
Article:
Smith, Daniel and Tiffin, Paul Alexander orcid.org/0000-0003-1770-5034 (2018) Evaluating the validity of the selection measures used for the UK's Foundation medical training programme:a national cohort study. BMJ Open. pp. 1-10. ISSN: 2044-6055

https://doi.org/10.1136/bmjopen-2018-021918

BMJ Open

Evaluating the validity of the selection measures used for the UK's Foundation medical training programme: a national cohort study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-021918.R1
Article Type:	Research
Date Submitted by the Author:	23-Apr-2018
Complete List of Authors:	Smith, Daniel; General Medical Council, Education and Standards Tiffin, Paul; University of York, Health Sciences
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Health informatics
Keywords:	MEDICAL EDUCATION & TRAINING, Foundation programme, Selection

Evaluating the validity of the selection measures used for the UK's Foundation medical training programme: a national cohort study

Daniel Smith (corresponding author): Education Data Development Manager, the General Medical Council, 350 Euston Road, London, NW1 3JN, UK.
Email: daniel.smith@gmc-uk.org
020 7189 5373

Dr Paul A Tiffin: Reader in Psychometric Epidemiology and NIHR Career Development Fellow, Health Sciences, University of York, YO10 5DD, UK.
Email: paul.tiffin@york.ac.uk

Abstract

Objectives

Currently relative performance at medical school (Educational Performance Measure - the EPM decile), additional educational achievements and the score on a Situational Judgement Test (SJT) are used to rank applicants to the UK Foundation Years postgraduate medical training programme. We sought to evaluate whether these three measures were predictive of subsequent successful completion of the programme, and thus were valid selection criteria.

Methods

Data were obtained from the UK Medical Education Database (UKMED) on 14,131 UK applicants to the foundation programme starting in 2013 and 2014. These data included training outcomes in the form of Annual Reviews of Competency Progression (ARCPs) which indicated whether the programme was successfully completed. The relationship between applicants' performance on the three selection measures to the odds of successful programme completion were modelled.

Results

On univariable analyses all three measures were associated with the odds of successful completion of the programme. Converting the SJT score to deciles to compare the effect sizes suggested that one decile increase in the EPM increased the odds of completing the programme by approximately 15% whereas the equivalent value was 8% for the SJT scores. On multivariable analyses (with all three measures included in the model) these effects were only independently and statistically significant for EPM decile (OR 1.14, 95% CI 1.10 to 1.18, p<0.001) and SJT z-score decile (OR 1.05, 1.01 to 1.09, p=0.02).

Conclusions

The EPM decile and SJT scores may be effective selection measures for the Foundation Programme. However, Educational Achievements does not add value to the other two measures when predicting programme completion. Thus, its usefulness in this context is less clear. Moreover, our findings suggest that the weighting for the EPM decile score, relative to SJT performance, should be increased.

Strengths and limitations of this study

- The study covers two complete cohorts of UK applicants to foundation training with ARCP outcomes for the first two years of their foundation training. Furthermore there is no 'range restriction' (i.e. outcomes can only be observed in selected candidates) as nearly all UK applicants get a place on the foundation training programme.
- There is very little variation captured in ARCP outcomes: most trainees complete the foundation programme. Thus, the measure, in this context, will only yield information on trainees at the lower end of the performance range.
- More sensitive criteria for assessing the validity of the selection measures would be useful. The study did not have access to more granular data on 'Doctors in Difficulty' captured locally by Foundation Schools and some of the doctors will have been assisted in order to complete the programme. Similarly additional information may be present in the end of placements reports completed by trainees' clinical supervisors in the e-portfolios each trainee is required to maintain.

Introduction

In the UK, the Foundation Programme is a two-year generic training programme which forms the bridge between medical school (four to six years depending on the school) and specialist medical/general practice training. It is equivalent to the internships in the US.[1]

In order to standardise recruitment to this stage of training the *Improving Selection to the Foundation Programme* (ISFP) project developed methods for selection into the Foundation Programme. These were the Situational Judgement Test (SJT) and the Educational Performance Measure (EPM). SJTs present a series of scenarios that depict hypothetical workplace inter-personal dilemmas. In the SJT format used in the Foundation Programme selection candidates must rank a number of possible behavioural responses, in order of appropriateness and/or effectiveness. A score is generated based on the similarity of the ordering to those previously agreed on by a subject matter expert panel. The content domains of the SJT are labelled as follows: *Coping with pressure*, *Working effectively as part of a team*, *Effective communication*, *Problem solving*, and *Commitment to professionalism*.[2 3,4]

The SJT scores have been in use for selection to the Foundation Programme since 2013. Such SJT scores have been shown to have predictive validity for subsequent workplace performance across a range of occupations.[5 6] However they have only been used in medical selection in more recent years and there are fewer validity studies in this context. Previously a study examining the validity of the SJT for selection into General Practice (GP) training reported that the scores accounted for 6% of the variation in end of GP training assessments. However, it is not clear if this finding relates to the knowledge (AKT) or the clinical (CSA) assessments.[7] A separate study examined recruitment to Core Medical Training (CMT).[8] The authors concluded that SJT performance was the best single predictor of interview scores. However the study did not use outcomes from medical training such as Membership of the Royal Colleges of Physicians of the United Kingdom (MRCP) exam results to explore the predictive validity of the SJT for selection into CMT programmes. The authors note that the study sample comprised only a subset of applicants – those who applied to both CMT and GP. Thus this sample may not have been representative of the CMT applicant population as a whole. A follow-up study of this same cohort of CMT trainees reported on the relationship between the selection methods used and subsequent performance at postgraduate membership exams.[9] The authors reported that the clinical problem solving test (CPST) and SJT scores explained an additional 6.3 to 21.6% of the variance in performance on the membership exams, after accounting for the variance predicted by the existing CMT selection methods.

In relation to the SJT for selection into Foundation Training; a report on the initial validation study, produced by Work Psychology Group (the company that developed the SJT) has been published.[10 11] The authors gathered data on the performance during the first year of the programme ('F1') for a sample of 391 F1 doctors across five foundation schools that entered foundation training in 2013 using a bespoke questionnaire, matched to the domains measured at application, completed by supervisors. Their sample specifically targeted doctors in the first year of the Foundation Programme who had received particularly high or particularly low SJT scores. The relationship between application scores (SJT and EPM) and the performance scores from the bespoke questionnaire and ARCP were analysed. They found that higher

SJT scores and higher EPM were associated with higher ratings of F1 performance on the questionnaires. They found the correlations were different when their sample was split into high and low SJT scorers. From this analysis they concluded that the EPM had stronger correlation with performance for the high scoring SJT group and the SJT a stronger correlation for the low scoring SJT group. This observation could be explained by the contrasting psychometric properties of the two measures: SJTs tend to yield maximal information at the lower range of ability (that is, most candidates find the items relatively easy[12]whilst EPM is likely to have been able to discriminate between average to high performing candidates (as in effect, the measure is based on rankings with peers in medical school). They found no differences on EPM or SJT scores for those who received unsatisfactory ARCP outcomes compared to those that received satisfactory outcomes. However they note this was likely to be due to their small sample size. The authors made the following recommendation: "*that further studies are undertaken to explore the relationship between performance at application and performance outcomes beyond F1 (for example at the end of F2 and into specialty training) and that application scores (particularly SJT scores) spanning the full range of scores are targeted. If the relationship between application scores and ARCP outcomes is to be examined further, a large population (ideally all schools) should be targeted, as incidences of unsatisfactory ARCP outcomes appear to be very rare (1.1% in the present sample)*."

The Education Performance Measure (EPM) has two elements which this study considers as separate variables: EPM decile score and Educational Achievements.

The EPM Decile score

Students in the graduating cohort are ranked on their medical school performance. Schools were free to decide which assessments to include, provided they met the following criteria:

- Summative (and hence subject to formal controls)
- Cover clinical knowledge, skills and performance
- Cover non-clinical performance
- Cover all aspects of the curriculum assessed up to the end of the penultimate year at medical school
- Represent the average performance of the applicants over time, rather than being limited to a snap-shot
- Include written and practical forms of assessment

Schools were required to consult with students and publish on their website which assessments they included in the score.[3]

Educational Achievements

These are scored by considering additional degrees (maximum of 5 points, scored 5 points for a PhD through to 1 point for a 2.2 class intercalated

degree which does not extend the degree programme) and publications (maximum 2 points, 1 point per publication).

There are no published studies on the predictive validity of the EPM decile scores for selection into the foundation years. However, Simon et al (2015) report on the relationship between EPM decile scores and SJT scores, but the data were obtained from trainees in a self-reported survey rather than directly from the UK Foundation Programme Office (UKFPO). [13] Their survey achieved a response rate of only 8% (N= 3,175 from 12 medical schools), so their results (showing no observed association between EPM decile and SJT score) are likely to have been subject to validity-threatening response bias. It is not clear why this approach to the study was taken rather than obtaining the data directly from the UKFPO.

These three selection measures are combined into an overall score that is used to rank applicants to the Foundation programme, with the EPM decile score and Educational Achievements combined into the EPM score and given equal weighting to the SJT score.[14] Each applicant is allocated in rank order to their highest preference Foundation School (a conceptual grouping of medical schools, deaneries and organisations delivering healthcare - Trusts or Boards depending on the Country within the UK) where a place is available. Highly ranked applicants are more likely to be placed in their first preference Foundation School.

