UNIVERSITY OF
LINCOLN

How is performance at selection to general practice related to performance at the endpoint of GP training?

Report to Health Education England

Dr Vanessa Botan, Dr Nicki Williams, Prof Graham R. Law, Prof Aloysius Niroshan Siriwardena

April 7, 2022

Contents

List of Figures

List of Tables

Study Advisory Group

Dr Vanessa Botan[1]

Dr Nicki Williams[2]

Professor Graham Richard Law[1]

Professor Aloysius Niroshan Siriwardena [13]

Dr Chis Elfes[3]

Dr Kim Emerson[3]

Dr Susan Bodgener[3]

Dr Fiona Kameen[3]

Dr Meiling Denney[3]

Dr Ryan Prince[2]

Professor Lindsey Pope[3]

Professor Rich Withnall[3]

1. Community and Health Research Unit, University of Lincoln
2. General Practice National Recruitment Office (GPNRO), Health Education England
3. MRCGP Examinations, Royal College of General Practitioners

Correspondence to:

Professor Aloysius Niroshan Siriwardena
Community and Health Research Unit,
University of Lincoln, College of Social Science
Lincoln Medical School (LMS3006)
Brayford Pool, Lincoln LN6 7TS
Email: nsiriwardena@lincoln.ac.uk
Tel: +44 (0)1522 886949

Acknowledgements

We would like to thank members of the Study Advisory Group for their advice and support. We are grateful to Health Education England for funding the study and providing selection data to Stuart Copus, Tom Anstey and the team at MRCGP examinations for linking and providing anonymised data for analysis.

Summary

Background
The selection process for entry to speciality training for general practice (GP) in the UK was changed in 2016. Doctors scoring above an agreed threshold in the computer-marked Multi-Specialty Recruitment Assessment (MSRA) were deemed appointable on that score alone and were offered a direct pathway (DP) to training, exempting them from further assessment at the final Selection Centre (SC). The SC was subsequently suspended in response to the COVID-19 pandemic and has yet to be reinstated. We aimed to evaluate the relationship between performance at selection and outcomes of GP training at licensing, to reassess the threshold score in MSRA used to bypass the SC, and to estimate the incremental predictive value of the SC after MSRA.

Methods
We used a longitudinal design linking selection, licensing and demographic data from doctors applying to enter GP specialty training in 2016. MSRA scores were divided into 12 score bands and SC scores into seven score bands to better identify MSRA or SC scores that corresponded to differing GP performance on licensing assessments. Multivariable logistic regression models were used to establish the predictive validity of the MSRA scores and score bands for passing or failing the Membership of the Royal College of General Practitioners (MRCGP) licensing assessments including the Applied Knowledge Test (AKT), Clinical Skills Assessment (CSA) or Recorded Consultation Assessment (RCA), Workplace Based Assessment - Annual Review of Competence Progression (WPBA-ARCP), and performance overall. The model adjusted for sex, ethnicity, country of qualification, and declared disability. Receiver Operating Characteristic (ROC) curves of MSRA scores against performance outcomes were constructed to determine the optimal MSRA threshold scores for achieving licensing.

Results
We included 3338 doctors who entered specialty training for general practice in 2016 of different sex (female 63.81% vs male 36.19%), ethnicity (White British 53.95%, minority ethnic 43.04% or mixed 3.01%), country of qualification (UK 76.76%, non-UK 23.24%), and declared disability (disability declared 11.98%, no disability declared 88.02%).
MSRA scores or score bands were highly predictive for all assessments of GP training outcome (AKT, CSA, RCA, and WPBA-ARCP). Lower SC score bands were predictive of lower pass rates on summative assessments and /or ARCP outcomes 2, 3, or 4. Adding SC scores did not change the predictive validity of the MSRA, and therefore the SC did not add further information to MSRA scores. An MSRA threshold of 500 (or, more precisely, 497) was optimal for correctly identifying pass/fail rates on the AKT, RCA, and CSA within the study period, and only standard outcomes on WPBA-ARCP. Thirty-five percent of candidates in the lowest two MSRA Bands (i.e., scores below 420) had at least one developmental outcome (2, 3) or outcome 4. Ethnicity did not reduce the chance of passing GP licensing tests once sex, place of primary medical qualification, declared disability and MSRA scores were taken into account.

Conclusion
MSRA scores predict licensing outcomes for AKT, CSA, RCA, and WPBA-ARCP within five years of starting training. The optimal MSRA threshold score for predicting an uncomplicated training pathway to licensing was around 500 in this large cohort. The SC added little to the predictive validity of the MSRA. Doctors scoring below this threshold may need additional support during training to maximise their chances of achieving licensing.

Introduction

Any doctor providing proof of eligibility for UK specialty training is entitled to take part in the selection process. This process is important to ensure that those chosen are likely to successfully complete postgraduate medical education and gain a licence to practice in their chosen speciality. Considerations include the financial cost of training, the challenges of providing extra support when needed and the impact on trainees who fail to complete in a specialty where there is no long term 'middle grade'. Failure to complete training is wasteful of training places and resources and detrimental to trainees and schemes (Plint & Patterson 2010, Davison et al. 2016).

The current system of selection to GP training in the United Kingdom (UK) has three phases:

- Stage 1 Administrative: Candidates provide proof of eligibility for UK specialty training, including proof of Foundation Level competence.
- Stage 2 Multi-Specialty Recruitment Assessment (MSRA): a computer-based Multiple Choice Question examination, developed and delivered by the National Recruitment Office (NRO) as a shortlisting tool for many medical specialties, including General Practice (GP). The MSRA contains both Clinical Problem-Solving items and Situational Judgement Tests.
- Stage 3 Selection Centre (SC): A GP-specific face-to-face assessment using Objective Structured Clinical Examination (OSCE) style simulations and a written test.

All three stages are centrally organised. Stage 2 is delivered through Pearson Vue centres at many locations. Stage 3, SC, is delivered by individual deaneries: each OSCE station is graded on multiple domains by a single assessor, usually a GP.

GP specialty training after this selection process is a planned three-year whole time equivalent programme of primary and secondary care placements. At exit, competence for independent practice is assessed in a three-part assessment, the MRCGP. The three strands are:

- Applied Knowledge Test, AKT (clinical problem solving).
- Clinical Skills Assessment CSA: in 2020, this was replaced by the Recorded Consultation Assessment RCA due to Covid measures.
- Work Place Based Assessment, WPBA, as judged by the Annual Reviews of Competence Progression, ARCP.

In 2020/21 the average time taken to complete GP training (including maternity leave, less than full time training and supportive extensions) was 3.8 years. Only 1.2% of those admitted to training left the programme without achieving a licence to practice. The current study followed a large cohort for 5 years.

The MSRA has been positively evaluated as a method of shortlisting (Patterson et al. 2009). Predictive validity was assessed using supervisor ratings of job performance after one year in addition to results of the AKT and CSA (Patterson et al. 2013). The SC was validated through trainer ratings of trainee job performance after three months (Patterson et al. 2005). Both MSRA and SC were considered to be fair tests for selection (Patterson et al. 2011).

There is limited evidence on outcomes in other assessments such as the Workplace Based Assessment Annual Review of Competence Progression (WPBA-ARCP), overall performance,

or the Recorded Consultation Assessment (RCA), an assessment which replaced the CSA from mid- 2020 in response to the Covid pandemic (Botan et al. 2021).

Prior to 2016, all candidates who achieved an agreed 'score' at MSRA progressed to the SC as the final step in selection. In 2016, a change was introduced, allowing candidates who achieved a threshold score of 575 or above on the MSRA (10-13% of top scorers) access to the Direct Pathway (DP). They were deemed appointable on MSRA score alone and exempt from attending the Stage 3 Selection Centre (Ooi & Ooi 2021). The MSRA threshold score for accessing the DP was based on statistical modelling which showed that 100% of candidates undertaking the MSRA in 2013, 2014 and 2015 who achieved a score of 575 or more in Stage 2 were subsequently deemed appointable at Stage 3.

By the time of data collection (2021), trainees from the 2016 selection cohort would be expected to have completed training in 2019, 2020, and 2021. This provided an opportunity for this study to evaluate the process of selection to identify outcomes of DP (Stage 1 and 2 alone) versus Stages 1-3 and to reassess the 575 threshold score at MSRA for determining eligibility for Direct Pathway.

Recent studies have suggested that factors such as specific learning difficulties, ethnicity (and place of primary medical qualification), and gender should be taken into consideration when developing predictive models (Asghar et al. 2019, 2018). Hence this study also considered those factors, particularly looking at any changes in differential attainment between entrance and endpoint assessments.

The research question was 'How is performance at selection to general practice related to performance at the endpoint of GP training?'

Our objectives were:

1. To determine the predictive validity of the MSRA for outcomes of GP training including performance at the MRCGP AKT, CSA, RCA and WPBA-ARCP.
2. To determine the predictive validity of the MSRA and the Selection Centre for those who were required to undertake both for outcomes of GP training.
3. To reassess the threshold score in MSRA for determining eligibility for Direct Pathway.

Methods

5.1 Design

We used a longitudinal design, which tracked doctors from selection to the end of GP training by linking selection, licensing and demographic data from UK doctors entering GP specialty training in 2016.

