Variation in Performance on Common Content Items at UK Medical Schools

David Hope (david.hope@ed.ac.uk)
University of Edinburgh
David Kluth
University of Edinburgh
Matthew Homer
University of Leeds
Avril Dewar
University of Edinburgh
Richard Fuller
University of Liverpool
Helen Cameron
Aston University

Research Article

Keywords: medical school, knowledge, performance

Posted Date: February 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-153247/v1

Abstract

Background

Due to the diverse approaches to medical school assessment, making meaningful cross-school comparisons on knowledge is difficult. Ahead of the introduction of national licensing assessment in the UK, we evaluate schools on “common content” to compare candidates at different schools and evaluate whether they would pass under different standard setting regimes. Such information can then help develop a cross-school consensus on standard setting shared content.

Methods

We undertook a cross-sectional study in the academic sessions 2016-17 and 2017-18. Sixty “best of five” multiple choice items were delivered each year, with five used in both years. In 2016-17 30 (of 31 eligible) medical schools undertook a mean of 52.6 items with 7,177 participants. In 2017-18 the same 30 medical schools undertook a mean of 52.8 items with 7,165 participants for a full sample of 14,342 medical students sitting common content prior to graduation. Using mean scores, we compared performance across items and carried out a “like-for-like” comparison of schools who used the same set of items then modelled the impact of different passing standards on these schools.

Results

Schools varied substantially on candidate total score. Schools differed in their performance with large (Cohen’s *d* around 1) effects. A passing standard that would see 5% of candidates at high scoring schools fail left low-scoring schools with fail rates of up to 40%, whereas a passing standard that would see 5% of candidates at low scoring schools fail would see virtually no candidates from high scoring schools fail.

Conclusions

Candidates at different schools exhibited significant differences in scores in two separate sittings. Performance varied by enough that standard setting approaches that produce realistic fail rates in one medical school may produce substantially different pass rates in other medical schools – despite identical content and the candidates being governed by the same regulator. Regardless of which hypothetical standards are “correct” as judged by experts, large institutional gaps in pass rates must be explored and understood by medical educators before shared standards are applied. The study results can assist cross-school groups in developing a consensus on standard setting future licensing assessment.

Introduction

Assessment in medical education should ensure doctors are competent, safe practitioners (1, 2). Typically, candidates approaching registration must sit an “exit” assessment to confirm suitability to work as a doctor (3). The reliability and validity of such assessments are of great importance in maintaining the quality of medical education and ensuring patient safety.

Evaluating such assessments can be difficult. In almost all regulatory environments doctors graduate from different institutions. Therefore, a range of institutional contexts, curricula, admissions policies, and resources produce doctors who are nominally equivalent, but differ in experiences (4). Regulators seek to ensure equivalence across institutions by monitoring and enforcing a shared set of values and requirements (5).

As the content, structure, and weighting of exit assessments vary, direct comparisons across institutions are very difficult to carry out. Several partial solutions have been tested. One approach is to compare candidates on later – usually postgraduate – assessment which can act as a comparative measure. Research has shown that graduates of different medical schools exhibit large differences in performance on postgraduate assessments (6). Relatedly, evidence has suggested that the performance of individual medical students and doctors exhibits at least moderate stability over time (7, 8) which suggests the type of candidates applying to medical schools, or their experiences at medical schools, may create meaningful differences between cohorts upon graduation. Importantly, performance on postgraduate assessment predicts not just technical skill, but professionalism – including the likelihood of being sanctioned while working as a doctor (9).

This research necessarily contains limitations. Postgraduate attainment can only measure capabilities some years after doctors begin work and cannot confidently identify the source of such differences. Postgraduate assessments are often specialised and sat by only a subset of doctors, and candidates who exit the profession soon after graduation will never sit them.

An alternative method lies in the use of “common content.” Here, a group of institutions pool resources and share assessment content across institutions. So, a group of medical schools may share stations in a clinical examination, or multiple-choice questions (MCQs) in a written examination, with the remaining content set locally and independently. By evaluating both the approach to standard setting and the attainment of different cohorts, it is possible to get a better sense of how variable institutions within a regulatory framework are. Research on common content has suggested that different medical schools set very different standards for identical content. Research on MCQ-type written assessment has shown significant differences in medical school standard setting with typically medium effects, with the attendant risk that candidates who passed at institutions with lenient standards would have failed – and therefore not graduated – at institutions with more stringent standards for the same content (10). A follow-up exploration of standard setting at some of the same schools described institutional, individual, and group factors combining to create highly unique standard setting procedures despite using the same content at all institutions (11).

