Does the Intercollegiate Membership of the Royal College of Surgeons (MRCS) examination predict 'on-the-job' performance during UK higher specialty surgical training?

Scrimgeour, D. S. G.; Brennan, P. A.; Griffiths, G.; Lee, A .J.; Smith, F. C. T.; Cleland, J.

Published in:
Annals of the Royal College of Surgeons of England

DOI:
10.1308/rcsann.2018.0153

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Royal College of Surgeons
ADVANCING SURGICAL CARE

EDUCATION

Ann R Coll Surg Engl 2018; **100:** 669–675
doi 10.1308/rcsann.2018.0153

Does the Intercollegiate Membership of the Royal College of Surgeons (MRCS) examination predict 'on-the-job' performance during UK higher specialty surgical training?

DSG Scrimgeour[1,2,3], PA Brennan[3], G Griffiths[4], AJ Lee[5], FCT Smith[3], J Cleland[1]

[1]Centre for Healthcare Education Research and Innovation, University of Aberdeen, Aberdeen, Scotland, UK
[2]Department of Colorectal Surgery, Aberdeen Royal Infirmary, Aberdeen, Scotland, UK
[3]Intercollegiate Committee for Basic Surgical Examinations, UK
[4]Department of Vascular Surgery, Ninewells Hospital, Dundee, Scotland, UK
[5]Medical Statistics Team, University of Aberdeen, Aberdeen, Scotland, UK

ABSTRACT

INTRODUCTION The Intercollegiate Membership of the Royal College of Surgeons (MRCS) is a mandatory examination to enter higher surgical specialty training in the UK. It is designed to help to ensure that successful candidates are competent to practice as higher surgical trainees. The annual review of competence progression (ARCP) assesses trainees' competence to progress to the next level of training and can be interpreted as a measure of 'on-the-job' performance. We investigated the relationship between MRCS performance and ARCP outcomes.

MATERIALS AND METHODS All UK medical graduates who passed MRCS (Parts A and B) from 2007 to 2016 were included. MRCS scores, attempts and sociodemographics for each candidate were crosslinked with ARCP outcomes (satisfactory, unsatisfactory and insufficient evidence). Multinomial logistic regression was used to identify potential independent predictors of ARCP outcomes.

RESULTS A total of 2570 trainees underwent 11,064 ARCPs; 1589 (61.8%) had only satisfactory outcomes recorded throughout training; 510 (19.9%) had at least one unsatisfactory outcome; and 471 (18.3%) supplied insufficient evidence. After adjusting for age, gender, first language and Part A performance, ethnicity (non-white vs white, OR 1.36, 95% CI 1.08 to 1.71), Part B passing score (OR 0.98, 95% CI 0.98 to 1.00) and number of attempts at Part B (two or more attempts vs one attempt, OR 1.50, 95% CI 1.16 to 1.94) were found to be independent predictors of an unsatisfactory ARCP outcome.

CONCLUSION This is the first study to identify predictors of ARCP outcomes during higher surgical specialty training in the UK and provides further evidence of the predictive validity of the MRCS examination.

KEYWORDS
MRCS – clinical performance – predictive validity – ARCP

Accepted 2 June 2018

CORRESPONDENCE TO
DSG Scrimgeour, E: duncan.scrimgeour@nhs.net

Introduction

The Intercollegiate Membership of the Royal College of Surgeons (MRCS) examination is one of the assessments used to determine whether trainees (residents) have acquired the knowledge, skills and attributes required to progress to higher specialty surgical training in the UK. We have previously examined the relationship between MRCS performance and national selection performance for general and vascular surgery, and have found a significant relationship between Part B MRCS and selection score.[1] However, to assess the predictive validity of the MRCS further the relationship between performance on both parts (A: knowledge and B: clinical) of this mandatory postgraduate surgical examination and subsequent performance in clinical practice needs to be investigated.

Since 2007, all surgical trainees in the UK have been assessed by an annual review of competence progression (ARCP),[2] a formally defined process in which a panel composed of a postgraduate dean or training programme director and at least two senior doctors working within the relevant specialty with a role related to postgraduate training decides whether a trainee is competent to progress to the next level of training. To inform this decision-making

process, trainees are required to provide the panel with a portfolio of evidence which includes a specific number of workplace-based assessments (as defined at the start of each training year), evidence of success at the relevant postgraduate examination (see below), an up-to date surgical logbook, a personal development plan for the year and a structured report from their educational and clinical supervisors. ARCPs can therefore be considered mostly as a measure of 'on-the-job' performance. Additional evidence can include multisource feedback, audits, presentations at regional, national and international meetings and peer reviewed publications.

