

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Older doctors and progression through specialty training in the United Kingdom: a cohort analysis of General Medical Council data.
<b>AUTHORS</b>	Pyne, Vicky; Ben-Schlomo, Yoav

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Kevin Hayes St George's University London, UK  I am a first author on one of the cited references in this paper but None declared
<b>REVIEW RETURNED</b>	04-Jun-2014

<b>GENERAL COMMENTS</b>	<p>There is a clear, important research question and testing of a Null hypothesis with an appropriate study design. The conclusions are clear and not overstated. Strengths and weaknesses are clearly stated.</p> <p>I would take issue with strength 3 at the beginning as written:</p> <p>"Results are counter to prevailing beliefs and research regarding mature medical students showing that despite doing better at university, they appear to 'do worse' once they have become doctors"</p> <p>The results do not necessarily mean these doctors "do worse" it is simply a measure of specific outcomes - this is actually acknowledged in the discussion section, so this assertion needs to be clarified in line with the discussion section.</p> <p>The study does offer potential reasons for the findings but clearly and correctly states that the study does not prove them just raises more potential qualitative research questions about this important cohort of doctors.</p> <p>There are several mentions of other "minority groups" having similar outcomes. They may have similar outcomes but they are different cohorts and cannot necessarily be compared to this cohort - it needs to be clearer in the discussion what "association" if any these findings have</p> <p>Overall a very good study - a few minor revisions as above only</p>
-------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<b>REVIEWER</b>	John C. Mclachlan Durham University
<b>REVIEW RETURNED</b>	14-Aug-2014

<b>GENERAL COMMENTS</b>	<p>This is an important topic which is likely to court controversy. I believe it deserves publication in principal but some important issues should be addressed prior to this.</p> <p>1) The manuscript should also be reviewed by someone with very high level understanding of previous work in this area. I suggest Professor Chris McManus, UCL and/or Dr Paul Tiffin, Durham. I believe more sophisticated analyses could be done, and would prove informative.</p> <p>2) The authors are correct that large numbers give small p values. What matters is the effect size, and the authors should calculate a value for this. An Odds Ratio, of course, is not an effect size.</p> <p>3) The authors should comment that the data is right-censored – there is no ‘excellent’ category above mere progression. If therefore older graduates showed greater variance than school leaver graduates, then this would be undetectable in the current study. It could be that mature students do generally make better doctors overall, as a variety of soft measures seem to suggest, but a small proportion of them struggle for the career reasons mentioned. These results therefore do not necessarily contradict previous understandings.</p> <p>4) The authors should refer more widely to previous work on age effects in doctors – for instance Norcini et al (2013) Medical Care 51;1034-1039.</p>
-------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<b>REVIEWER</b>	Chris McManus UCL, UK
<b>REVIEW RETURNED</b>	04-Sep-2014

<b>GENERAL COMMENTS</b>	<p>This is an interesting paper, which looks at career progression of UK trained doctors through ARCP in relation to age at graduation. It makes the strong claim that doctors who are older at qualification perform less well. That however does seem to contradict other evidence (and, for instance, older graduates on accelerated undergraduate courses seem to perform substantially better in examinations in medical school, and therefore presumably are likely to do better on postgraduate training; see BMC Med Ed, Mahesan et al, 2011, 11:76). That raises a number of questions about the present study which need resolving. In view of the recent paper by Tiffin et al in the BMJ on ARCP I have used that in comparison with the present study, not least as the datasets seem to overlap substantially but the conclusions potentially seem incompatible.</p> <p>1. ARCP/RITA classifications are complex, and Tiffin et al chose as ‘satisfactory codes 1, 6, C and G, whereas the present chooses as unsatisfactory 3,4,7.3, D and E, which is not the complement of the Tiffin classification. Tiffin et al also use ordinal regression, since the classifications can be classified in some form of hierarchy. There is an argument for also carrying out the current study using the Tiffin approach.</p> <p>2. Tiffin et al also excluded “ARCP outcomes related to examination failure”. That raises important questions about whether age is related to academic or non-academic problems, and it would be useful to have similar analyses for the present data.</p> <p>3. Age at qualification is not an easy variable. The current authors use a cut point of 29+ at graduation. However they then divide older graduates into 29 to 31 and 32+ and find a dose-response effect.</p>
-------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	<p>However the classification of those of 28 or less is far from obvious, and it is not clear that they are homogenous (i.e. without a dose-response effect). 32+ are different from 29-31 and &lt;29, and it therefore seems possible that the so-called "Normal age" group [surely an unfortunate bit of phrasing?] is also heterogenous. If 23-24 is a typical post-school leaving age, and 26-27 is a typical age for graduates to qualify, then there could well be variation here. Given the large Ns then surely the data needed dividing up into something like &lt;-23, 24,25,26,27,28, 29-31, 32+ in order to see what is going on. At present the classification is too simplistic.</p> <p>4. Age and age at qualification are separate, and confounded with cohort. Tiffin et al found no effect of age at all which seems difficult to reconcile with the current data. Some exploration/explanation is required.</p> <p>5. I like the idea of 'graduate-friendly' medical schools, but a single cut-off seems too easy. Could we please see a plot of ARCP problem rates by percentage of 'older' graduates from each medical school. Medical schools are known to differ strongly in their success rates at MRCP/MRCGP (see the McManus paper in the same BMJ as the Tiffin paper), and in a proper analysis there would be dummy variables for medical schools in the analysis.</p> <p>6. The tables would benefit from including Ns on a systematic basis.</p>
--	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

### VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name Kevin Hayes

Institution and Country St George's University London, UK

Please state any competing interests or state 'None declared': I am a first author on one of the cited references in this paper but None declared

There is a clear, important research question and testing of a Null hypothesis with an appropriate study design. The conclusions are clear and not overstated. Strengths and weaknesses are clearly stated.

I would take issue with strength 3 at the beginning as written:

"Results are counter to prevailing beliefs and research regarding mature medical students showing that despite doing better at university, they appear to 'do worse' once they have become doctors"  
The results do not necessarily mean these doctors "do worse" it is simply a measure of specific outcomes - this is actually acknowledged in the discussion section, so this assertion needs to be clarified in line with the discussion section.

We thank Mr. Hayes for this comment and agree with his suggestion. We have amended this accordingly.

The study does offer potential reasons for the findings but clearly and correctly states that the study does not prove them just raises more potential qualitative research questions about this important cohort of doctors. There are several mentions of other "minority groups" having similar outcomes. They may have similar outcomes but they are different cohorts and cannot necessarily be compared to this cohort - it needs to be clearer in the discussion what "association" if any these findings have We are sorry if this was not clear. We were referring to the similarly increased risk of problems with progression as assessed by the ARCP process. We have clarified this.

Overall a very good study - a few minor revisions as above only

We thank Mr. Hayes for his positive comments and suggestions.

---

Reviewer: 2

Reviewer Name John C. Mclachlan

Institution and Country Durham University

Please state any competing interests or state 'None declared': None Declared

This is an important topic which is likely to court controversy. I believe it deserves publication in principal but some important issues should be addressed prior to this.

1) The manuscript should also be reviewed by someone with very high level understanding of previous work in this area. I suggest Professor Chris McManus, UCL and/or Dr Paul Tiffin, Durham. I believe more sophisticated analyses could be done, and would prove informative.

2) The authors are correct that large numbers give small p values. What matters is the effect size, and the authors should calculate a value for this. An Odds Ratio, of course, is not an effect size.

We generally prefer to use the simplest statistical methods if we can as this is easier to interpret for the average reader. In our experience, more sophisticated analyses can occasionally be more informative but in general support the conclusions from simpler analyses. We agree about the issue of effect estimates rather than p-values however we are confused by the statement that an odds ratio is not an effect estimate. We note that this is the same effect estimate used by Tiffin et al (BMJ 2014).

3) The authors should comment that the data is right-censored – there is no 'excellent' category above mere progression. If therefore older graduates showed greater variance than school leaver graduates, then this would be undetectable in the current study. It could be that mature students do generally make better doctors overall, as a variety of soft measures seem to suggest, but a small proportion of them struggle for the career reasons mentioned. These results therefore do not necessarily contradict previous understandings.

This is a very cogent point and we agree with Dr. Mclachlan that because of a potential ceiling effect, we could be missing a bimodal distribution so that mature doctors could be both having problems and doing excellently. We have added this to the limitations and discussion.

4) The authors should refer more widely to previous work on age effects in doctors – for instance Norcini et al (2013) Medical Care 51;1034-1039.

We thank Prof. Mclachlan for highlighting this reference and note that this paper relates to a doctor's years since graduation as opposed to their age.