Following entry to the Foundation Programme each medical trainee's progress is reviewed at an Annual Review of Competency Progression (ARCP). The rating at ARCP is based on a portfolio of evidence collated in the e-portfolio including reviews from supervisors. ARCPs were introduced into foundation training in 2012.[15]

This approach to Foundation Programme selection, and in particular the equal weighting given to the EPM and SJT scores, has attracted criticism. In particular, Naim et al note that an applicant could jeopardise five to six years of hard work at medical school through underperformance on a single, two-hour test. [16] With the advent of UKMED it is now possible to link information from the various databases used to administer medical education. [17] Thus, there was an opportunity to assess the extent to which the selection scores predicted foundation ARCP outcomes for two cohorts of applicants.

This study is thus important in adding to the growing body of evidence in relation to the use of SJTs in medical selection. To the authors' knowledge it is the first investigation into whether performance on this measure predicts successful completion of an early postgraduate medical training programme. Moreover, developing a better understanding of how SJT scores should be weighted within the selection process, in relation to academic achievement, is crucial. This need is especially pressing given the recent, rapid and international implementation of SJTs into medical selection over a range of career stages.

Methods

Study population

All applicants making their final application (i.e. the application that led to the applicant starting on the foundation programme) in 2013 and 2014 were eligible for inclusion in the analysis (N = 15,249). Figure 1 outlines which cases had enough data points for inclusion in the analysis that follows. There were data on 14,131 doctors that met the criteria for inclusion in the main analyses.

<INSERT Figure 1 >

The final sample available for analysis comprised 7,134 doctors who started their foundation training in 2013 (50.5%) and 6,997 who started in 2014 (49.5%). All doctors had attended a UK medical school, as at this point in its development UKMED did not include graduates from non-UK institutions. The mean age on the 1st August in the year the doctor started foundation training was 24.88, with a standard deviation of 2.73 (N = 14,131). Their characteristics are further described in table 1.

Table 1 Sample demographics

Demographic	Group	N	%
Sex	Man	6,258	44.29
	Woman	7,873	55.71
Ethnicity	White	9,595	67.90
	Asian or Asian British	3,122	22.09
	Black or Black British	371	2.63
	Mixed	552	3.91
	Other Ethnic Groups	432	3.06
	Missing	59	0.42
Nationality	Other nationalities	1,634	11.56
	British	12,497	88.44

Patient and Public Involvement

Neither patients nor the public were involved in this study. Medical students and trainees are represented on the UKMED Advisory Board which approves UKMED research projects.

Data management

The UKMED includes data from Foundation Programme applications obtained from the UKFPO's Application System and MSC Assessment. Educational achievement scores were capped at seven as per the guidance in the UKFPO Applicant Handbook.[11]

SJT scores are only equated across papers within the application year, so to allow scores from more than one year to be used together, the SJT raw score was converted to a z-score based on the operational statistics for the applicant's paper and year in the technical reports published by the Work Psychology Group on the Improving Selection to the Foundation Programme website.[3, 4] To aid interpretation of the odds ratios in the models used to predict successful completion, we converted this z-score into deciles based on their rank within the year of application to allow direct comparison with the EPM decile. Thus, both predictors were placed on a similar metric, though the locally derived nature of the EPM decile must be borne in mind.

The UKMED receives ARCP outcome data from the GMC's annual collection of ARCP outcomes from postgraduate training providers.[18 19] The outcome variable used was obtaining an 'outcome 6'.This is defined by the UKFPO as "Satisfactory completion of F2 – Recommendation for the award of the Foundation Achievement of Competence Document" at the end of their F2 year in 2015 (2013 starters) or 2016 (2014 starters).[14]

Medical schools were defined using the values in the Higher Education Statistics Agency (HESA) data. Cases were grouped as follows:
The medical school on entry and exit was used when the admitting school and graduating school were the same. Where the graduating medical school was not the same as the admitting medical school and the case was part of a group that formed a distinct cohort with more than 20 cases, such as students who started at St Andrews, Durham or Oxbridge but complete their clinical undergraduate years at another medical school we used first and last medical school combined. EPM decile scores are calculated for each medical school's graduating cohort.

We used the foundation school that awarded the outcome 6. In 12,188 (86.3%) cases the foundation school awarding the outcome 6 was the same as the foundation school awarding the ARCP outcome in the first year of the foundation programme. Some of the 14% of changes reflected renaming and merging of foundation schools. For example 74 cases started in Staffordshire Foundation School and received their outcome 6 from West Midlands North Foundation School. There were also 906 doctors who started at one of the following foundation schools- North Central, North East Thames or North West Thames Foundation School- but received their outcome 6 from North Thames Foundation Schools. HESA records disability for each academic year the student has a record.[20] For these analyses we used the disability value from the final year. The disability categories recorded by HESA were collapsed into 'no disability' versus 'disability'.

Age was calculated as age at the start of the foundation programme: age on 1st August 2013 or 1st August 2014, depending on the year or application.

Analysis

Data management and analysis was conducted using SPSS version 24 and MLwiN 2.32.[21] Multi-level logistic regression models with trainees cross-classified and nested within medical schools and foundation schools were fitted in MLwiN using the method described by Leckie.[22]

In addition to testing which of the three selection measures predicted completion of the Foundation Programme, additional analyses was performed to understand if any socio-demographic variables were related to foundation outcomes. This model was fitted using stepwise backwards elimination: at each step, a non-significant variable was eliminated from the model, until only significant variables remained at the p=0.05 level. The stepwise method used a reduced dataset where no missing data on the covariates existed, to ensure 'true nesting' whilst model building (N = 7,539). The following variables were not statistically significant when entered in the multivariable model: sex, ethnicity, nationality, school type, receipt of a UKCAT bursary, graduate on entry, parental education, IMD - quintile Socio-economic-classification, Young participation (POLAR3) quintile, Income Deprivation Affecting Children Index (IDACI) rank, Course type grouped, year commencing medical school and SJT paper number. Please see the *Supplementary material- Results from univariable analyses* - for more information on these variables.

Figures 2, 3, and 4 display the mean proportions with 95% confidence intervals. These were calculated using the method recommended by Newcombe and Altman.[23]

Results

13,788 (97.6%) of cases in the sample had successfully achieved the required competencies (ARCP 'outcome 6') by the end of their two year programme. The results from univariable logistic regressions (Table 2) show that all three measures, the SJT z-score (mean = 0.10, SD=0.86) EPM (mean = 5.62, SD=2.83) and Educational Achievements (mean=2.47, SD=2.01) predict obtaining an outcome 6. For each decile increase in the EPM an applicant achieves the odds of an 'outcome 6' increase by roughly 15%; for each decile increase in the SJT decile score an applicant achieves the odds of an outcome 6 increase by approximately 8%. For each additional point in their Educational Achievement score a foundation applicant's odds of an outcome 6 increase by roughly 7%.

Figures 2 and 3 illustrate the relationship between completion of the Foundation Programme and the SJT z-score and EPM deciles: for both measures those applicants achieving below the lowest decile have a reduced proportion of 'outcomes 6' than those above the top decile, but the relationship for the intermediate deciles is non-linear.

The distribution of the Educational Achievements score does not allow splitting into decile groups, so the relationship between the proportion of 'outcome 6s' and the score is presented separately on figure 4. The groups with scores of 0, 2, 3 and 7 have the same mean proportion of outcome 6.

<INSERT Figures 2 ,3 and 4 >

Table 2 Univariable relationships between each foundation selection measure and successful completion (ARCP 'outcome 6')

Predictor variable (n)	OR	95% Confidence intervals for OR	p
EPM - decile score (n=14,131)	1.152	1.107 to 1.198	<0.001
SJT Z- score decile (n=14,128)	1.076	1.036 to 1.118	<0.001
SJT Z - score (n=14,128)	1.294	1.152 to 1.452	<0.001
SJT Equated score (n=14,131)	1.008	1.005 to 1.012	<0.001
Educational Achievements (n=14,131)	1.065	1.010 to 1.124	0.021

The three selection scores intercorrelated to some degree: the highest correlation being between EPM decile and SJT Z score (spearman's rho=0.30, N 14,128) and the lowest between Educational Achievements and SJT Z score (rho=0.12, N = 14,131), with the correlation between EPM decile and Educational Achievements at rho = 0.28 (N=14,131).

In the first model (Model 1 - table 3) a two-level multivariable logistic regression was fitted to predict achievement of an 'outcome 6' with trainees nested within a cross-classification of medical schools and foundation schools. The EPM decile score and the SJT z-score were statistically significant predictors at the $p<0.05$ level. In terms of interpreting the results; for an increase of one EPM the odds of successful completion of the Foundation Programme increased by approximately 14%. Similarly, for an increase of one SD on the SJT z-score the odds of an outcome 6 increased by approximately 18%. The Education Achievement measure is not

independently and statistically significantly associated with successful completion when all three measures are included in a model.

Table 3 Results from multilevel multivariable logistic regression analyses predicting an 'outcome 6' with the foundation selection measures (N=14,128 applicants nested in medical schools and foundation schools (cross-classified).

	Variable	OR	95% C.I. for OR	p
Model 1	EPM - decile score	1.140	1.095 to 1.185	<0.001
	SJT z-score	1.183	1.008 to 1.259	0.011
	Educational Achievements	0.998	0.937 to 1.059	0.944
Model 2	EPM - decile score	1.141	1.098 to 1.184	<0.001
	SJT z-score decile	1.049	1.010 to 1.088	0.017
	Educational Achievements	0.995	0.932 to 1.058	0.867
Model 3	EPM - decile score	1.113	1.072 to 1.154	0.000
	SJT z-score	1.122	0.996 to 1.247	0.075
	Educational Achievements	1.061	0.998 to 1.123	0.068
	No disability	1.660	1.345 to 1.976	0.002
	Number of applications	0.544	0.154 to 0.934	0.002
	Age	0.904	0.873 to 0.935	0.000

In the second model (Model 2 - table 3) the SJT z-score was converted to the same metric as the EPM (decile ranks within year of application to foundation) to allow the adjusted odds ratios obtained for the two measures to be directly compared. As before, one decile increase in the EPM increases the odds of an outcome 6 by 14%; one decile increase in the rank of the SJT z-score independently increases the chance of an outcome 6 by 5% (p=0.02).