Overviews of the surgical training pathway in the UK have been presented elsewhere.[3] For the purposes of the current study, it is important to highlight that trainees must demonstrate, through the ARCP process, successful completion of the core surgical training curriculum before entering the third year of training. Part of the evidence required for this is success at the MRCS examination.

In its current format, the MRCS consists of two parts, A and B. Part A is divided into two papers (three hours and two hours) and is designed to assess knowledge of both the principles of surgery in general and applied basic sciences. Part A uses a combination of extended matching multiple-choice questions and single best answers.[1]

Part B is an objective structured clinical examination that tests two broad areas: knowledge (eight stations including anatomy and applied surgical pathology, applied surgical science and critical care) and clinical and technical skills (10 stations including clinical examination, communication skills and technical skills). Each station is nine minutes long and is manned. There are two additional rest stations and two preparation stations.[1]

The aim of this study was to investigate the predictive validity of MRCS in relation to clinical practice. Specifically, our aim was to evaluate whether the results of this mandatory examination can predict how trainees will perform during higher specialty surgical training in the UK as assessed by performance at ARCP.

Materials and methods

All UK medical graduates who had attempted both parts of the MRCS (Parts A and B) since its origin in September 2007 to February 2016 were included. MRCS scores, number of attempts at each part of MRCS, date of medical school graduation, date of birth and the self-declared demographics of gender, first language and ethnicity were extracted and linked to ARCP outcomes for specialty training (ST) years 3–8 for each surgical trainee up to 16 January 2018. All data were anonymised before release to the research team. ARCP outcomes (Table 1) were divided into three groups: satisfactory (a trainee with only ARCP outcomes 1 and 6 throughout their surgical training), unsatisfactory (trainee had received at least one ARCP outcome 2, 3 or 4) and insufficient evidence (trainee had at least one ARCP outcome 5 recorded during training but none at 2, 3 or 4). All FTSTA, fixed term specialty training appointments; LAT, locum appointed for training and out-of-programme research ARCP outcomes (outcomes 7 and 8, respectively) were excluded from the overall analyses as these posts do not represent standard training in the UK.

Except for MRCS passing scores, all variables were subsequently dichotomised. Self-classified ethnicity was coded as 'white' or 'non-white', self-declared first language was categorised as 'English' or 'not English' and number of attempts at each part of the MRCS was grouped as 'one attempt' or 'two or more attempts'.

Older doctors (which we defined in our previous studies as 29 years or over at graduation from medical school)[1,4]

Table 1 Possible outcomes awarded to surgical trainees at annual review of competence progression.

ARCP Outcome	Description
1	Satisfactory progress; competences achieved as expected
2	May progress but requires specific/targeted training to achieve certain competences
3	Inadequate progress – additional training required
4	Released from training programme with or without competencies
5	Incomplete evidence presented – additional training time may be needed
6	Recommendation for completion of training having gained all required competences
7/LAT or FTSTA trainees:	
7.1	Satisfactory progress or completion of LAT placement
7.2	Development of specific competences required; additional training time not required
7.3	Inadequate progress by trainee; additional training time required
7.4	Incomplete evidence presented
8	Out-of-programme research, approved clinical time or a career break

ARCP, annual review of competence progression; FTSTA, fixed term specialty training appointments; LAT, locum appointed for training.

have been shown to face more challenges as they progress through specialty training.[5] We therefore included age group at graduation in our analyses.

Since MRCS pass marks vary from each examination sitting, performance at Part A and B of the MRCS was described in terms of percentage above the pass mark; for example a candidate scoring zero per cent has achieved the minimum pass mark for the examination.