This is well worth pursuing!

We thank Prof. Mclachlan for his positive comments and suggestions.

---

Reviewer: 3

Reviewer Name Chris McManus

Institution and Country UCL, UK

Please state any competing interests or state 'None declared': None declared

This is an interesting paper, which looks at career progression of UK trained doctors through ARCP in relation to age at graduation. It makes the strong claim that doctors who are older at qualification perform less well. That however does seem to contradict other evidence (and, for instance, older graduates on accelerated undergraduate courses seem to perform substantially better in examinations in medical school, and therefore presumably are likely to do better on postgraduate training; see BMC Med Ed, Mahesan et al, 2011, 11:76). That raises a number of questions about the present study which need resolving. In view of the recent paper by Tiffin et al in the BMJ on ARCP I have used that in comparison with the present study, not least as the datasets seem to overlap substantially but the conclusions potentially seem incompatible.

We thank Prof. McManus for his positive comments and agree that there is overlap between the datasets we have used and that by Tiffin et al. (we were unaware of this work at the time we were conducting our analyses) however we do not believe that the findings are contradictory (see comments below).

1. ARCP/RITA classifications are complex, and Tiffin et al chose as 'satisfactory codes 1, 6, C and G, whereas the present chooses as unsatisfactory 3,4,7.3, D and E, which is not the complement of the Tiffin classification. Tiffin et al also use ordinal regression, since the classifications can be classified in some form of hierarchy. There is an argument for also carrying out the current study using the Tiffin approach.

We have compared our codes with Tiffin et al and for RITA they are complementary as the missing code F reflects out of programme experience. There are discrepancies for the ARCP codes. We specifically chose to not include code 2 and we justified our reasons for this in the paper. In the experience of one of the authors as an ARCP assessor, code 2 is most often used for trainees with inadequate documentation not poor progression. The panel will give an outcome 2 with the proviso that this is converted to a 1 if the trainee provides this within a reasonable time frame.

2. Tiffin et al also excluded "ARCP outcomes related to examination failure". That raises important questions about whether age is related to academic or non-academic problems, and it would be useful to have similar analyses for the present data.

Tiffin et. al. included international graduates which we specifically excluded and hence looks at PLAB examination results. Our hypothesis was to test whether more mature doctors have greater difficulty in progressing through postgraduate training and clearly examination failure is one potential reason. We would not wish to exclude this reason for failure to progress and specifically mention this potential explanation in the discussion. Furthermore, the ARCP dataset does not include the reason for failure (whether academic or non-academic).

3. Age at qualification is not an easy variable. The current authors use a cut point of 29+ at graduation. However they then divide older graduates into 29 to 31 and 32+ and find a dose-response effect. However the classification of those of 28 or less is far from obvious, and it is not clear that they are homogenous (i.e. without a dose-response effect). 32+ are different from 29-31 and <29, and it therefore seems possible that the so-called "Normal age" group [surely an unfortunate bit of phrasing?] is also heterogenous. If 23-24 is a typical post-school leaving age, and 26-27 is a typical age for graduates to qualify, then there could well be variation here. Given the large Ns then surely the data needed dividing up into something like <-23, 24,25,26,27,28, 29-31, 32+ in order to see what

is going on. At present the classification is too simplistic.

Our hypothesis was for “mature students” for which there is no standard definition so we chose one that we feel has strong face validity. We agree with Prof. McManus that there may be heterogeneity in the baseline group. We have therefore repeated the analyses with this group sub-divided as he has suggested. These additional analyses have been informative and we have now added a new supplemental table. The data show that, as before, the oldest group ( $\geq 32$  years) are markedly worse than the younger groups but in fact there is little difference in the 26-31 year group in the unadjusted analysis though adjustment slightly increase the odds ratio for the 29-31 year group. If anything the threshold for increased problems with progression appears to be at 26 years and above; lower than we have previously shown. We do not feel this in anyway invalidates our a priori definition of “mature” student but this post-hoc analysis is of interest and also demonstrates the non-linearity of the relationship with age. We have added some discussion of this finding into the paper.

We have also changed the name of the ‘Normal Age’ group to ‘Younger group’.

4. Age and age at qualification are separate, and confounded with cohort. Tiffin et al found no effect of age at all which seems difficult to reconcile with the current data. Some exploration/explanation is required.