The variables that were statistically significant predictors at the p<0.05 level in a series of univariable analyses (See *Supplementary material - Results from univariable analyses*) were included in a third multivariable model (Model 3- table 3). As there were no missing data for the final variables included, it was possible to use all cases. Trainees with no disability recorded by HESA in their final year of medical school had an approximately 66% higher odds of an

outcome 6 compared to those with a disability recorded. Older trainees were less likely to be awarded an outcome 6: each additional year reduced the odds by approximately 9%. Trainees who had made additional applications were less likely to get an outcome 6, with each additional application reducing the odds by roughly 54%.

After controlling for age, disability, the number of applications made and the medical school the EPM decile score remained a significant predictor of obtaining an outcome 6 whilst SJT z-score did not.

Discussion

In this study we were able to show that the three main selection measures (SJT score, EPM and Educational Achievements) used in ranking applicants to the UK Foundation Programme were related to the odds of successful completion of this stage of training. However, the relationship between Educational Achievements and completion was not independent of the other two measures. In contrast, the SJT score appeared to offer some degree of incremental predictive validity over that provided by the EPM deciles, suggesting that it is capturing additional, and relevant, information on applicants as intended by its developers. As such, our findings add to the emerging evidence of the potential usefulness and validity of SJTs in medical selection across different career stages and for different clinical specialisations in relation to a number of educationally and clinically relevant outcomes. [24 25]

Our findings in relation to trainee age are in keeping with those reported by an earlier study by Pyne and Ben-Shlomo[26] who reported that older doctors in their sample of specialty trainees were more likely to have problems with progression at ARCP than their younger colleagues. The relationship between disability and ARCP has not previously been reported on: here we found that those trainees who had a disability (any category) as recorded by HESA were less likely to obtain an outcome 6.

Reflecting on our key findings it is perhaps unsurprising that Educational Achievements was not independent of the other measures. This metric would have been skewed by the relatively small number of students who had participated in research projects, often as part of intercalated degrees. Such students would also have been likely to have been ranked relatively highly, according to their medical school EPMs. It is also worth commenting that, at first glance, both the EPM deciles and the SJT scores appear equally predictive of completion of the foundation programme. However, when we attempt to place both measures, albeit be crudely, on the same scale (i.e. divided into deciles) it is clear that EPM deciles are more predictive of this outcome, compared to SJT scores, with a ratio of roughly 2:1. Of course, the situation is complicated by the fact that EPM deciles are locally derived measures, whilst SJTs are nationally standardised tests. There may be ways in future research that the EPMs can be adjusted to make them more nationally comparable using 'Peer-Competition Rescaling.'[27] Nevertheless, despite the local nature of EPM deciles this finding remains relevant to policy

as both SJT scores and EPM deciles are used in the national selection process. It is also known that SJTs are generally encountered as relatively easy tests by candidates, and therefore most of the information is available on those below the average level of performance.[12] This infers that the SJT scores are likely to be relatively poor at differentiating more highly performing candidates from each other. We noted that once the scores were divided into deciles, this would have inevitably led to some loss of information, and resulted in a reduced degree of statistical significance in the relationship between this predictor and the outcome of interest, with the p value reducing from p = 0.01 to p = 0.02 when entered into a multivariable model with the other selection measures. In contrast EPM deciles may be able to differentiate between both low and high performing candidates. This proposition is supported by our findings illustrated in figure 2: the centre portion of the graph is relatively flat with the slopes being steepest at the extreme ends of lower and highly EPM decile ranked applicants. We also noted that, in contrast to EPM, SJT scores were not independently and statistically significant predictors of outcome 6s once three other background variables were put into a multivariable model (disability, number of applications and age). However, as applicant ranking does not take into account these latter three factors this finding does not have direct importance to policy, though it does suggest that at least some of the variance in SJT scores may be associated with these variables, though the directionality of influence could not be established from these data. Nevertheless, it is known that some demographic factors (such as female sex and ethnicity) are associated with SJT performance in general [28] and therefore the use of such selection measures may have an indirect effect on the advantage (or disadvantage) experienced by certain subpopulations of applicant.

This was a relatively complete national dataset with few missing data. However, a number of limitations of the study are worth noting. Firstly, the UKMED phase 1 cohort does not include graduates from non-UK medical schools who apply to Foundation training. However, from 2015 onwards the non-UK students have been included in the data UKMED receives on foundation applications, permitting research into this area. Secondly, suboptimal ARCP outcomes are relatively rare in the Foundation Programme period of training. Unlike later periods of postgraduate medical training there are fewer categories of ARCP outcome, and therefore at Foundation stage they are a relatively information poor variable, picking up only cases where issues impacting performance were severe enough to prevent the doctor completing the programme. Moreover there were no data on which trainees received remedial support during the foundation programme were available for this study. Thus it is possible that a number of doctors who received outcome 6s had required, sometimes considerable, support in order to achieve this. To establish whether the SJT and EPM deciles predict which trainees required additional support from their foundation schools would require UKMED to have identifiable data on doctors in difficulty on the foundation programme. The UKFPO Annual reports give figures for the numbers of doctors in difficulty: in 2014 there were 186 F1 and 163 F2s from UKMED medical school monitored via foundation schools' doctors in difficulty (DiD) policies and processes. In 2015 there were 251 F1s and 210 F2s

monitored via these policies[29]. We do not know which ARCP outcomes these doctors were finally awarded. Similarly at the other end of the scale the outcome 6 category does not capture performance that exceeds that required to achieve the foundation competencies. Thus, in psychometric terms, the outcome would have generally yielded information on trainees at the lower end of performance. That is, the results may tell us relatively little about candidates who score relatively highly on the three selection measures.

More variance may also be present in the end of placements reports completed by trainees' clinical supervisors in the e-portfolio each trainee is required to maintain. A score derived by summing across the six to eight clinical supervisor reports that are completed over the course of the foundation programme may provide a more granular outcome measure. Such an approach may be considered closer to the criterion used by Patterson et al who employed supervisor ratings of trainee job performance at one year into training as an outcome to validate selection measures for entry to GP training against.[7]

Some variation in ARCP outcomes across foundation schools was noted. Moreover, EPM is a local rather than national measure. However, the use of a cross-classified multilevel model should have controlled for these potential clustering effects on the outcome of interest.

The present study relied on data that were already routinely collected and not specifically designed as a criterion to assess the validity of the selection measures. As Austin and Villanova note using a general overall performance construct such as successful completion of the programme may not allow adequate matching to the predictive measures and it may be more fruitful to match selection measures to particular criteria representing particular aspects of job performance.[30] Nevertheless trainees undertaking the foundation programme and those responsible for managing the programmes are likely to regard it as an important criterion even if fails to capture the full range of foundation doctors' performance. Other possible criteria include the Multi-Specialty Recruitment Assessments (MSRAs) used for selection into the next stage of training including GP training programmes.[31]

Our findings have clear indications for selection policy, into the Foundation Programme. In particular they provide evidence, to support the intuitive sense, expressed by some, that excessive weight may be given to the SJT scores within the ranking process[16] Rather our results would suggest a 2:1 weighting ratio between the EPM decile and the SJT score may be the optimum, if the aim is to rank and advantage candidates most likely to successfully complete the Foundation Programme. Policy wise, our findings provide less evidence to support using Educational Achievements in calculating applicant rankings for the Foundation Programme. If the aim is to select candidates most likely to complete this phase of training then our results suggest that Educational Achievements does not add any incremental value beyond the EPM deciles or SJT scores. However, we are aware we have used a somewhat circumscribed outcome- i.e. completion of the programme. It may be that the

selectors wish to also advantage applicants, according to their academic records, that are most likely to be educationally, and perhaps clinically successful in their longer term careers. If removal of the Educational Achievements and an increase in the weighting of the EPM deciles were to be considered then an exercise could be undertaken to model whether such a change would have a significant impact on the numbers of applicants whose rank changes to an extent that they would have been offered a different unit of application. Such weightings would be more in-line with research in selection outside of medicine, where measures of ability are more strongly predictive than personality-type measures. Schmidt and Hunter reviewed meta-analyses of selection methods: when predicting performance in job training programmes they report a mean correlation coefficient of 0.56 for cognitive ability; whereas for integrity and conscientiousness tests they report correlation coefficients of 0.38 and 0.30.[32] One might argue that The EPM measure will inevitably reflect cognitive ability whilst the SJT is more akin to an integrity or personality test.

Further research could focus on the extent to which these Foundation selection measures predict long-term success and career choices in medicine. Moreover there is an immediate intention to further explore and describe the types of disability reported by medical graduates in a forthcoming UKMED project: *UKMEDP54 Declared disability in the UKMED dataset 2002-2016 an exploratory descriptive analysis.*[33]

In conclusion, the continued use of the SJT in selection into the Foundation Programme is justified by these findings, though it may be that excessive weight is being placed on the score, relative to the other two measures. Moreover, the UKFPO could consider discontinuing the use of Educational Achievements for ranking applicants to the foundation programme, unless it is shown to be associated with more distal markers of success in medical training.

