

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

TITLE (PROVISIONAL)	Differences in progression by surgical specialty: a national cohort study
AUTHORS	hope, carla; Lund, Jonathan; Griffiths, Gareth; Humes, David

VERSION 1 – REVIEW

REVIEWER	Robinson, David Cardiff & Vale University Health Board, General Surgery
REVIEW RETURNED	15-Jul-2021

GENERAL COMMENTS	<p>This paper describes the variation in outcomes at ARCP across nine of the ten surgical subspecialties. Overall, it is well written, the methodology is robust, well described and uses techniques previously demonstrated to be reliable and the study is based on a very large population of trainees/ARCPs. There are a few minor changes that I would recommend -</p> <p>Abstract Add the associated p-value to the statistics documented in the abstract</p> <p>Introduction Line 7 page 9 – comma not needed after “a”</p> <p>Methods Should patient and public involvement state that it was not sought as no patients were included in the study?</p> <p>Results Median age (and IQR) at ARCP seems slightly higher than I would have envisaged, is this the overall median across all ARCPs or is this the median age of final ARCP/CCT? I think this is worth rechecking.</p> <p>Table 2 – Shouldn't the cohort populations described under gender match the cohort numbers relating to the PMQ region descriptors? I appreciate the percentages are rounded to the nearest whole number but the crude numbers should already be whole numbers? This is not the case for GS and ENT e.g. total cohort for GS gender = 2630 (1780+850) vs. cohort for PMQ = 2620 (2035+90+495)</p> <p>The figures for Max Fax used in table 2 need to be looked at again as they do not correlate correctly with the cohort population/percentages - the figures in parentheses do not correlate with crude number for each cohort. i.e. it states that 30</p>
-------------------------	--

	<p>individuals are EEA which is described as 10%, yet 85 are IMG described as 4%?</p> <p>Table 3 – I feel that the p-values established on chi square test should be included if you are going to state that all specialties were significant except paediatric surgery.</p> <p>I would recommend separating outcome 3's (extended training) from the outcome 4's as I think the number of outcome 4's by specialty is an interesting finding and should be highlighted. It would also be interesting to know whether there is any significance statistically between specialties in relation to outcome 4's.</p> <p>The three paragraphs from page 16, line 35 starting with "On univariable analysis..." describes what is shown in table 4. I would recommend removing the descriptions outlined in these paragraphs and just highlight that univariable and multivariable figures can be found in table 4. The key points from this table can then be highlighted in the summary of findings as the first paragraph of the discussion as you have already done.</p> <p>Discussion While I feel that a valid explanation (supported by evidence) is given relating to the declining operative numbers which may negatively impact ARCP outcomes, I am not sure I agree that higher standards have been sought by ARCP panels between 2010 and 2017, is there any evidence to support this? In the time frame you have studied there have been multiple iterations of the surgical syllabuses for each of the specialties described in this study, is there anything in these that evidence that requirements have progressively become tougher?</p> <p>Given we are on the brink of the new surgical curricula, I feel that this warrants mention when you state that "The aim of the surgical training programmes is to produce competent day one consultants, this study should prompt a review of these requirements across specialties." on page 21, line 54. This is the focus of the new curricula.</p> <p>Page 22, line 38 – remove the word "than" after burnout</p>
--	---

REVIEWER	Ellis, Ricky University of Aberdeen Institute of Applied Health Sciences
REVIEW RETURNED	23-Aug-2021

GENERAL COMMENTS	<p>bmjopen-2021-053391 Differences in progression by surgical specialty: a national cohort study</p> <p>Thank you for the opportunity to review this interesting and important paper. The paper is clear, concise and is well written. I have made several suggestions for minor revisions and I hope that this paper is published promptly. This manuscript addresses a very important topic and is timely given recent changes to the surgical curriculum.</p> <p>1. Abstract:</p>
-------------------------	--

