data on ARCP outcomes recorded for the years 2010-2012 were obtained from the GMC. In 2010 and 2011 the Conference of Postgraduate Medical Deans (COPMeD) collected ARCP data from all the UK deaneries; from 2012 the GMC has managed the data collection process. The ARCP data is collected at event level, where one record is one ARCP outcome, so, for example, dual-CCT (Certificates of Completion of Training) trainees have one outcome recorded per specialty. All ARCP events between 3 August 2011 and 31 July 2012 were collected from the UK deaneries' databases in an annual retrospective report that asked the returning officer at each deanery to account for all outcomes awarded during the previous training year (August to July) and all trainees in training programmes at the deanery. Some trainees had more than one outcome awarded in a given reporting year.

The data collection process used an Excel spreadsheet file that contained validation rules. The data were further verified on receipt and queries were raised with the returning officers in deaneries. Data were checked against the list of permitted values for each field and for compliance with rules to ensure consistent recording across a given record, for example, the possible outcomes for a particular type of trainee. The data collection process is described in a series of briefing notes for deanery staff published on

the GMC website.[1] The GMC and the COPMeD worked with the Deanery Data Managers Group (DDMG) to ensure consistency of recording.[2]

Ethnicity is classified as sensitive personal data by the Data Protection Act. Following legal advice that consent would be required for deaneries to pass this to the GMC, ethnic status was collected directly from trainees via the National Training Survey conducted annually by the GMC. Additional demographic data such as sex and date of birth came from the List of Registered Medical Practitioners (LRMP).[3]

There were a few data relating to the Record of In-Training Assessment (RITA), which was the annual appraisal system for doctors in training prior to the ARCP being introduced. Where applicable RITA grades were converted into the equivalent ARCP outcomes (e.g. a RITA 'C' outcome was recoded as ARCP outcome '1': 'satisfactory progression'). For the purposes of analysis, satisfactory progression at ARCP was also defined to include successful completion of training (ARCP outcome 6) with eligibility for application to be placed on the specialist register. The structure of training varies across medical specialities. For example, obtaining royal college membership is sometimes a pre-requisite for entering 'higher specialist training', as in psychiatry, whereas in other specialities, such as pathology, it is not. Thus, for the purposes of this study we did not discriminate between 'core training' and 'higher specialist training'.

The GMC provided the data in anonymous format: each GMC number was transformed into a unique study identifier. The research team therefore could link each doctor in the dataset to the demographic and educational data held by the GMC but not derive the original registration number for the doctor. Data on ARCP outcomes were available for 60654 doctors. However, the present analysis was concerned only with those relating to educational performance in UK graduates and those international medical graduates that had obtained registration with the GMC via the PLAB system, as opposed to alternative routes. In total 125208 ARCP outcomes relating to educational progress and performance were available for 53436 doctors who were either UK graduates or international medical graduates registering via the PLAB test. The flow of data in the study is depicted in Figure 1 of the main paper.

The PLAB exam performance data, on both parts 1 and 2, along with information on the dates and number of sittings, were available for all 27726 international medical graduate candidates who passed the exam from July 2000 to the end of 2010. As pass marks vary between diets (sittings) the raw PLAB scores were converted to the score obtained by the candidate *relative to the pass mark for that sitting*. Marks obtained at a first attempt at postgraduate medical examinations may be considered the optimum metric of underlying ability (and hence the best predictor of future performance).[4] Thus this was the score that was used when exploring the ability of PLAB performance to predict later ARCP performance. However, PLAB mark at pass was used when modelling the potential impact of altering the pass mark thresholds for the two parts of the exam. Previous research on postgraduate medical examinations suggests that multiple resits may be associated with disproportionately poor performance.[5] Therefore we also categorised international medical graduates based on the number of resits at parts 1 and 2 of the PLAB in order to explore any association with later ARCP performance. For part 1, candidates were categorised as having one, two, three, or four or more sittings; for part 2 candidates were categorised according to whether they had taken the exam one, two, or three or more times before passing. These categories were decided upon as few doctors had taken the PLAB part 1 more than four times and the part 2 more than three times.