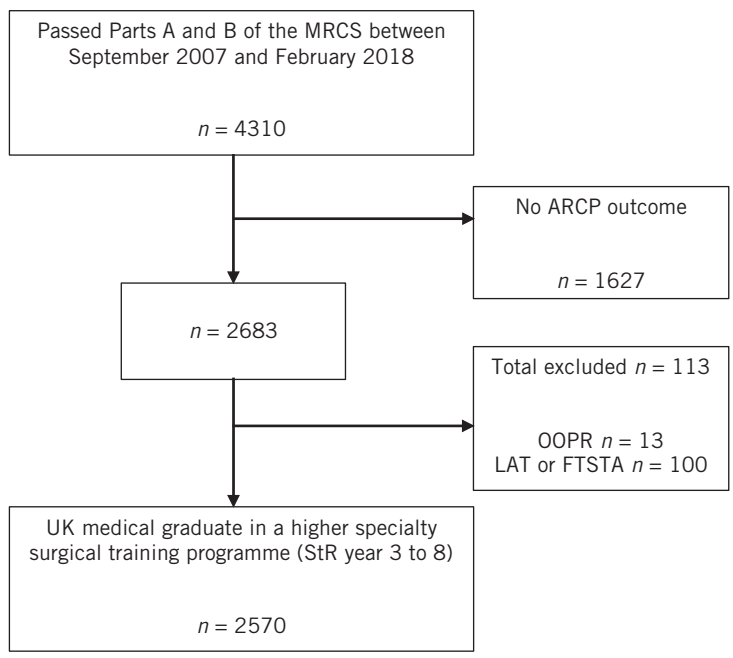
There is no specific ethics committee for MRCS, but both the Intercollegiate Committee for Basic Surgical Examinations and its internal quality assurance subcommittee, approved this study. The Data Analysis, Audit and Research Group of the Joint Committee on Surgical Training approved the release of ARCP outcomes.

Statistical analysis

All analyses were conducted using SPSS version 24.0. The Chi-squared test was initially employed to determine any significant associations with ARCP outcomes. The relationship between Part A and Part B MRCS score and ARCP outcome was examined using Kruskal–Wallis (owing to skewed data) and analysis of variance, respectively. Multinomial logistic regression models were created to identify potential independent predictors of an unsatisfactory and an insufficient evidence ARCP outcome. Any variable with $P < 0.10$ on univariate analysis was entered into the logistic regression model. All potential predictors with $P > 0.05$ in the full model were subsequently removed until only statistically significant predictors remained in the final model. Potential interactions between the remaining significant predictors were also examined.

Results

A total of 4310 UK medical graduates passed both parts of the MRCS between September 2007 and February 2016 (Fig 1). No ARCP outcomes were available for 1627 MRCS candidates as these doctors had either yet to complete core surgical training, had not been appointed to a surgical training programme, were at the beginning of their ST3 post at the time of ARCP data collection or had decided not to enter or apply for a surgical training post. Of the remaining 2683 trainees, 13 were undertaking out-of-programme research and 100 were in either a LAT or FTSTA post. A total of 11,064 ARCP outcomes were linked to the remaining 2570 surgical trainees. Of these, 1589 (61.8%) trainees had satisfactory ARCP outcomes recorded throughout their entire training period;



ARCP, annual review of competence progression
FTSTA, fixed-term specialty training appointments
LAT, locum appointed for training
MRCS, Membership of the Royal College of Surgeons
OOPR, out-of-programme research
StR, specialty registrar

Figure 1 Flow diagram of surgical trainees in the study.

Table 2 Annual review of competence progression outcomes during specialty surgical training for 2570 UK graduates by sociodemographic factors and number of attempts at Part A and B of the Intercollegiate Membership of the Royal College of Surgeons.

Variable	ARCP outcome						Missing data		*P*-value[a]
	Satisfactory		Unsatisfactory		Insufficient evidence				
	(*n*)	(%)	(*n*)	(%)	(*n*)	(%)	(*n*)	(%)	
Gender:							6	0.2	0.042
Women (*n* = 824)	538	65.3	146	17.7	140	17			
Men (*n* = 1740)	1047	60.2	363	20.9	330	19			
First language:							524	20.4	0.133
English	1191	61.6	391	20.2	350	18.1			
Non-English	63	55.3	32	28.1	19	16.7			
Age at graduation:							2	0.1	0.013
< 29 years (*n* = 2358)	1472	62.4	452	19.2	434	18.4			
≥ 29 years (*n* = 116)	116	55.2	58	27.6	36	17.1			
Ethnicity:							539	21.0	0.001
White	788	63.6	218	17.6	233	18.8			
Non-white	464	58.6	194	24.5	134	16.6			
Attempts at Part A MRCS:									0.061
One (*n* = 1981)	1249	63.0	378	19.1	354	17.9			
Two or more (*n* = 589)	340	57.7	132	22.4	132	22.4			
Attempts at Part B MRCS:									< 0.001
One (*n* = 2022)	1276	63.1	367	18.2	379	18.7			
Two or more (*n* = 548)	313	57.1	143	26.1	92	16.8			

