Journal of Public Health Advance Access published September 19, 2007

Journal of Public Health | pp. 1–7 | doi:10.1093/pubmed/fdm060

Selection into specialty training in public health: performance of the Medical Training Application Service shortlisting

Nora Pashayan[1], Celia Duff[2], Brendan W. Mason[3]

[1]Department of Public Health and Primary Care, Institute of Public Health, University Forvie Site, Robinson Way, Cambridge CB2 0SR, UK
[2]East of England Multi-Professional Deanery, Block 3, Ida Darwin Site, Fulbourn, Cambridge CB21 5EE, UK
[3]National Public Health Service for Wales, The Temple of Peace and Health, Cathays Park, Cardiff CF10 3NW, UK
Address correspondence to Brendan W. Mason, E-mail: brendan.mason@nphs.wales.nhs.uk

ABSTRACT

Objective To assess the performance of shortlisting against appointability to public health specialty training under the Medical Training Application Service (MTAS) selection methodology using multiple modality in person assessment.

Methods Candidates who had applied to public health specialty training programme in Wales and East of England and shortlisted were assessed in the first assessment round. Further to MTAS review, candidates not previously short listed were offered assessment in the second round. Receiver operating characteristic (ROC) analysis was done.

Results In both the programmes, the shortlisting scores of candidates considered appointable were substantially higher than those considered not appointable, a score difference of 13.0 (95% confidence interval (CI) 3.0–23.0) and 13.5 (95% CI 3.4–23.5) respectively. The area under the ROC curve (ROC_{AUC}) was 0.88 (95% CI 0.63–1.00) in Wales and 0.77 (95% CI 0.57–0.97) in East of England. The shortlisting scores of the two programmes that gave an optimum performance (maximum sum of the sensitivity and specificity) were comparable (scores of 62 and 63 respectively).

Conclusion MTAS shortlisting undertaken in two independent public health specialty training programmes discriminated well between appointable and not appointable candidates.

Keywords medical training application service (MTAS), specialty training, postgraduate medical education, assessment centre, receiver operating characteristic analysis

Introduction

Under the auspices of Modernising Medical Careers,[1] a new on-line application system, Medical Training Application Service (MTAS),[2] for recruitment and selection into specialty training in the UK was activated in January 2007. Since then, the operation of MTAS has been highly scrutinized and changes introduced. After the first round (round 1a) of shortlisting and interviewing, the offer of guaranteed assessment for a preferred choice was extended to candidates not previously shortlisted (round 1b).[3,4]

Under MTAS, shortlisting did not take account of medical school and country of training. The majority of the weighting for shortlisting was based on 150-word answers to shortlisting questions and less weighting was given to past experience, examinations passed in the specialties, higher qualifications or publications. This was in sharp contrast to the previous CV-led process where past experience, achievements and higher specialty examinations passed were used to support shortlisting.

The new system was devised in such a way as to remove possible bias due to candidates' past experience and achievements, to be uniform and fair in identifying individuals with potential for successful performance in that specialty and giving equal opportunity to all applicants.[5,6] A key principle for recruitment to specialty training laid down by Postgraduate Medical Education and Training Board (PMETB) is the maintenance of the generic nature of UK Foundation Training and thus no previous post should be a requirement for appointment or advantage applicants for appointments at Specialty Training Year 1 level.[7]

Nora Pashayan, Specialist Registrar in Public Health

Celia Duff, Associate Dean and Head of School, Public Health

Brendan W. Mason, Consultant Epidemiologist and Training Programme Director for Public Health

MTAS has been criticized for not taking full account of the past experience and qualifications of applicants. This raised concerns that shortlisting might discriminate poorly between appointable and not appointable candidates.[8,9] Despite the widespread dissatisfaction with MTAS, no objective evidence has been published on the performance of the selection process. The decision to offer at least one guaranteed assessment to all candidates who applied for specialty training provided a unique opportunity to investigate the performance of the shortlisting process. We examine the performance of this process in identifying the best candidates, taking as reference the assessment centre, a selection process using a combination of assessment modalities and multiple assessors.[10] The analysis is based on the shortlisting and recruitment experience of two public health training programmes.

Methods

The public health training programmes in Wales and East of England used standard national shortlisting but different assessment methods.

