Title: Unhappiness and dissatisfaction in doctors cannot be predicted by selectors from medical school application forms: A prospective, longitudinal study

Authors:

Ic I McManus (i.mcmanus@ucl.ac.uk)
Sheeraz S Iqbal (sheeraziqbal@gmail.com)
Amuthan A Chandrarajan (zchaxc2@ucl.ac.uk)
E E Ferguson (Eamonn.Ferguson@nottingham.ac.uk)
Joanna J Leaviss (jcl@psychology.nottingham.ac.uk)

Version: 2 Date: 25 October 2005

Author's response to reviews: see over
Unhappiness and dissatisfaction in doctors cannot be predicted by selectors from medical school application forms: A prospective, longitudinal study

Responses to Editorial and Referees’ Comments: Responses are shown in italics.

Editorial Comments

We will also need you to clarify the ethical approval of this study. Did you obtain the doctors’ consent to distribute the forms they had submitted as medical students with their personal statements and referees’ opinions? It is mentioned that the forms were anonymised, but we would expect that personal information of this nature may still identify the clinicians.

Response: Under the provisions of the UCL Research Ethics Committee, this study was exempt from the need for formal application (http://www.grad.ucl.ac.uk/ethics/index.php); see also http://www.grad.ucl.ac.uk/ethics/exemptions.php. However all research carried out in the Department of Psychology is also covered by the Departmental Research Ethics Committee, and this study was seen and approved by the Committee.

We did not obtain the specific consent of the doctors for the use of the personal statements and referees’ reports, but when returning the 2002 questionnaire we included on the last page the following statement:

“This questionnaire is entirely for the purposes of research. Both for your own reassurance and as a normal part of research ethics, we would be grateful if you would acknowledge the declaration by initialling it.

“DECLARATION: This questionnaire is entirely for the purposes of educational research, its contents will be kept strictly confidential, will not be made known to anyone outside of the research study, and will not otherwise be disclosed or published except in an aggregated form in which individuals cannot be identified.

I C McManus MA MD PhD FRCP”

By initialing the form the doctors had therefore explicitly agreed to the general use of information in the survey for research purposes. The application forms were not provided for research purposes but as part of the normal process of university application (and in many cases the referees’ reports would not have been known to the applicants).

We took great care in anonymising the personal statements and referees’ reports, and we do not honestly believe that “personal information of this nature may still identify the clinicians”. We are happy to provide examples of the forms should this be doubted.

We have modified the statement under Ethical Permission to make the situation clearer, and have provided the web addresses for the UCL Research Ethics Committee.

Could you also please go through the manuscript formatting checklist, the link to which is provided at the bottom of this e-mail, and ensure that your revised manuscript conforms to
all of the points. It is important that your files are correctly formatted.

Response: We hope that the manuscript meets all the specific requirements.

Reviewer's report 1

David Alan Powis

General

This is a concise and well-written paper that addresses an important issue in the context of medical student selection: what is the predictive value of a (candidate’s) personal statement and the referee report on the application form originally submitted in connection with medical school admission? The conclusion that the statements do not predict whether the doctors, when graduated and into their careers, will be happy and satisfied with their chosen profession will be of considerable interest to medical student selectors. The text flows well and leads the reader logically through the story. However there are four issues that I consider should be addressed by the authors before an editorial decision on publication is made. I have also listed a few minor points that should be considered by the authors.

Response: We thank the referee for these kind comments, and hope that the responses below will make the paper easier to comprehend.

-------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

1. There are two statements made at the bottom of page 7 that at first appear to be quite convincing: ‘there was no correlation between the difference in Big Five personality scores…..Assessors are not therefore implicitly assessing aspects of personality’, and ‘The difference in educational achievement between the doctors in a pair…..it seems likely that the assessors are judging that the doctor……’.