ARCP, Annual Review of Competence Progression; MRCS, Membership of the Royal College of Surgeons.
[a] All *P*-values are from chi-squared analysis.

510 (19.9%) had at least one unsatisfactory ARCP outcome; and 471 (18.3%) had supplied insufficient evidence to at least one ARCP panel, but were otherwise progressing satisfactorily.

Univariate analysis found that women were more likely to achieve a satisfactory ARCP outcome (65.3% vs 60.2%) and less likely to achieve an unsatisfactory (17.7% vs 20.9%,) or insufficient evidence ARCP outcome (17.0% vs 19.0%) throughout surgical training compared to men ($P = 0.042$; Table 2). No statistically significant differences in ARCP outcomes were found between trainees who had declared English as their first language compared with those who had declared another language ($P = 0.133$). Mature medical graduates were less likely to achieve a satisfactory ARCP outcome (55.2% vs 62.4%) or insufficient evidence (17.1% vs 18.4%) outcome and more likely to be awarded an unsatisfactory (27.6% vs 19.2%) ARCP outcome than younger graduates ($P = 0.013$). Self-classified white trainees were more likely to receive a satisfactory (63.6% vs 58.6%) or insufficient evidence (18.8% vs 16.6%) ARCP outcome compared with non-white trainees, but white trainees were less likely to achieve an unsatisfactory outcome (17.6% vs 24.5%) throughout their surgical training ($P = 0.001$). No statistically significant relationship was found between ARCP outcome and number of attempts at Part A MRCS ($P = 0.061$). Those who required two or more attempts to pass Part B MRCS were less likely to achieve a satisfactory (57.1% vs 63.1%) or insufficient evidence (16.8% vs 18.7%) ARCP outcome compared with those who passed the examination at their first attempt. Trainees who made multiple attempts at Part B were also more likely to be awarded an unsatisfactory ARCP outcome (26.1% vs 18.2%) during their surgical training ($P < 0.001$).

There was no statistically significant relationship found between Part A MRCS score and ARCP outcome ($P = 0.071$). Trainees who received a satisfactory ARCP outcome throughout their training had achieved a higher Part B MRCS score compared with those who were awarded an unsatisfactory ARCP outcome (mean difference in score 1.71%, 95% confidence interval, CI, 0.74 to 2.69; Table 3). Those who achieved an unsatisfactory ARCP outcome scored lower in Part B MRCS compared with those who were awarded an insufficient evidence ARCP outcome

Table 3 Relationship between Part B of the Intercollegiate Membership of the Royal College of Surgeons score (percentage above the pass mark) and annual review of competence progression outcome.

ARCP outcome	Mean Part B score (%)	Mean difference (%) in Part B score between ARCP outcomes		
				95% CI
Satisfactory (1)	20.53	1 vs 2	1.71	0.74 to 2.69
Unsatisfactory (2)	18.82	1 vs 3	0.37	–0.63 to 1.38
Insufficient evidence (3)	20.16	2 vs 3	–1.34	–2.56 to –0.12

CI, Confidence Interval

Table 4 Multinomial logistic regression models for obtaining an unsatisfactory or an insufficient evidence outcome in the annual review of competence progression during higher specialty surgical training for UK medical graduates.