Shortlisting

The national electronic application form covered biographical information (such as qualifications, eligibility and employment) and shortlisting questions with space for 150-word answers related to commitment to specialty, technical skills (clinical, academic and research), personal skills and probity. The application forms were independently scored by two public health consultants in each programme using national shortlisting scoring indicators. The scores from the two assessors were added together to produce a total shortlisting score (maximum score of 90).

The number of candidates shortlisted for the first round of the interviews was determined by the number of available posts to fill. Initially, 3 applicants to the Wales programme and 20 applicants to the East of England programme with the highest shortlisting scores were invited to be assessed for the available one and eight training placements, respectively (round 1a). As a result of the MTAS review, all the remaining applicants to the Wales programme and applicants who then expressed East of England public health training programme as their preferred choice were invited to an assessment centre (round 1b).

Assessment centre—Wales Public Health Training Programme

The Wales assessment centre offered three 20-min interviews, each of which included four or five set of questions covering the nationally agreed person specification. All interviewers on each panel independently scored each question based on positive and negative indicators. The maximum possible score was 124. The cut point for the assessment scores used to separate appointable and not appointable candidates was agreed by the whole panel when all interviews were completed.

Assessment centre—East of England Public Health Training Programme

The East of England programme was part of a cluster of four programmes running a joint assessment centre. The first component of the assessment process comprised numerical and verbal reasoning tests. Candidates who secured a score of $\geq$50% on both tests were eligible to proceed. The assessment centre combined six stations: four 14-min competency-based interviews, a presentation and a group exercise. The maximum possible assessment score was 104. Candidates were considered appointable if their aggregate score exceeded the cutoff point set by the panel.

For both the programmes, all the assessors had received training in behavioural observation, recording and scoring. All the assessors were blind to the information provided in the application and to the shortlisting score.

Statistical analysis

The association between the shortlisting and assessment scores was measured by the Spearman's rank correlation coefficient. The difference between the mean shortlisting score of candidates who were or were not appointable was compared with paired Student's *t*-test, after assessing the assumption of normality of the dependent variable by tests for skewness and kurtosis. Variation in the assessment scores by assessment day was examined using the Kruskal-Wallis test.

Receiver operating characteristic (ROC) analysis was performed to study the ability of shortlisting to discriminate between candidates deemed appointable or not into specialty training in public health, on the basis of their assessment score. The ROC curve was used to identify the cutoff value of the shortlisting score that maximizes the sum of sensitivity and specificity of the shortlisting process. The area under the curve (AUC) was used as a measure of the overall performance of the ROC curve as it reflects, in this case, the probability that shortlisting will correctly classify candidates as appointable or not appointable. AUC can take values between 0 and 1, where 1 is a perfect screening test and 0.5 is a test equal to chance.[11] All statistical tests were two-tailed. $P < 0.05$ was considered statistically significant. Data were analysed using Stata 9.0.[12]

Results

Wales Public Health Training Programme

Twenty-five eligible individuals applied to the public health training programme in Wales. All were eventually offered an interview and 17 accepted it (Fig. 1a).

Performance of shortlisting

The rank correlation between the shortlisting and interview scores was 0.60 ($P = 0.01$). Six candidates were judged to be appointable, with a mean shortlisting score of 64 (range 55–73); the mean shortlisting score of the 11 deemed not