5.2 Ethical approval

Independent ethical review was sought, and approval gained from the University of Lincoln Human Ethics Committee (reference 2020_3645).

5.3 Data collection and processing

MSRA (Multi-Specialty Recruitment Assessment) and Selection Centre (SC) scores (available only for those scoring less than 575 on the MSRA) for doctors undertaking selection tests in 2016 were linked with their AKT (Applied Knowledge Test), CSA (Clinical Skills Assessment), RCA (Recorded Consultation Assessment), WPBA - ARCP (Workplace Based Assessment - Annual Review of Competence Progression) outcomes up to 2021. Demographic data were pseudonymised and securely transferred to the research team under data sharing agreements for analysis.

The datasets comprised individual candidate data. Candidates were assigned a unique (non-personally identifiable) number which allowed linkage between the various assessments, demographic data including gender, ethnicity, country of graduation, and declared disability with results for each assessment including overall scores, scores for assessments' sub-domains and outcomes of pass (1) or fail (0). The declared disability data included mainly declared specific learning difficulties (SpLDs), but there was also a smaller proportion (about 10%) which included other disabilities such as hearing impairments or other physical disabilities. Ethnicity was divided into three categories: white ethnicity background, ethnic minority background, and mixed minority background. Binary variables included: country of graduation (UK or non-UK graduates), gender (male or female), and declared disability (not having a declared disability or having a declared disability).

Evidence for WPBA is collected by the trainee throughout the year and a summative judgement of progress is made annually at the ARCP. The following Outcomes are available to the panel:

1 Achieving progress and competences at the expected rate.

2 Development of specific competences required - additional training time not required.

3 Inadequate progress by the trainee - additional training time required.

4 Released from training programme with or without specified competences.

5 Incomplete evidence presented - additional training time may be required.

6 Has gained all the required competences for the completion of training.

7 N1/2 - Sick or maternity/paternity leave.

8 Out of Programme Pause (OOP).

9 N21/22 - Resignation.

10 Satisfactory progress.

Outcomes 2 and 3 were considered developmental outcomes and Outcome 4 occured when the trainee was released from training. A successful completion of the ARCP was considered if only standard outcomes had been achieved after removing outcomes 2, 3, and 4 determined by a failure on one of the licensing tests (AKT, CSA, or RCA).

MSRA scores were divided into 12 score bands and SC scores were divided into 7 score bands to better identify specific MSRA or SC scores that may have corresponded to differing GP training outcomes. The score bands were chosen based on the distribution of the data and in order to achieve bands that were narrow enough to precisely identify candidates with differing performance.

5.4 Statistical analysis

Multivariable linear regression models were used to establish the predictive validity of the MSRA scores for each assessment's overall score. The variables sex, ethnicity, country of primary medical qualification (medical graduation), and declared disability were used as predictors together with the MSRA scores. Assumptions of no multicollinearity and no outliers were checked.

Multivariable logistic regression models were used to establish the predictive validity of the MSRA scores and score bands for each assessment's outcome (i.e., pass or fail). The variables sex, ethnicity, country of graduation, and declared disability were used as predictors together with the MSRA scores or score bands.

For the multivariable linear regression models, the unstandardised Beta coefficients (B) representing the slope of the regression line, the standard error (S.E.) of B representing the spread of the observations from the fitted regression line, the r-squared (R^2) representing the proportion of the variance in the outcome explained by the predictor(s), and 95% confidence intervals (CI) representing the range of values that contains the true mean of the population with 95% confidence were reported. For the multivariable logistic regression models, the odds ratio (OR) representing the odds that the outcome will occur given a predictor, compared to the odds of the outcome occurring in the absence of that predictor (i.e., at baseline) and pseudo r-squared (pseudo R^2) representing the certainty with which the model can predict the dichotomous outcome (y=0 or y=1) were reported.

To determine changes in the predictive validity of the MSRA for those who had gone through the Selection Centre, further multivariable logistic models were run with the SC scores added to the statistical models.

To reassess the threshold score in MSRA for determining eligibility for Direct Pathway to GP training, receiver operating characteristic (ROC) curve analysis was used. MSRA scores were plotted against the outcome (i.e., pass or fail) for AKT, CSA, RCA, and ARCP (where a failure represented the presence of at least one developmental outcome or outcome 4 after removing those corresponding to exam failure) and against the overall outcome of all these assessments (i.e., pass – if all the assessments had been passed; fail – if at least one assessment had been failed). To interpret the ROC curve results, both the maximum specificity approach and the equivalence between specificity and sensitivity approach were considered. Descriptive statistics indicating number and percentages of candidates were used to describe the demographic data for each assessment.

Results

In total, 3,429 candidates took the MSRA, 2,883 the AKT, 2,313 the CSA, and 545 the RCA.

6.1 The predictive validity of MSRA scores for outcomes of GP training

Table 1: *MSRA scores as predictors of AKT, CSA, and RCA scores and number of ARCP outcomes 2, 3, and 4.*

Outcomes	Predictor: MSRA Score			
	Unstandardised beta (B)	**Standard error of B (S.E.)**	**95% Confidence Interval (CI) of B**	**p-value**
AKT Score	0.14	0.01	0.13, 0.16	<0.001
CSA Score	0.05	0.00	0.04, 0.06	<0.001
RCA Score	0.07	0.01	0.04, 0.10	<0.001
ARCP Outcomes 2, 3, and 4	-0.01	0.00	-0.01, -0.00	<0.001

N.B. These results are the summary of multivariable regression models which accounted for the demographic covariates sex, ethnicity, country of qualification, and disability.

MSRA scores were significant predictors for all GP training outcome scores and for the number of in-training ARCP outcomes 2, 3, and 4. As indicated by the unstandardised beta coefficient (B), an increase of 1 mark on the MSRA corresponded to an increase of 0.14 on the average mark on the AKT, of 0.05 on the average mark of CSA, of 0.07 on the average mark of RCA. After excluding outcomes 2, 3, and 4 due to failures in AKT, CSA, or RCA, an increase of 1 mark on the MSRA was also associated with an average increase of 0.01 on the total number of standard ARCP outcomes. These were all statistically significant ($p<0.001$) as seen in Table 1. Moreover, the standard error of B indicated a low spread of the observations from the fitted regression line. There was also a significant positive correlation between MSRA scores and AKT ($r=0.61$, $p<0.001$), CSA ($r=0.54$, $p<0.001$), and RCA ($r=0.57$, $p<0.001$) and a negative correlation between MSRA scores and the number of ARCP outcomes 2, 3, and 4 ($r=-0.35$, $p<0.001$).

Table 2: *MSRA scores as predictors of AKT, CSA, and RCA pass rates and presence of at least one of the ARCP outcomes 2, 3, or 4*

Outcomes	Predictor: MSRA Score			
	Odds Ratio (OR)	**Standard error (S.E.)**	**95% Confidence Interval (CI) of OR**	**p-value**
AKT Pass	1.03	0.00	1.02, 1.04	<0.001
CSA Pass	1.01	0.00	1.00, 1.01	<0.001
RCA Pass	1.01	0.20	0.27, 1.16	=0.087
Presence of ARCP Outcomes 2, 3, or 4	1.01	0.00	1.01, 1.01	<0.001

N.B. These results are the summary of multivariable regression models which accounted for the demographic covariates sex, ethnicity, country of qualification, and disability.

MSRA scores had predictive validity for the pass rates on the AKT and CSA, but not the RCA. As shown by the odds ratio (OR), an increase of one mark on the MSRA increased the

likelihood of passing the AKT by 0.03 and the likelihood of passing the CSA by 0.01. It also increased the likelihood of passing the RCA by 0.01, but this was non-significant. MSRA scores were a significant predictor for ARCP outcomes, increasing the likelihood of having only standard (satisfactory) outcomes on the ARCP by 0.01. These results can be seen in Table 2.

Overall, MSRA scores had predictive validity for all GP training outcomes scores, for AKT and CSA pass rates, and ARCP in-training performance.

MSRA scores were divided into 12 score bands and fail rates for each of the outcomes of GP training were calculated for each band as seen in Figure 1. MSRA score bands had predictive validity for AKT pass rates. Bands 2 and above significantly increased the likelihood of passing the AKT compared to Band 1, which corresponded to MSRA scores lower than 400, as seen in Table 3. Bands 10, 11, and 12 perfectly predicted passing the AKT thus they were not included in the model. MSRA Bands 2 and above increased the likelihood of passing the CSA, but only Bands 7 and 10 reached significance, whilst Bands 9 and 12 perfectly predicted pass rates as seen in Table 4. Bands 2 and above significantly increased the likelihood of passing the RCA, except for Bands 5,8, 10 and 11 as seen in Table 5. Odds ratios for Bands 8, 10, and 11 indicated an eight to nine times higher likelihood of passing the RCA compared to Band 1, but the low numbers of observations in these bands led to high variability and lack of statistical significance.

