

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)	Variation in quality of acute stroke care by day and time of admission: prospective cohort study of weekday and weekend centralised hyperacute stroke unit care and non-centralised services
AUTHORS	Melnychuk, Mariya; Morris, Stephen; Black, Georgia; Ramsay, Angus I. G.; Eng, Jeannie; Rudd, Anthony; Baim-Lance, Abigail; Brown, Martin; Fulop, Naomi; Simister, Robert

VERSION 1 - REVIEW

REVIEWER	Lawrence Best University College London UK
REVIEW RETURNED	10-Aug-2018

GENERAL COMMENTS	<p>I would first like to congratulate the authors on a well executed study on an area of medical care which requires much more good research. The manuscript is generally well written and the data appropriately analysed. My only comments would be as follows: For the patient characteristics comparisons multiple comparisons adjustment should ideally be performed, although it would be down to the authors to determine which groups should be considered as one 'family' of statistical tests. The gold standard and easiest is bonferroni correction though it is extremely conservative.</p> <p>Ideally uncontrolled analyses should also be run and at least summarily reported as this aids study replication and comparison although I can imagine space constraints may limit the ability to do this.</p> <p>Something else which came to mind was whether the difference in hospital stay is related to the rhythms of discharge in a hospital. The average hospital length of stay seems to correspond with the majority of patients being discharged towards the end of the working week, perhaps all the required organisation such as physio SALT referral etc. leads to a natural discharge date for these patients and that this is responsible for the increased length of stay for patients hospitalised at weekends. Of course this is simply conjecture and I would defer to the collective experience of the authors as to whether or not this may be a factor worthy of mention in the manuscript.</p>
-------------------------	---

REVIEWER	Sandy Middleton Australian Catholic University, Australia
REVIEW RETURNED	18-Sep-2018

GENERAL COMMENTS	<p>This is an important paper presenting data that will be of great interest within the UK and also internationally. The main issue I had was suggestions to improve the clarity of the results to make the paper more useful to non-researchers such as health service planners and policy developers ie those who will be looking to translate your findings into practice.</p> <p>The title of the manuscript is clear regarding the purpose and does not say you will be directly comparing the two cohorts (eg London V's rest of England). However, language used in other sections of the paper do use the word 'comparison' or compared' (eg Page 9, line22) and as a researcher I was looking for a direct statistical comparison which I could not find eg between-group analysis rather than just the within-group analysis presented eg in the abstract, 'Setting: ... London HASUs compared with the rest of England'; eg your hypothesis, Page 5, Line 43. Can you please make it very clear throughout that you are not presenting between group comparisons – or alternatively, undertake these analyses. If you choose to undertake these additional analyses, you could consider separating out your aim into two parts eg i) to investigate variations in the quality of acute stroke care and outcomes by day and time of admission in London HASUs and the rest of England; ii) to compare variations in these indicators and outcomes between London HASUs and the rest of England. If you choose not to add in these additional analyses can you make it explicit early in the manuscript that you are not doing this comparison so the reader is not expecting to find this.</p> <p>Could you also clarify early in the manuscript whether time of admission means admission to Emergency Department/ hospital, or admission to the HASU/ Stroke Unit.</p> <p>Some specific comments around this and other issues are given below.</p> <p>Abstract:</p> <p>Page 5, line 29-32: Can you give the time periods of the improved quality of stroke care referenced by citations 28 and 29.</p> <p>Page 6, Line 22: Please provide a reference for the 'estimated to be 90%'.</p> <p>Page 6, line 39-40: Can you make it clearer that 'assessment by a physiotherapist, OT, SLT within 24 hours' are three separate indicators here please; as written it looks like one combined variable.</p> <p>Page 6, Line 54: Please replace 'spell' with 'stay' or 'episode of care'.</p> <p>Page 7, Line 4: You define LOS but then write 'length of stay' in full in same sentence.</p> <p>Page 7, Line 14: Please define RCP in full.</p>
-------------------------	--