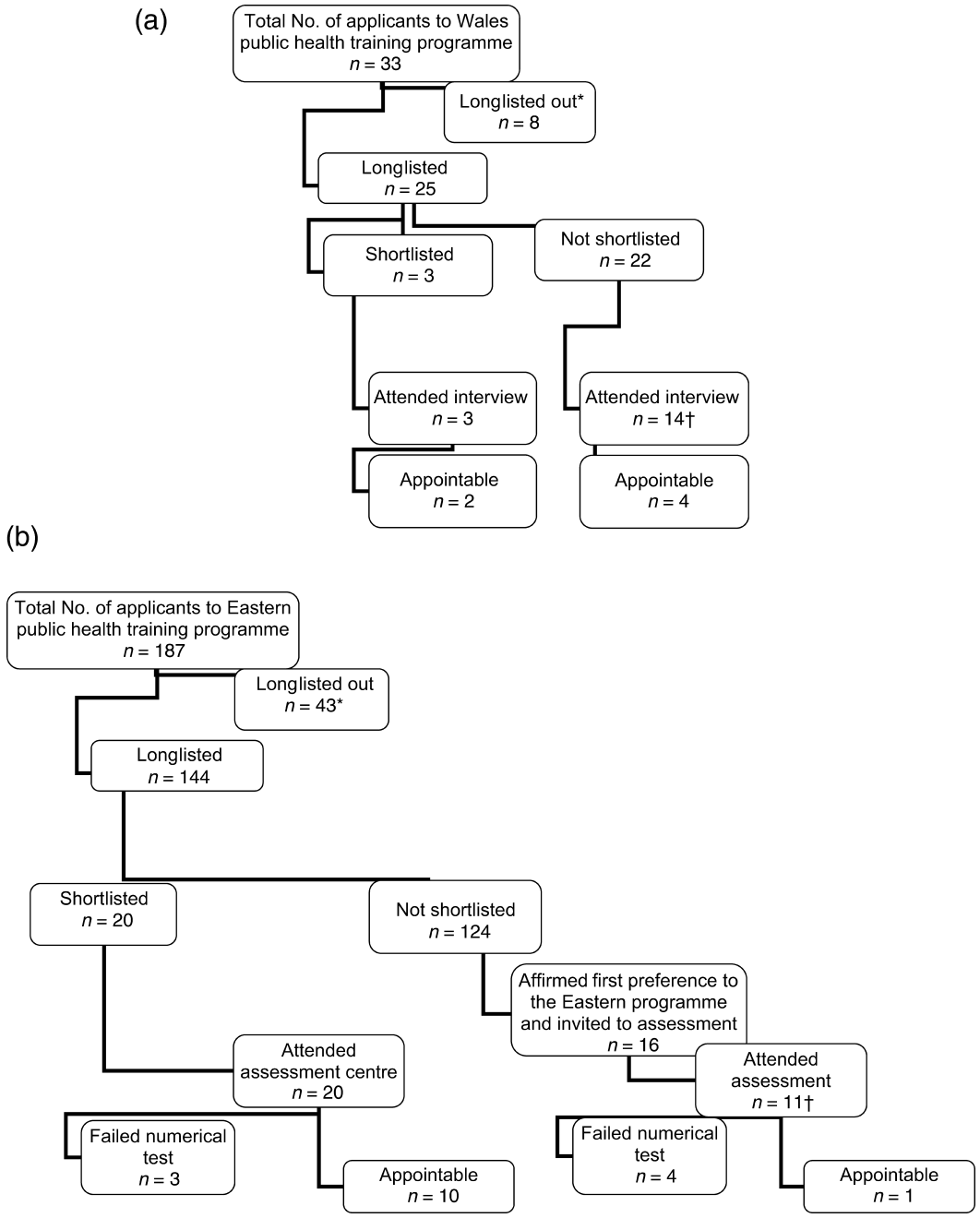

Fig. 1 (a) Flow diagram of applicants to Wales public health training programme. *Not meeting the eligibility criteria for medical qualifications, language skills, GMC registration, work in the UK. †All eligible candidates who have applied to Wales were offered interview. (b) Flow diagram of applicants to East of England public health training programme. *Not meeting the eligibility criteria for medical qualifications, language skills, GMC registration, work in the UK. † 11 candidates took the numerical and verbal test. Those who failed either test did not progress to the rest of the assessment centre.

appointable was 51 (range 34–68), a difference of 13 points (95% CI 3–23; $P = 0.02$).

ROC analysis for the selection of appointable candidates from their shortlisting scores demonstrated acceptable performance of shortlisting [AUC = 0.83 (95% CI 0.63–1.00)] (Fig. 2a). With a shortlisting score cutoff of 68 and an assessment score cutoff of 98 (79% of the total assessment score), the shortlisting process had a sensitivity of 33%, a specificity of 91%, a positive predictive value (PPV) of 67% and a negative predictive value (NPV) of 71%. The shortlisting score cutoff that maximizes the sum of sensitivity and specificity was derived from the ROC analysis. This cutoff was 62, giving a sensitivity of 67%, a specificity of 91%, a PPV of 80% and an NPV of 83% (Supplementary data are available at additional file 1.doc).