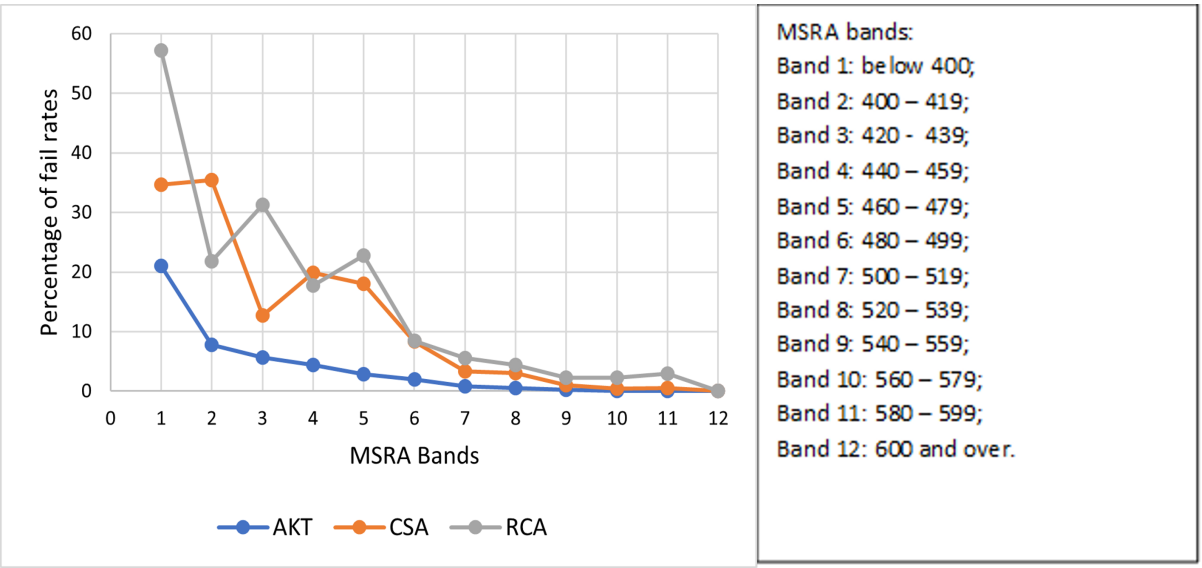

Figure 1: *Percentage of fail rates per MSRA score band for AKT, CSA, and RCA*

MSRA score bands were the main predictors of AKT pass rates. MSRA scores of 480 and above increased more than 15 times the chances of passing the AKT. When MSRA scores were taken into account, sex, country of qualification and declaration of disability were not independently associated with AKT performance as seen in Table 3. When MSRA scores and other demographic factors were taken into account, ethnic minority doctors were more likely than white British candidates to pass the AKT within five years of entering training (OR 2.05, 95% CI 1.03, 4.10, p=0.042.) This result indicates that, although the overall AKT pass rate is approximately 1% lower in ethnic minority doctors, adjustment for MSRA scores, which are affected by prior education and training, eliminates and reverses this difference.

MSRA score bands corresponding to scores higher than 400 increased the chances of passing the CSA, but this reached significance for scores ranging from 500-519 and 560-579, whilst score bands between 540-559 and over 600 perfectly predicted pass rates (i.e., all candidates

Table 3: *MSRA score bands and demographic characteristics as predictors of AKT pass rates*

Predictors	**AKT pass rates**			
	OR	**S.E.**	**95% CI of OR**	**p-value**
Sex (Female)	1			
Male	1.29	0.40	0.70, 2.36	=0.411
Ethnicity (White)	1			
Ethnic Minority	2.05	0.72	1.03, 4.10	=0.042
Mixed	1.20	1.30	0.14, 10.00	=0.865
Qualification Country (UK)	1			
Not-UK	1.17	0.46	0.54, 2.54	=0.686
Declared Disability (No)	1			
Yes	0.86	0.32	0.42, 1.77	=0.687
MSRA Bands (under 400)	1			
400-419	3.47	1.76	1.28, 9.36	=0.014
420-439	4.29	2.42	1.42, 12.94	=0.010
440-459	6.86	3.68	2.40, 19.11	<0.001
460-479	9.93	5.77	3.18, 31.03	<0.001
480-499	15.34	9.86	4.35, 54.08	<0.001
500-519	37.53	28.75	8.37, 168.40	<0.001
520-539	53.30	46.67	9.58, 296.52	<0.001
540-559	104.06	117.95	11.28, 959.69	<0.001
Cons	1.69	0.97	0.55, 5.21	<0.001
	$pseudoR^2 = 0.13, X^2(13) = 56.78, p<0.001$			

that obtained those scores passed the CSA) as seen in Table 4. In general, scores higher than 500 increased the likelihood of passing the CSA. When MSRA scores were taken into account, ethnicity was not associated with differential CSA performance. However, male sex (OR 0.58, 95% CI 0.39, 0.86, p=0.007), country of qualification outside the UK (OR 0.27, 95% CI 0.16, 0.45, p<0.001) and declaration of disability (OR 0.38, 95% CI 0.24, 0.61, p<0.001) were associated with a reduced chance of passing the CSA.

MSRA bands corresponding to scores higher than 480 increased the chances of passing the RCA five times or more. Scores between 480 and 519 significantly increased the likelihood of passing the RCA as seen in Table 5. When MSRA scores were taken into account, sex, ethnicity or declaration of disability were also not associated with RCA performance, but a non-UK country of qualification was associated with a reduced chance of passing the RCA (OR 0.30, 95% CI 0.11, 0.80, p=0.017).

MSRA score bands predicted both the number of ARCP in-training outcomes 2, 3, and 4 and the presence of at least one of these outcomes. MSRA score Bands 1 and 2 (i.e., scores lower than 420) corresponded to a higher average number of ARCP outcomes 2, 3, or 4 per person and to a higher percentage of individuals with at least one of these outcomes as seen in Figures 2 and 3. MSRA Band 6 and above (scores higher than 480) were significant predictors of a lower number of ARCP outcomes 2, 3, and 4 and MSRA Band 8 and above (scores higher than 520) significantly increased the likelihood of having only positive ARCP outcomes compared to Band 1 as seen in Tables 6 and 7. Importantly, being a male, not graduating in the UK, and having a disability were significant predictors of a poorer performance on the ARCP.

MSRA scores were good predictors of the number of ARCP outcomes 2, 3, and 4. MSRA

Table 4: *MSRA score bands and demographic characteristics as predictors of CSA pass rates*

Predictors	CSA pass rates			
	OR	**S.E.**	**95% CI of OR**	**p-value**
Sex (Female)	1			
Male	0.58	0.12	0.39, 0.86	=0.007
Ethnicity (White)	1			
Ethnic Minority	0.72	0.19	0.43, 1.20	=0.201
Qualification Country (UK)	1			
Not-UK	0.27	0.07	0.16, 0.45	<0.001
Declared Disability (No)	1			
Yes	0.38	0.09	0.24, 0.61	<0.001
MSRA Bands (under 400)	1			
400-419	0.92	0.39	0.40, 2.10	=0.848
420-439	2.58	1.29	0.97, 6.88	=0.059
440-459	1.04	0.43	0.47, 2.33	=0.915
460-479	0.99	0.41	0.44, 2.22	=0.972
480-499	1.48	0.67	0.61, 3.60	=0.389
500-519	4.00	2.28	1.31, 12.23	=0.015
520-539	2.47	1.34	0.85, 7.15	=0.097
560-579	11.58	12.67	1.36, 98.83	=0.025
580-599	6.86	7.53	0.80, 58.98	=0.080
Cons	17.76	8.66	6.83, 46.20	<0.001
	pseudo$R^2 = 0.21, X^2(13) = 178.87, p<0.001$			

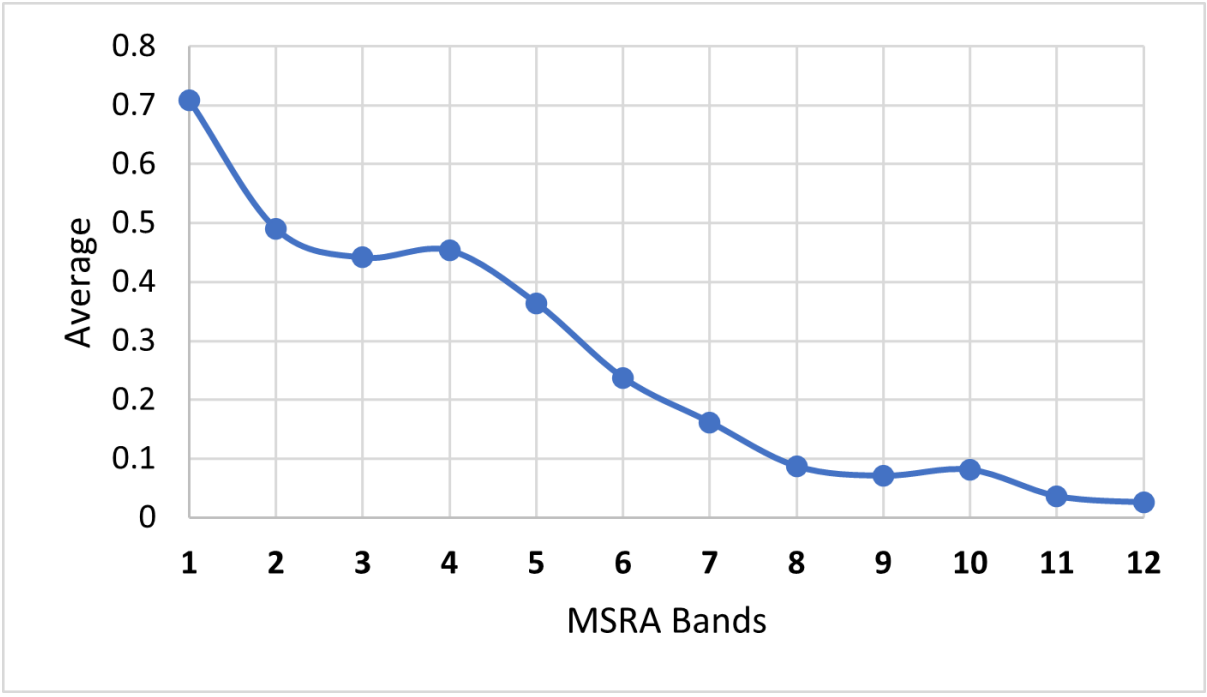

Figure 2: *Average number of ARCP outcomes 2, 3, and 4 per person (y-axis) for each MSRA band (x-axis)*

scores above 480 significantly reduced the number of these ARCP outcomes as seen in Table 6. Analogously, MSRA score bands corresponding to scores above 520 significantly increased by at least three times the likelihood of having only standard ARCP outcomes as seen in Table 7.