The International English Language Testing System (IELTS) scores for international medical graduates were also linked to the larger dataset via a unique identifier. These scores, available for 25896 international medical graduates who completed the PLAB test, were given as both the subtest scores (graded 7.0 to 9.0) and an overall score.

Demographic data were obtained from the GMC LRMP. These included the date of birth and the date of first registration (or provisional registration if present) that were used to calculate the age and the duration of UK-based experience for the doctor at each ARCP taken (in years). The research team only had access to a dichotomised version of ethnicity (white/non-white) though this variable was originally categorised according to the Office of National Statistics classification system.[6] This was to ensure that data were anonymous: collapsing categories is recommended by the information commissioner to avoid identification using combinations of variables.[7]

The sample was drawn from all medical specialities. For the purposes of descriptive analysis these were grouped into twelve specialities based on the nature of the clinical work expected to be undertaken during training. These were generally classified according to the UK royal colleges associated with the training scheme [8] with a number of exceptions; *public health* was treated as a separate speciality, having its own distinctive training; *clinical oncology* was grouped with medicine, rather than radiology (which is the royal college it is linked to). All clinical science specialities (e.g. immunology, medical virology etc.) were grouped with *pathology*. The validity of this grouping strategy was evaluated by assessing the number of doctors who fell into more than one speciality; in the event only 2472 (5%) of the sample doctors relating to 3328 'codable' ARCP outcomes (i.e. those related to performance) straddled groups. On closer exploration these were often doctors doing highly specialised training or changing speciality, for example between core medical training and general practice. Speciality was not available in eight cases. The proportions of doctors falling into each group are shown in table 3 of the main paper, along with the proportion of international medical graduates, compared to UK graduates. Table 3 of the main paper also displays the proportion of ARCPs taken in each speciality group that were categorised as satisfactory (i.e. ARCP outcome 1 or 6 or equivalent RITA rating). This proportion is inversely correlated with the percentage of international medical graduates in each speciality group (Spearman's rho =-0.61, p=0.04).

Model building

The relationship between ARCP outcome category and PLAB status was explored initially using multinomial logistic regression analyses to determine the extent to which the ARCP categories could be reliably distinguished (*discrimination*) and considered to be related to the same underlying construct (*dimensionality*). Categories were then collapsed and models for ordered categorical (ordinal) dependent variables were used if deemed appropriate. A multi-level modelling framework was used which allowed the intercept for each model to randomly vary across doctors. Thus the data were considered hierarchical in nature, with ARCPs conceptualised as being nested within individual doctors. This approach also conveniently controlled for the fact that doctors had each taken varying numbers of ARCPs during the study period.

Age and UK-based experience duration varied little within each cluster of observations by doctor. Consequently these values were averaged to obtain an estimate of age and UK experience for each individual during the study period that could be used as a variable at the doctor-level for the purposes of multi-level model building. Model building proceeded in a forward stepwise manner with variables contributing most to improved fit being entered first. Exploratory analyses indicated complex curvilinear relationships between ARCP outcome and both age and years of UK-based experience. However, we selected not to enter polynomial (or fractional polynomial) terms into the model for two reasons; firstly this would make direct interpretation of the coefficients extremely challenging, and; secondly the addition of polynomial terms to the model made relatively little difference to the main effect we were interested in (i.e. the effect of international medical graduates status on ARCP outcomes). All combinations of potential interactions between predictor variables were tested for statistical significance at the $p<0.05$ level.

'Simulation' of a change on pass mark was performed by ranking the PLAB candidates with available data (i.e. candidates who passed the exam from July 2000 to the end of 2010, not just those with ARCP outcomes recorded) and placing them in roughly equal twelfths. The duodecile boundaries were based on their exam performances, relative to the pass mark at the sitting in which they passed. These groupings were then used in the analyses involving PLAB graduates with ARCP outcomes recorded. These groups of international medical graduates, categorised by the duodecile boundaries, were then compared to UK graduates by generating a dummy variable coded as 1=PLAB international medical graduate in that particular twelfth (or lower) and 0=UK graduate.