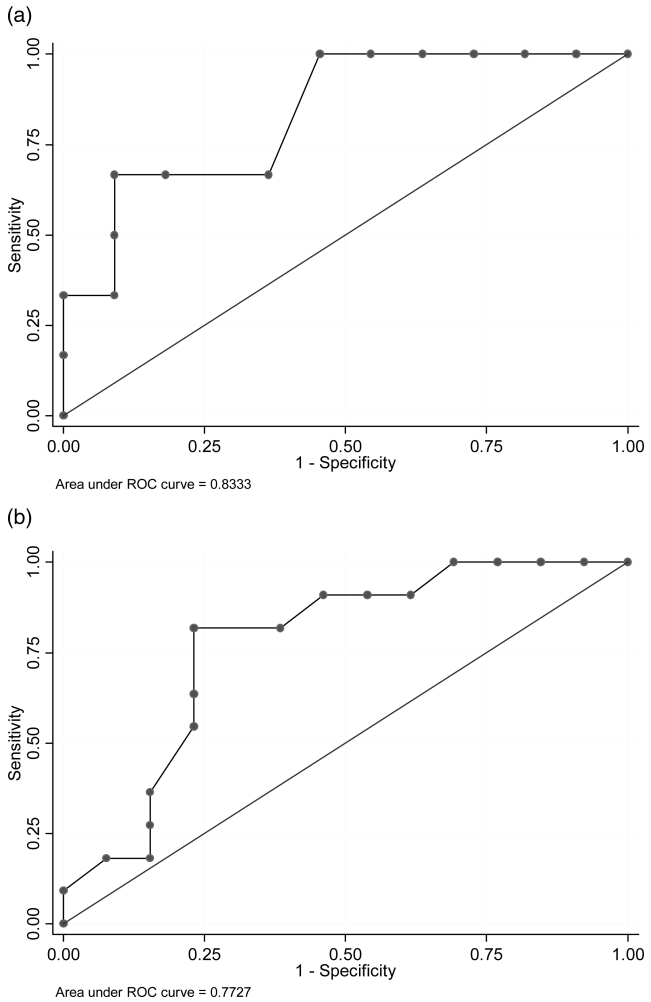

Fig. 2 (a) ROC curve for selection of appointable candidates from the shortlisting scores, Wales public health training programme. (b) ROC curve for selection of appointable candidates from the shortlisting scores, East of England public health training programme.

East of England Public Health Training Programme

The East of England programme had 144 eligible applicants. Combining rounds 1a and 1b, 36 candidates were invited to take the numerical and verbal reasoning tests, of whom 31 attended. Twenty-four of 31 passed the tests and completed the assessment process. A total of 11 candidates, 10 from the shortlisted group and 1 from the non-shortlisted group, were considered appointable (Fig. 1b).

Shortlisting score

Out of 124 candidates, who were not shortlisted for round 1a but offered the opportunity to declare the East of England as the preferred choice in round 1b, 11 attended the assessment. There was no statistically significant difference in the mean shortlisting score of those who attended the round 1b assessment ($n = 11$) and those who did not ($n = 113$); mean score difference was 1.8 (95% CI −6.0–9.6).

Candidates who were considered appointable from the combined rounds 1a and b ($n = 11$) had higher mean shortlisting score than those considered not appointable ($n = 20$); 65.5 (range 53–82) versus. 52.0 (range 5–71), respectively, a difference of 13.8 (95% CI 3.7–23.8; $P = 0.01$).

Assessment score

Fig. 3 presents the mean score and 95% CI for each assessment component by day of assessment. The mean (SE) assessment score of day 3 (round 1b) [47.6 (8.20)] was consistently lower than those of day 1 (round 1a) [71.6 (4.49)] and day 2 (round 1a) [75.6 (2.44)], $P = 0.03$.

Performance of shortlisting

The rank correlation of shortlisting and assessment scores was 0.50 ($P = 0.013$). With a shortlisting score cutoff of 60 and an assessment score cutoff of 70 (67% of the total assessment score), shortlisting had a sensitivity of 91%, a specificity of 46%, a PPV of 59% and an NPV of 86%.

Shortlisting was significantly better than chance in identifying appointable candidates (AUC = 0.77; 95% CI 0.57–0.97, $n = 24$) (Fig. 2b). The optimum shortlisting cutoff score, derived from the ROC analysis, was 63 (Supplementary data are available at additional file 1.doc). Increasing the cutoff score to 63, gave a sensitivity of 82%, a specificity of 77%, a PPV of 75% and an NPV of 83%.