	<p>Page 9, Line 24-30: Can you please give the directions for those indicators with statistical differences eg mRS pre-stroke, level of consciousness on arrival at hospital, method of admission and time from onset of symptoms. From the previous sentence this reads as a between group differences which I don't think it was, so can you make this clearer please.</p> <p>Page 16, line 46 and 48: To improve the readability can you not ask people to remember or refer back to the 'first' and 'second' hypothesis and improve the clarity of these sentences, particularly the one commencing 'The first hypothesis'.</p> <p>For all the tables, can you please provide exact P values. Can you please report the P values to a consistent number of significant figures and three decimal points may be more helpful.</p> <p>Can you reference the supplementary tables/ figures within the manuscript at the relevant points please and explain any relevant statistical testing undertaken as presented in these tables/ figures.</p> <p>Table S3: I found this table confusing. It is not possible to establish the direction of the findings ie if the level of consciousness on arrival or the NIHSS score is lower or higher in relation to those outcomes that are significant (eg Rest of England: brain scan within one hour). This will make it difficult for those wishing to improve these indicators in terms of knowing which of the stroke population to concentrate on. Can you please present these results more clearly. Please provide exact P values.</p> <p>With the supplementary figures, I read it as if a P value <0.05 would mean the variation is significant. In some of the figures (eg Figure S1b MRS 3-6) this is significant for the rest of England yet the graph shows a more flattened line for the rest of England when compared with the non-significant HASU. This was a bit confusing. Could you clarify please.</p> <p>The Figures were very small and hard to read, is it possible to improve the clarity?</p>
--	---

REVIEWER	Bichun Ouyang RUMC, US
REVIEW RETURNED	18-Dec-2018

GENERAL COMMENTS	<p>The statistical methods used seem to be appropriate. I have few questions regarding the stats in the paper.</p> <ol style="list-style-type: none"> 1. For testing the variation across the 42 time periods, the authors seemed to create 41 dummy variables. Which time period was used as the reference time period (the omitted period)? Is the joint p-value (mentioned in page 8 line 12) from the likelihood ratio test? 2. When variation across the time periods was detected, the authors tried to investigate the patterns (page 11 and 12) across the time periods. Were those patterns tested formally? If so, what analyses were performed to test those patterns? 3. The authors performed quite many regression analyses of quality indicators and outcomes. Did the authors consider the adjustment for multiple testing?
-------------------------	--

	<p>4. In table 1, I can see that most variables were significant, but the actual differences were quite small as the sample size is very large. In this case, effect size may be helpful.</p> <p>5. This question may not be statistically relevant. Why did the authors group the time of admission into those 6 four-hour periods? Is there any clinical rationale to do so?</p>
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

My only comments would be as follows:

1) For the patient characteristics comparisons multiple comparisons adjustment should ideally be performed, although it would be down to the authors to determine which groups should be considered as one 'family' of statistical tests. The gold standard and easiest is bonferroni correction though it is extremely conservative.

The reviewer makes an important point for which we are grateful. We have amended the p-value column of Table 1 providing the four digits of p-values that can be contrasted with the threshold p-value of 0.0038, corrected for multiple comparisons using Bonferroni approach ($0.05/13=0.0038$, where 0.05 is the reference p-value and 13 is the number of patient characteristics used). We added these threshold p-value in the footnotes of Table 1.

2) Ideally uncontrolled analyses should also be run and at least summarily reported as this aids study replication and comparison although I can imagine space constraints may limit the ability to do this.

We thank the reviewer for this suggestion. We have reported the analyses without adjustment on patient characteristics and other controls in Table 1 in the Supplementary materials.

3) Something else which came to mind was whether the difference in hospital stay is related to the rhythms of discharge in a hospital. The average hospital length of stay seems to correspond with the majority of patients being discharged towards the end of the working week, perhaps all the required organisation such as physio SALT referral etc. leads to a natural discharge date for these patients and that this is responsible for the increased length of stay for patients hospitalised at weekends. Of course this is simply conjecture and I would defer to the collective experience of the authors as to whether or not this may be a factor worthy of mention in the manuscript.