Research on “common content” clinical examination stations have found similar problems, with standards for the same station varying by up to 13% between the most lenient and stringent school (3). Evidence on attainment, rather than standards, remains very sparse but some research on clinical examinations showed medical school cohorts scoring significantly differently on common content stations, in a pool of four medical schools (12).

This is extremely important as it suggests that, even if the content tested in different medical schools is equivalent, the local variability of standards may lead to candidates passing in some environments when they would have failed in others. Indeed, research has suggested that across many measures – content, type, duration, and standard setting – medical schools have a widely varying range of approaches (6, 13). The fear that monitoring systems do not ensure comparability across institutions has led to recommendations for a “licensing” assessment which acts as a single point of measurement for all candidates within a regulatory framework (14). The utility of this proposal remains contested. To some it represents the advance of a test-centric culture where learning is devalued (15) and educational diversity reduced (16). To others, there are potentially significant benefits to patient safety by harmonising standards (14, 17).

The practical and theoretical challenges of implementing any multi-site assessment is significant. In the Netherlands, a progress test delivered across institutions has led to a more effective use of resources and enabled cross-school research, but also disagreements over item quality and logistical difficulties in organising the new assessment (18, 19). In the United States, students have responded to the United States Medical Licensing Examination (USMLE) Step 1 with a range of effective self-directed learning behaviours to maximise the likelihood of passing (20). However, the focus on the candidate’s USMLE score has led authors to claim other aspects of performance – including achievements during medical school – have been under-valued, which has in turn led to reporting changes whereby only the candidate’s pass/fail status is reported (21). Such research demonstrates that cross-school assessment inevitably has serious implications for curriculum design and student learning even in areas which the assessment does not directly assess.

Despite the potentially significant impacts of a new licensing assessment on passing rates at medical schools, little is known about how such assessment might influence standard setting and pass rates. As a first step, medical educators at all affected schools should be aware of the relative performance of their students and the potential impact of different standard setting regimes, which can in turn help develop a consensus on how to standard set national licensing assessment in a way that recognises educational diversity while also ensuring patient safety.

To develop better evidence in this area, we used “common content” MCQs developed by the Medical Schools Council Assessment Alliance (MSCAA(10)) to compare candidates at 30 medical schools, evaluate performance differences and estimate the impact of different standards on pass rates ahead of the implementation of a licensing assessment in the United Kingdom.

Methods

Study Design

We undertook a cross-sectional study in academic sessions 2016-17 and 2017-18. The MSCAA organised 60 core items for participating schools in 2016-17 and 60 in 2017-18, with five used in both years. These were all “single best answer” multiple choice questions with one correct option and four distractors. Items were developed and standard set by subject area experts and designed to test knowledge relevant to new doctors. Further details of the standard setting process can be found in previous publications (10, 11).

Participants

All UK medical schools were offered the opportunity to participate in the common content project. The items were embedded within a larger local assessment, and medical schools varied in how many items they selected. All items were delivered within an exit examination sat near the end of medical school. In 2016-17 30 medical schools undertook a mean of 52.6 common content items, with a total candidate number of 7,177. In 2017-18 30 medical schools undertook a mean of 52.8 common content items, with a total candidate number of 7,165, making for 14,342 sittings evaluated within this study. Full details can be found in Table 1.

Ethics

Ethical approval was granted by the University of Edinburgh Medicine and Veterinary Medicine ethics committee. All participant details – both schools and candidates – were anonymised, and the research team had no access to deanonymized data.

Data collection

Following the completion of assessment, each school reported on the common content items to the MSCAA. This included notes from staff or candidates expressing concerns over item quality and a report of performance per candidate per question. The MSCAA then evaluated the reliability of the assessment using a combination of Classical Test Theory (CTT) and Rasch analysis to test whether items were of acceptable quality for analysis. Where a candidate failed to answer a question, this was coded as 0 (incorrect). An exploration of missing responses identified no pattern that would call into question the validity of any items or candidate response patterns.

