PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The COVID-19 pandemic interim foundation year 1 post and confidence in core skills and competencies: a longitudinal survey
AUTHORS	Gatti, Cristina; Parker-Conway, Kathryn; Okorie, Michael

VERSION 1 – REVIEW

REVIEWER	Martins, Vera Center for Health Technology and Services Research
REVIEW RETURNED	13-Mar-2022

GENERAL COMMENTS	I would like to thank you for the opportunity I have been given to review this paper. The topic is interesting, and the findings may help guide future educational interventions, however there are some questions and concerns that I would like to address. Comments: The main conclusions of your study are related with the objective of your study, but could you please explain how these results are related with the COVID 19 pandemic in a clear way, as the title is related with COVID 19 pandemic. Could you please explain clearly how the medical educational program is organized in your country and in what the FY1 consists before the COVID 19 pandemic. Could you please mention in the Introduction section some studies related with the FY1 and the main conclusions? I have some concerns related with the research approval. Could you please explain why this questionnaire was exempted from a formal ethics review? You complete a self-appraisal questionnaire of the NHS Health Research Authority, could you please clarify how this document works and the content? Page 2, line 20- participants: you should describe the number of participants entering and completing the study. You have different information on item Participants and item Results. Page 3, line 32-35 Do you have some more references to fundament this statement? Could you please rewrite this using those references? You use abbreviations during the article, most of them are described but there are sections where the larger number of abbreviations may be difficult for the reader to understand some sections of the article, for example: page 5, line 27-29. In page 5, line 11 the abbreviations CG, KPC and MO related with the authors names are not described, for example Cristina Gatti (CG)… In page 5, line 10-12 you refer that the questionnaire was formulated by the authors and piloted amongst nine doctors. Could you please clarify how this questionnaire has been validated?

REVIEWER	Ohn, May

	University Malaysia Sabah, medicine
REVIEW RETURNED	26-Mar-2022

GENERAL COMMENTS	Thank you for giving opportunity to review this educational study. As the authors had declared sample size is too small and this's the first comparative study, I am not so sure was it meant to be done as pilot study. I am intrigued in selection process of cohort group (please describe in more detail). What are the difference in intervention and control group activities? As shown in the table, non-FiY group showed decrease in confidence in some of the outcome, is it happened by chance or any probable mechanism behind the regression of confidence level? I am confused with the job description of 13 non-FiY1 who participated in the study? I suppose Paired-t test (if parametric data, however, sample size doesn't fit to use it) or other suitable test should be used before regression study. Linear regression study was used to adjust university attendance, however, all non-FiYs were BSMS graduates. How the adjustment was calculated as it can be major confounding factor in contributing the differences. Table 4 should mention breakdown detail of each intervention and control group as comparison rather than shown as overall cohort group. And, P-value difference between 2 groups should be mentioned as last column in table 4. Please attach local R&D team's ethic approval and agreement form.

REVIEWER	McNiece, Rosemary Kingston University, Mathematics
REVIEW RETURNED	30-Jul-2022

GENERAL COMMENTS	This is an useful report on an important aspect of evaluating and ultimately enabling the transition experiences of medical graduates to junior doctors and as such is worthy of publication for wider dissemination. While the study is based on a small sample, it has been well planned and features some commendable design and validation practices, e.g. seeking input on preliminary questionnaire designs from medics in order to improve access and interpretability. The questionnaires themselves are clear and set at an appropriate length for maximising participation. Although the overall response rate was only 20% this is not atypical of response rates to this type of 'voluntary' survey. My main concerns with this paper lie in the analysis of the data collected and in the interpretations of the results. I believe the authors are in danger of drawing conclusions from the data analyses that are not statistically valid and hence should not report the results as being as statistically significant without clarification of the metrics underlying the analyses. Responses collected on a 5 point Likert scale, as used here for self assessment of confidence, are not generally appropriate for analysis using parametric statistics such as linear regression. While this does happen in practice it is statistically dubious. Specifically, linear regression methodology is based wholly on the assumption that the underlying data is normally distributed, while deviations from this assumption are tolerated, even for small