The reviewer makes an important point for which we are grateful. It is true that patients are less likely to be discharged until they have been accessed by all relevant professionals. Based on our results from analyses on therapists' assessments and evidence from qualitative observations (Black et al. 2019, R&R in BMJ Open), patients admitted out of working hours are less likely to be seen by therapists within 24h in London and within 24h and 72h in RoE. Thus it is very likely that the organisational factors, e.g. therapists' rotas, would have important impact of the length of hospital stay. Due to data limitations, we are unable, unfortunately, to investigate this further and formally test

the exact mechanism for this result. In our discussion on page 20, lines 33-37 we suggested that this would be a potential avenue for the future research that could have quite important implications for the HASUs functioning.

Reviewer 2

1) The title of the manuscript is clear regarding the purpose and does not say you will be directly comparing the two cohorts (eg London V's rest of England). However, language used in other sections of the paper do use the word 'comparison' or compared' (eg Page 9, line22) and as a researcher I was looking for a direct statistical comparison which I could not find eg between-group analysis rather than just the within-group analysis presented eg in the abstract, 'Setting: ... London HASUs compared with the rest of England'; eg your hypothesis, Page 5, Line 43. Can you please make it very clear throughout that you are not presenting between group comparisons – or alternatively, undertake these analyses. If you choose to undertake these additional analyses, you could consider separating out your aim into two parts eg i) to investigate variations in the quality of acute stroke care and outcomes by day and time of admission in London HASUs and the rest of England; ii) to compare variations in these indicators and outcomes between London HASUs and the rest of England. If you choose not to add in these additional analyses can you make it explicit early in the manuscript that you are not doing this comparison so the reader is not expecting to find this.

The reviewer is correct. We should have provided more clarity on the purpose of our analysis. As the reviewer rightfully points out, the goal of the paper is to investigate the variation in the quality of acute stroke care and outcomes by day and time of admission in London HASUs and the rest of England. The focus of the analysis was not in the formal comparison of levels of these measures or the differences in levels between London and the rest of England, but rather in the comparison of within-region variation, i.e. whether within-London HASUs variation is different from that of the Rest of England. On the extensive margin side, we could have observed four possibilities: (1) significant variation in the care quality measures in London HASUs and in the Rest of England, or (2) significant variation in care quality measures in London HASUs, but not in the Rest of England, or (3) significant variation in care quality measures in the Rest of England, but not in London HASUs, or (4) no significant variation in care quality measures either in London HASUs or the Rest of England.

Our statistical tests are based on the within-group variation, i.e. we are testing whether outcomes and care processes vary by the day/time of admission in London HASUs, and in parallel whether the same measures vary by the day/time of admission in the Rest of England. In this paper, we hypothesised that there would be less variation across the week in care quality measures in London HASUs compared with the rest of England, and that this would also translate into less variation in outcomes in London HASUs, which holds if the condition (3) above is binding. To clarify this further in the text, we amended the corresponding wording in the Abstract on page 1, line 22, and Introduction on page 4, lines 41-47.

2) Could you also clarify early in the manuscript whether time of admission means admission to Emergency Department/ hospital, or admission to the HASU/ Stroke Unit.

We have amended the wording to "admission to the hospital" in the Abstract, page 1, line 37.

- 3) Some specific comments around this and other issues are given below.

Abstract:

Page 5, line 29-32: Can you give the time periods of the improved quality of stroke care referenced by citations 28 and 29.

In the referred sources, the time of admission was not available in HES-ONS data. These sources used difference-in-difference design with the cut-off point of July 2010 for London, as the beginning of the 'after' period, and followed patients for 21 month afterwards.

- 4) Page 6, Line 22: Please provide a reference for the 'estimated to be 90%'.