Data analysis

As medical schools varied in the common content items they selected, making like-for-like comparisons was challenging. We utilised a two-part approach. In part 1, we compared means/facility scores, standard deviations, and discrimination indices for every item for every medical school that used the item. This allowed us to compare the homogeneity of medical schools in terms of both their average score and their variability. We sought to identify where (and how frequently) a given medical school significantly varied compared to other schools to see whether variability could be explained by small deviations across many

items, or large deviations in a small number of items. This analysis was intended to be primarily descriptive, though we carried out a formal test of significance (via t-tests) for completeness.

In part 2, we selected a subsample of schools who had all sat a large proportion of the items. 13 schools sat the same 41 items in 2016-17, and 14 schools sat the same 48 items in 2017-18. For these schools we carried out a like-for-like analysis of their within-year performance, tested whether performance of the top and bottom tertiles (representing “high scoring” and “low scoring” schools) differed significantly and modelled the impact of different passing standards. An a-priori power calculation showed that analyses used were able to detect small effect sizes at 80% power (22). School codes were not re-used, so the same code referred to a different school in each year.

Part 1 – item performance

We report here a Classical Test Theory (CTT) analysis of the data. While there are advantages to alternative methods – especially Rasch analysis (23) – the comparative simplicity and familiarity of CTT methods were considered desirable given the objective of maximising accessibility for the largest possible audience (24). While we analysed the data using both a CTT and Rasch framework, only the CTT values are reported here.

For each item, we calculated the overall mean (or facility) score (between zero, indicating no candidate answered the item correctly, and one, indicating all candidates answered correctly), the Standard Deviation (SD) and the discrimination index (a measure of whether the item could discriminate between candidates who performed well or poorly on the assessment as a whole (25)). Facility and discrimination values did not differ significantly between the two study years, and so we repeated the same analysis on each cohort. We calculated mean item performance on items ($M = 0.74$, $SD = 0.18$) and mean item discrimination on items ($M = 0.20$, $SD = 0.10$). We then calculated mean item performance (and associated SDs) for each school, per year. We then identified the proportion of items where the school was one or two SDs above the mean score, and one or two SDs below the mean score as a measure of the school’s overall performance against all medical schools.

To further explore this, we compared the frequency of items where the school scored two SDs below the mean. For the analysis, we compared the bottom and top tertiles and ran the analysis for each session. This gave a percentage measurement from zero (the school had no items 2 SDs below the mean) to 100% (the school’s cohort scored 2 SDs below the mean for every item). We calculated tertiles by the school’s mean mark across all the items they used, and so compared the bottom tertile (the ten lowest performing medical schools on this assessment) against the top tertile (the ten highest performing medical schools).

The main goal of this was not to provide a precise comparison – because schools did not sit exactly the same items this was not possible – but to explore whether differences between schools could be explained by some schools exhibiting much higher rates of incorrect responses across a range of

domains. We chose to use 2 SDs as a cutoff as this generally indicated a notably lower score compared to the average school.

Part 2 – modelling standards

By comparing item usage across all schools, we identified schools which shared many items. We modelled the interaction of school numbers vs. item numbers: at one extreme it would be possible to compare all schools on a very small number of items, and at the other extreme a very small number of schools on all items. After modelling options, we were able to identify 13 schools from the 2016-17 cohort that had used the same 41 items, and a further 14 schools from the 2017-18 cohort that had used the same 48 items.

This gave us two samples of medical schools sitting identical content with an acceptable pool of items. For both years, Cronbach's alpha = 0.7, indicating an acceptable level of internal consistency for the two sets of items. We compared the bottom and top third of medical schools (rounded for uneven group sizes) in each sample on mean score. As in part 1, the sample size was adequate to test for small effects at 80% power.

We then modelled the effect of different passing standards. We identified the pass score that would give a score as close as possible to a 5% fail rate at (a) the four highest-scoring schools ("stringent") and (b) the four lowest scoring schools ("lenient"). This number was chosen to match the typical fail rate of the Prescribing Safety Assessment (PSA), an assessment sat by candidates across UK medical schools with similar features to future potential licensing assessments (26). We then estimated the impact of imposing these passing standards on the medical schools.