All data management and analysis, including multiple imputation, was performed in STATA 13 MP. The only exception to this was that analysis of multiply imputed data was performed in Mplus 7.11 (STATA does not currently support multi-level ordinal logistic regression analysis of multiply imputed datasets).

In order to explore the extent to which the ARCP outcomes could be treated as ordered (i.e. ordinal in nature) a univariable multi-level multinomial logistic regression was performed with international medical graduate status as the sole predictor variable. The model was implemented within the generalised structural equation modelling (GSEM) framework of STATA. This model allowed for the model intercept to randomly vary

across individual doctors. The model was constrained so that individual-level effects were assumed to be equal across all pairs of outcome categories and fixed at one. The base category was repeatedly swapped so that the effect on international medical graduate status between all ten combinations of outcome categories could be evaluated. The Relative Risk Ratios (RRRs) estimated from this model indicated that, allowing for the tendency of ARCP outcomes to correlate within individual doctors, the categories relating to '*extended training required*' and '*released from programme*' could not be discriminated between reliably (i.e. the coefficient representing the effect of international medical graduate status on the risk of being in the latter category rather than the former was not statistically significant at the $p<0.05$ level). This is understandable given that individual doctors in training are likely to have their training extended (as remedial action) before they are released from training due to performance issues. Consequently these two latter ARCP outcome categories were collapsed into recoded outcomes as follows:

1- *satisfactory progression*
2- *insufficient evidence presented*
3- *targeted training required* (but training time not extended)
4- *extended training time required/left programme*

The multi-level multinomial regression was performed again using these modified outcomes. The results are depicted in supplementary table 1 and reveal that the RRR estimates for contiguous outcome categories are very similar, and are certainly not significantly different at the $p<0.05$ level. For example, the RRR for being in ARCP outcome categories that are spaced two thresholds apart (i.e. outcomes '1' vs '3' and '2' vs '4') are approximately two in both cases (2.07 and 2.13 respectively) with relatively narrow 95% confidence intervals for the estimates. This suggests that the modified ARCP outcome categories listed above could be plausibly treated as ordinal in nature. Treating the ARCP outcomes as ordinal in this way allows the additional information present in the ordering of the categories to be incorporated into the modelling process, providing additional study power. Moreover, the interpretation of ordinal logistic regression models is more straightforward than in the case of multinomial regression, where outcomes are treated as nominal (unordered).

Consequently, a series of univariable multi-level ordinal logistic regression models were tested. Again, the intercept was allowed to vary randomly across individual doctors. The results are depicted in the main paper. The multi-level models in STATA were estimated using the Gauss-Hermite approach to integration. Twelve integration points were used (increasing the number of points did not substantially alter the estimates). Missing observations were dealt with via listwise deletion. Note that in the multi-level ordinal regression analyses the three cut-points, representing the thresholds between the four categories of ARCP outcome (analogous to the intercept in a binary logistic regression) are not shown.

Missing data

Whether data on ethnicity were missing or not was unrelated to whether the doctor was a UK or an international medical graduate (p=0.15).

Those international medical PLAB graduates who were recorded as having undergone at least one ARCP in the study period (n=11022) were no less likely to be male or non-white than those without a performance related outcome reported (n=15704) (p>0.05 for inter-group differences in both cases). However, those with ARCP outcomes recorded had higher scores at first sitting, relative to the pass mark, at both part 1 (β=3.47; 95% confidence interval 2.99 to 3.94) and part 2 (β=1.16; 95% confidence interval 1.05 to 1.27) of the PLAB, compared to those without ARCP outcomes. Moreover, those international medical graduates with ARCPs recorded had significantly fewer resits at both part 1 and part 2 of the PLAB exam. There was also a slight trend, of borderline statistical significance, for IELTS scores to be higher in those international medical graduates with ARCP outcomes (p=0.06).