Many thanks for pointing this out. The reference "Royal College of Physicians. Sentinel Stroke National Audit Programme (SSNAP) Clinical Audit. 2015." [35] has been added to the text.

- 5) Page 6, line 39-40: Can you make it clearer that 'assessment by a physiotherapist, OT, SLT within 24 hours' are three separate indicators here please; as written it looks like one combined variable.

Many thanks for spotting this. We have amended the wording to "assessments by a Physiotherapist within 24 hours* and within 72 hours; by Occupational Therapist within 24 hours* and within 72 hours; and by Speech and Language Therapist within 24 hours* and within 72 hours"

- 6) Page 6, Line 54: Please replace 'spell' with 'stay' or 'episode of care'.

Many thanks for spotting this. We have replaced 'spell' with 'stay'.

- 7) Page 7, Line 4: You define LOS but then write 'length of stay' in full in same sentence.

Many thanks for spotting this. We have replaced 'length of stay' with 'LOS' in this sentence.

- 8) Page 7, Line 14: Please define RCP in full.

Many thanks for this suggestion. We have amended this to 'Royal College of Physicians'.

9) Page 9, Line 24-30: Can you please give the directions for those indicators with statistical differences eg mRS pre-stroke, level of consciousness on arrival at hospital, method of admission and time from onset of symptoms. From the previous sentence this reads as a between group differences which I don't think it was, so can you make this clearer please.

We thank the reviewer for this suggestion. In Table 1 we provided p-values for the between-region differences, e.g. London HASUs vs Rest of England. We have amended the wording on page 8, lines 27-46.

10) Page 16, line 46 and 48: To improve the readability can you not ask people to remember or refer back to the 'first' and 'second' hypothesis and improve the clarity of these sentences, particularly the one commencing 'The first hypothesis'.

We have amended these to the following: "The lower variation in care quality measures across the week in London HASUs was confirmed, but only with respect to "front door" measures of acute stroke care. With respect to the health outcomes: there was no variation in mortality at three days and disability at hospital discharge by day and time of admission across the week in London HASUs".

11) For all the tables, can you please provide exact P values. Can you please report the P values to a consistent number of significant figures and three decimal points may be more helpful.

We thank the reviewer for this suggestion. We provided p-values with four decimal points in both, Table 1 and Table 2 (also in supplementary materials).

12) Can you reference the supplementary tables/ figures within the manuscript at the relevant points please and explain any relevant statistical testing undertaken as presented in these tables/ figures.

We thank the reviewer for this suggestion. We have revised and added relevant references on page 7, lines 43-44, and page 12, lines 43-47.

13) Table S3: I found this table confusing. It is not possible to establish the direction of the findings ie if the level of consciousness on arrival or the NIHSS score is lower or higher in relation to those outcomes that are significant (eg Rest of England: brain scan within one hour). This will make it difficult for those wishing to improve these indicators in terms of knowing which of the stroke population to concentrate on. Can you please present these results more clearly. Please provide exact P values.

We thank the reviewer for this suggestion. It is difficult to unambiguously establish the direction of the differences in the results for 42 time periods, therefore, we amended Table S3 that is now equivalent to the Table 2 in the main text. Table S3 provides the results for four time periods controlling for the NIHSS score instead of level of consciousness on arrival, so the reader could compare the results with those in Table 2. We have also reported p-values with four decimal points. The p-values for 42 time periods are, as before, reported at each corresponding figure.

14) With the supplementary figures, I read it as if a P value <0.05 would mean the variation is significant. In some of the figures (eg Figure S1b MRS 3-6) this is significant for the rest of England yet the graph shows a more flattened line for the rest of England when compared with the non-significant HASU. This was a bit confusing. Could you clarify please.

We thank the reviewer for this observation. The variation in London HASUs is not statistically significant because of the larger uncertainty at each time point. So, while based on the point estimates in each period it appears there is more variation in London, the 95% CIs in each period are wider. We have added the following to the text on page 11, lines 50-56: "It is worth noting that, based on the point estimates in each period, it appears there is more variation in mRS scores in London HASUs. One reason why the variation in London HASUs was not statistically significant might be because of the larger uncertainty at each time point.