Results

Part 1 – item performance

In 2016-17, schools in the lowest tertile (that is, their total score on the common items placed them in the lowest third when ranked by performance) had a number of items with facility scores two SD below the mean (M = 7.81%, SD = 4.4%) whereas the top tertile (upper third) had none, a significant difference (t(9)=5.61, p = .001). This pattern was repeated in 2017-18 with the bottom tertile having some (M = 6.62%, SD = 4.19%) and the top tertile again having none, a significant difference (t(9)=5, p = .001). This meant that for both years, schools in the bottom tertile reported significantly higher rates of items with facility scores two SD below the mean, indicating a different level of knowledge among those medical school students compared to the top tertile cohorts. This suggests that differences in scores may reflect differences in knowledge across a range of areas.

A full summary of the medical schools, the number of items they used, their scores relative to other medical schools, and their local sample size can be found in Table 1.

Part 2 – modelling standards

In 2016-17, comparing the bottom (*M* = 0.76, *SD* = 0.1) and top (*M* = 0.85 *SD* = 0.08) tertiles identified a statistically significant difference ($t(1570.1) = -20.82$, $p = .001$) with a large effect size ($d = 1.01$). This pattern was repeated in 2017-18 where comparing the bottom (*M* = 0.68, *SD* = 0.1) and top (*M* = 0.78, *SD* = 0.09) tertiles identified a statistically significant difference ($t(1562.5) = -20.5$, $p = .001$), again with a large effect size ($d = 1.02$).

The passing standards diverged with important practical consequences. In 2016-17, the stringent standard was 29.5 (71.95%) and the lenient standard 24.5 (59.76%), out of a total of 41. In 2017-18 the stringent standard was 29.73 (61.94%) and the lenient standard 24 (50%), out of a total of 48. Table 2 summarises the impact of these illustrative standards on pass rates: applying the most stringent standards to the lowest-scoring medical school would lead to a fail rate of 39.52% in 2016-17 and 31.98% in 2017-18. Conversely, applying the lenient standard would lead to one medical school in 2016-17 and four in 2017-18 having no failing candidates at all.

Discussion

This paper explores the use of “common content” in high-stakes assessment. We show that candidates from different medical schools exhibit significant differences in scores on common content, and that these differences are partly generalisable – with schools differing across many domains. Importantly, a like-for-like comparison shows scores vary by enough that standard setting approaches that produce realistic fail rates – that is, fail rates that match those reported in similar assessments and for medical schools (26, 27) –may produce substantially different fail rates despite identical content and candidates being governed by the same regulatory environment. It is important for all medical educators – including not just assessment specialists but those responsible for clinical teaching – to be aware of such trends and to contribute to ongoing discussions on how to reach a consensus on standard setting for national licensing assessment. Even if the standards here are taken as illustrative only, the observed gaps in hypothetical passing rates emphasises the need for medical educators to agree whether standards should be uniformly applied, or locally-determined – as either approach will have substantial practical implications for any cross-institutional assessment.

These findings extend and support previous research. They suggest that gaps found in postgraduate attainment (6) may be partly attributed to differences in undergraduate medical education or attainment. The limited previous evidence of attainment gaps on common content has been reinforced (12). The emerging consensus that standard setting is a highly localised and subjective process influenced by context (10, 11) may be an explanation for the attainment gaps found here. Schools may be emphasising different areas and levels of knowledge, which then leads to significant differences on a shared assessment.

The evidence suggests that a common set of passing standards would impose very high (or low) pass rates on some schools. That this is not happening currently could be explained by standard setters being heavily influenced by the performance of their local students rather than applying a more objective national standard. Alternatively, it could be that material outside the common content is unique – implying less equivalence across schools. Differences in attainment between schools may be due to differences in cohort ability, or variations in the format and emphasis of assessment at each institution.

If medical schools have divergent standards due to “localisation,” significant disruption may occur if a single national standard is imposed. This may have substantial effects on passing rates and may disrupt workforce supply or affect stakeholder confidence in the exit assessment unless all stakeholders can work together to develop a sufficiently flexible approach that is acceptable to everyone.

This work shows that a shared regulatory environment alone does not necessarily develop homogeneity of performance, though it may have set an effective minimum standard if the standards of the lowest-performing medical school were found to be acceptable to all stakeholders. Importantly, however, given the known passing rates of UK medical schools, were such a “minimum standard” acceptable it would raise the concern that high-performing medical schools may be failing candidates who would be considered of passing quality by that minimum standard.