Acknowledgments

Members of the UKFPO Rules Group as at Thursday 24th November 2016 kindly commented on an earlier version of this paper which only used applicants to the 2013 Foundation Programme. The group advised re-running the study with another cohort of applicants.

Ben Griffith, Andy Knapton and Kirsty White at the GMC commented on the paper. Olga Sierocinska King at the MSC collated information on the EPM and commented on the paper. Peter Tang at the MSC commented on the paper.

Peter McNair and Ganesan Gurusamy contribute to the maintenance of the UKMED Database held by the GMC.

Source of data was the UK Medical Education Database ("UKMED") [project number] extract generated on Approved for publication on 3 November 2016. I am grateful to UKMED for the use of these data. However, UKMED bears no responsibility for their analysis or interpretation.
The data includes information derived from that collected by the Higher Education Statistics Agency Limited ("HESA") and provided to the GMC ("HESA Data"). Source: HESA Student Record [2002/2003 and 2012/2013] Copyright Higher Education Statistics Agency Limited. The Higher Education Statistics Agency Limited makes no warranty as to the accuracy of the HESA Data, cannot accept responsibility for any inferences or conclusions derived by third parties from data or other information supplied by it."

Contributors

DS wrote the first draft of the paper and completed the analysis
PAT contributed to the supervision of the analyses, edited and reviewed and critically appraised the content of the paper.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.
DS is employed by the GMC as a data analyst working on the UKMED project. The views expressed here are his views and not the views of the GMC.

PAT is supported in his research by an NIHR Career Development Fellowship. This paper presents independent research part-funded by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health." PAT also wishes to acknowledge funding from the Worldwide Universities Network (WUN) that supports an academic collaboration furthering research into selection into the professions.

Competing interests:

DS is employed by the GMC. PAT has previously received research funding from the ESRC, the EPSRC, the DH for England, the UKCAT Board, and the GMC. In addition, PAT has previously performed consultancy work on behalf of his employing University for the UKCAT Board and Work Psychology Group and has received travel and subsistence expenses for attendance at the UKCAT Research Group.

Ethics approval

The authors did not need to seek formal NHS ethical approval for this study as it was a secondary data analysis of existing data. UKMED has received a letter from Queen Marys University of London Ethics of Research Committee on behalf of all UK medical schools to confirm ethics exemption for projects using exclusively UKMED data.

Provenance and peer review: Not commissioned; externally peer reviewed.

Data sharing statement: researchers wishing to re-analyse the data used for this study can apply for access to the same data via UKMED.

References

[1] The UK Foundation Programme Office (2017) http://www.foundationprogramme.nhs.uk/pages/home/about-the-foundation-programme Accessed 19 November 2017.

[2] Medical Schools Council on behalf of the cross-stakeholder Project Group (May 2012). *Improving Selection to the Foundation Programme Final Report of the Parallel Recruitment Exercise*. Available at https://isfporguk.files.wordpress.com/2017/04/final-report-of-pre-full-appendices.pdf . Accessed 19 November 2017.

[3] Patterson F, Ashworth H, Murray, H, Empey, L and Aitkenhead A (May 2013)
Analysis of the Situational Judgement Test for Selection to the Foundation Programme 2013
Technical Report. Available at https://isfp.org.uk/fp-technical-reports/ . Accessed 19 November 2017.

[4]. Patterson F, Murray, H, Baron, H, Aitkenhead A and Flaxman, C. (June 2014) *Analysis of the Situational Judgement Test for Selection to the Foundation Programme 2014 Technical Report.* Available at https://isfp.org.uk/fp-technical-reports/ . Accessed 19 November 2017

[5] Patterson F, Zibarras L, Ashworth V. Situational judgement tests in medical education and training: Research, theory and practice: *AMEE Guide No 100. Med Teach*. 2016; 38 (1):3–17.
[6] McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. III. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60(1), 63-91.
[7] Patterson F., Lievens F, Kerrin M., Munro N, Irish B. . The predictive validity of selection for entry into postgraduate training in general practice: evidence from three longitudinal studies. *British Journal of General Practice* 2013 ; 63: 734 – 741
[8] Patterson F, Carr V, Zibarras L, et al (2009). New machine-marked tests for selection into core medical training: Evidence from two validation studies. *Clinical Medicine*; 9(5): 1–4
[9] Patterson F, Lopes S, Harding S, Vaux E, Berkin L, Black D. (2017) The predictive validity of a situational judgement test, a clinical problem solving test and the core medical training selection methods for performance in specialty training. *Clinical Medicine (Lond)*. 2017 Feb;17(1):13-17. doi: 10.7861/clinmedicine.17-1-13.
[10] Patterson, F., Kerrin ,M., Edwards, H, Ashworth, V and Baron, H. (February 2015). Validation of the F1 Selection Tools. Available here: http://www.foundationprogramme.nhs.uk/download.asp?file=Validation_of_the_F1_selection_tools_report_FINAL_for_publication.pdf . Accessed 19 November 2017.
[11] Cousans F, Patterson F, Edwards H, Walker K, McLachlan J, Good (2017). Evaluating the complementary roles of an SJT and academic assessment for entry into clinical practice.
Adv Health Sci Educ Theory Pract. 2017 May;22(2):401-413. doi: 10.1007/s10459-017-9755-4. Epub 2017 Feb 8.)
[12] Tiffin, P and Carter, M. (May 2015) Understanding the measurement model of the UKCAT Situational Judgment Test: Summary Report. Available at https://www.ukcat.ac.uk/media/1183/understanding-the-measurement-model-of-the-ukcat-sjt.pdf. Accessed 17 April 2018.
[13] Simon E1, Walsh K, Paterson-Brown F, Cahill D. (Feb 2015). Does a high ranking mean success in the Situational Judgement Test? *Clinical Teacher*. 2015 Feb;12(1):42-5
[14] UK Foundation Programme (June 2017) *UKFP 2018 Applicants' Handbook.* Available at: http://www.foundationprogramme.nhs.uk/download.asp?file=UKFP_2018_Applicants_Handbook_-_FINAL.pdf . Accessed 19 November 2017.
[15] The UK Foundation Programme Office (2012). The UK Foundation Programme Reference Guide. Available at: http://www.foundationprogramme.nhs.uk/download.asp?file=FP_REFERENCE_GUIDE_2012_(updated_for_2014)_with_updated_links_April_2015.pdf Accessed 19 November 2017
[16] Najim M, Rabee R, Sherwani Y, Ahmed M, Ashraf M, Al-Jibury O, Najim R. The situational judgement test: a student's worst nightmare. *Adv Med Educ Pract.* 2015 Oct 5;6:577-8. doi: 10.2147/AMEP.S95023.
[17] Dowell J, Cleland J, Fitzpatrick S, McManus C; Nicholson S; Oppé T; Petty-Saphon K, Sierocinska-King O; Smith D, Thornton S, White K. The UK

Medical Education Database (UKMED) What is it? Why and how might you use it? *BMC Medical Education (*2018) 18:6
DOI 10.1186/s12909-017-1115-9

[18] GMC (4 July 2016) *Collection of ARCP and RITA outcomes awarded in 2015/2016 – Briefing Note 1*. Available at: http://www.gmc-k.org/ARCP2016BriefingNote1_Final_20160703.pdf_66790612.pdf accessed 27/10/2017

[19] GMC (3 July 2015) ARCP and RITA Data Collection 2015 – Briefing Note 1. Available at: https://www.gmc-uk.org/ARCP2015BriefingNote1_Final03072015.pdf_61654575.pdf accessed 27/10/2017

[20] HESA Student Record (2017) Disability. Available at: https://www.hesa.ac.uk/collection/c16051/a/disable Accessed 7 January 2018.

[21] MLwiN: A software package for fitting multilevel models. Available at: http://www.bristol.ac.uk/cmm/software/mlwin/ Accessed 22 November 2017.

[22] Leckie, G. (2013). Cross-Classified Multilevel Models. LEMMA VLE Module 12, 1-60. (http://www.bristol.ac.uk/cmm/learning/course.html).

[23] Newcombe R.G. and Altman D. G. Proportions and their differences in Altman D G, Machin D, Bryant T N and Gardner M J (1989) *Statistics with Confidence 2nd Edition*. BMJ Books.

[24] Patterson F1, Cousans F, Edwards H, Rosselli A, Nicholson S, Wright B. The Predictive Validity of a Text-Based Situational Judgment Test in Undergraduate Medical and Dental School Admissions. *Acad Medicine*. 2017 Sep;92(9):1250-1253. doi: 10.1097/ACM.0000000000001630.

[25] Patterson F, Rowett, E, Hale R, Grant M, Roberts C, Cousans F, Martin S. The predictive validity of a situational judgement test and multiple-mini interview for entry into postgraduate training in Australia*. BMC Med Educ*. 2016; 16: 87. doi: 10.1186/s12909-016-0606-4.

[26] 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. doi:10.1136/bmjopen-2014-005658

[27] Tiffin, P A and Paton L W (March 2017) *Exploring the validity of the 2013 UKCAT SJT- prediction of undergraduate performance in the first year of medical school: Summary Version of Report.* Available at: https://www.ukcat.ac.uk/media/1119/exploring-the-validity-of-the-2013-ukcat-sjt-prediction-of-ug-performance-in-1st-yr-of-med-school-summary-version-posted-27032017.pdf. Accessed 14 January 2018.