ARCP outcome comparison pair	(1) Satisfactory progress vs (2) Incomplete evidence	(2) Incomplete evidence vs (3) Targeted training	(3) Targeted training vs (4) Extended training/left programme	(1) Satisfactory progress vs (3) Targeted training	(2) Incomplete evidence vs (4) Extended training/left programme	(1) Satisfactory progress vs (4) Extended training/left programme
Relative Risk Ratio	1.59 (1.50 to 1.68)	1.28 (1.19 to 1.38)	1.57 (1.45 to 1.71)	2.07 (1.94 to 2.21)	2.13 (1.99 to 2.28)	3.45 (3.26 to 3.65)
Intercept	0.06	0.71	2.40	0.032	0.81	0.04

Supplementary Table 1. Results of a multi-level multinomial logistic regression analysis with ARCP outcome treated as a nominal outcome variable and region of primary medical qualification (PLAB international medical graduate vs UK graduate) as a predictor. The p is <0.001 for the coefficients in all cases. The modelling was performed using data from 125208 ARCPs conducted with 53436 doctors.

Data imputation

The second multi-level multivariable ordinal regression model built and tested included ethnicity. Analysis suggested that ethnic status (white vs non-white) was related to both the outcome and the predictor variables (sex, age, UK experience and place of graduation). Consequently missing ethnic status was multiply imputed using a logistic regression whereby ethnicity was treated as a binary dependent variable with lowest ARCP outcome for that doctor, sex, age, years of UK practice, international medical graduate status and whether they had ever had 'out of programme' experience. Thus missing ethnic status was imputed in all but two cases (where age was also missing for two observations in the final dataset to be analysed). The imputation was implemented in STATA to create five imputed data sets, using the logistic regression to specify the posterior distribution from which the imputed values were randomly drawn from.[9] The datasets, which included the imputed ethnicity values, were then analysed and the results combined using the software package Mplus 7.11. A comparison of results from imputed and non-imputed datasets was conducted as a sensitivity analysis and indicated that the results were very similar; the mean difference in the values of the regression coefficients recovered was 9% (sd 15%; median 3%). This would tend to support (though not conclusively) the assumption that values for ethnic status were missing at random (MAR- i.e. the missing values are related to the values of the observed variables). However, there is no way of robustly testing the MAR assumption and this represents the primary limitation of multiple imputation in such circumstances.[10]

Supplementary results

Here we include supplementary tables from the univariable analyses. We also present additional results from the multivariable models which now **include ARCP outcomes associated with royal college exam failure**. This allows comparison with the findings from the multivariable models developed and portrayed in the main paper. Such comparison provides additional information regarding the degree to which differential pass rates at royal college exams may influence any observed disparities in performance between UK and international medical graduates. As in the main paper, p values relating to coefficients are less than 0.001, unless otherwise stated.

PLAB candidate vs UK graduate status as a predictor of ARCP outcome

The results from the univariable multi-level ordinal logistic regression model are depicted in supplementary table 2. As can be seen, international medical graduates who registered via the PLAB system had more than twice the odds of a UK graduate for receiving a less satisfactory ARCP outcome.

Predictor	Odds ratio	95% CIs
IMG status	2.46	2.35 to 2.58
Male sex	1.42	1.37 to 1.48
Age*	1.04	1.03 to 1.04
UK experience	0.91	0.90 to 0.92
Non-white ethnicity**	1.98	1.88 to 2.08

Supplementary Table 2. Results from a series of multi-level univariable ordinal regressions with ARCP outcome category as the dependent variable and PLAB International Medical Graduate (IMG) status, sex, age, UK-based experience and ethnicity as the predictor variables. Higher categories are associated with a less satisfactory ARCP outcome. All odds ratios are significant at the $p<0.001$ level. Note the numbers of doctors included in the analyses varies by predictor variable due to missing observations with n=53436 unless indicated otherwise.