The extent to which educational diversity in content knowledge and topic specialisation is a desirable outcome (16) or a problem requiring regulation needs further discussion among educators and stakeholders. Either way, the experience of national assessment elsewhere suggests inevitable disruption during the implementation period (18–21).

The underlying ambiguity around current standard setting processes emphasises a challenge to medical education itself. If ongoing research on standard setting and empirical evidence suggests standard setting is not objective(10, 11) we must consider defensibility. We cannot judge from this work whether highly scoring medical schools are too stringent (reducing workforce supply when they fail a candidate) or whether lower scoring medical schools are too lenient (graduating those who are not fit to practise) or whether they are simply different in ways current regulatory processes fail to identify. It is extremely difficult to establish if there is a “correct” approach in a complex environment, and involvement of stakeholders throughout institutions affected by national licensing assessment is necessary.

Strengths and limitations

This study has several methodological strengths. The items have been reviewed and audited by experts then sat by many candidates across many institutions. This led to a high-quality dataset covering almost all candidates within a single regulatory environment. Our ability to compare schools on shared subsets of items allowed for a rigorous estimation of the impact of different standard setting regimes using empirical data. As such it serves as a valid model for a future licensing assessment. Importantly, we have opted for a widely understood, simple analytical approach via Classical Test Theory to make the results accessible to the largest possible audience.

Despite this, there were limitations. The pool of items is smaller than would be expected in a full-sized examination, and candidates also sat locally developed items which could not be included in this analysis. Some schools used relatively few common content items and the mechanism by which schools select or reject items – or how they are integrated into wider assessment and teaching – remains underexplored. Finally, while the accessibility of the work is a positive, more advanced methods such as Rasch inevitably offer additional analytic tools not employed in this analysis (24). However, it should be noted that the Rasch model of this dataset did not contradict any of the conclusions set out here.

Future research

Future research should explore the stability of these trends and expand the availability of common content material to better compare medical schools. It is important to identify the mechanisms behind these differences, and to ensure that a broad range of medical schools across the spectrum of performance are involved in standard setting any proposed licensing assessment. More generally, the subjectivity of standard setting methods suggests we must more thoroughly explore the link between performance at medical school and performance in the workplace – to see how graduates of different ability levels perform in work. Doing so will help ensure undergraduate medical education are appropriate to the role(s) candidates are trained for.

Conclusions

This study has highlighted differences in performance across UK medical schools. It is essential all stakeholders work together to better understand these differences and determine the extent to which the differences reflect desirable educational diversity – or indicate a need for change.

Declarations

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations. Approval for the work was granted by the University of Edinburgh Medicine and Veterinary Medicine ethics committee. All participants provided informed consent to participate in the research via their institutions.

Consent for publication

Not applicable.

Availability of data and materials

Due to the confidentiality and sensitivity of high-stakes assessment data, the datasets described in this study are not publicly available. If you wish for more information about the dataset or study, please contact David Hope (david.hope@ed.ac.uk).

Competing interests

We declare that the authors have no competing interests as defined by BMC, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

Funding

The Medical Schools Council Assessment Alliance funded this research.

Authors' contributions

David Hope wrote the manuscript text and developed the main analysis.

Avril Dewar and Matthew Homer contributed to the analysis and reporting of results.

David Kluth, Richard Fuller, and Helen Cameron provided expertise on assessment, advice on analyses, and interpreting results.

All authors contributed to the initial grant application that supported this work and the ethics application that allowed it to progress.

All authors contributed to and revised the manuscript.

Acknowledgements

Not applicable.

References

1. Cox M, Irby DM, Epstein RM. Assessment in medical education. N Engl J Med. 2007;356:387–96.
2. Norcini JJ, McKinley DW. Assessment methods in medical education. Teach Teach Educ. 2007;23:239–50.
3. Boursicot KA, Roberts TE, Pell G. Standard setting for clinical competence at graduation from medical school: a comparison of passing scores across five medical schools. Adv Health Sci Educ. 2006;11(2):173–83.
4. Devine OP, Harborne AC, McManus IC. Assessment at UK medical schools varies substantially in volume, type and intensity and correlates with postgraduate attainment. BMC Med Educ. 2015;15(1):146.
5. General Medical Council. Outcomes for Graduates. Manchester: General Medical Council; 2015.
6. McManus I, Elder AT, de Champlain A, Dacre JE, Mollon J, Chis L. Graduates of different UK medical schools show substantial differences in performance on MRCP(UK) Part 1, Part 2 and PACES examinations. BMC Med. 2008;6:5.
7. McManus I, Woolf K, Dacre J, Paice E, Dewberry C. The Academic Backbone: longitudinal continuities in educational achievement from secondary school and medical school to MRCP(UK) and the

specialist register in UK medical students and doctors. BMC Med. 2013 Nov 14;11(1):242.