Discussion

Main findings of this study

The analysis of public health selection data from two public health training programmes shows that shortlisting

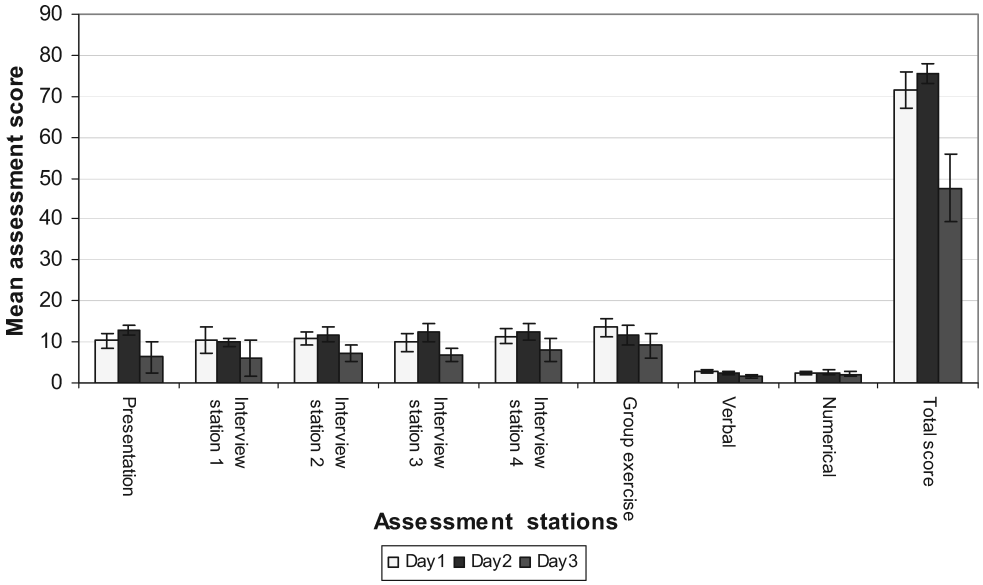

Fig. 3 Mean assessment score and 95% confidence interval of each assessment station by assessment day*, East of England public health training programme. *Round 1a assessments were on days 1 and 2, and round 1b on day 3.

discriminated well between candidates who were subsequently deemed suitable for specialty training in public health. Shortlisting identified two of six appointable candidates in Wales and 10 of 11 appointable candidates in the East of England.

A pragmatic approach to managing volume of interviews in each deanery dictated the total number of candidates invited to assessment from ranked shortlisted applicants. The single post available in Wales led to a high shortlisting cutoff resulting in high specificity at the expense of low sensitivity (1 of 11 not appointable candidates was shortlisted and 4 of 6 appointable candidates were not shortlisted). In the East of England, with eight available posts, 20 of 144 eligible applicants were shortlisted. The lower shortlisting cutoff score improved the sensitivity of the process. The sensitivity of a screening test is central to the rationale of screening. Shortlisting can be conceptualized as a screening tool to separate those who are probably appointable from those who are probably not appointable. Shortlisting does not provide a perfect assessment of a candidate's suitability for appointment. If it did, it would not be necessary to interview candidates.

An overall performance of shortlisting as a screening test is provided by the area under the ROC curve. The AUC of shortlisting in the two public health training programmes (0.77 and 0.83, respectively) demonstrates that the process had good ability to classify candidates into relevant subgroups, as defined by the reference test.[13]

Having determined that the process provides good discrimination, what is the best cutoff point for shortlisting? This can be determined by considering the consequences of false-negative classifications (failing to offer an interview to an appointable candidate) and false-positive classifications (offering an interview to a candidate who is not appointable). ROC analysis was used to estimate the optimum cutoff value of the shortlisting score. Optimum cutoff was taken as the one that maximizes the sum of the sensitivity and specificity of the shortlisting process. This keeps a fair balance between the trade-off of sensitivity and specificity. High sensitivity should be balanced with the implications of high proportion of false-positive classification.

The absolute number of true negatives and true positives can only be determined by offering an assessment centre to all candidates. However, the shortlisting scores of those who were not shortlisted but completed the assessment centre in round 1b and those who did not attend assessment were comparable (East of England programme, $P = 0.65$). This might suggest that performance at the interview would be comparable and that the observed sensitivity and specificity would be applicable to this cohort of applicants.

The performance of a screening test is measured against a reference procedure or the gold standard. In this study, the assessment centre was taken as the reference to identify suitable candidates for training placements. Studies have shown that assessment centres, with multiple and varied assessment opportunities, are successful in predicting future

job performance across a wide range of occupations,[7] including the medical domain.[14,15]

The choice of assessment score cutoff to define true positives, true appointable candidates, determines the results of this analysis and might be used as an argument to reject the results. Although there is as yet no agreed assessment cutoff score, a large panel of trained and experienced assessors collectively agreed this score.