Predictor	ARCP outcome			
	Insufficient evidence		Unsatisfactory	
	Odds ratio	95 % CI	Odds ratio	95 % CI
Model 1:[a]				
MRCS Part A score (% above the pass mark)	1.01	1.00 to 1.03	1.00	0.99 to 1.02
MRCS Part B score (% above the pass mark)	0.99	0.98 to 1.00	0.98	0.97 to 1.00
Female sex	0.81	0.63 to 1.04	0.81	0.63 to 1.04
Age at graduation (< 29 years at graduation)	0.88	0.57 to 1.37	0.72	0.49 to 1.07
Non-white ethnicity	0.94	0.73 to 1.20	1.33	1.05 to 1.68
Part A MRCS ≥ 2 attempts	0.85	0.63 to 1.16	0.92	0.69 to 1.22
Part B MRCS ≥ 2 attempts	1.03	0.76 to 1.39	1.51	1.15 to 1.97
Model 2:[b]				
MRCS Part B score (% above the pass mark)			0.98	0.97 to 1.00
Non-white ethnicity			1.36	1.08 to 1.71
Part B MRCS ≥ 2 attempts			1.50	1.16 to 1.94

CI, Confidence Interval; MRCS, Membership of the Royal College of Surgeons; ARCP, annual review of competence progression.
[a] Model 1 included all potential predictors with P < 0.10 in univariate analysis.
[b] Model 2 excluded all potential predictors from model 1 with P > 0.05.

(mean difference in score –1.34%, 95% CI –2.56 to –0.12). No significant difference in mean Part B MRCS score was identified between those with unsatisfactory and insufficient evidence ARCP outcomes ($P = 0.664$).

Ethnicity (non-white vs white trainees, odds ratio, OR, 1.36, 95% CI 1.08 to 1.71; Table 4), number of attempts at Part B MRCS (two or more attempts vs one attempt, OR 1.50, 95% CI 1.16 to 1.94) and Part B MRCS score (OR 0.98, 95% CI 0.97 to 1.00) were all found to be independent predictors of an unsatisfactory ARCP outcome. No significant interactions were found between ethnicity, number of Part B MRCS attempts or Part B MRCS score on failure to progress via an unsatisfactory outcome. There were no statistically significant independent predictors of an insufficient evidence ARCP outcome.

Discussion

Since MRCS is a prerequisite for entering higher specialty surgical training in the UK, it is important to investigate the potential relationship between MRCS performance and subsequent performance during higher surgical training. This is the first study to report on overall 'on-the-job' performance of UK medical graduates in higher specialty surgical training, as evaluated by ARCP. We found that performance in Part B MRCS and ethnicity were independent predictors of an unsatisfactory ARCP outcome during higher surgical training in the UK. No significant independent predictors of an insufficient evidence ARCP outcome were found.

Our results echo findings from studies investigating the relationship between other high-stakes medical

examinations around the world and performance in clinical practice and other medical assessment processes. For example, doctors with low scores in both the Part I knowledge exam and the clinical Part II examination of the Medical Council of Canada Qualifying Examinations are more likely to achieve an unacceptable outcome in a peer assessment on quality of care.[6] Similarly, low scores in the knowledge-based Step I[7] and the clinical Step 2[8] of the United States Medical Licensing Examination (USMLE) have been found to be associated with poor performance on the annual in-training examination for general surgery, at any time during residency. In the UK, candidates who score higher in all parts of the Membership of the Royal College of Physicians examination are more likely to do better in workplace-based assessments than those who underperform.[9]

We found that after adjusting for the influence of Part A MRCS performance, gender and age group, trainees who failed Part B MRCS at first attempt were 50% more likely to be awarded an unsatisfactory outcome during their surgical training compared with those who passed at first attempt. Although statistically significant, the odds of predicting an unsatisfactory ARCP outcome for Part B passing score was almost 1 (95% CI 0.97 to 1.00) and its real significance should therefore be interpreted with caution. We found a similar relationship in our previous study, which identified that multiple attempts at Part B MRCS was an independent predictor of selection score for entry into general and vascular higher surgical training.[1] These findings support the predictive validity of MRCS and provide strong evidence that doctors who require multiple attempts are more likely to require targeted support through surgical training.