[28] Whetzel , D L McDaniel M A and Nguyen N T. Subgroup Differences in Situational Judgment Test Performance: A Meta-Analysis *Human Performance*, 21:291–309, 2008. Human Performance, 21:291–309, 2008

[29] UKFPO reports available at http://www.foundationprogramme.nhs.uk/pages/resource-bank/general. Accessed 23 November 2017.

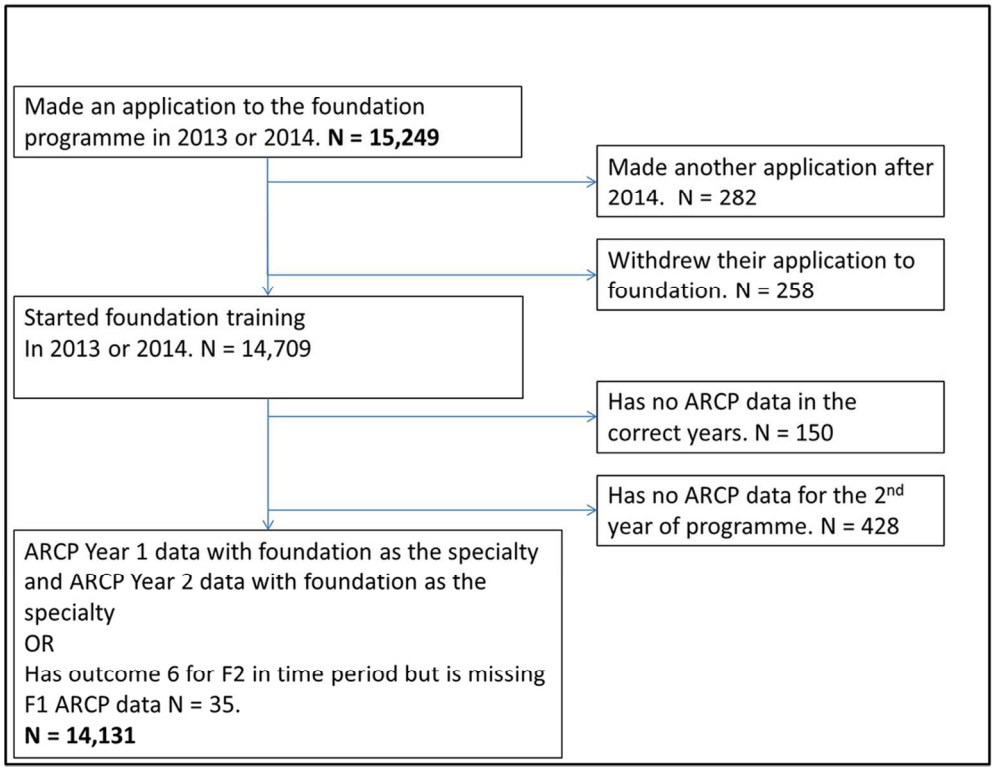
[30] Austin, J. T., & Villanova, P. (1992). The criterion problem: 1917–1992. *Journal of Applied Psychology*, 77(6), 836-874.

[31] Work Psychology Group (22/08/2017) *Multi-Specialty Recruitment Assessment Test Blueprint & Information.* Available at:

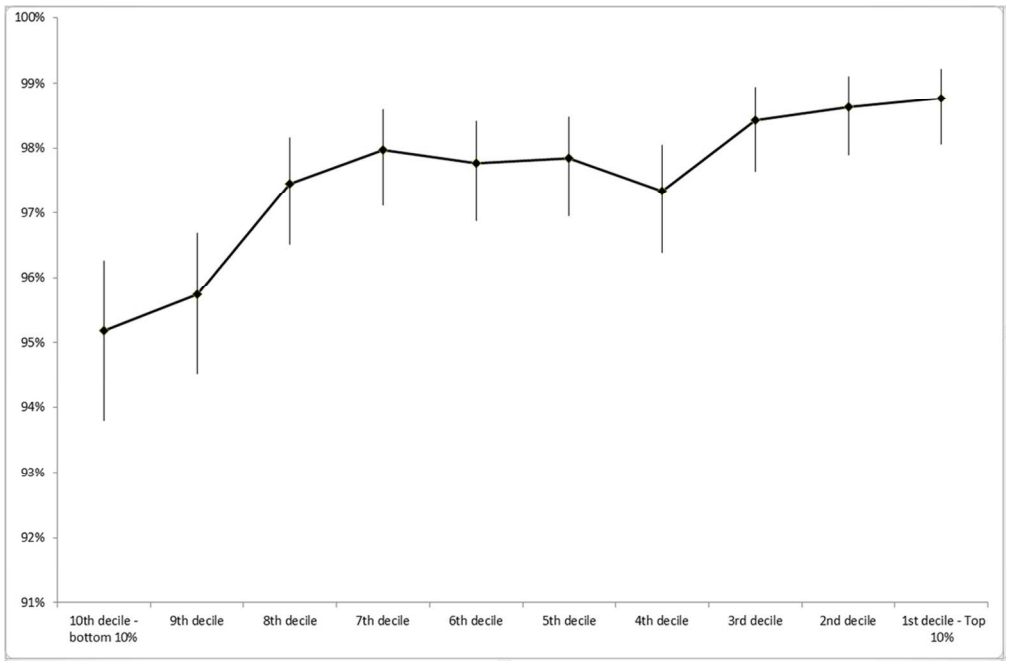
https://gprecruitment.hee.nhs.uk/LinkClick.aspx?fileticket=8pprMZn6aPE%3d&portalid=8 Accessed 3 April 2018.
[32] Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.
[33] UKMED Accepted Applications. Available at: https://www.ukmed.ac.uk/accepted_applications. Accessed 7 January 2018.

Figure Legends

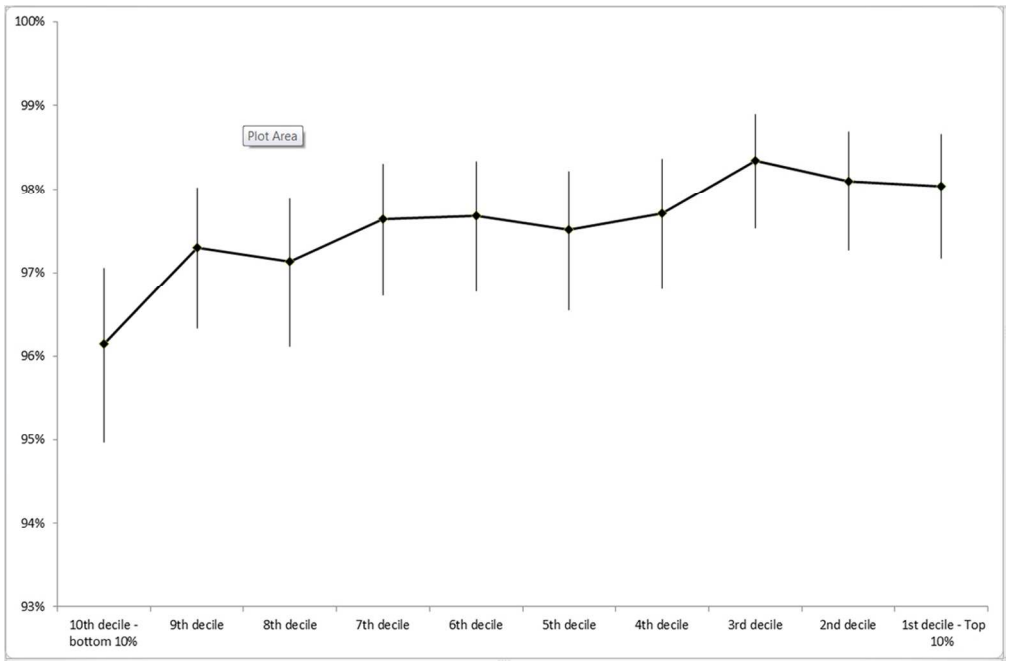
Figure 1: Data flow through study
Figure 2: Percent Awarded an outcome 6 by EPM deciles (N = 14,131)
Figure 3: Percent awarded an outcome 6 by SJT z-score deciles (N = 14,128)
Figure 4: Percent Awarded an outcome 6 by Educational Achievements Score (N = 14,131)



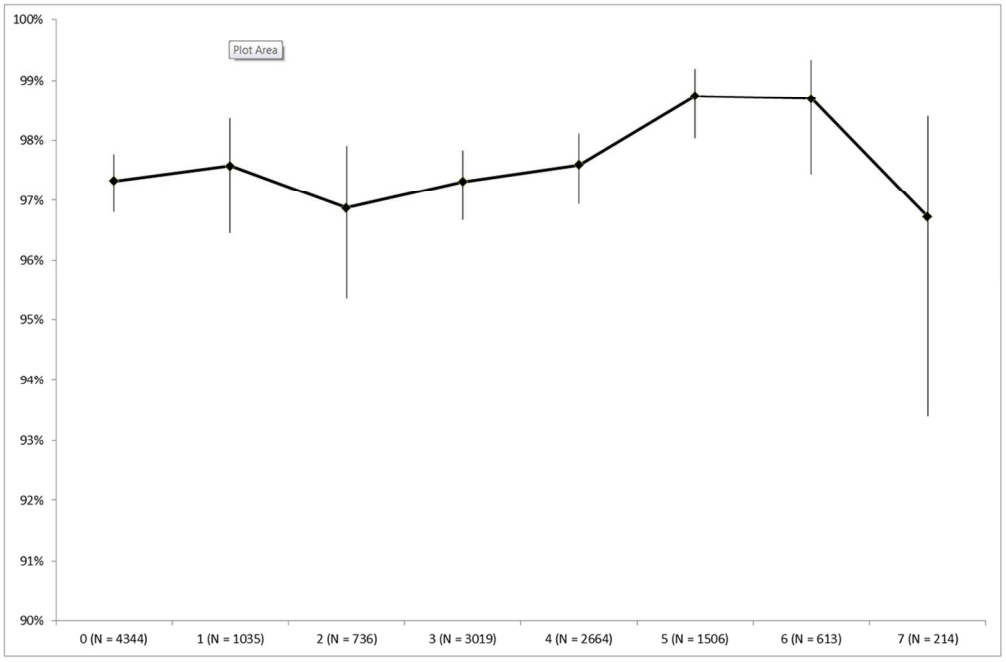
99x76mm (300 x 300 DPI)



99x65mm (300 x 300 DPI)



99x65mm (300 x 300 DPI)



109x71mm (300 x 300 DPI)

Evaluating the validity of the selection measures used for the UK's Foundation medical training programme: a national cohort study - Supplementary material

Table 1 Variable definitions and sources

Note: not all values held on the GMC register are available on the public facing version.