	<p>May I suggest the location of surgical training is added to the second line of the abstract for context e.g., “Progression through training in the United Kingdom is assessed by the Annual Review of Competency Progression (ARCP)”.</p> <p>The line from the abstract conclusion “Across all specialities, female sex and older age were associated with non-standard outcomes.” Should be moved to results section (with odds ratios and confidence intervals).</p> <p>Aim: “This study aimed to examine variation in ARCP outcomes within surgical training and identify differences between specialities.” By “differences between specialties” do the authors mean variation in ARCP outcomes or other differences (e.g., socio-demographics)? This may need clarification.</p> <p>2. Introduction:</p> <p>Overall, the Introduction is good.</p> <p>Line 42 page 7 “It can be used as a marker of progression” is a repetition of the previous two sentences.</p> <p>Consider adding a definition for differential attainment in the last paragraph of the conclusion</p> <p>3. Methods</p> <p>Line 48 page 10: It appears the word “No’ is missing from the start of the patient and public involvement statement.</p> <p>Am I right in understanding that if a trainee chooses to leave a programme, they are awarded an outcome 4? If so, should ARCP outcome 4 be included in group 4 given that leaving trainees may have met all requirements for progression at ARCP?</p> <p>I note with interest that ethnicity was not included as a predictor variable. Please consider adjusting for ethnicity given the weight of evidence for differential attainment in the postgraduate setting. Ethnicity is known to be associated with group-level differences in performance in almost all postgraduate medical examinations (including MRCS and FRCS; surgical examinations). If ethnicity is associated with these key markers of performance and career progression, then the association with progression through ARCP’s must also be explored.</p> <p>Also consider adjusting for first language as a confounding variable in analyses (if data is available) when including international medical graduates if significant differences in outcomes are found on univariate analysis (as has been found elsewhere: DOI 10.1002/bjs.5.50212 and DOI 10.1016/j.surge.2017.10.001)</p> <p>4. Results</p> <p>Consider including standard deviations when quoting mean scores, n/n when using descriptive percentages of cohorts and state Odds ratio (OR) and Confidence Interval (CI) in abstract and results sections the first time that they are used.</p>
--	--

	<p>Please describe the degree of missing data for each variable before multiple imputation was performed.</p> <p>Age row in Table 2 requires “Mean” before “years”.</p> <p>In table 2 the authors have marked all specialties with significant differences in the number of satisfactory outcomes compared to general surgery in column 2. It would also be interesting to see which columns also showed statistically significant differences to General Surgery i.e., whether T+O have a significant difference in ARCP's with insufficient evidence compared to General Surgery.</p> <p>Page 17 line 40 is missing “OR” in the results for Age</p> <p>5. Discussion</p> <p>Page 21 line 11: “The likelihood of a non-standard outcome has increased throughout time”. Perhaps it is worth reiterating here that this is despite the numbers of non-standard outcomes reducing over time as you have stated on page 14 line 28 which is a very interesting finding.</p> <p>On page 23 the authors discuss variation in academic elements of the curriculum between specialties as a possible reason behind the differences in rates of non-standard outcomes. Perhaps this discussion could be extended to highlight the need for further work looking at all possible reasons behind these differences that include non-academic aspects of the curriculum. If trainees are struggling to meet work-based assessment requirements or operative numbers in some specialties more than others then this should prompt urgent investigation into the quality and provision of training opportunities.</p> <p>Page 23 line 31: “Reassuringly, we have found a low proportion of trainees are asked to leave the programme.” Do we know how many of this small cohort were asked to leave vs. how many chose to leave voluntarily from this data? 1.3% being asked to leave is reassuring, but does this represent the full extent of trainee attrition in higher specialist training?</p> <p>Page 23 paragraph 3: This is a very interesting and topical point of discussion. Perhaps this could be extended slightly to discuss the impact that higher rates of non-standard outcomes for some groups of trainees (e.g., female trainees, older trainees and general surgical trainees) may have on career satisfaction and attrition.</p> <p>Overall, the discussion is very good. I feel it would benefit from using these important results to discuss the importance of exploring why the variation in ARCP outcomes may exist and what further work is required to investigate this. In addition, it feels like the recent changes to the surgical curriculum are a relevant topic, worthy of mention. It will be interesting to see this study repeated in a few years to see whether ARCP outcomes change following the introduction of multiple consultant reports and after removing or altering many of the old requirements for operative numbers, academic and management and leadership aspects of the curriculum.</p>
--	---

	<p>6. Miscellaneous</p> <p>References 10 and 11 are the same.</p> <p>Needs the blanket UKMED statement regarding ethical approval.</p>
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Abstract

Add the associated p-value to the statistics documented in the abstract
Added

Introduction

Line 7 page 9 – comma not needed after “a”
Apologies- thank you.