These statements and their conclusions fit in very well with the line the argument is taking and, as such, tend not to attract a critical gaze. However read with a good memory of what was written in the methods section the picture suddenly becomes confused. (The methods describe that the assessors completed a ‘brief assessment of personality…..Big Five etc’ but the methods make no explicit mention that the doctors being assessed had earlier completed the Big Five. Therefore one begins to wonder why the similarities and differences between the assessors’ characteristics should give rise to the conclusions drawn! (It would help to have specific sub-sections in both the methods and results sections dealing with the psychometric tests, and who took them. in the context of this study)

Response: We agree that this is confusing and have attempted to clarify matters. In the present study we asked the assessors to complete a Big Five personality measure. However we have also, in the 1991 cohort study, asked the doctors to complete a Big Five Personality measure – on two occasions in fact, and in the present study we are referring to the measure completed in the 2002 follow-up. We have added a note to the end of the Method section to emphasise the potential for confusion, and we have tried always to use the terms 'assessors' and 'doctors' in a consistent way. We have also explicitly described in the Methods section that we had Big Five measures in the doctors.
2. Accepting now that it is the doctors’ and not the assessors’ characteristics that underpinned the conclusions drawn, why does table S1 not show any significant difference between the educational achievement of happy and unhappy doctors?

Response: Happy and unhappy doctors do not differ in terms of their academic achievement (and that is what is shown in table S1). However doctors that the assessors think will be the happier do have higher academic achievement (and that is the basis for the correlation reported in the last paragraph of the Results section). We have added a sentence to the final paragraph of the results section to make this clearer.

3. The personal statement (of the candidate) and the referee’s report (about the candidate) would be presumed to provide different information yet the two were apparently read by the assessors as a single data source. There are several possibilities of relevance to the present study, for example the PS ad RR could give a consistent picture of the candidate. Alternative they might emphasise different aspects of the candidate’s character, personality and achievements. Or the RR might contradict parts of the PS. I presume that the authors did not have the assessors look at the PS and RR independently, but I would like to read some discussion of why it was decided that both together would be the better option in view of the objective of the study. (Incidentally, do the authors have any idea of the relative weighting the assessors gave to PS and RR in their determinations?)

Response: We had originally thought of a study design in which some assessors saw only personal statements, some saw only referees’ reports, and some saw both. However we realised that a more efficient design in terms of participants’ time was to use personal statements and referees’ reports together, and only if assessors could carry out the task using both sets of information would we then need to carry on and do additional studies in which just one form of information was presented. We agree that it is theoretically possible that in some way the two forms of statement contradict one another, but we feel it is unlikely in practice. Others who disagree are, of course, welcome to do additional studies to demonstrate such a process. As far as the relative weighting made by the assessors, we have no formal data, but can say that when assessors commented they referred to both sets of information.

4. A reader could draw a wrong conclusion from a sentence in the discussion (bottom, p8): ‘…it might be more beneficial to use psychometrically assessed personality scores than……’. A reader might get the impression that you are saying that the Big 5 data reported here for the doctors is of predictive value. (Incidentally what do you mean by the word ‘beneficial’ in this context?)

Response: We agree that ‘beneficial’ was not a happy choice of phrase, and have replaced it with, “Even should structured personal statement coding be correlated with personality, it might still be more effective to use psychometrically assessed personality scores than personal statements, if it were desired to select on the basis of personality.” We do believe that Big Five measures are predictive of career satisfaction, and have reported previously (BMC Medicine, 2004) that Big Five measures during PRHO posts predict satisfaction five or more years later in entirely different posts. We do not, as yet, have data showing that Big Five measures at selection can predict satisfaction five years after qualifying. However, given the stability of the Big Five across the lifespan, and the predictive validity of the Big Five for happiness and satisfaction in life in general, we have little doubt that it will be predictive in medical school applicants and doctors.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1. ‘satisfied’ and ‘happy’ are not, strictly, synonyms, but they are used in this paper as if they are. However, in the present context it is reasonable to use them as such, but a statement should be
made to this effect. (I wonder if the assessors, or the doctors, made a distinction between the terms. Also when the authors were drafting the paper what determined the choice of ‘happy’ or ‘satisfied’ or ‘happy/satisfied’ in the text?)

Response: We agree that we use the terms as if synonymous, and we have not attempted to differentiate between them in the present paper. Our use of the terms in the paper is intended to show that we regard the terms as equivalent, and the term “happy/satisfied” is precisely the one used in the question answered by assessors (see method section). We have added to the end of the appropriate paragraph in the Method section the sentence, “It should be emphasised that we made no attempt to differentiate ‘happiness’ and ‘satisfaction’, and for the purposes of this paper these words should be regarded as synonyms.”.