The success of the assessment centre in predicting successful progress through specialty training depends on the design and conduct of assessment. Follow-up of the recruits will be required to establish whether performance at assessment is a good indicator of performance in training. Research on the longer-term predictive validity of the selection is required.

What is already known on the topic

There is no objective evidence on the performance of shortlisting under the new MTAS in discriminating between candidates appointable or not to specialty training.

Application through a central portal with local selection, similar to the MTAS, has been used successfully in USA, Canada, Australia and New Zealand for at least 30 years.[16]

Research has shown that competency-based selection process using assessment centres improves the validity of selection, compared with traditional selection methods.[14,15,17]

What this study adds

To the best of our knowledge, this is the first study that has assessed MTAS shortlisting against appointability to specialty training. For public health training programmes in Wales and in East of England, the MTAS shortlisting discriminated acceptably between candidates subsequently deemed appointable and not appointable as measured by the performance in the assessment process.

Limitations of the study

It is not possible to be certain that the findings of this study can be generalized to other specialties or other parts of the UK. No comparable data are available from shortlisting used in previous years to enable a systematic comparison of the two systems.

In conclusion, this study has shown that shortlisting undertaken in two independent public health specialty training programmes using the same national scoring template discriminated acceptably between candidates subsequently deemed appointable or not. There were a number of problems with MTAS and its implementation. It is essential that any further review of selection methodology identifies and separates the constituent parts of these problems. If we are to learn from the experience and develop a system that has the confidence of the profession, it is crucial that we start the process from an objective systematic evidence-based analysis of the available data.

Authors' contributions

All authors contributed to the conception and design of the study and interpretation of the data. B.M. and N.P. analysed the data for Wales and East of England, respectively. All authors contributed to the writing of the manuscript. B.M. and C.D. are the guarantors.

Acknowledgment

We would like to thank Dr John Powles for helpful comments.

Conflict of interest

None declared.

References

1 Modernising Medical Careers. http://www.mmc.nhs.uk/pages/home (12 June 2007, date last accessed).

2 Medical Training Application Service. http://www.mtas.nhs.uk (27 March 2007, date last accessed).

3 Review of Round 1 Applications to ST 2007. Modernising Medical Careers. http://mmmc.nhs.uk/pages/review (12 June 2007, date last accessed).

4 MMC Recruitment in Wales. 2007. Modernising Medical Careers. http://www.mmcwales.org/news/mmc-recruitment-in-wales-30-march-2007 (12 June 2007, date last accessed).

5 Delamothe T. Centralised application services for specialist training. *BMJ* 2007;**334**(**7607**):1285–6.

6 Black C, Maxwell P, Marshall M *et al.* MTAS: which way now? Interview by Rebecca Coombes. *BMJ* 2007;**334**(**7607**):1300.

7 Principles for Entry to Specialist Training. 2006. Postgraduate Medical Education and Training Board. http://www.pmetb.org.uk/index.php?id=456 (20 August 2007, date last accessed).

8 Baseley S. Junior doctors driven abroad by new system. *Guardian*. 2 March 2007.

9 Hall C. Doctors' training system 'a shambles'. *Daily Telegraph*. 3 March 2007.

10 Patterson F, Norfolk T. What is? An assessment centre. http://esrccoi.group.shef.ac.uk/pdf/whatis/ass_centre.pdf (12 June 2007, date last accessed).

11 Hanley JA. The robustness of the 'binormal' assumptions used in fitting ROC curves. *Med Decis Making* 1988;**8**(**3**):197–203.

12 Stata Statistical Software release. Texas: Stata Corporation, 2005.

13 Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. *BMJ* 1994;**309**(**6948**):188.

14 Patterson F, Ferguson E, Norfolk T *et al*. A new selection system to recruit general practice registrars: preliminary findings from a validation study. *BMJ* 2005;**330**(**7493**):711–4.

15 Randall R, Davies H, Patterson F *et al*. Selecting doctors for postgraduate training in paediatrics using a competency based assessment centre. *Arch Dis Child* 2006;**91**(**5**):444–8.

16 Jefferis T. Selection for specialist training: what can we learn from other countries? *BMJ*; 2007;**334**(**7607**):1302–4

17 Patterson F, Lane P, Ferguson E *et al*. Competency-based selection system for general practitioner registrars. *BMJ Career Focus* 2001;**323**:S2–7311.