	samples, the type of data collected here does not fit with the general acceptance of normally distributed data (or at least is not shown to). The report states that linear regressions were adjusted for age, gender and university, how was this achieved, were any differences observed or not? Further, there is no detail about how the responses have been scored and hence how the change in score between surveys is calculated (e.g. one might assume that the responses are allocated an ordered score such as 1,2,3,4 & 5 and change is calculated as a numeric difference between responses at each survey – or similar method). This mechanism is not specified and should be clarified. Similarly the results report a ' change in score' with corresponding CI but again this is meaningless without definition of the scoring mechanism and statistically CIs also assume underlying normality in the data. To proceed to publication there are several options. You may choose a different, more appropriate statistical procedure, either parametric or non parametric for analysing your data, I think this would be the best option. Alternatively you could proceed with the linear regression analyses used but this would need to be qualified against the concerns and limitations of interpreting the results along the lines of the issues discussed above. I would say that this type of error is common in the application of statistical analysis to survey based data - it is quite easily remedied and does not diminish the value of the data collected when interpreted correctly. Other minor points – relating to the abstract Participants : this summary of participants should include the fact that out of all those approached, usable responses were obtained from 39 participants. Conclusion : - here is the first mention of this study having taken place during the Covid pandemic (other than in the title) – I think this should be stated earlier perhaps under Objectives or Setting.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Prof. Vera Martins, Center for Health Technology and Services Research
Comments to the Author: (Responses in blue)

The main conclusions of your study are related with the objective of your study, but could you please explain how these results are related with the COVID 19 pandemic in a clear way, as the title is related with COVID 19 pandemic.

Many thanks for this comment. We have amended the manuscript appropriately - pages 2, 3, 11 and 15.

Could you please explain clearly how the medical educational program is organized in your country and in what the FY1 consists before the COVID 19 pandemic. Could you please mention in the Introduction section some studies related with the FY1 and the main conclusions?

Many thanks for this comment. We have amended the manuscript appropriately - page 3.

I have some concerns related with the research approval. Could you please explain why this questionnaire was exempted from a formal ethics review? You complete a self-appraisal questionnaire of the NHS Health Research Authority, could you please clarify how this document works and the content?

We liaised with the Research and Development department at University Hospitals Sussex NHS Foundation Trust (Sponsor Representative for research) from the conception of this project. We were informed that the project fit into the category of 'service evaluation,' and as a result formal review by the NHS ethics committee was not required. In addition, we completed an NHS ethics self-assessment process (Do I need NHS Ethics approval? [hra-decisiontools.org.uk]) and this confirmed that formal ethics approval was not required.

Page 2, line 20- participants: you should describe the number of participants entering and completing the study. You have different information on item Participants and item Results.

Many thanks. We have amended the manuscript appropriately - page 2.

Page 3, line 32-35 Do you have some more references to fundament this statement? Could you please rewrite this using those references?

Many thanks. We have amended the manuscript appropriately - page 4. For clarity, we have highlighted which references correspond to the sentence sections for your review: "Assistantships have been positively received as an intervention in preparing students for practice (Braniff, 2015), with students feeling increased confidence in managing acute situations, gaining responsibility, on-call work (Fullbrook, 2015), integrating into a team (Braniff 2015), administrative skills and duties of an FY1 (Jones 2016), and in therapeutics (16-20)."

You use abbreviations during the article, most of them are described but there are sections where the larger number of abbreviations may be difficult for the reader to understand some sections of the article, for example: page 5, line 27-29. In page 5, line 11 the abbreviations CG, KPC and MO related with the authors names are not described, for example Cristina Gatti (CG)…

Many thanks. We have amended the manuscript appropriately - page 5.

In page 5, line 10-12 you refer that the questionnaire was formulated by the authors and piloted amongst nine doctors. Could you please clarify how this questionnaire has been validated?
Many thanks. We have amended the manuscript appropriately - pages 5 and 6.

Reviewer: 2
Dr. May Ohn, University Malaysia Sabah
Comments to the Author:
Thank you for giving opportunity to review this educational study. As the authors had declared sample size is too small and this's the first comparative study, I am not so sure was it meant to be done as pilot study.

Many thanks for your comments. It was not intended as a pilot study. Given the nature of the covid-19 pandemic, the interim FY1 post was commenced rapidly so we sought to take the opportunity to evaluate it at the time. However, our data can also be used to inform future studies.