Table 5: *MSRA score bands and demographic characteristics as predictors of RCA pass rates*

Predictors	RCA pass rates			
	OR	**S.E.**	**95% CI of OR**	**p-value**
Sex (Female)	1			
Male	0.74	0.25	0.37, 1.45	=0.377
Ethnicity (White)	1			
Ethnic Minority	0.48	0.25	0.18, 1.32	=0.156
Mixed	0.14	0.13	0.20, 0.94	=0.043
Qualification Country (UK)	1			
Not-UK	0.30	0.15	0.11, 0.80	=0.017
Declared Disability (No)	1			
Yes	0.58	0.22	0.27, 1.23	=0.156
MSRA Bands (under 400)	1			
400-419	5.46	3.40	1.61, 18.51	=0.006
420-439	5.98	4.73	1.27, 28.18	=0.024
440-459	5.00	3.07	1.50, 16.65	=0.009
460-479	2.60	1.53	0.81, 8.24	=0.107
480-499	6.24	4.54	1.50, 25.95	=0.012
500-519	5.95	4.96	1.16, 30.47	=0.032
520-539	9.89	12.15	0.89, 109.88	=0.062
560-579	9.97	13.52	0.71, 142.06	=0.090
580-599	8.03	10.16	0.67, 95.92	=0.100
Cons	7.69	6.30	1.55, 38.28	=0.013
	pseudo$R^2 = 0.18, X^2(14) = 54.75, p<0.001$			

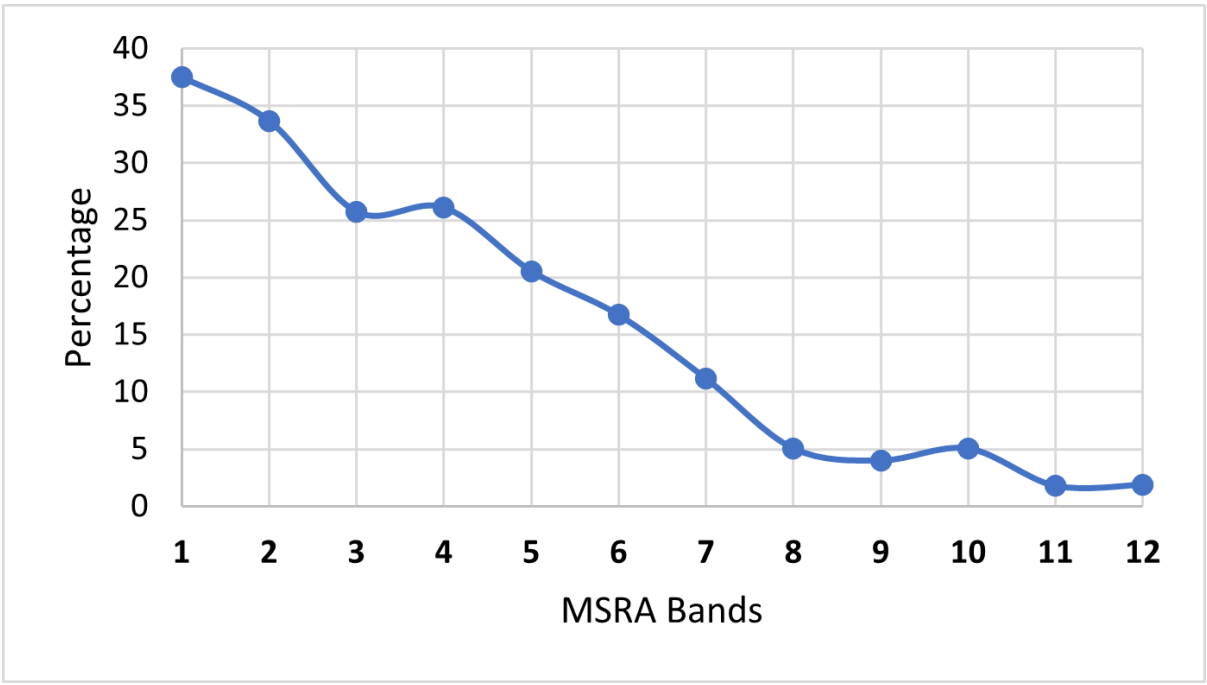

Figure 3: *Percentage of individuals with at least one of the ARCP outcomes 2, 3, or 4 (y-axis) for each MSRA score band (x-axis)*

When MSRA scores were taken into account, ethnicity was not associated with presence of ARCP outcomes 2, 3, or 4, but male sex (OR 0.50, 95% CI 0.37, 0.68, p<0.001), country of

Table 6: *MSRA score bands and demographic characteristics as predictors of the number of ARCP outcomes 2, 3, and 4*

Predictors	**Number of ARCP outcomes 2, 3, 4**			
	B	**S.E.**	**95% CI of B**	**p-value**
Sex (Female)	0			
Male	0.13	0.03	0.08, 0.19	<0.001
Ethnicity (White)	0			
Ethnic Minority	0.07	0.03	0.01, 0.14	=0.025
Mixed	0.05	0.07	-0.09, 0.19	=0.499
Qualification Country (UK)	0			
Not-UK	0.12	0.04	0.04, 0.21	=0.004
Declared Disability (No)	0			
Yes	0.35	0.04	0.26, 0.44	<0.001
MSRA Bands (under 400)	0			
400-419	-0.09	0.12	-0.32, 0.14	=0.434
420-439	-0.06	0.12	-0.29, 0.16	=0.585
440-459	-0.11	0.11	-0.33, 0.10	=0.300
460-479	-0.17	0.11	-0.39, 0.04	=0.104
480-499	-0.27	0.11	-0.48, -0.06	=0.012
500-519	-0.25	0.11	-0.47, -0.05	=0.017
520-539	-0.35	0.11	-0.57, -0.14	=0.001
540-559	-0.29	0.11	-0.52, -0.08	=0.007
560-579	-0.32	0.11	-0.54, -0.10	=0.004
580-599	-0.35	0.12	-0.58, -0.12	=0.003
600 and over	-0.34	0.13	-0.58, -0.09	=0.008
Cons	0.31	0.11	0.10, 0.52	<0.001
	$R^2 = 0.12, F(16, 1, 754) = 15.06, p<0.001$			

qualification outside the UK (OR 0.50, 95% CI 0.34, 0.74, p=0.001) and declaration of disability (OR 0.33, 95% CI 0.23, 0.49, p<0.001) were associated with a reduced chance of having only standard ARCP outcomes.

Table 7: *MSRA score bands and demographic characteristics as predictors of the presence of only standard ARCP outcomes*

Predictors	**Presence of only ARCP standard outcomes**			
	OR	**S.E.**	**95% CI of OR**	**p-value**
Sex (Female)	1			
Male	0.50	0.08	0.37, 0.68	<0.001
Ethnicity (White)	1			
Ethnic Minority	0.70	0.13	0.49, 1.01	=0.057
Mixed	0.62	0.27	0.26, 1.46	=0.274
Qualification Country (UK)	1			
Not-UK	0.50	0.10	0.34, 0.74	=0.001
Declared Disability (No)	1			
Yes	0.33	0.07	0.23, 0.49	<0.001
MSRA Bands (under 400)	1			
400-419	0.81	0.36	0.34, 1.94	=0.639
420-439	1.14	0.51	0.47, 2.76	=0.775
440-459	1.14	0.48	0.50, 2.61	=0.754
460-479	1.56	0.67	0.67, 3.60	=0.303
480-499	1.58	0.68	0.68, 3.69	=0.292
500-519	1.72	0.76	0.72, 4.07	=0.217
520-539	4.18	2.07	1.59, 11.01	=0.004
540-559	3.30	1.66	1.24, 8.83	=0.017
560-579	3.32	1.73	1.20, 9.21	=0.021
580-599	12.06	9.96	2.39, 60.87	=0.003
600 and over	5.65	4.72	1.10, 29.06	=0.038
Cons	8.04	3.54	3.39, 19.05	<0.001
	$\text{pseudoR}^2 = 0.23, X^2(16) = 455.88, p<0.001$			

6.2 The predictive validity of the MSRA and the selection centre (SC)

Lower SC score bands corresponded to poorer GP training outcomes as seen in figures 4, 5 and 6.