Variable	Source	Definition
Sex	GMC Register[1]	
Age at start of Foundation Programme	GMC Register	Calculated at 1st August 2013 or 2014 depending on year of entry to the Foundation Programme.
Ethnicity - higher level ONS groups	GMC Register	GMC Register uses the ONS Ethnicity Groupings[2].
Disability recorded in final year by HESA	HESA Student Record[3]	HESA Disability value for the final year. The disability categories recorded by HESA[4] were collapsed into 'no disability' versus 'disability'.
Nationality	GMC Register	Grouped into 'All other nationalities' and 'British'.
School Type	HESA Student Record	Taken from HESA State School Marker, which groups the last provider attended into: Privately funded school From state-funded school or college Unknown school type.
Bursary	UKCAT data	Student had applied for a bursary for UKCAT test costs.[5]
Graduate	HESA Student Record	Based on HESA_QUALENT. The highest qualification on entry. The following values are defined as graduate entry 'First degree of UK institution', 'First degree with honours leading to Qualified Teacher Status (QTS)/registration with a General Teaching Council (GTC)', 'Higher degree of UK institution', 'Non-UK doctorate degree', 'Non-UK first degree', 'Non-UK masters degree', 'PGCE with QTS/GTC registration', 'PGCE without QTS/GTC registration', 'Postgraduate diploma or certificate, excluding PGCE', 'UK doctorate degree',

Variable	Source	Definition
		'UK first degree with honours', 'UK masters degree', 'UK ordinary (non-honours) first degree', 'Postgraduate Certificate in Education or Professional Graduate Diploma in Education', 'Graduate of other overseas institution', 'Professional Graduate Certificate in Education', 'Integrated undergraduate/postgraduate taught masters degree on the enhanced/extended pattern', 'Graduate of EU institution'
Parental Education	HESA Student Record	Taken from[6]: The following question is about your parents' level of education. This includes natural parents, adoptive parents, step-parents or guardians who have brought you up. Do any of your parents (as defined above) have any higher education qualifications, such as a degree, diploma or certificate of higher education? Yes No Don't know Information refused
IMD - Quintile	HESA Student Record	Index of Multiple Deprivation. Based on the postcode on application to medical school.
Socio-economic-classification	HESA Student Record or UKCAT Test Registration form if missing from HESA record.	Taken from the HESA value for Occupation code. These are collected from this question: “If you are under 21, please give the occupation of your parent, step-parent or guardian who earns the most. If he or she is retired or unemployed, give their most recent occupation. If you are 21 or over, please give your own occupation." These occupations are then grouped using a coding frame developed by the Office of National Statistics.[7 8]
Young participation (POLAR3) quintile 1(low) - 5(high)	HESA Student Record Higher Education Funding Council for England (HEFCE)	Based on the postcode on application to medical school linked to HEFCE data: Young participation quintile 1(low)-5(high) and (unclassified) The young participation classification (POLAR3) is based the participation in HE of

Variable	Source	Definition
	data on POLAR[9]	young people who reached 18 between 2005-2009. .
Income Deprivation Affecting Children Index (IDACI) rank	HESA Student Record	Based on the postcode on application to medical school linked to Income Deprivation Affecting Children Index (IDACI) constructed by the Social Disadvantage Research Centre at the University of Oxford as part of the Indices of Deprivation 2007[10]. Rank of 1 is the most deprived, and 32482 the least deprived, on this overall measure.
Course type grouped	HESA Student Record	HESA CTITLE[11] AND HESA COURSE_ID[12] values were manually mapped to the course types described by the Medical School Council in their annual Entry Requirements publication[13]. These mapping were confirmed with the medical schools.
Year commencing medical school	HESA Student Record	Year extracted from the HESA Commencement Date[14].
N Foundation programme applications	Foundation Applications	Number of applications present in the application data from 2012 onwards.
SJT Paper number	MSC Assessment data linked to the foundation application data on application ID.	The version of the SJT sat by the applicant.
Medical school	HESA Student Record	Derived from the first and last medical attended. First and last medical schools are derived from the UKPRN[15] in the HESA Student Record.

Results from univariable analyses

HESA rounding rules have been applied to the Ns in these tables[16].

Table 2 Univariable analyses: predictor variables against demographic and application variables for cases with available data N = 14,131

Factor	Category	N doctors in sample	Mean SJT raw Z-score	SD SJT raw Z-score	Median EPM Decile 1 = 10th decile - bottom 10% 10 = 1st decile - Top 10%	Median Educational Achievement	N doctors in sample for EPM and EA	Mean SJT Z-score	Median EPM Decile	Median Educational Achievement
Sex	Man	6,255	-0.015	0.858	5.0	3	6,260	F= 216.261, P < 0.001	U = 22016438.0, P < 0.001	U = 23124514.5, P < 0.001
	Woman	7,870	0.197	0.841	6.0	3	7,875			
Age at start of Foundation Programme	Correlation	14,128	Pearson		Spearman		14,131	-0.064, P < 0.001,	0.028, P < 0.001	0.455, P < 0.001
Ethnicity - higher level ONS groups	White	9,590	0.236	0.799	6.0	3	9,595	F = 214.889 P < 0.001	H = 760.089, P < 0.001	H = 11.878, P = 0.018
	Asian or Asian British	3,120	-0.197	0.875	4.0	3	3,120			
	Black or Black British	370	-0.418	0.987	4.0	3	370			
	Mixed	550	0.109	0.863	5.0	3	550			
	Other Ethnic Groups	430	-0.209	0.980	4.0	3	430			

Factor	Category	N doctors in sample	Mean SJT raw Z-score	SD SJT raw Z-score	Median EPM Decile 1 = 10th decile - bottom 10% 10 = 1st decile - Top 10%	Median Educational Achievement	N doctors in sample for EPM and EA	Mean SJT Z-score	Median EPM Decile	Median Educational Achievement
Disability recorded in final year by HESA	Disability reported	1,305	0.022	0.901	5.0	3	1,305	F = 13.022 P < 0.001	U = 7236728.5 P< -0.001	U = 8180937.0 P = -0.170
	No disability	12,825	0.111	0.850	6.0	3	12,825			
Nationality	Other nationalities	1,635	-0.213	0.912	5	1	1,635	F = 256.850, P = 0.003	U = 8563853.50, P < 0.001	U = 8692711.00, P < 0.001
	British	12,495	0.144	0.839	6	3	12,490			
School Type	State-funded school of college	9,295	0.129	0.851	6.0	3	9,295	F = 1.015, P = 0.314	U = 15966780.0, P< -0.001	U = 16412200.5, P < 0.001
	Privately funded school	3,690	0.146	0.827	5.0	3	3,695			
Bursary	No record	13,775	0.108	0.856	6.0	3	13,775	F = 17.300 P < 0.001	U = 2253252.50 P = 0.014	U = 1823441.50, P< -0.001
	Record of UKCAT Bursary application	355	-0.084	0.810	5.0	1	355			
Graduate	Not Graduate on entry	11,530	0.109	0.849	5.0	2	11,530	F = 1.015, P = 0.314	U = 13105070.5, P< -0.001	U = 8606441.00, P< -0.001
	Graduate on entry	2,600	0.075	0.884	6.0	4	2,600			
Parental Education	No higher ed. qualifications	2,420	0.058	0.857	6.0	3	2,420	F = 24.933 P < 0.001	U = 9569120.0, P = 0.026	U = 9534798.50, P = 0.011
	Have higher ed. qualifications	8,140	0.155	0.835	6.0	3	8,145			
IMD - Quintile	1 – Least deprived	4,285	0.176	0.820	6.0	3	4,285	F= 39.111 P < 0.001	H = 66.534, P < 0.001	H = 4.961, P = 0.291
	2	2,840	0.182	0.817	6.0	3	2,840			
	3	2,210	0.142	0.836	6.0	3	2,210			