Despite evidence of a relationship between multiple attempts at high-stakes medical examinations and other performance indicators, most studies have focused on the relationship between examination scores and future performance, without distinguishing between those who initially passed and those who required multiple attempts.[10–12] However, when number of attempts have been investigated, a clear pattern emerges. For example, passing the Professional and Linguistic Assessments Board test at first attempt is independently predictive of more satisfactory outcomes at ARCP compared with requiring more than one attempt.[13] Similarly, multiple attempts at the Medical College Admission Test are associated with an increased risk of failing USMLE Step 2.[14] The relationship between sitting a postgraduate UK medical examination on several occasions and subsequent performance during specialty training in the UK was previously unknown, but McManus *et al.*[15] did find that as the number of attempts at each part of MRCP increased, final passing score decreased. The same pattern exists for candidates requiring multiple attempts to pass Part A MRCS.[4]

In the United States, USMLE Step I failure has been repeatedly found to be associated with underperformance in other medical assessment processes.[16,17] More recently, Ogunyemi et al.[18] reported that students failing the USMLE Step I at first attempt were 3.8 times more likely to fail USMLE Step 2. Another study found that the relative risk of not being specialty board certified after failing USMLE Step I, was 2.2.[19] Overall, these findings suggest that doctors who require multiple attempts to pass mandatory examinations at the beginning of their career are more likely to have similar difficulties with other standardised examinations throughout their medical training. Identifying these individuals at the earliest opportunity may help trainers to make appropriate remedial action plans and give appropriate careers advice.

We found no significant relationship between age group, gender and the likelihood of a trainee being awarded an unsatisfactory ARCP outcome. Although this seems, at face value, to contradict the conclusions of the recent General Medical Council report *How Doctors Progress Through Key Milestones*, this report was descriptive and did not include any formal statistical analysis.[20] After adjusting for several potential confounders, Pyne and Ben-Shlomo found that older medical graduates from all specialties were more likely to receive an unsatisfactory ARCP outcome.[5] However, unlike our study, these authors included doctors in temporary posts (LAT and FTSTA), as well as those at the start of their training (years 1 and 2) and those in higher specialty training (years 3–8) without distinguishing between the two groups. Given that we found no significant association between age at graduation and higher surgical training ARCP outcomes in our cohort, it may be that age is more important in predicting ARCP outcomes earlier in training. This suggestion merits further investigation.

Ethnic minority doctors have been frequently found to be more likely to underperform in UK medical postgraduate examinations, irrespective of whether they graduated in the UK or overseas, compared with white doctors, and our results confirm that this relationship continues in to higher surgical training.[4,5,10,13] These observations support the growing evidence that a true differential attainment exists between different ethnic groups throughout postgraduate medical training in the UK.

One of the major strengths of this study was the size of the study sample. We included all UK graduates who passed both parts of the MRCS and were therefore eligible, in principle, to apply for higher specialty surgical training in the UK. Unlike in some previous studies,[5,13] trainees in temporary posts (LAT and FTSTA) were excluded as they do not represent standard training in the UK. These previous studies also categorised ARCP outcome 5 (insufficient data) as an example of an unsatisfactory ARCP outcome or excluded the group altogether from analysis. We included ARCP outcome 5 as a separate group allowing for a more meaningful interpretation of the results.

The observational nature of the current study meant that we could not control for the effects of unmeasured variables that were not captured in the dataset. Another potential issue is that little is known about the reliability of the ARCP process. However, the ARCP panel considers a wide range of evidence collected by the trainee each year, and this is the closest measurable entity of 'on-the-job' performance currently available. Data on first language and ethnicity were less complete than data on demographic

factors such as gender, age and date of birth. However, the rate of missing data in our study was similar to that quoted in other educational studies (around 20%).[21]

Conclusions

This study provides further evidence in favour of the predictive validity of the MRCS examination and, for the first time, identifies the characteristics of individuals at risk of a less than satisfactory ARCP outcome during higher surgical training in the UK. Trainees who fail Part B MRCS at first attempt are more likely to achieve an unsatisfactory ARCP outcome during higher surgical training compared with those who pass at first attempt. This information may help to identify doctors who will require additional support through their training and may help to guide ARCP panels when considering borderline trainees. Understanding the relationship between ethnicity and differential attainment in postgraduate medical education and training remains a significant but important challenge facing the medical profession.

Acknowledgements

The authors acknowledge I Targett from the Royal College of Surgeons of England and C Santos from the Joint Committee on Surgical Training and Intercollegiate Surgical Curriculum Programme for help with data collection, and both L Smith and G Ayre from the Intercollegiate Committee for Basic Surgical Examinations for their support during this project. This study was funded by the Royal Colleges of Surgeons of England, Ireland and Edinburgh, and the Royal College of Physicians and Surgeons of Glasgow.