Note: *n=52429, **n=41352

Analysing the sample as a whole, the raw odds ratios for male sex was 1.42, indicating the odds of being in a less satisfactory category of ARCP outcome was almost 50% greater for men compared to women. Overall, increasing age was also generally associated with less satisfactory outcomes at ARCP. However, when analysed separately, mean age at ARCP was associated with lower odds of a less satisfactory ARCP outcome in UK graduates (odds ratio 0.97; 95% confidence interval 0.97 to 0.98). UK-based experience (as estimated by mean time since registration at ARCP) was predictive of better ARCP performance for both groups, with the odds of being in a more satisfactory category, overall, increasing by approximately 10% for every year of practice. Whilst non-white ethnicity is, overall, associated with being in a less satisfactory category of ARCP outcome this effect was not observed when PLAB graduates were analysed separately (odds ratio 0.90; 95% confidence interval; 0.74 to 1.09, p=0.3). In contrast the odds ratio for being in a lower ARCP category if ethnicity was reported as non-white was 1.48 for a UK graduate (95% confidence interval 1.40 to 1.57). In

supplementary table 3 it can be seen that increasing IELTS and PLAB scores reduce the odds of receiving a less satisfactory ARCP outcome.

Predictor	Odds ratio	95% CIs
Overall IELTS score (n=27043)	0.59	0.54 to 0.65
PLAB part 1 score, relative to pass mark (n=28189)	0.98	0.98 to 0.98
PLAB part 2 score, relative to pass mark (n=28189)	0.92	0.91 to 0.93

Supplementary Table 3. Results from a series of multi-level univariable ordinal regression models. ARCP outcome category is the dependent variable and IELTS and PLAB scores are the predictors. Lower categories are associated with more satisfactory ARCP outcomes (i.e. odds ratio of <1.00 indicate that the predictor is associated with better ARCP ratings). All odds ratios are significant at the $p<0.001$ level.

Regarding the multivariable models; interpreting the results of regression analyses that include a large number of interaction variables (relative to the main effects) can be extremely challenging, especially in the present case where the coefficients relating to the main effect being examined are substantially altered. For this reason the results, both with and without interaction terms are presented here.

The results portrayed in supplementary tables 4 and 5 can be interpreted as follows: from supplementary table 4 it can be seen that international medical graduates status is independently predictive of a less satisfactory ARCP outcome, even after controlling for the effects of age, sex, experience and ethnicity. Four significant interactions were identified and entered into the model (supplementary table 5). Firstly, there is an interaction between age and years of UK experience. The odds ratio for this term is greater than 1.00 suggesting that, overall, the beneficial effects of experience on the probability of obtaining a more satisfactory outcome at ARCP are attenuated in older doctors. There are also significant interactions between international medical graduate status and the other three predictor variables. These suggest that for international medical graduates UK-based experience has an additional positive effect on ARCP outcome but that being male and/or being older are disproportionately worse in this respect, than for UK graduates. Thus it would be erroneous to simply conclude that international medical graduate status is not a significant and independent predictor of ARCP outcome; rather the interpretation is more subtle; the effects of being an

international medical graduate, compared to a UK graduate, on ARCP may be mediated by the complex and differential effects that sex, age and experience have on the two groups in relation to postgraduate performance.

Predictor	Odds ratio	95% CIs
Male sex	1.32	1.27 to 1.38
Age	1.07	1.06 to 1.07
UK experience	0.87	0.86 to 0.87
Non-white ethnicity	1.36	1.29 to 1.44
IMG status	1.37	1.28 to 1.46

Supplementary Table 4. Results of a multi-level ordinal logistic regression analysis with ARCP category as the outcome. Lower categories are associated with more satisfactory ARCP outcomes (i.e. odds ratios of <1.00 indicate that the predictor is associated with better ARCP ratings). The predictor variables are sex, age, UK-based experience, ethnicity and place of primary medical qualification (UK vs PLAB international medical graduate [IMG]) from the sample of doctors (N=53429 relating to 125193 ARCPs). Significant interactions terms were excluded. All odds ratios are significant at the p<0.001 level.