I am intrigued in selection process of cohort group (please describe in more detail). What are the difference in intervention and control group activities?

Many thanks for your comments. The cohort groups selected were available for the study within the very short time frame, as rapid adjustments were made during the COVID-19 pandemic. These comprised those working in the hospital at the time and graduates from our local medical school. We have amended the manuscript appropriately in order to clarify the point - pages 4 and 5.

As shown in the table, non-FiY group showed decrease in confidence in some of the outcome, is it happened by chance or any probable mechanism behind the regression of confidence level?

Many thanks. We have amended the manuscript appropriately -page 14

I am confused with the job description of 13 non-FiY1 who participated in the study?

Many thanks for your comments. We have amended the manuscript in order to clarify this point - page 5.

I suppose Paired-t test (if parametric data, however, sample size doesn't fit to use it) or other suitable test should be used before regression study.

Many thanks. We sought advice from a medical statistician who assisted with this work. Statistician's advice: Performing univariable modelling (e.g. a paired t-test) before multivariable modelling does not add further information and can be misleading. The univariable results are redundant after multivariable modelling. Choosing to only perform the multivariable modelling if the univariable modelling is "significant" will lead to bias.

Linear regression study was used to adjust university attendance, however, all non-FiYs were BSMS graduates. How the adjustment was calculated as it can be major confounding factor in contributing the differences.

Many thanks. Statistician's advice: This is adjusted for in the multivariable models. Incidentally, this is an example of why the univariable comparisons would be misleading – this is important to adjust for.

Table 4 should mention breakdown detail of each intervention and control group as comparison rather than shown as overall cohort group. And, P-value difference between 2 groups should be mentioned as last column in table 4.

Many thanks. We have amended Table 4 to include intervention and control group comparison.

Statistician's advice: The p-values are not required in this table. Some people select variables for multivariable models based on p-values comparing the groups, but this should never be done as it leads to overfitting and bias. Variables should be selected a priori based on their clinical/biological/whatever importance. P-values in this table would distract and mislead, especially with the small sample size (a high p-value would not necessarily tell you much about the lack of difference). Important differences between intervention/control are already accounted for in the multivariable analysis, so low p-values here would be moot.

Please attach local R&D team's ethic approval and agreement form.

Many thanks for your comments. We liaised with the Research and Development department at University Hospitals Sussex NHS Foundation Trust (Sponsor Representative for research) from the conception of this project. We were informed that the project fit into the category of 'service evaluation,' and as a result formal review by the NHS ethics committee was not required. In addition, we completed an NHS ethics self-assessment process (Do I need NHS Ethics approval? [hra-decisiontools.org.uk]) and this confirmed that formal ethics approval was not required.

Please see email response from the Head of Research and Development, University Hospitals Sussex NHS Foundation Trust:
'I think the easiest thing is to point the reviewer to the link Do I need NHS Ethics approval? (hra-decisiontools.org.uk) It will be fairly self-explanatory then.
Irrespective of this research involving staff as subjects no longer requires ethical review, but in some case may require HRA approval. You can also note that in my opinion, as the Sponsor Representative for University Hospitals Sussex, this was a service evaluation project, and therefore does not require formal approval by the HRA or any other body.'

Scott Harfield
Head of Research & Development
University Hospitals Sussex NHS Foundation Trust
Clinical Research Facility
2nd Floor Sussex House
1 Abbey Road, Brighton, BN2 1ES
01273 696955 ext 7497'

Reviewer: 3
Mrs. Rosemary McNiece, Kingston University
Comments to the Author:
This is an useful report on an important aspect of evaluating and ultimately enabling the transition experiences of medical graduates to junior doctors and as such is worthy of publication for wider dissemination.
While the study is based on a small sample, it has been well planned and features some commendable design and validation practices, e.g. seeking input on preliminary questionnaire designs from medics in order to improve access and interpretability. The questionnaires themselves are clear and set at an appropriate length for maximising participation. Although the overall response rate was only 20% this is not atypical of response rates to this type of 'voluntary' survey.

My main concerns with this paper lie in the analysis of the data collected and in the interpretations of the results. I believe the authors are in danger of drawing conclusions from the data analyses that are not statistically valid and hence should not report the results as being as statistically significant without clarification of the metrics underlying the analyses.