Adjusting for SC scores did not change the predictive validity of MSRA score bands as seen in Table 8. SC score bands do not have predictive validity for AKT nor RCA (p>0.05 for all score bands). SC score bands 2 and above predicted better outcomes on the CSA when compared to Band 1 (p<0.05). Being successful on ARCP (i.e., having only standard outcomes after removing outcomes 2, 3, and 4 due to failure on other examinations) was not predicted by SC scores (OR: 1.00, 95%CI: 1.00, 1.01, p=0.06). Lower numbers of ARCP outcomes 2, 3, and 4 were predicted only by SC band 6 which corresponded to scores between 48-50 (r= - 0.11, 95%CI -0.22, -0.01, p=0.04). Adding SC scores to MSRA scores did not explain additional variance in AKT scores (variance remained at 26%) or in the number of ARCP outcomes 2, 3, and 4 (variance remained at 12%). It slightly increased the variance explained in CSA scores by 2% and in RCA scores by 1%.

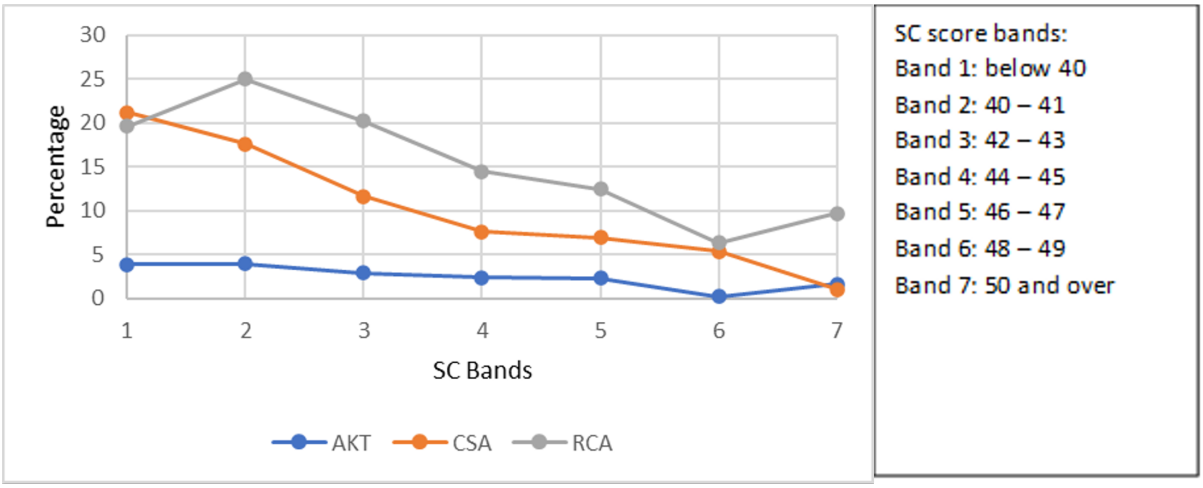

Figure 4: *Fail rates at the 5 year data collection point expressed as percentages (x-axis) per SC score bands (y-axis) for each outcome of GP training*

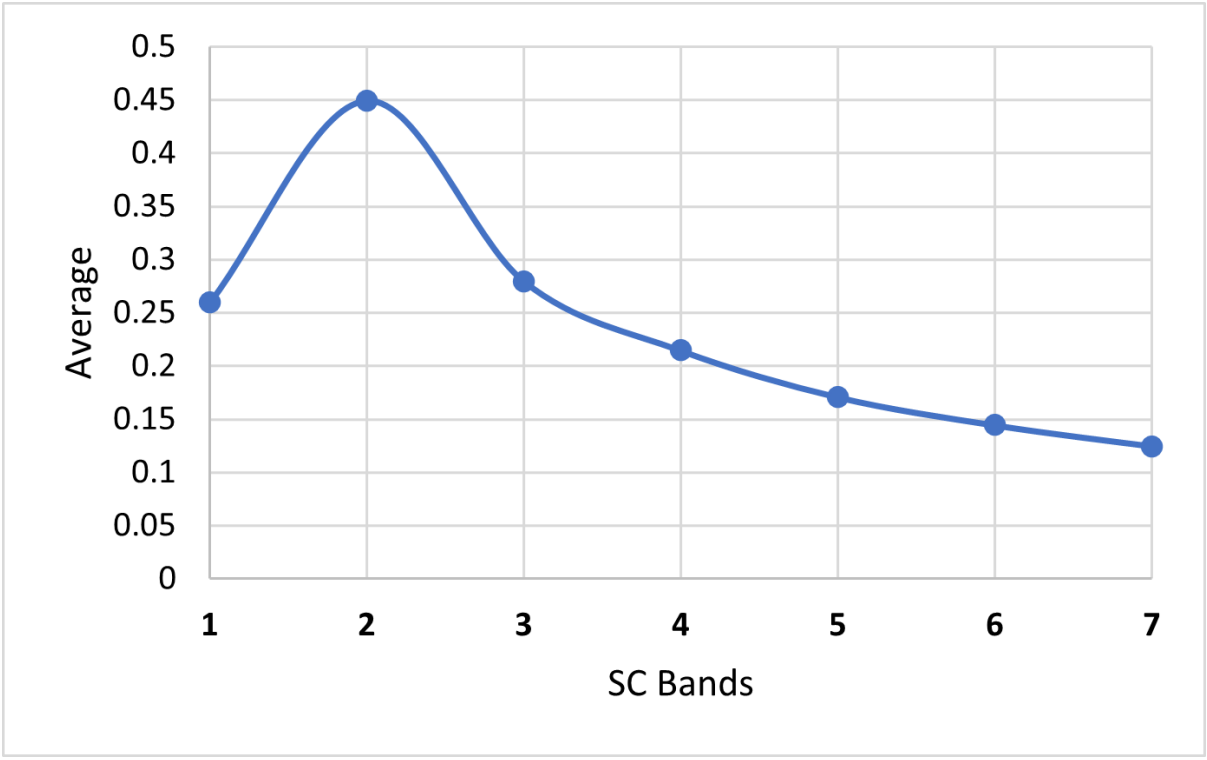

Figure 5: *Average Number of ARCP outcomes 2, 3, and 4 per person (y-axis) for each SC score band (x-axis)*

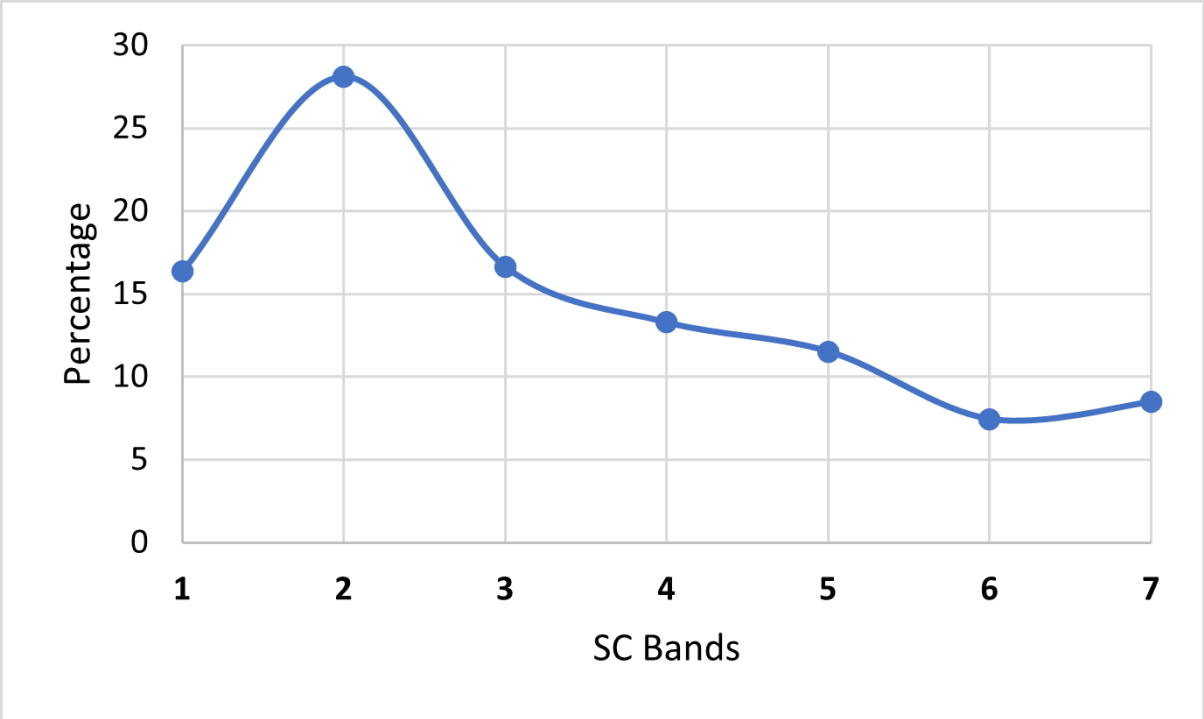

Figure 6: *Percentage of individuals with at least one of the outcomes 2, 3, or 4 (y-axis) for each SC score band (x-axis)*