2. ‘data were’ rather than ‘data was’ (page 2 methods, and possibly elsewhere).

Response: It’s fair cop, guv! We have now checked all uses of the term ‘data’, and are pleased to say that all were used appropriately.

3. Is it grammatically correct to say ‘none of the groups of assessors were above chance expectations…….’? (p 2 results).

Response: It certainly doesn’t read well, we agree, and have replaced it with, “were significantly different from chance expectations”.


Response: We didn’t use this phrase, and neither should we have, since at no point do we pair our assessors. Perhaps we have misunderstood the force of the comment?

5. ‘with a fifth or more of junior doctors….’ (p 3 background).

Response: Corrected.

6. ‘but because of time constraints…..’ Whose time constraints? Do you mean to say because we didn’t wish to overload the assessors? (p5)

Response: We did indeed, and we have now said that.

7. Results, p6 and elsewhere: p rather than P for probability.

Response: Journals differ on this one. We have no idea of house style on this one, since our 2004 paper contains ‘p’ and our 2005 paper contains, in one table, a mixture of ‘p’ and ‘P’. We will leave it to the sub-editors to sort out.

8. Eighty four rather than 84 at start of sentence (p7)

Response: Done, but again, we have no idea of house style.

9. Middle para, p10 should be in results section rather than discussion section.

Response: Agreed, except that no results are actually presented here, and we merely wanted to flag this up as an approach, rather than have to present formal measures. We have therefore left it as it is.
10. Would it help to have an example PS and RR in the supplementary information section to show non UK readers what these typically contain?

Response: It would be helpful, but while we are happy to use this information for research purposes, we are not convinced that it is appropriate to publish it (and see the editorial comments above about possible identifiability).

11. Give n=40 (or whatever) for the happy and unhappy group data given in table S1

Response: Done.

12. what is the value for ‘n’ assessors reported in Table S3?

Response: This has now been added.

Discretionary Revisions (which the author can choose to ignore)

None

Reviewer's report 2

Ed Peile

General

This study is wholly original in the question that it addresses and in the methodology adopted. The dichotomised study population may or may not inform us reliably on the middle-of-the-road students and doctors. Overall, I see this as a useful study, debunking some of the myths on which selectors make judgements and causing us to rethink how we can select students who are likely to find contentment in medicine.

Response: We thank the referee for these generous comments. We also acknowledge that our study may not be able to detect a non-linear effect involving individuals in the middle of the distribution, but feel that is sufficiently unlikely in theoretical terms for it to be ignored on a first attempt at the question.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

There is a central point which needs consideration. The authors get a bit tied up in relating intellectual ability to happiness. The second para, Page 3 asserts that since emotional achievement and intellectual ability have been found to be unrelated to stress, burnout and dissatisfaction, assessors are likely to avoid these measures and to use personal statement or referee’s report. I doubt this is true, because I doubt that assessors are mindful of the research evidence (confirmed in this study) on this point and I think that intuitively many assessors would
presume that brighter students might thrive better than less able ones in such a highly academic course. Contrast this with Para 1 on page 8 where it says, “It seems likely that assessors are judging that the doctor with the higher predicted educational achievement will be the happier, more satisfied doctor.” Thus the authors' findings undermine a key premise of study design.

Response: The referee is correct that we are confusing what is known about in the literature and what it is likely that the referees themselves think about the process (and we agree that it is unlikely that they are well-informed about current research). We have therefore replaced the entire last sentence in this paragraph with, “Empirically, educational achievement and intellectual ability have been found to be unrelated to stress, burnout and dissatisfaction with medicine [2], and therefore the most likely source for information which might be able to predict happiness with a medical career is the applicants’ personal statement and their referees’ reports.” We hope this clarifies matters and

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

None

Discretionary Revisions (which the author can choose to ignore)

None

Reviewer's report 3

Michael J Goldacre

General:

This is an interesting and important paper. The conclusions - that the assessors did not predict who would be happy or unhappy doctors - deserve to be widely known. They are so important that it might be prudent for the authors to recommend that the findings should be confirmed or refuted in a larger study (though I see no reason to expect that they would be different).