Responses collected on a 5 point Likert scale, as used here for self assessment of confidence, are not generally appropriate for analysis using parametric statistics such as linear regression. While this does happen in practice it is statistically dubious. Specifically, linear regression methodology is based wholly on the assumption that the underlying data is normally distributed, while deviations from this assumption are tolerated, even for small samples, the type of data collected here does not fit with the general acceptance of normally distributed data (or at least is not shown to).

Many thanks for the comments. We sought advice from a medical statistician who assisted with this work. Statistician's advice: It's debatable whether parametric or non-parametric approaches are more appropriate to analyse Likert scales (or perhaps, which is least inappropriate). However, in reality the approach chosen doesn't really matter that much in that the conclusions reached are broadly the same, even if the distribution over the scale is not normal. But this is in the case of directly analysing the scale – not the difference between pre and post values. In this study we took pre and post values and calculated the difference. When you calculate the difference in this way, you expect a normal distribution.

We assumed the distribution is normal enough to use a parametric method and this was felt to be acceptable by our medical statistician.

The report states that linear regressions were adjusted for age, gender and university, how was this achieved, were any differences observed or not?

Many thanks for the comments. Statistician's advice: This is adjusted for in the multivariable models. If age, gender or university make a difference, they're accounted for, if they don't, their presence in the model doesn't matter.

As stated on page 10: "There was generally no effect seen for the variables adjusted for, except in the outcome of being 'trained and managing cardiac and respiratory arrest,' in which females were less confident than males by 0.80 (95% CI -1.374 to -0.218, p=0.008)."

Further, there is no detail about how the responses have been scored and hence how the change in score between surveys is calculated (e.g. one might assume that the responses are allocated an ordered score such as 1,2,3,4 & 5 and change is calculated as a numeric difference between responses at each survey – or similar method). This mechanism is not specified and should be clarified. Similarly the results report a ' change in score' with corresponding CI but again this is meaningless without definition of the scoring mechanism and statistically CIs also assume underlying normality in the data.

Many thanks for this comment. We have amended the manuscript appropriately- page 8.

To proceed to publication there are several options. You may choose a different, more appropriate statistical procedure, either parametric or non parametric for analysing your data, I think this would be the best option. Alternatively you could proceed with the linear regression analyses used but this would need to be qualified against the concerns and limitations of interpreting the results along the lines of the issues discussed above.
I would say that this type of error is common in the application of statistical analysis to survey based data - it is quite easily remedied and does not diminish the value of the data collected when interpreted correctly.

Many thanks for your comment. We sought advice from a medical statistician when planning and conducting the analysis, and in addition most recently in relation to reviewer's comments. As stated, we have been advised the multivariable model used in the analysis was appropriate. However, we acknowledge that the analysis is exploratory (performed without an a priori sample size calculation on a limited sample, increasing the risk that real differences could have been missed) and that p-values were not adjusted for multiple testing – increasing the risk that low-ish p-values could be due to chance. As such we have adjusted our interpretation to account for this in the strengths and limitations section and discussion (see page 3, 12, 13, 15).

Other minor points – relating to the abstract

Participants : this summary of participants should include the fact that out of all those approached, usable responses were obtained from 39 participants.
Many thanks. We have made appropriate amendments to the manuscript - page 2.

Conclusion : - here is the first mention of this study having taken place during the Covid pandemic (other than in the title) – I think this should be stated earlier perhaps under Objectives or Setting.
Many thanks. We have made appropriate amendments to the manuscript - pages 2, 3 and 11.

VERSION 2 – REVIEW

REVIEWER	Martins, Vera Center for Health Technology and Services Research
REVIEW RETURNED	11-Sep-2022

GENERAL COMMENTS	I would like to thank you for the opportunity I have been given to review this paper again. The topic is interesting, and the findings may help guide future educational interventions. The authors responded to the revisions suggested.

REVIEWER	Ohn, May University Malaysia Sabah, medicine
REVIEW RETURNED	20-Sep-2022

GENERAL COMMENTS	The authors had done the thorough revision and I am hereby to confirm acceptance the manuscript.

REVIEWER	McNiece, Rosemary Kingston University, Mathematics
REVIEW RETURNED	21-Sep-2022

GENERAL COMMENTS	no further comments