Methods

Should patient and public involvement state that it was not sought as no patients were included in the study?
This has been amended.

Results

Median age (and IQR) at ARCP seems slightly higher than I would have envisaged, is this the overall median across all ARCPs or is this the median age of final ARCP/CCT? I think this is worth rechecking.

This is the median age across all ARCPs for that specialty. Which will encompass ST3 to ST8, given that a fair proportion will take time out of programme for research (potentially additionally 3 years) this seems reasonable and in line with our other studies. The earliest age at which a typical undergraduate would start ST3 would be 27-28 years if no breaks were taken since school.

Table 2 –

Shouldn't the cohort populations described under gender match the cohort numbers relating to the PMQ region descriptors? I appreciate the percentages are rounded to the nearest whole number but the crude numbers should already be whole numbers? This is not the case for GS and ENT e.g. total cohort for GS gender = 2630 (1780+850) vs. cohort for PMQ = 2620 (2035+90+495)

The crude numbers have to be rounded to the nearest 0 or 5 when displaying in a table as per the UKMED requirements. We have added this to the figure legend. This explains the totals not adding up.

The figures for Max Fax used in table 2 need to be looked at again as they do not correlate correctly with the cohort population/percentages - the figures in parentheses do not correlate with crude number for each cohort. i.e. it states that 30 individuals are EEA which is described as 10%, yet 85 are IMG described as 4%?

This has been corrected, sorry for the error.

Table 3 –

I feel that the p-values established on chi square test should be included if you are going to state that all specialties were significant except paediatric surgery.

Added to table 3.

I would recommend separating outcome 3's (extended training) from the outcome 4's as I think the number of outcome 4's by specialty is an interesting finding and should be highlighted. It would also be interesting to know whether there is any significance statistically between specialties in relation to outcome 4's.

Unfortunately we are not able to separate outcome 4s as there are too few outcomes per specialty and the trainees would be potentially identifiable. This was restricted by HESA who will not allow any figures of less than 22.5 headcounts to be published. Although I agree this group is interesting and it had been our original intention to do a subgroup analysis of the Outcome 4s.

The three paragraphs from page 16, line 35 starting with “On univariable analysis...” describes what is shown in table 4. I would recommend removing the descriptions outlined in these paragraphs and just highlight that univariable and multivariable figures can be found in table 4. The key points from this table can then be highlighted in the summary of findings as the first paragraph of the discussion as you have already done.

This has been cut down as suggested and Table 4 signposted for these values.

Discussion

While I feel that a valid explanation (supported by evidence) is given relating to the declining operative numbers which may negatively impact ARCP outcomes, I am not sure I agree that higher standards have been sought by ARCP panels between 2010 and 2017, is there any evidence to support this? In the time frame you have studied there have been multiple iterations of the surgical syllabuses for each of the specialties described in this study, is there anything in these that evidence that requirements have progressively become tougher?

We have reworded the ‘higher standards of the ARCP panel’. What we meant to convey was that over time the ARCP panel have gained more experience implementing the ARCP criteria since the change from RITA and old curricula.

Given we are on the brink of the new surgical curricula, I feel that this warrants mention when you state that “The aim of the surgical training programmes is to produce competent day one consultants, this study should prompt a review of these requirements across specialties.” on page 21, line 54. This is the focus of the new curricula.

We have added information about the new curriculum to the discussion and acknowledged that this has now come into effect since the writing of our article.

Page 22, line 38 – remove the word “than” after burnout
Corrected.

Reviewer: 2

1. Abstract:

May I suggest the location of surgical training is added to the second line of the abstract for context e.g., "Progression through training in the United Kingdom is assessed by the Annual Review of Competency Progression (ARCP)".

United Kingdom added.

The line from the abstract conclusion "Across all specialities, female sex and older age were associated with non-standard outcomes." Should be moved to results section (with odds ratios and confidence intervals).