Table 8: *MSRA score bands as predictors of the number of ARCP outcomes 2, 3, and 4*

	AKT		**CSA**		**RCA**		**ARCP negative outcome**	
	without SC OR (95% CI)	with SC OR (95%CI)	without SC OR (95% CI)	with SC OR (95% CI)	without SC OR (95% CI)	with SC OR (95% CI)	without SC OR (95% CI)	with SC OR (95% CI)
MSRA scores	1.03** (1.02, 1.04)	1.03** (1.02, 1.03)	1.01** (1.00, 1.01)	1.01** (1.00, 1.01)	1.01* (1.01, 1.02)	1.01** (1.01, 1,02)	1.01** (1.01, 1.01)	1.01** (1.00, 1.01)
MSRA Bands (under 400)								
400-419	3.47* (1.28, 9.36)	3.09* (1.14, 8.40)	0.92 (0.40, 2.10)	0.81 (0.35, 1.91)	5.46* (1.61, 18.51)	5.35* (1.50, 19.11)	0.81 (0.34, 1.93)	0.82 (0.34, 1.97)
420-439	4.29* (1.42, 12.94)	4.33* (1.41, 13.24)	2.58 (0.97, 6.88)	2.59 (0.93, 7.13)	5.98* (1.27, 28.18)	5.36* (1.10, 26.04)	1.14 (0.47, 2.76)	1.10 (0.45, 2.67)
440-459	6.86** (2.40, 19.66)	6.36** (2.21, 18.31)	1.04 (0.47, 2.33)	1.00 (0.44, 2.31)	5.00* (1.50, 16.64)	5.03* (1.45, 17.42)	1.14 (0.50, 2.61)	1.07 (0.47, 2.46)
460-479	9.93** (3.18, 31.03)	9.58* (3.03, 30.27)	0.99 (0.44, 2.22)	0.91 (0.39, 2.13)	2.59 (0.81, 8.25)	2.41 (0.72, 8.06)	1.55 (0.67, 3.69)	1.53 (0.66, 3.58)
480-499	15.34** (4.35, 54.08)	14.60** (4.11, 51.85)	1.48 (0.61, 3.60)	1.43 (0.57, 3.59)	6.24* (1.51, 25.95)	6.30* (1.46, 27.22)	1.58 (0.68, 3.69)	1.46 (0.62, 3.44)
500-519	37.53** (8.37, 168.40)	35.72** (7.84, 162.79)	4.00* (1.31, 12.23)	3.67* (1.17, 11.50)	5.95* (1.16, 30.47)	6.42* (1.16, 35.39)	1.72 (0.73, 4.07)	1.61 (0.68, 3.84)
520-539	53.30** (9.58, 296.52)	52.15** (9.23, 294.50)	2.47 (0.85, 7.15)	2.28 (0.77, 6.78)	9.89 (0.89, 109.88)	11.03 (0.96, 127.27)	4.18* (1.59, 11.01)	4.13* (1.57, 10.91)
540-559	104.06 (11.28, 959.69)	90.99** (9.70, 853.30)	/	/	/	/	3.30* (1.24, 8.82)	3.00* (1.11, 8.06)
560-579	/	/	11.58* (1.36, 98.83)	7.31 (0.84, 63.46)	9.98 (0.70, 142.06)	10.08 (0.59, 172.28)	3.32* (1.20, 9.21)	2.49 (0.89, 6.96)

6.3 Reassessing threshold scores in MSRA for access to the Direct Pathway using ROC curve analysis

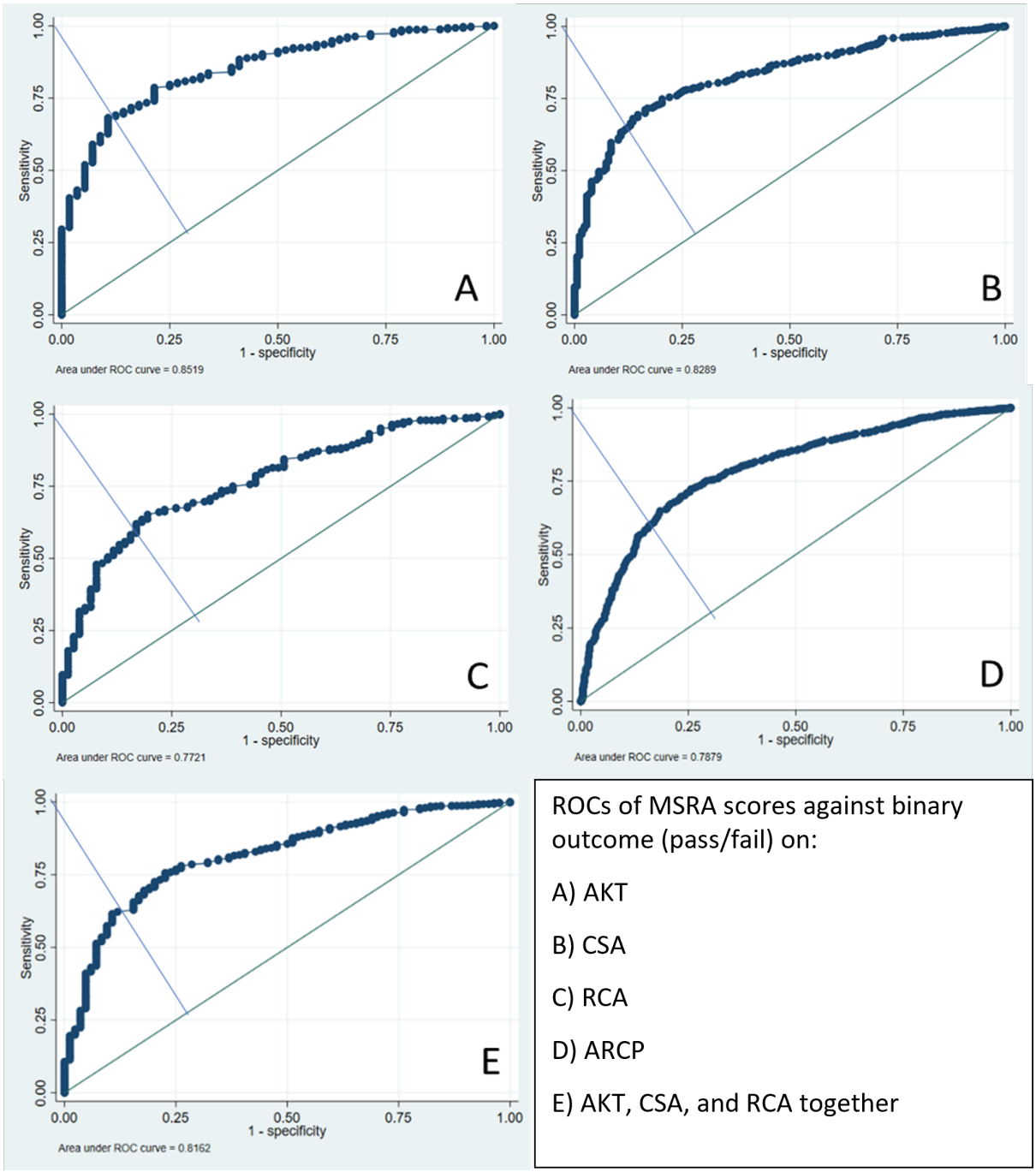

Figure 7: *ROCs of MSRA scores against binary outcome (pass/fail) on each assessment*

An MSRA threshold of 555 maximises specificity for AKT pass within a five-year period of data collection , whereas an MSRA threshold of 585 maximises specificity for CSA and/or RCA. This means that all candidates scoring above this threshold passed these training assessments during this time period. If we choose to maximise specificity for ARCP (only standard in training outcomes) the MSRA threshold would be 650.

If we chose equivalence between specificity (Sp) and sensitivity (Se), then an optimal

threshold score for DP of 472 should be selected for AKT, 483 for CSA, 466 for RCA, and 495 for ARCP. An MSRA threshold score of 483 should be chosen for correctly identifying pass/fail rates for AKT, RCA, and CSA combined.

A threshold of 497 (or a round score of 500) should be chosen for correctly identifying pass/fail rates for AKT, RCA, and CSA and only standard outcomes on the ARCP during the 5 years after entering training.

In this cohort, 93.5% of candidates that scored 500 or more performed well on all the assessments within five years whilst only 65.1% of those who scored below 500 achieved the same.

6.4 Pass rates and demographics

Pass rates were the highest for AKT, with 98% of candidates passing this assessment within the study period. This was followed by the CSA (92% pass rate) and the RCA (86% pass rate, but candidates only had a maximum of three attempts of a maximum allowed 4) as seen in Figure 8. Pass rates were lowest for the RCA, but the number of possible attempts was the lowest (3 compared to 5 for AKT and CSA) and the circumstances were different due to the new context of the COVID-19 pandemic. Pass rates decreased as the number of attempts increased as seen in Figure 9.