References

1. Scrimgeour DSG, Cleland J, Lee AJ *et al*. Impact of performance in a mandatory postgraduate surgical examination on selection into specialty training. *BJS Open* 2017; **1**: 1–8.
2. Conference of Postgraduate Medical Deans. *A Reference Guide for Postgraduate Specialty Training in the UK: the Gold Guide*. 6th ed. London: COPMED; 2016.
3. Brennan PA, Sherman KP. Editorial: the MRCS examination e an update on the latest facts and figures. *Br J Oral Maxillofac Surg* 2014; **52**: 881–883.
4. Scrimgeour DSG, Cleland J, Lee AJ, Brennan PA. Which factors predict success in the mandatory UK postgraduate surgical exam: the Intercollegiate Membership of the Royal College of Surgeons (MRCS)? *Surgeon* 2018; **16(4)**: 220–226.
5. Pyne Y, Ben-Shlomo Y. Older doctors and progression through specialty training in the UK: a cohort analysis of General Medical Council data. *BMJ Open* 2015; **5**: 1–5.
6. Wenghofer E, Klass D, Abrahamowicz M *et al*. Doctor scores on national qualifying examinations predict quality of care in future practice. *Med Educ* 2009; **43**: 1,166–1,173.
7. de Virgilio C, Yaghoubian A, Kaji A *et al*. Predicting performance on the American Board of Surgery qualifying and certifying examinations: a multi-institutional study. *Arch Surg* 2010; **145**: 852–856.
8. Spurlock DR, Holden C, Hartranft T. Using United States Medical Licensing Examination (USMLE) examination results to predict later in-training examination performance among general surgery residents. *J Surg Educ* 2010; **67**: 452–456.
9. Levy JB, Mohanaruban A, Smith D. The relationship between performance in work place based assessments, MRCP exam and outcome from core medical training. *Med Educ* 2011; **45(Suppl 2)**: 5.
10. Wakeford R, Denney M, Ludka-Stempien K *et al*. Cross-comparison of MRCGP & MRCP(UK) in a database linkage study of 2284 candidates taking both examinations: assessment of validity and differential performance by ethnicity. *BMC Med Educ* 2015; **15**: 1–12.
11. Tamblyn R, Abrahamowicz M, Dauphinee WD *et al*. Association between licensure examination scores and practice in primary care. *JAMA* 2002; **288**: 3,019–3,026.
12. Dewhurst NG, McManus C, Mollon J *et al*. 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: 1–8.
13. Tiffin PA, Illing J, Kasim AS, McLachlan JC. Annual Review of Competence Progression (ARCP) performance of doctors who passed Professional and Linguistic Assessments Board (PLAB) tests compared with UK medical graduates: national data linkage study. *BMJ* 2014; **348**: g2622.
14. Gauer JL, Wolff JM, Jackson JB. Do MCAT scores predict USMLE scores? An analysis on 5 years of medical student data. *Med Educ Online* 2016; **21**: 1–7.
15. McManus I, Ludka K. Resitting a high-stakes postgraduate medical examination on multiple occasions: nonlinear multilevel modelling of performance in the MRCP (UK) examinations. *BMC Med* 2012; **10**: 1–19.
16. Myles T, Galvez-Myles R. USMLE Step 1 and 2 scores correlate with family medicine clinical and examination scores. *Fam Med* 2003; **7**: 510–513.
17. Biskobing DM, Lawson SR, Messmer JM, Hoban JD. Study of selected outcomes of medical students who fail USMLE Step 1. *Med Educ Online* 2006; **11**: 4,589.
18. Ogunyemi D, Taylor-Harris DS. Factors that correlate with the US medical licensure examination Step-2 scores in a diverse medical student population. *J Natl Med Assoc* 2005; **97**: 1,258–1,262.
19. McDougle L, Mavis BE, Jeffe DB *et al*. Academic and professional career outcomes of medical school graduates who failed USMLE Step 1 on the first attempt. *Adv Health Sci Educ Theory Pract* 2013; **18**: 279–289.
20. General Medical Council. How do Doctors Progress Through Key Milestones During Training? London: GMC; 2016. https://www.gmc-uk.org/education/standards-guidance-and-curricula/projects/differential-attainment/research (cited August 2018).
21. Enders CK. Using the exception maximization algorithm to estimate coefficient alpha for scales with item-level missing data. *Psychol Methods* 2003; **8**: 322–337.