Factor	Category	N doctors in sample	Mean SJT raw Z-score	SD SJT raw Z-score	Median EPM Decile 1 = 10th decile - bottom 10% 10 = 1st decile - Top 10%	Median Educational Achievement	N doctors in sample for EPM and EA	Mean SJT Z-score	Median EPM Decile	Median Educational Achievement
	4	1,425	0.052	0.849	5.0	3	1,425			
	5	860	-0.183	0.976	5.0	3	860			
Socio-economic-classification	Semi-routine and routine occupations	1,195	0.045	0.862	6.0	3	1,195	F= 7.499, P < 0.001	H = 6.568, P = 0.161	H = 18.486, P = 0.001
	Lower supervisory and technical occupations	275	0.082	0.798	5.0	3	275			
	Small employers and own account workers	600	0.000	0.899	5.0	3	600			
	Intermediate occupations	1,310	0.136	0.842	6.0	3	1,310			
	Managerial and professional occupations	9,735	0.145	0.838	6.0	3	9,735			
Young participation (POLAR3) quintile 1(low) - 5(high)	1 - low	515	0.023	0.883	6.0	3	515	F= 6.336, P < 0.001	H = 7.647, P = 0.105	H = 22.671, P < 0.001
	2	1,115	0.081	0.879	6.0	3	1,115			
	3	1,905	0.101	0.874	5.0	3	1,905			
	4	3,010	0.138	0.845	6.0	3	3,010			
	5 - high	6,275	0.166	0.820	6.0	3	6,280			
Income Deprivation Affecting Children Index (IDACI) rank	Correlation	10,401	Spearman		Spearman		10,401	0.082, P < 0.001,	0.077, P < 0.001,	-0.021, P = 0.33
Course type grouped	Standard Entry Medicine	12,060	0.101	0.854	6.0	3	12,060	F = 10.997, P < 0.001	H = 112.659, P < 0.001	H = 688.658, P < 0.001

Factor	Category	N doctors in sample	Mean SJT raw Z-score	SD SJT raw Z-score	Median EPM Decile 1 = 10th decile - bottom 10% 10 = 1st decile - Top 10%	Median Educational Achievement	N doctors in sample for EPM and EA	Mean SJT Z-score	Median EPM Decile	Median Educational Achievement
	Gateway/Prelim/Foundation	645	-0.015	0.897	5.0	3	645			
	Graduate Entry Programme	1,425	0.174	0.843	7.0	4	1,425			
Year commencing medical school	2004 and earlier	90	-0.057	0.925	4.0	4	90	F= 18.347 P < 0.001	H = 315.087, P < 0.001	H = 1632.435 P < 0.001
	2005	115	-0.086	0.885	2.0	3	115			
	2006	385	-0.124	0.934	3.0	3	385			
	2007	3,150	0.176	0.848	6.0	4	3,150			
	2008	5,815	0.115	0.848	6.0	3	5,815			
	2009	3,680	0.071	0.844	6.0	1	3,680			
	2010	775	0.133	0.847	7.0	4	775			
	2011	95	-0.553	1.004	4.0	1	95			
	2012	20	-0.788	0.573	3.0	0	20			
N Foundation programme applications	1	13,770	0.112	0.851	6.0	3	13,775	F= 28.767 P < 0.001	H = 356.188 P < 0.001	H = 8.909 P = 0.012
	2	350	-0.234	0.930	2.0	2	350			
	3	5	-0.312	0.816	1.0	3	5			

Factor	Category	N doctors in sample	Mean SJT raw Z-score	SD SJT raw Z-score	Median EPM Decile 1 = 10th decile - bottom 10% 10 = 1st decile - Top 10%	Median Educational Achievement	N doctors in sample for EPM and EA	Mean SJT Z-score	Median EPM Decile	Median Educational Achievement
SJT Paper number	1	11,420	0.092	0.870	6.0	3	11,420	F = 5.896, P = 0.003	H = 0.003, P = 0.998	H = 11.595, P = 0.003
	2	2,690	0.150	0.787	6.0	3	2,690			
	3	15	0.397	0.890	5.5	4	15			
Medical school	Aberdeen	340	-0.033	0.851	6.0	1	340	F = 11.420, P < 0.001	H = 115.362, P < 0.001	H = 2691.699, P = 0.105
	Barts	615	0.027	0.833	6.0	3	615			
	Birmingham	760	0.053	0.845	6.0	1	760			
	Bradford_Leeds	70	-0.153	0.732	4.0	0.5	70			
	Brighton and Sussex	275	0.064	0.932	6.0	3	275			
	Bristol	440	0.229	0.704	6.0	3	440			
	Cambridge	290	0.380	0.860	6.0	4	290			
	Cambridge_Imperial	40	0.182	0.542	6.0	3	40			
	Cambridge_King's	50	0.230	0.942	5.0	3	50			
	Cambridge_Oxford	30	0.688	0.764	9.0	5	30			
	Cambridge_UCL	105	0.658	0.667	8.0	3	105			

Factor	Category	N doctors in sample	Mean SJT raw Z-score	SD SJT raw Z-score	Median EPM Decile 1 = 10th decile - bottom 10% 10 = 1st decile - Top 10%	Median Educational Achievement	N doctors in sample for EPM and EA	Mean SJT Z-score	Median EPM Decile	Median Educational Achievement
	Cardiff	545	0.133	0.892	6.0	1	545			
	Dundee	300	0.090	0.835	6.0	3	300			
	Durham_Newcastle	185	0.084	0.877	6.0	3	185			
	Edinburgh	440	0.369	0.811	6.0	2	440			
	Glasgow	450	0.115	0.843	5.0	1	450			
	Hull York	275	-0.018	0.921	6.0	1	275			
	Imperial	610	0.155	0.849	6.0	4	610			
	Keele	240	0.048	0.855	6.0	0	240			
	King's	705	0.107	0.892	6.0	3	705			
	Lancaster	85	0.024	0.775	5.0	0	85			
	Leeds	410	0.187	0.684	6.0	3	410			
	Leicester	435	-0.067	0.881	6.0	3	435			
	Liverpool	555	-0.132	0.911	6.0	0	555			
	Manchester	660	0.041	0.847	6.0	1	660			

Factor	Category	N doctors in sample	Mean SJT raw Z-score	SD SJT raw Z-score	Median EPM Decile 1 = 10th decile - bottom 10% 10 = 1st decile - Top 10%	Median Educational Achievement	N doctors in sample for EPM and EA	Mean SJT Z-score	Median EPM Decile	Median Educational Achievement
	Newcastle	485	0.252	0.810	6.0	1	485			
	Norwich	265	-0.140	0.884	6.0	0	265			
	Nottingham	600	0.004	0.892	6.0	3	600			
	Oxford	260	0.600	0.725	5.0	4	260			
	Oxford_Imperial	20	0.244	0.550	5.0	3	20			
	Peninsula	385	-0.130	0.863	5.0	0	385			
	Queen's	500	0.044	0.914	6.0	0	500			
	Sheffield	455	0.139	0.852	6.0	0	455			
	Southampton	455	-0.006	0.926	6.0	2	455			
	St Andrews_Edinburgh	40	0.101	0.869	6.0	4	40			
	St Andrews_Glasgow	50	-0.051	0.642	5.0	3	50			
	St Andrews_Manchester	145	-0.021	0.838	6.0	3	145			
	St George's	535	0.121	0.776	6.0	3	535			
	Swansea	65	-0.122	0.886	5.5	4	65			

Factor	Category	N doctors in sample	Mean SJT raw Z-score	SD SJT raw Z-score	Median EPM Decile 1 = 10th decile - bottom 10% 10 = 1st decile - Top 10%	Median Educational Achievement	N doctors in sample for EPM and EA	Mean SJT Z-score	Median EPM Decile	Median Educational Achievement
	Swansea_Cardiff	70	0.274	0.800	5.0	4	70			
	UCL	590	0.295	0.735	5.0	3.5	590			
	Warwick	305	0.157	0.824	6.0	4	305			

Table 3 Univariable analyses: outcome 6 against demographic and application variables for cases with available data N = 14,131

Factor	Category	N doctors in sample	% Outcome 6	Mean	Bivariate Association Outcome 6
Sex	Man	6,260	98.00%		F = 0.116 P = 0.733
	Woman	7,875	98.00%		
Age at start of Foundation Programme	No outcome 6	345		26.122	F = 71.979 P < 0.000
	Has outcome 6	13,790		24.858	
Ethnicity - higher level ONS groups	White	9,595	97.58%		F = 0.450 P = 0.772
	Asian or Asian British	3,120	97.63%		
	Black or Black British	370	97.84%		
	Mixed	550	96.74%		
	Other Ethnic Groups	430	97.69%		
Disability recorded in final year by HESA	Disability reported	1,305	95.56%		F = 24.741 P = 0.000
	No disability	12,825	97.78%		
Nationality	Other nationalities	1,635	96.7%		F = 6.009 P = 0.014
	British	12,495	97.7%		
School Type	State-funded school of college	9,295	97.34%		F = 5.523

Factor	Category	N doctors in sample	% Outcome 6	Mean	Bivariate Association Outcome 6
	Privately funded school	3,695	98.05%		P = 0.019
Bursary	No record	13,775	97.58%		F = 0.020 P = 0.887
	Record of UKCAT Bursary application	355	97.46%		
Graduate	Not Graduate on entry	11,530	97.76%		F = 9.542 P = 0.002
	Graduate on entry	2,600	96.73%		
Parental Education	No higher ed. qualifications	2,420	97.00%		F = 0.629 P = 0.428
	Have higher ed. qualifications	8,145	98.00%		
IMD - Quintile	1 – Least deprived	4,285	97.95%		F = 3.611 P = 0.006
	2	2,840	98.03%		
	3	2,210	96.92%		
	4	1,425	96.63%		
	5	860	97.32%		
Socio-economic-classification	Semi-routine and routine occupations	1,195	96.15%		F = 3.646 P = 0.006
	Lower supervisory and technical occupations	275	97.45%		
	Small employers and own account workers	600	97.17%		
	Intermediate occupations	1,310	97.25%		
	Managerial and professional occupations	9,735	97.85%		
Young participation (POLAR3) quintile 1(low) - 5(high)	1 - low	515	97.00%		F = 2.155 P = 0.071
	2	1115	97.00%		