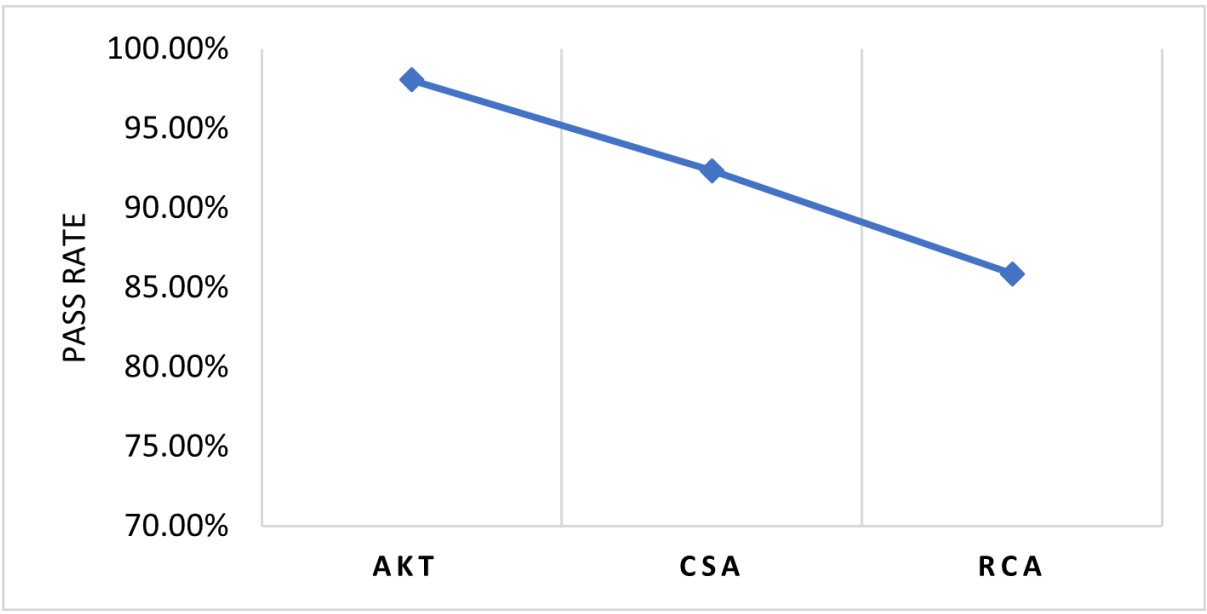

Figure 8: *Pass rates for each assessment*

Demographic data showed that a higher percentage of female candidates, candidates of white ethnicity, UK medical graduates and those without a declared disability took the MSRA and accessed the Direct Pathway as seen in table 9.

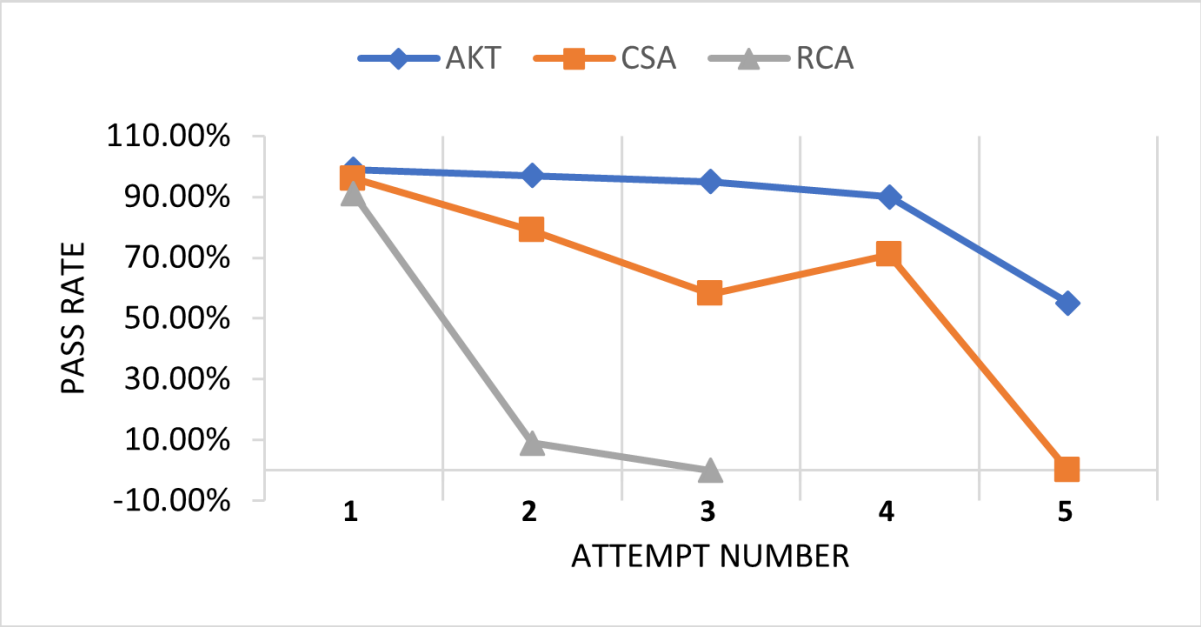

Figure 9: *Pass rates per number of attempts for each assessment*

Table 9: *Characteristics of candidates taking the MSRA and accessing the Direct Pathway*

	MSRA		
	Direct Pathway		
	Yes	**No**	**Total**
Gender			
Female	402 (18.87%)	1,728 (81.13%)	2,130 (63.81%)
Male	136 (11.26%)	1,072 (88.74%)	1,208 (36.19%)
Ethnicity			
White	444 (25.30%)	1,311 (74.70%)	1.755 (53.95%)
Ethnic Minorities	71 (5.07%)	1,329 (94.93%)	1,400 (43.04%)
Mixed	15 (15.31%)	83 (84.69%)	98 (3.01%)
Qualification			
UK	547 (20.98%)	2,085 (79.22%)	2,632 (76.76%)
non-UK	3 (0.40%)	794 (99.60%)	797 (23.24%)
Disability			
No	248 (13.61%)	1,574 (86.39%)	1,822 (88.02%)
Yes	14 (5.65%)	234 (94.35%)	248 (11.98%)

Discussion

7.1 Main findings

MSRA scores or score bands were predictive for all GP licensing outcomes at the five-year point (AKT, CSA, RCA, and WPBA-ARCP). Lower SC score bands corresponded to poorer GP training outcomes but adding SC scores did not change the predictive validity of the MSRA, and therefore the SC did not add further information to MSRA scores.

MSRA Bands 1 and 2 were most susceptible to ARCP outcomes 2, 3, and 4 with about 35% of candidates that fall within these bands having at least one of these outcomes that was not due to a failure on the AKT or CSA/RCA.

Our analysis suggests that an MSRA threshold of 500 (or, more precisely, 497) is the optimal for correctly identifying successful outcomes on the AKT, RCA, and CSA within the study time frame and only standard outcomes on ARCP at the five-year point. Importantly, 93.5% of candidates that scored 500 or more performed well on all the assessments.

Whatever the future of DP system, this figure might be useful for educational planning purposes and as indicator for individuals more (or less) likely to need extra support in training. This has implications for the training community and might open opportunities to explore the efficacy of different interventions at an early stage.

An interesting finding from this study is that differences in licensing assessments for ethnic minority doctors were no longer apparent when scores at selection and other demographic factors (sex, country of qualification and declared disability) were taken into account.

7.2 Comparison with previous research

This study confirmed that the MSRA scores were predictive of performance at licensing assessments, not only for AKT and CSA as previous investigators have found (Patterson et al. 2013), but also for RCA total score (not pass/fail rates) and for WPBA ARCP standard outcomes. Furthermore, it revealed that the SC added little to the predictive validity of the MSRA.

A previous study combined MSRA and SC scores to investigate prediction of performance in AKT and CSA in UK and International Medical Graduates in one deanery, and found good prediction for the combined score (Wakeford 2012) whereas this study was able to separate MSRA and SC for all candidates and found little additional predictive value of the SC.

This study also justified the use of the threshold score employed to bypass the SC from 2016 to 2020, when the SC was suspended because of COVID-19. No previous publications have looked at this issue.

There has been a longstanding debate about ethnicity as a factor in GP licensing in the UK (Wakeford et al. 1992) culminating in a study which asserted, despite lack of evidence to support this, that the 'subjective bias due to racial discrimination' in the clinical skills assessment may be a cause of failure for UK trained ethnic minority candidates (Esmail & Roberts 2013). This conclusion was at odds with a judicial review which found against evidence of bias (Rendel et al. 2015) and a narrative review which also concluded that evidence of racial bias was lacking (de Bere et al. 2015). Qualitative studies have found more complex educational and social factors affecting the training pathway and performance (Woolf et al. 2016).

A problem with previous studies is that, although gender and place of primary medical

qualification have been considered, other important factors which are also related to ethnicity, such as declared disability, have not been accounted for in regression models (Asghar et al. 2019, 2018). Importantly, the multivariable predictive models in this study showed that the MSRA was one the main factors that influenced endpoint assessments' outcomes and that ethnicity ceased to be a significant predictor for all endpoint assessments when taking MSRA scores into account. When looking at country of qualification, International Medical Graduates (IMGs) were just as likely to pass AKT but less likely to pass CSA, RCA or achieve only standard ARCP outcomes once the MSRA scores and other demographic factors were accounted for.