The OR and CI for female sex and older age are already in the results section of the abstract (last line). We feel summarising this in the conclusion as a major finding is important.

Aim: "This study aimed to examine variation in ARCP outcomes within surgical training and identify differences between specialties." By "differences between specialties" do the authors mean variation in ARCP outcomes or other differences (e.g., socio-demographics)? This may need clarification.

This has been clarified. 'Differences in outcomes between specialties'.

2. Introduction:

Line 42 page 7 "It can be used as a marker of progression" is a repetition of the previous two sentences.

This has been deleted.

Consider adding a definition for differential attainment in the last paragraph of the conclusion.

A definition and reference has been added to the last paragraph of the introduction

3. Methods

Line 48 page 10: It appears the word "No" is missing from the start of the patient and public involvement statement.

Thank you, amended.

Am I right in understanding that if a trainee chooses to leave a programme, they are awarded an outcome 4? If so, should ARCP outcome 4 be included in group 4 given that leaving trainees may have met all requirements for progression at ARCP?

Yes- trainees that voluntarily leave the programme are also awarded an outcome 4. The method used for the groupings of ARCP outcome using UKMED data has previously been validated and we have replicated this in our paper. Within outcome 4 there will be a mix of those asked to leave and those that choose to leave as there is no data separating the two available we have chosen to group them together as per previous papers. Despite the mix in outcome 4s we feel that we cannot exclude this group from the analysis as this is the most severe outcome awarded.

I note with interest that ethnicity was not included as a predictor variable. Please consider adjusting

for ethnicity given the weight of evidence for differential attainment in the postgraduate setting. Ethnicity is known to be associated with group-level differences in performance in almost all postgraduate medical examinations (including MRCS and FRCS; surgical examinations). If ethnicity is associated with these key markers of performance and career progression, then the association with progression through ARCP's must also be explored.

We did not include ethnicity as a variable due to large amounts (>50%) of missing data. This is because for our cohort we included non-UK graduates and no ethnicity data was available for this group. The surrounding variables had considerable missing data as not to allow multiple imputation. Therefore, we chose place of primary medical qualification as we had a complete dataset for this. It was important for us to include non-UK graduates in our analysis as often they are excluded from research using UKMED due to incomplete data and they make up a large proportion of the surgical workforce.

Also consider adjusting for first language as a confounding variable in analyses (if data is available) when including international medical graduates if significant differences in outcomes are found on univariate analysis (as has been found elsewhere: DOI 10.1002/bjs.5.50212 and DOI 10.1016/j.surge.2017.10.001)

We did not have first language as a variable in the dataset so could not adjust for this.

4. Results

Consider including standard deviations when quoting mean scores, n/n when using descriptive percentages of cohorts and state Odds ratio (OR) and Confidence Interval (CI) in abstract and results sections the first time that they are used.

We have reported median and IQR as not normally distributed. OR and CI added to Methods section where they are first used.

Please describe the degree of missing data for each variable before multiple imputation was performed.

We did not perform multiple imputation. Our method states how we handled missing data (coded as a separate category). There was no missing data for the demographics in Table 2. The only variable with missing data was LTFT, however due to restrictions on publishing data with less than 22.5 headcounts as per HESA we cannot include LTFT in Table 2 as so few trainees per speciality are LTFT and would potentially be identifiable. Apologies for this.

Age row in Table 2 requires "Mean" before "years".
Added- median.

In table 2 the authors have marked all specialties with significant differences in the number of satisfactory outcomes compared to general surgery in column 2. It would also be interesting to see which columns also showed statistically significant differences to General Surgery i.e., whether T+O have a significant difference in ARCP's with insufficient evidence compared to General Surgery.

This is depicted in the caterpillar plot (Figure 1). We chose to use a graphical representation as it is quite cumbersome to describe these findings in text format given the nature of the multi-level model and the difficulty in interpreting the statistical output for the majority of our audience. We hope that Figure 1 allows our readers to see the significant difference in individual outcomes by speciality. And

has the advantage that all specialities can be compared to each other as opposed to just general surgery.

Page 17 line 40 is missing "OR" in the results for Age
Corrected.