Factor	Category	N doctors in sample	% Outcome 6	Mean	Bivariate Association Outcome 6
	3	1,905	97.00%		
	4	3,010	97.00%		
	5 - high	6,280	98.00%		
Income Deprivation Affecting Children Index (IDACI Rank)	No outcome 6	230		18581.632	F = 11.514 P = 0.001
	Has outcome 6	10,170		20640.535	
Course type grouped	Standard Entry Medicine	12,060	97.70%		F = 2.849 P = 0.058
	Gateway/Prelim/Foundation	645	97.20%		
	Graduate Entry Programme	1,425	96.70%		
Year commencing medical school	2004 and earlier	90	93.48%		F = 4.538 P = 0.000
	2005	115	92.24%		
	2006	385	94.82%		
	2007	3,150	97.75%		
	2008	5,815	97.75%		
	2009	3,680	97.80%		
	2010	775	97.04%		
	2011	95	97.92%		
	2012	20	100.00%		
N Foundation programme applications	1	13,775	97.70%		F = 19.114 P = 0.000
	2	350	92.84%		

Factor	Category	N doctors in sample	% Outcome 6	Mean	Bivariate Association Outcome 6
	3	5	85.71%		
SJT Paper number	1	11,420	97.56%		F = 0.706 P = 0.494
	2	2,690	97.66%		
	3	15	92.86%		
Medical school	Aberdeen	340	97.08%		F = 2.226 P = 0.000
	Barts	615	97.72%		
	Birmingham	760	98.69%		
	Bradford_Leeds	70	94.12%		
	Brighton and Sussex	275	97.10%		
	Bristol	440	98.87%		
	Cambridge	290	98.62%		
	Cambridge_Imperial	40	100.00%		
	Cambridge_King's	50	96.00%		
	Cambridge_Oxford	30	96.55%		
	Cambridge_UCL	105	100.00%		
	Cardiff	545	97.79%		
	Dundee	300	98.33%		
	Durham_Newcastle	185	96.74%		
	Edinburgh	440	98.41%		

Factor	Category	N doctors in sample	% Outcome 6	Mean	Bivariate Association Outcome 6
	Glasgow	450	97.10%		
	Hull York	275	98.18%		
	Imperial	610	95.74%		
	Keele	240	97.11%		
	King's	705	98.01%		
	Lancaster	85	96.39%		
	Leeds	410	98.05%		
	Leicester	435	96.77%		
	Liverpool	555	98.20%		
	Manchester	660	98.03%		
	Newcastle	485	98.35%		
	Norwich	265	97.36%		
	Nottingham	600	98.17%		
	Oxford	260	98.45%		
	Oxford_Imperial	20	95.45%		
	Peninsula	385	97.92%		
	Queen's	500	93.43%		
	Sheffield	455	97.37%		
	Southampton	455	96.92%		

Factor	Category	N doctors in sample	% Outcome 6	Mean	Bivariate Association Outcome 6
	St Andrews_Edinburgh	40	97.44%		
	St Andrews_Glasgow	50	97.92%		
	St Andrews_Manchester	145	94.52%		
	St George's	535	97.01%		
	Swansea	65	93.75%		
	Swansea_Cardiff	70	97.10%		
	UCL	590	99.15%		
	Warwick	305	97.72%		

References

[1] https://www.gmc-uk.org/doctors/register/LRMP.asp
[2]Office for National Statistics, (May 2015*). Harmonised Concepts and Questions for Social Data Sources Primary Principles Ethnic Group* Available at: https://gss.civilservice.gov.uk/wp-content/uploads/2016/03/P3-Ethnic-Group-June-16-1.pdf. Accessed 21 January 2018.
[3] HESA *Definitions: Students*. Available at: https://www.hesa.ac.uk/support/definitions/students Accessed 21 January 2018
[4] HESA Student Record (2017) Disability. Available at: https://www.hesa.ac.uk/collection/c16051/a/disable. Accessed 7 January 2018
[5] UKCAT Bursary Scheme. Available at: https://www.ukcat.ac.uk/ukcat-test/bursary-scheme/ Accessed 21 January 2018
[6] HESA Student Record - Parental education https://www.hesa.ac.uk/collection/c16051/a/pared . Accessed 21 January 2018
[7] HESA Student Record - Occupation code https://www.hesa.ac.uk/collection/c16051/e/soc2000 Accessed 21 January 2018
[8] HESA Student Record - Socio-economic classification https://www.hesa.ac.uk/collection/c16051/e/sec Accessed 21 January 2018
[9] Higher Education Funding Council for England (HEFCE) POLAR – Participation of Local Areas Available at http://www.hefce.ac.uk/analysis/yp/POLAR/POLAR3,data/ Accessed 21 January 2018
[10]Communities and neighbourhoods Indices of Deprivation 2007 Downloaded from http://webarchive.nationalarchives.gov.uk/20100410180038/http://communities.gov.uk/communities/neighbourhoodrenewal/deprivation/deprivation07/), Accessed 6 April 2016.

[11] HESA Course title. Available at: https://www.hesa.ac.uk/collection/c16051/e/ctitle Accessed 15 January 2018.
[12] HESA Course identifier. Available at: https://www.hesa.ac.uk/collection/c16051/a/course_courseid Accessed 15 January 2018.
[13] Medical Schools Council *Entry requirements for UK medical schools* https://www.medschools.ac.uk/media/2032/msc-entry-requirements-for-uk-medical-schools.pdf. Accessed 15 January 2018.
[14] HESA CommencmentDate. Available at https://www.hesa.ac.uk/collection/c16051/a/comdate Accessed 15 January 2018.
[15] HESA UK Provider Reference Number. Available at https://www.hesa.ac.uk/collection/c16051/a/ukprn Accessed 15 January 2018.
[16] HESA Rounding and suppression to anonymise statistics. Available at: https://www.hesa.ac.uk/about/regulation/data-protection/rounding-and-suppression-anonymise-statistics. Accessed 15 January 2018.

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract >> **>> "cohort" and "validity" in title**
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found >>**Yes states the measures, the analysis and the ORs**
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported **>>Yes we give the background to recruitment to the foundation programme and ARCP in foundation on pages 2 through to 5**
Objectives	3	State specific objectives, including any prespecified hypotheses **"there was an opportunity to assess the extent to which the selection scores predicted foundation ARCP outcomes for two cohorts of applicants." Page 5**
Methods		
Study design	4	Present key elements of study design early in the paper **Yes in Study population section on page 5 and Figure 1**
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection **"All applicants making their final application (i.e. the application that led to the applicant starting on the foundation programme) in 2013 and 2014 were eligible for inclusion in the analysis (N = 15,249). Page 5** **And** **"UKMED receives ARCP outcome data from the GMC's annual collection of ARCP outcomes from postgraduate training providers. The outcome variable used was obtaining an 'outcome 6'.This is defined by the UKFPO as "Satisfactory completion of F2 – Recommendation for the award of the Foundation Achievement of Competence Document" at the end of their F2 year.12 in 2015 (2013 starters or 2016 (2014 starters)." Page 6**
Participants	6	*(a)* Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up **>>In Figure 1**
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed >>N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable **>>Yes in Data Management section pages 6 - 7**

Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group **>>Yes in Data Management section pages 6 -7**
Bias	9	Describe any efforts to address potential sources of bias **Yes in Discussion section we mention the cases not available, not possible to address these in the study design as the data were not available to us.** **"However, a number of limitations of the study are worth noting. Firstly, the UKMED phase 1 cohort does not include graduates from non-UK medical schools who apply to Foundation training. However, from 2015 onwards the non-UK students have been included in the data UKMED receives on foundation applications, permitting research into this area. Secondly, suboptimal ARCP outcomes are relatively rare in the Foundation Programme period of training." Page 11**
Study size	10	Explain how the study size was arrived at **"All applicants making their final application (i.e. the application that led to the applicant starting on the foundation programme) in 2013 and 2014 were eligible for inclusion in the analysis (N = 15,249). "** **Page 5**
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why **Yes in Data Management section pages 6 and 7**
Statistical methods	12	*(a)* Describe all statistical methods, including those used to control for confounding **>>Multivariate analysis will control for confounding of the variables. Page 7**
		(*b*) Describe any methods used to examine subgroups and interactions **>>Subgroups and interactions not analysed as the not relevant to main study hypotheses.**
		(*c*) Explain how missing data were addressed **>>Nothing/very little missing on main variables used**
		(*d*) If applicable, explain how loss to follow-up was addressed **>>N/A**
		(*e*) Describe any sensitivity analyses **>N/A**
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed **>>Figure 1**

		(b) Give reasons for non-participation at each stage	
		(c)	**>>Figure 1**
		(d) Consider use of a flow diagram	
		(e)	**>>Figure 1**
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	
		>>**Yes Table 1**	
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) Summarise follow-up time (eg, average and total amount) ***	
Outcome data	15*	Report numbers of outcome events or summary measures over time	
		>>Yes only one outcome successful completion	
Main results	16	*(a)* Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	
		>>95% Cis used in table 2 and table 3	
		(b) Report category boundaries when continuous variables were categorized	
		Data Management section pages 6 and 7	
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	
		>>See Model 3	
Discussion			
Key results	18	Summarise key results with reference to study objectives **>> Yes Page 10 -12**	
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias**>> Yes Page 10 -12**	
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence **>> Yes Pages 10 -12**	
Generalisability	21	Discuss the generalisability (external validity) of the study results **>>N/A except to other years of foundation programme****	
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	
		>Yes in Funding section	

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.