Predictor	Odds ratio	95% CIs
Male sex	1.33	1.27 to 1.40
Age	1.01*	1.00 to 1.02
UK experience	0.68	0.65 to 0.71
Age/experience interaction	1.01	1.01 to 1.01
IMG status/experience interaction	0.94	0.92 to 0.96
IMG status/age interaction	1.03	1.02 to 1.04
IMG status/male interaction	1.13**	1.02 to 1.24
IMG status	0.88†	0.59 to 1.32

Supplementary Table 5. Results of a multi-level ordinal logistic regression analysis with ARCP category as the outcome. Lower categories are associated with more satisfactory ARCP outcomes (i.e. odds ratios of <1.00 indicate that the predictor is associated with better ARCP ratings). The predictor variables are sex, age and UK-based experience and place of primary medical qualification (UK vs PLAB international medical graduate [IMG]) from the sample of doctors (N=53429 relating to 125193 ARCPs). Statistically significant interactions terms were included. All odds ratios are significant at the p<0.001 level except where indicated.

Note: *p=0.19, **p=0.01, †p=0.54

PLAB performance as a predictor of ARCP outcome

Although the full results are not shown here, when a model for predicting ARCP outcome from PLAB part 1 performance is estimated without interaction terms the scores (relative to pass at first attempt) are independently predictive of a more satisfactory ARCP performance (odds ratio 0.98; 95% confidence interval 0.98 to 0.98). From supplementary table 6 we see that when interaction terms are included that, paradoxically, increasing PLAB part 1 score is predictive of *less satisfactory* performance at ARCP. The interaction terms have completely changed the model and the relationship of the main effect of interest with the outcome. The interaction terms can be interpreted as follows: the benefits of higher PLAB part 1 scores seem to be enhanced by male sex and more UK-based experience. It would be erroneous to conclude that PLAB part 1 scores were independently predictive of *less* satisfactory ARCP outcome; rather there are complex interactions with sex and UK-based experience that are likely to mediate their impact on later ARCP performance.

Predictor	**Odds ratio**	**95% CIs**
Male sex	1.50	1.38 to 1.64
Age	1.06	1.05 to 1.07
UK experience	0.88	0.86 to 0.90
Overall IELTS score	0.75	0.69 to 0.82
PLAB part 1/experience interaction	0.99	0.99 to 0.99
PLAB part 1/male sex interaction	0.99*	0.99 to 0.99
PLAB part 1 score	1.01**	1.00 to 1.01

Supplementary Table 6. Results of a multi-level ordinal logistic regression analysis with ARCP category as the outcome. Lower categories are associated with more satisfactory ARCP outcomes (i.e. odds ratios of <1.00 indicate that the predictor is associated with better ARCP ratings). The predictor variables are sex, age, UK-based experience, IELTS performance and PLAB part 1 scores (relative to pass mark) in 27043 ARCPs taken by 10945 international medical graduates. All odds ratios are significant at the p<0.001 level except where indicated.

Note: *p=0.01, **p=0.03

Predictor	Odds ratio	95% CIs
Male sex	1.33	1.23 to 1.45
Age	1.07	1.06 to 1.08
UK experience	0.88	0.85 to 0.92
Overall IELTS score	0.77	0.71 to 0.84
PLAB part 2/experience interaction	0.99*	0.99 to 0.99
PLAB part 2 score	0.98**	0.95 to 1.01

Supplementary Table 7. Results of a multi-level ordinal logistic regression analysis with ARCP category as the outcome. Lower categories are associated with more satisfactory ARCP outcomes (i.e. odds ratios of <1.00 indicate that the predictor is associated with better ARCP ratings). The predictor variables are sex, age, UK-based experience, IELTS performance and PLAB part 2 scores (relative to pass mark) in 27043 ARCPs taken by 10945 international medical graduates. All odds ratios are significant at the p<0.001 level except where indicated.