5. Discussion

Page 21 line 11: "The likelihood of a non-standard outcome has increased throughout time". Perhaps it is worth reiterating here that this is despite the numbers of non-standard outcomes reducing over time as you have stated on page 14 line 28 which is a very interesting finding.

This actually reads standard outcomes reduced over time (page 14, line 28).

On page 23 the authors discuss variation in academic elements of the curriculum between specialities as a possible reason behind the differences in rates of non-standard outcomes.

Perhaps this discussion could be extended to highlight the need for further work looking at all possible reasons behind these differences that include non-academic aspects of the curriculum. If trainees are struggling to meet work-based assessment requirements or operative numbers in some specialities more than others then this should prompt urgent investigation into the quality and provision of training opportunities. Page 23 line 31: "Reassuringly, we have found a low proportion of trainees are asked to leave the programme." Do we know how many of this small cohort were asked to leave vs. how many chose to leave voluntarily from this data? 1.3% being asked to leave is reassuring, but does this represent the full extent of trainee attrition in higher specialist training?

There is no data on those that voluntarily leave the programme in UKMED. As far as we are aware this data is only held locally by individual deaneries as opposed to nationally available so we cannot answer the above.

Page 23 paragraph 3: This is a very interesting and topical point of discussion. Perhaps this could be extended slightly to discuss the impact that higher rates of non-standard outcomes for some groups of trainees (e.g., female trainees, older trainees and general surgical trainees) may have on career satisfaction and attrition.

This is a good point and we added this thought to the discussion.

Overall, the discussion is very good. I feel it would benefit from using these important results to discuss the importance of exploring why the variation in ARCP outcomes may exist and what further work is required to investigate this. In addition, it feels like the recent changes to the surgical curriculum are a relevant topic, worthy of mention. It will be interesting to see this study repeated in a few years to see whether ARCP outcomes change following the introduction of multiple consultant reports and after removing or altering many of the old requirements for operative numbers, academic and management and leadership aspects of the curriculum.

We have added some text to the discussion regarding the new curriculum and learning from good practice in well performing specialities to improve the quality of surgical training for all.

6. Miscellaneous

References 10 and 11 are the same.

This has been amended- thank you.

Needs the blanket UKMED statement regarding ethical approval.

This was uploaded as an ethical statement separate to the main manuscript which I believe will be included with the published article as an appendix.

VERSION 2 – REVIEW

REVIEWER	Robinson, David Cardiff & Vale University Health Board, General Surgery
REVIEW RETURNED	19-Oct-2021

GENERAL COMMENTS	<p>Thank you for addressing my comments. The only minor change I would consider making is putting the p-values that have been added to the table in the text in parentheses as while I suggested for them to be added they have reduced the aesthetics of the table by having a single column for p-values. Maybe consider something along the lines of - "All specialities (p<0.001) except paediatric surgery (p=0.162) had a significantly greater number of satisfactory outcomes than general surgery on Chi square test (Table 3). The table could then have this column deleted. I think a column would have been more valuable if the p-values were varied and not all so strong. I have no further concerns and support the publication of this paper.</p>
-------------------------	--

REVIEWER	Ellis, Ricky University of Aberdeen Institute of Applied Health Sciences
REVIEW RETURNED	20-Oct-2021

GENERAL COMMENTS	<p>Thank you for the opportunity to review this paper again. The previous suggestions have all been addressed and I believe that the manuscript is now much stronger. It is an important paper and I hope it is published in a timely fashion as I feel this adds valuable evidence to the body of work investigating differences in progression through surgical training.</p> <p>I have one comment and one minor suggestion for improving the analysis.</p> <p>With regards to separating Outcome 3 and Outcome 4 as suggested, I understand that HESA data rules prevent the sub-analysis of candidates with Outcome 4 due to small cohort numbers to ensure data anonymity. Ideally, trainees with Outcome 4 should be removed from the analysis, leaving a clearer picture for individuals with Outcome 3. The analysis of trainees with Outcome 4 could be suggested as a future study which would need a longer data collection period to give a larger cohort size for meaningful statistical analysis. I understand that previous studies have previously grouped these together and therefore the authors wish to continue to use this grouping. In which case I suggest that the authors add this to the study limitations and cite previous studies that have used this grouping to justify the decision. Not separating these groups is a justifiable choice but one that should be highlighted to readers.</p>
-------------------------	---