It is important to notice that differential attainment (DA) is present throughout the training journey, being similar in both selection and licensing assessments. The explanations are complex and multiple but due most 'likely to difference in training experience and other cultural factors between candidates trained in the UK and abroad' (Esmail & Roberts 2013). These may include: differences in experiences of recruitment, training and assessments, more limited professional networks (lack of mentorship or peer support), cultural barriers (language difficulties, lack of understanding of cultural norms, bias against seeking support or additional training), social challenges (poor work-life balance, separation from family and social support outside the work setting), and psychological difficulties (stress, anxiety, burnout) (Woolf et al. 2016). Another factor affecting performance of non-UK graduates in clinical licensing tests may be differences in initial medical training, where a doctor-centred rather than patient-centred approach to consulting may be taught and learnt (Jalal et al. 2019).

Doctors declaring disability (mostly commonly Specific Learning Difficulty [SpLD]) were not different in AKT pass rates but were less likely to achieve a CSA pass and an ARCP standard outcome once MSRA and other demographic factors were taken into account. These confirm findings from previous studies where no differences in AKT outcomes were found for candidates with SpLD once other factors were accounted for, but where differences were found in the CSA (Asghar et al. 2019, 2018).

7.3 Strengths and limitations

The study benefitted from accurate data linkage using GMC numbers to link selection and outcome data for AKT, CSA, RCA and WPBA (ARCP) performance and high rates of completeness for demographic data including sex, ethnicity, place of primary qualification and declared disability.

The data allowed us to separately explore the impact of MSRA and SC on licensing outcomes which has been a criticism of previous approaches (Wakeford 2014).

The 'Standard' GP training programme is three years whole time equivalent. This study followed the 2016 cohort to 2021 as it was anticipated that most participants would by then have attempted their licensing assessments. However, in that time period, not all participants that attempted the licensing tests and were unsuccessful had the opportunity to take them the permitted four times. Whilst for AKT and CSA this number was small (only 6% of candidates), it involved all participants for the RCA who could only attempt this assessment three times by the end of the study. Thus, this was a factor that influenced the lower pass rates particularly for RCA. Candidates on training extensions, maternity leave etc may have successfully completed after the study end.

The absence of significant differences for IMGs and those with declared disabilities in the RCA may have been due to the smaller numbers of candidates who were able to take this assessment. Only 545 attempted the RCA compared to 2,883 for AKT and 2,313 for the CSA.

7.4 Implications for policy, practice and future research

The SC is currently suspended (2020-2022) because of the COVID-19 pandemic. This study suggests that it adds little predictive validity to the MSRA in terms of future performance in GP licensing assessments. So, unless the SC can be changed to increase predictive validity (for all candidate groups), the MSRA used alone may be sufficient. It could also be argued that the balance of the two components of the MSRA (SJT and CPS) may widen access by reducing the inherent disadvantage to candidates of lower socioeconomic backgrounds seen in purely cognitive assessments (Lievens et al. 2016).

Our analysis, using a large cohort of doctors, suggests an optimal score at MSRA to predict an uncomplicated training journey without need for additional support. Although pressure to recruit to GP specialty training schemes will vary with workforce needs, cut scores for selection need to take into consideration the likelihood of a doctor selected for training passing their licensing examinations against the costs to trainees and deaneries, as educational providers, of failure with remediation or extensions to training. Doctors admitted to training with lower scores may need additional support during training to maximise their chances of successful licensing.

Further research is needed to investigate the initial cause of the DA and the effect of possible interventions to mitigate this (e.g., exam preparation, information available to candidates etc.).

The MSRA exam is currently used as a shortlisting tool by many different specialties. It might be of use to repeat a similar scrutiny comparing entrance and exit assessments in other specialties.

This study examined the relationship between two adjacent steps in the educational journey. It could be useful to extend that enquiry in the other direction to the Foundation Programme. Findings in GP Equalities Impact Reports (2009-11) indicate that candidates attending Foundation Programmes (FP) were more successful in GP selection. In addition, past literature suggests overseas doctors who enter speciality training without having attended a UK FP have greater difficulty adapting to the UK healthcare system (Jalal et al. 2019). Links between these two findings could be explored to determine how the FP contributes to more effective performance in selection.

7.5 Conclusions

MSRA scores predict licensing outcomes for AKT, CSA, RCA, and WPBA-ARCP within five years of starting training.

The optimal MSRA threshold score for predicting an uncomplicated training pathway to licensing was around 500 in this large cohort.

Doctors admitted to GP speciality training who score below this threshold may need additional support during training to maximise their chances of achieving licensing.

SC scores did not add further information to MSRA scores which supports the decisions both to institute the Direct Access process in 2016 and the subsequent suspension of the SC in 2020 during the pandemic.

Ethnicity did not reduce the chance of passing GP licensing tests once gender, place of primary medical qualification, declared disability and MRSA scores were taken into account. Comparing candidate scores by ethnicity creates a false impression of differential attainment which should be addressed by routinely taking these factors into account.

References

Asghar, Z. B., Siriwardena, A. N., Elfes, C., Richardson, J., Larcombe, J., Neden, K. A., Salim, A., Smalley, D. & Blow, C. (2018), 'Performance of candidates disclosing dyslexia with other candidates in a UK medical licensing examination: cross-sectional study', *Postgraduate Medical Journal* **94**(1110), 198–203.

Asghar, Z., Williams, N., Denney, M. & Siriwardena, A. N. (2019), 'Performance in candidates declaring versus those not declaring dyslexia in a licensing clinical examination', *Medical Education* **53**(12), 1243–1252.

Botan, V., Laparidou, D., Phung, V.-H., Cheung, P., Freeman, A., Wakeford, R., Denney, M., Law, G. R. & Siriwardena, A. N. (2021), 'Candidate perceptions of the UK recorded consultation assessment: cross-sectional data linkage study', *Education for Primary Care* pp. 1–9.

Davison, I., McManus, C. & Taylor, C. (2016), 'Evaluation of GP specialty selection'.

de Bere, S. R., Nunn, S. & Nasser, M. (2015), 'Understanding differential attainment across medical training pathways: a rapid review of the literature'.

Esmail, A. & Roberts, C. (2013), 'Academic performance of ethnic minority candidates and discrimination in the MRCGP examinations between 2010 and 2012: analysis of data', *BMJ* **347**.

Jalal, M., Bardhan, K. D., Sanders, D. & Illing, J. (2019), 'Overseas doctors of the nhs: migration, transition, challenges and towards resolution', *Future Healthcare Journal* **6**(1), 76.

Lievens, F., Patterson, F., Corstjens, J., Martin, S. & Nicholson, S. (2016), 'Widening access in selection using situational judgement tests: evidence from the UKCAT', *Medical Education* **50**(6), 624–636.

Ooi, S. Z. Y. & Ooi, R. (2021), 'Impact of the recent changes of the multi-specialty recruitment assessment (MSRA) weightage in specialty training recruitment during the covid-19 pandemic', *Postgraduate Medical Journal* .

Patterson, F., Baron, H., Carr, V., Plint, S. & Lane, P. (2009), 'Evaluation of three short-listing methodologies for selection into postgraduate training in general practice', *Medical Education* **43**(1), 50–57.

Patterson, F., Ferguson, E., Norfolk, T. & Lane, P. (2005), 'A new selection system to recruit general practice registrars: preliminary findings from a validation study', *BMJ* **330**(7493), 711–714.

Patterson, F., Lievens, F., Kerrin, M., Munro, N. & Irish, B. (2013), 'The predictive validity of selection for entry into postgraduate training in general practice: evidence from three longitudinal studies', *British Journal of General Practice* **63**(616), e734–e741.

Patterson, F., Zibarras, L., Carr, V., Irish, B. & Gregory, S. (2011), 'Evaluating candidate reactions to selection practices using organisational justice theory', *Medical Education* **45**(3), 289–297.

Plint, S. & Patterson, F. (2010), 'Identifying critical success factors for designing selection processes into postgraduate specialty training: the case of uk general practice', *Postgraduate Medical Journal* **86**(1016), 323–327.

Rendel, S., Foreman, P. & Freeman, A. (2015), 'Licensing exams and judicial review: the closing of one door and opening of others?', *British Journal of General Practice* **65**(630), 8–9.

Wakeford, R. (2012), 'International medical graduates' relative under-performance in the MRCGP AKT and CSA examinations', *Education for Primary Care* **23**(3), 148–152.

Wakeford, R. (2014), 'GMC's planned study into GP selection scheme's value in predicting MRCGP performance is flawed', *BMJ* **348**.

Wakeford, R., Farooqi, A., Rashid, A. & Southgate, L. (1992), 'Does the MRCGP examination discriminate against Asian doctors?', *British Medical Journal* **305**(6845), 92–94.

Woolf, K., Rich, A., Viney, R., Rigby, M., Needleman, S. & Griffin, A. (2016), 'Fair training pathways for all: understanding experiences of progression', *London: UCL, Education RDfM* .