Note: *p=0.02, **p=0.12

Likewise, without the significant interaction term included, the multivariable model developed for PLAB part 2 scores demonstrate these scores are significantly and independently predictive of better outcomes at ARCP (odds ratios 0.94; 95% confidence interval 0.93 to 0.95- full results not shown). However, once the interaction between PLAB part 2 scores and UK-based experience is accounted for this effect is no longer significant at the p<0.05 level (see supplementary table 7). This may suggest that achievement at the PLAB part 2 test is generally only translated into superior ARCP performance (including those that take into account outcome at royal college exams) by obtaining longer periods of UK-based clinical experience.

The interpretation of the results relating to PLAB resits shown in supplementary table 8 is identical to that provided in the main paper: compared to passing the PLAB parts after multiple attempts, passing at first sittings are independently and significantly predictive of more satisfactory outcome at ARCP. For PLAB part 1 this effect appears to level off at three attempts (i.e. two resits) whilst for part 2 these effects level off at two attempts (i.e. one resit). These 'levelling off' effects are apparent when all possible combinations of pairs of resit categories are compared. This is achieved by repeatedly swapping the resit category serving as the comparator (base) category (full results not shown).

Part 1 of the PLAB test			**Part 2 of the PLAB test**		
Predictor	Odds ratio	95% CIs	Predictor	Odds ratio	95% CIs
Male sex	1.44	1.32 to 1.56	Male sex	1.40	1.29 to 1.52
Age	1.06	1.05 to 1.07	Age	1.07	1.06 to 1.08
UK experience	0.86	0.84 to 0.88	UK experience	0.84	0.82 to 0.86
Overall IELTS score	0.70	0.64 to 0.76	Overall IELTS score	0.69	0.63 to 0.75
Two attempts at PLAB part 1	1.55	1.40 to 1.71	Two attempts at PLAB part 2	1.46	1.32 to 1.62
Three attempts at PLAB part 1	2.07	1.76 to 2.44	Three or more attempts at PLAB part 2	1.59	1.24 to 2.03
Four or more attempts at PLAB part 1	1.95	1.60 to 2.37	-	-	-

Supplementary Table 8. Results of a multi-level multivariable ordinal logistic regression model estimating the effect of the number of attempts at PLAB parts 1 and 2 on 27043 ARCP outcomes taken by 10945 PLAB international medical graduates. Passing the exam at first sitting was used as the base (comparator) category. All odds ratios are significant at the $p<0.001$ level.

References

1. General Medical Council. ARCP RITA data collection 2013- Briefing Note 2. London: General Medical Council, 2013. http://www.gmc-uk.org/ARCP_2013_Briefing_Note_2___amended.pdf_53743694.pdf

2. National Deanery Data Group. Deanery Data Manual. The National Deanery Data Group (NDDG) of the Conference of Postgraduate Medical Deans of the United Kingdom, 2011. http://www.wessexdeanery.nhs.uk/pdf/NDDG%2000Draft%20Report%20V1%2003May2011.pdf

3. General Medical Council. List of registered medical practitioners. General Medical Council, 2014. www.gmc-uk.org/doctors/register/LRMP.asp

4. McManus IC, Katarzyna L. Resitting a high-stakes postgraduate medical examination on multiple occasions: nonlinear multilevel modelling of performance in the MRCP(UK) examinations. *BMC Medicine* 2012;10:60.

5. McManus IC, Ludka K. Resitting a high-stakes postgraduate medical examination on multiple occasions: nonlinear multilevel modelling of performance in the MRCP(UK) examinations. *BMC medicine* 2012;10:60.

6. Office for National Statistics. Ethnic Group. London: ONS, 2013. http://www.ons.gov.uk/ons/guide-method/measuring-equality/equality/ethnic-nat-identity-religion/ethnic-group/index.html

7. Information Comissioner's Office. Anonymisation: managing data protection risk code of practice. Information Comissioner's Office, 2012. http://ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation

8. General Medical Council. Approved specialty and subspecialty training curricula by Royal College. General Medical Council, 2014. www.gmc-uk.org/education/approved_curricula_systems.asp

9. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research* 2007;16:219-42.

10. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338:b2393.