	<p>Given the importance and timeliness of this paper in the body of work identifying differential attainment in UK surgical training and career progression, I do feel that there should be a sub-analysis of UK graduates (owing to the completeness of the data for this group) that includes ethnicity in the regression model. I suggest this as not only is there differential attainment in markers of success and progression in surgery by place of qualification but also by ethnicity as highlighted in the wider literature, GMC reports and in the recent report into diversity and inclusivity at the RCS led by Baroness Kennedy. If there is differential attainment in ARCP's for some groups of individuals even after adjusting for sex, LTFT, place of PMQ and age at ARCP then this data will be incredibly important and can be used to advocate for investigation into the possibility of discrimination and bias in ARCP processes and in the training environment. This could be presented in a Table 5 at the end of the results section. I feel that this would add great value to the study. Of course, this could be addressed in a separate study, in which case the authors could state their intentions to do this in the discussion.</p>
--	--

VERSION 2 – AUTHOR RESPONSE

Reviewer 1

Thank you for addressing my comments. The only minor change I would consider making is putting the p-values that have been added to the table in the text in parentheses as while I suggested for them to be added they have reduced the aesthetics of the table by having a single column for p-values. Maybe consider something along the lines of - "All specialities (p<0.001) except paediatric surgery (p=0.162) had a significantly greater number of satisfactory outcomes than general surgery on Chi square test (Table 3). The table could then have this column deleted. I think a column would have been more valuable if the p-values were varied and not all so strong. I have no further concerns and support the publication of this paper.

Thank you. Table 3 has been amended to remove the p value column.

Reviewer: 2

I have one comment and one minor suggestion for improving the analysis.

With regards to separating Outcome 3 and Outcome 4 as suggested, I understand that HESA data rules prevent the sub-analysis of candidates with Outcome 4 due to small cohort numbers to ensure data anonymity. Ideally, trainees with Outcome 4 should be removed from the analysis, leaving a clearer picture for individuals with Outcome 3. The analysis of trainees with Outcome 4 could be suggested as a future study which would need a longer data collection period to give a larger cohort size for meaningful statistical analysis. I understand that previous studies have previously grouped these together and therefore the authors wish to continue to use this grouping. In which case I suggest that the authors add this to the study limitations and cite previous studies that have used this grouping to justify the decision. Not separating these groups is a justifiable choice but one that should be highlighted to readers.

Thank you. Page 9, last paragraph, references the paper that validated the ARCP groupings, Reference 16. We have added to the discussion (page 21) the potential limitations of grouping outcome 3 and 4. Given that this method has previously been validated and used frequently in studies using UKMED data we have chosen not to remove Outcome 4s from the analysis.

Given the importance and timeliness of this paper in the body of work identifying differential attainment in UK surgical training and career progression, I do feel that there should be a sub-analysis of UK graduates (owing to the completeness of the data for this group) that includes ethnicity in the regression model. I suggest this as not only is there differential attainment in markers of success and progression in surgery by place of qualification but also by ethnicity as highlighted in the wider literature, GMC reports and in the recent report into diversity and inclusivity at the RCS led by Baroness Kennedy. If there is differential attainment in ARCP's for some groups of individuals even after adjusting for sex, LTFT, place of PMQ and age at ARCP then this data will be incredibly important and can be used to advocate for investigation into the possibility of discrimination and bias in ARCP processes and in the training environment. This could be presented in a Table 5 at the end of the results section. I feel that this would add great value to the study. Of course, this could be addressed in a separate study, in which case the authors could state their intentions to do this in the discussion.

The surgical workforce relies heavily on non-UK graduates and too often studies only include UK graduates in the analysis. We wanted to ensure that the entire surgical training cohort was represented in our paper. While ethnicity data was more complete for UK graduates there was still considerable missing data. We feel that further studies should incorporate ethnicity into any further analysis however this may require data from alternate sources or analysis of a 'younger' cohort in which data collection is more complete. Therefore, while we agree ethnicity is a very important area for research we are choosing not to perform a sub group analysis in this instance. Lack of ethnicity data has been added to the limitations section (page 21).