15) The Figures were very small and hard to read, is it possible to improve the clarity?

The reviewer makes an important point for which we are grateful. The requirements of the journal were to combine the figures in the way they are included in the text, for this reason they look a bit so small. However, we have been able to improve the resolution of the images, so now when the reader zoom in, the numbers are much easier to read.

Reviewer 3

The statistical methods used seem to be appropriate. I have few questions regarding the stats in the paper.

1) For testing the variation across the 42 time periods, the authors seemed to create 41 dummy variables. Which time period was used as the reference time period (the omitted period)? Is the joint p-value (mentioned in page 8 line 12) from the likelihood ratio test?

We thank the reviewer for this clarification. We have used factor variables instead of dummies, and in the median time period – 21st time period which corresponds to Thursday 08:00-11:59 – was chosen as the reference. Further we reported the joint significance from Wald test, this was chosen as less computationally expensive. For comparison, we calculated joint p-values from the likelihood ratio test, please see Table S2A in the Supplementary Materials. It is worth noting that p-values from both tests are very similar, and using either produce identical conclusions.

2) When variation across the time periods was detected, the authors tried to investigate the patterns (page 11 and 12) across the time periods. Were those patterns tested formally? If so, what analyses were performed to test those patterns?

We thank reviewer for pointing this out. We did not formally test the patterns, but rather documented what can be observed from plotting the average predicted probabilities of each outcome in each of 42 time periods controlling for the covariates.

3) The authors performed quite many regression analyses of quality indicators and outcomes. Did the authors consider the adjustment for multiple testing?

We thank reviewer from this suggestion. We have amended the p-value column of Table 2 providing the four digits of p-values that can be contrasted with the p-value threshold of 0.0024 for London HASUs and 0.0025 for the rest of England, corrected for multiple comparisons using Bonferroni approach ($0.05/21=0.0024$ for London HASUs and $0.05/20=0.0025$ for the rest of England, where 0.05 is the reference p-value and 21 and 20 are the number of models). We added these threshold p-values in the footnotes of Table 2.

4) In table 1, I can see that most variables were significant, but the actual differences were quite small as the sample size is very large. In this case, effect size may be helpful.

We thank reviewer for this suggestion. We have added the effect sizes to Table 1.

5) This question may not be statistically relevant. Why did the authors group the time of admission into those 6 four-hour periods? Is there any clinical rationale to do so?

Following Bray et al (2016), ref [16], we chose 4-hour periods because they were the shortest time periods that provided sufficient numbers of patients in each block for model fitting. Further, larger time periods corresponding to weekends and weekdays and regular in-hours or out-of-hours were chosen to aid comparison with previous literature of weekend effects: Monday to Friday 0800–1959 h, Saturday to Sunday 0800–1959 h, Monday to Friday 2000–0759 h, and Saturday to Sunday. This latter specification was also consistent with HASUs rota organisation.

VERSION 2 – REVIEW

REVIEWER	Lawrence Best University College London UK
REVIEW RETURNED	02-Apr-2019

GENERAL COMMENTS	<p>This is a well written piece of important research. I believe it is acceptable for publication. Other things which may be beneficial would be to perform a basic power calculation. This would allow more definitive comments about whether you have proven there is no association vs there is no evidence of association. However this is a minor point which is a general flaw in the literature. Performing follow-up studies to monitor attainment of these outcomes would also be beneficial in order to bring it closer to real time considering shift in NHS strategy which have occurred in the past few years.</p>
-------------------------	---

REVIEWER	Sandy Middleton Australian Catholic University, Australia
REVIEW RETURNED	22-Mar-2019