8. Hope D, Cameron H. Academic performance remains predictive over a five year medical degree. Innov Educ Teach Int. 2018;501–10.
9. Wakeford R, Ludka K, Woolf K, McManus IC. Fitness to practise sanctions in UK doctors are predicted by poor performance at MRCGP and MRCP (UK) assessments: data linkage study. BMC Med. 2018;16(1):230.
10. Taylor CA, Gurnell M, Melville CR, Kluth DC, Johnson N, Wass V. Variation in passing standards for graduation-level knowledge items at UK medical schools. Med Educ. 2017;51(6):612–20.
11. Yeates P, Cope N, Luksaite E, Hassell A, Dikomitis L. Exploring differences in individual and group judgements in standard setting. Med Educ. 2019;53(9):941–52.
12. Chesser A, Cameron H, Evans P, Cleland J, Boursicot K, Mires G. Sources of variation in performance on a shared OSCE station across four UK medical schools. Med Educ. 2009;43:526–32.
13. MacDougall M. Variation in assessment and standard setting practices across UK undergraduate medicine and the need for a benchmark. Int J Med Educ. 2015 Oct 31;6:125–35.
14. Rimmer A. GMC will develop single exam for all medical graduates wishing to practise in UK. BMJ. 2014 Oct 1;349:g5896.
15. Van Der Vleuten CP. The assessment of professional competence: developments, research and practical implications. Adv Health Sci Educ. 1996;1:41–67.
16. Allawi L, Ali S, Hassan F, Sohrabi F. UKMLA: American dream or nightmare? Med Teach. 2016;38(3):320.
17. Archer J, Lynn N, Coombes L, Roberts M, Gale T, Bere SR de. The medical licensing examination debate. Regul Gov. 2017;11(3):315–22.
18. Schuwirth L, Bosman G, Henning RH, Rinkel R, Wenink ACG. Collaboration on progress testing in medical schools in the Netherlands. Med Teach. 2010 Jan 1;32(6):476–9.
19. Tio RA, Schutte B, Meiboom AA, Greidanus J, Dubois EA, Bremers AJA. The progress test of medicine: the Dutch experience. Perspect Med Educ. 2016 Feb;5(1):51–5.
20. Burk-Rafel J, Santen SA, Purkiss J. Study Behaviors and USMLE Step 1 Performance: Implications of a Student Self-Directed Parallel Curriculum. Acad Med. 2017 Nov;92(11S):S67.
21. Pershing S, Co JPT, Katznelson L. The New USMLE Step 1 Paradigm: An Opportunity to Cultivate Diversity of Excellence. Acad Med. 2020 Sep 1;95(9):1325–8.
22. Faul F, Erdfelder E, Lang A-G, Buchner A. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007;39(2):175–91.
23. Schumacker RE, Smith EV. A Rasch Perspective. Educ Psychol Meas. 2007 Jun 1;67:394–409.
24. Tavakol M, Dennick R. Psychometric evaluation of a knowledge based examination using Rasch analysis: An illustrative guide: AMEE Guide No. 72. Med Teach. 2013 Jan 1;35(1):e838–48.
25. Allen MJ, Yen WM. Introduction to measurement theory. Monterey, CA: Brooks/Cole;

26. Maxwell SRJ, Coleman JJ, Bollington L, Taylor C, Webb DJ. Prescribing Safety Assessment 2016: Delivery of a national prescribing assessment to 7343 UK final-year medical students. Br J Clin Pharmacol. 2017;83(10):2249–58.
27. Arulampalam W, Naylor RA, Smith JP. A hazard model of the probability of medical school drop-out in the UK. J R Stat Soc Ser A Stat Soc. 2004;167:157–78.

Tables

Due to technical limitations, table 1,2 is only available as a download in the Supplemental Files section.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.jpg
- Table2.jpg