Response: We thank the referee for his comments, and are glad that the results are seen as important. We have followed his advice, and added a sentence at the end of the first paragraph of the discussion, saying, “In view of the implications of the present findings for current methods of student selection in the UK (and elsewhere), it would be reassuring if other studies, in other settings, and with different outcome criteria, were to find similar results, and we hope that such studies will be carried out.”

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct):

1. I found it took time, and two or three careful readings, to work out the distinction between the findings in Figure 2 and those in Figure 3, partly, I think, because the legends are so similar. Figure 2 is about the predictive power of the assessors' judgements in correctly identifying the happy doctor, and Figure 3 is about between-observer agreement in the judgements that they made. It might help if the opening phrases of the legend for Figure 2 said something like: "Predictive power of the individual assessors' judgements in correctly identifying the happy doctor in each of 20 pairs". And if the opening of the legend to Figure 3 said: "Between-assessor agreement in the number of pairs in which both assessors correctly identified the happy doctor". Whatever the detail, I think that the descriptions both on the legend and in the associated text need to be made clearer for ease of understanding. I also think that it would be helpful to the reader if the Abstract clarified that there were two measures of study outcome - first, the predictive power of the assessors' judgements, and, second, the extent of agreement between the assessors on each pair.

Response: We agree that the results are somewhat confusing, because of the two ways of looking at the same data (across the rows and down the columns, as it were). In response to Referee 1 we have tried to clarify other aspects of the presentation of results, and in particular have tried to be consistent about when we are referring to judgements grouped by assessors and when to judgements grouped by pairs of doctors. We have re-written much of the figure captions and hope that things are now somewhat clearer. We’ve avoided terms such as 'between-assessor agreement' mainly because we did not want to look at pair-wise measures of agreement, and nor is this quite like kappa or more traditional measures of agreement. Neither does any equivalent phrase come to mind for the results of figure 2.

3. There are some minor typing errors:

- Background, 2nd line: "or" (not "of")
- Background, first para, penultimate line: delete comma after whether
- Figure 2, legend, penultimate line: should read "the dark blue bar shows a pair".

Response: All corrected.

Discretionary Revisions (which the author can choose to ignore)

1. In Figure 2, there appears to be a peak at 10 in the blue bar. Using the authors' data, and a chi-squared test, I think that the assessors' judgements are more closely centred around 10 than would be expected by chance. Do the authors agree? This doesn't change their conclusions - if anything, it reinforces them.

Response: This is an intriguing comment. Although at first sight it might seem that the excess of points at 10 might reinforce the conclusion, I do worry when data appear to be underdispersed relative to binomial variation. I therefore carried out a Monte Carlo simulation, using SPSS, to see how unlikely it was to have 25 cases of exactly 10 out of 20, when there are 96 binomial replications. I replicated the situation 10,000 times and looked at the distribution of the number of times out of 96 in which there were exactly 10 out of 20 cases. The mode was 17, of which there were 1064 cases (10.64%) – and this is exactly what the expected distribution shows in figure 2). However there were 105 (1.05%) cases where exactly 25 of the 96 were ten, and a
further 120 in which 26 or more were ten, giving 225 (2.25%) in which 25 or more cases were exactly ten. This is of course a one-tailed test (and very a posteriori at that), and hence a two-tailed probability of obtaining a result as extreme as 25 or more in which exactly ten our of 20 are correct (given an N of 96) is p=.045). This is close enough to the conventional significance threshold, coupled with the fact that the event was only noticed because of an eye-ball test, to mean that it can probably be safely ignored (and we would not, for instance, have been as interested if the peak had been at 9 or 11). Overall, it is an interesting conclusion, but on balance I think it is probably not worth taking it further. Having said that, I am aware, but have not simulated, the expectations for a Lexian distribution, which this probably is, and which would be expected to be more overdispersed than a pure binomial.

2. In the Abstract, Methods, penultimate line, I think that it would be useful to replace the word "judged" with the phrase "used the doctors' original applications to judge".

Response: Agreed; and we have also taken the opportunity to clarify a number of other minor confusing points in the abstract, which we hope is now improved.