GENERAL COMMENTS	<p>Thank you for making the changes to this interesting manuscript. All of my comments were adequately addressed.</p> <p>Some small points below: Page 6, last sentence: 'We hypothesized were that there would be' Please fix grammar. Also in that sentence delete the word 'in' prior to 'within London' as it is redundant..</p> <p>Page 9, Patient and public involvement statement, 'of this study and ensure that ...'. Should this be 'ensured'?</p> <p>Table 1, Column labelled 'Effect size'. Should this column be labelled 'Difference'? It seems unconventional to label demographic differences 'effect size' as this usually relates to differences between an intervention ie the effect of an intervention rather than just the differences between proportions. Please check with a statistician.</p> <p>Table 2 and relevant ones in the supplement, it would be very useful to have a superscript symbol after the words 'P-Value' in the column headings, that would lead one's eye down to the comment that you were not using $P < 0.05$ as your measure of statistical significance as it's a bit buried in the footnote of the Tables.</p>
-------------------------	--

REVIEWER	Bichun Ouyang RUMC, US
REVIEW RETURNED	01-Apr-2019

GENERAL COMMENTS	The author has addressed all my questions.
-------------------------	--

VERSION 2 – AUTHOR RESPONSE

Reviewer 1

1) This is a well written piece of important research. I believe it is acceptable for publication. Other things which may be beneficial would be to perform a basic power calculation. This would allow more definitive comments about whether you have proven there is no association vs there is no evidence of association. However this is a minor point which is a general flaw in the literature.

The reviewer makes an important point about the power calculation for which we are grateful. Unfortunately, this might be not feasible in our analysis, because (1) there is no clear primary outcome, but rather a combination of twenty one different measures for processes of care and health outcomes that characterise the quality of acute stroke care, and if we do the power calculation, we would need to do it for every outcome; (2) we are not running the analyses for two groups, that would make this calculation straightforward, but for multiple groups, and in the extreme case we have forty-two groups for different periods across the week, so it is not clear how we do the comparison.

2) Performing follow-up studies to monitor attainment of these outcomes would also be beneficial in order to bring it closer to real time considering shift in NHS strategy which have occurred in the past few years.

We thank reviewer for this suggestion. We have included a following sentence to the Discussion section, Further research sub-section on 21 lines 39-47 : Performing follow-up studies to monitor attainment of key quality indicators and outcomes, complementary to the SSNAP clinical audit annual reports,[33,43] would also be beneficial in order to get an overall picture of national trends and dynamics over time, and look in detail at underlying reasons for that to understand what amendments to clinical guideline for stroke care ought to be proposed in the future.

Reviewer 2

1) Thank you for making the changes to this interesting manuscript. All of my comments were adequately addressed. Some small points below: Page 6, last sentence: 'We hypothesized were that there would be' Please fix grammar. Also in that sentence delete the word 'in' prior to 'within London' as it is redundant..

Many thanks for pointing this out. We have amended the wording to "We hypothesised that there would be less variation across the week in care quality measures within London HASUs compared with the variation in the rest of England, and that this would also translate into less variation in outcomes in London HASUs"

2) Page 9, Patient and public involvement statement, 'of this study and ensure that ...'. Should this be 'ensured'?

Many thanks for spotting this. We have corrected this to 'ensured'.

3) Table 1, Column labelled 'Effect size'. Should this column be labelled 'Difference'? It seems unconventional to label demographic differences 'effect size' as this usually relates to differences between an intervention ie the effect of an intervention rather than just the differences between proportions. Please check with a statistician.

We thank the reviewer for this suggestion. We have changed 'Effect size' to 'Difference' in column 4, Table 1.

4) Table 2 and relevant ones in the supplement, it would be very useful to have a superscript symbol after the words 'P-Value' in the column headings, that would lead one's eye down to the comment that you were not using $P < 0.05$ as your measure of statistical significance as it's a bit buried in the footnote of the Tables.

We thank the reviewer for this suggestion. We have added a superscript symbol '†' after words 'p-value' in ALL tables.

Reviewer 3

The author has addressed all my questions.

We are happy that we were able to address all reviewer's questions and thank very much for helping to improve our paper.