There are two issues that need to be considered here. One is the differences in the datasets and analysis strategy and the other is the interpretation of our results with respect to Tiffin et al.

Though the two datasets are from the same primary source and do indeed overlap temporally, they are different as Tiffin et al includes all non-UK graduates whilst these were explicitly excluded from our analyses. The Tiffin paper uses age at ARCP (parameterised as a continuous variable) whilst we have used age at provisional registration in much larger categorical groupings. As our hypothesis relates to age at registration and not age at ARCP, the Tiffin results are less relevant to this, though clearly they will be correlated to some degree.

In the univariable analysis section of the paper by Tiffin et al they state that “Increasing age (odds ratio 1.04, 1.03 to 1.04)... with increased odds of obtaining a less satisfactory outcome at ARCP.” The smaller effect estimate reflects the way that age was parameterised but is consistent with our results (though as we have now shown the assumption of linearity may or may not be valid). The “Age” coefficient in table 4 (Odds ratio 1.00, 95% CI 0.98 to 1.01) which appears null is harder to interpret as this from a multivariable model which also includes two interaction terms with age (age and UK experience, age and non-white ethnicity).

5. I like the idea of ‘graduate-friendly’ medical schools, but a single cut-off seems too easy. Could we please see a plot of ARCP problem rates by percentage of ‘older’ graduates from each medical school. Medical schools are known to differ strongly in their success rates at MRCP/MRCGP (see the McManus paper in the same BMJ as the Tiffin paper), and in a proper analysis there would be dummy variables for medical schools in the analysis.

Unfortunately the GMC aggregated some of the London medical schools into one group before releasing the data but we have now rerun the analysis as suggested with a dummy variable for each school and this makes almost no difference to the results. We have added a sentence to the results section to inform the readers of this analysis.

6. The tables would benefit from including Ns on a systematic basis.

We agree with this suggestion and have made this amendment.

We thank Prof. McManus for his positive comments and suggestions.

Finally, the authors would like to highlight an additional point added to the discussion after this review. This was suggested following a presentation to the Severn Deanery Foundation Programme Away Day (Sept. 2014).

#### VERSION 2 – REVIEW

<b>REVIEWER</b>	Mr Kevin Hayes MRCOG, Senior Lecturer and Consultant in O+G and Medical Education  St George's University London, UK  I am the first author on one of the quoted references but do not believe i have any conflict of interests
<b>REVIEW RETURNED</b>	02-Oct-2014

<b>GENERAL COMMENTS</b>	The responses to all of the reviewer queries have been well thought through and appropriate and the study does not try to overstate the conclusions or the reasons for the observed effects. The addition of the suggested age sub-analyses is informative and further backs up the original findings.  Many thanks for an interesting paper about an important cohort of trainees - in the future it would be extremely interesting to look at qualitative reasons to try and triangulate why the observed effect has been seen
-------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<b>REVIEWER</b>	Chris McManus Univeristy College London, UK
<b>REVIEW RETURNED</b>	12-Oct-2014

<b>GENERAL COMMENTS</b>	This paper has been much improved in the revision, and the authors have provided helpful and detailed comments on their responses to the various questions.  I am happy to see the paper go forward, and I have said "Accept". However in the past few days I have seen the GMC's new State of Medical Education and Practice in the UK for 2014 (available at the GMC website). The authors might be interested in figure 46, which suggests that UK graduates joining the register over the age of 30 have a higher rate of sanctions and warnings. The GMC do not test for significance but it appears to be significant in females but not males (and I haven't tested the interaction). The authors may however care to make a reference to this recent work.  THE GMC report also shows both the strength of the GMC in publishing such data, and its weakness as it merely reports percentages rather than asking about the details of effects, or indeed their significance. The present paper shows that merely
-------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	dividing at 30 is less than optimal.
--	--------------------------------------

## **VERSION 2 – AUTHOR RESPONSE**

Response to Reviewer 1: We thank Mr. Hayes for his positive comments and hope to do further work, both quantitative and qualitative to determine the reasons for these results.

Response to Reviewer 3: We thank Prof. McManus for his positive comments and have noted the paper and specific figure in question. We observe that, when non-UK graduates are excluded (as we have done in our study), the effect is less apparent and, as he states, may or may not be statistically significant. We have added a reference to this in the discussion of the paper and will contact the GMC for more information on this data for future publications in this area.