

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)	Disentangling quality and safety indicator data: a longitudinal, comparative study of hand hygiene compliance and accreditation outcomes in 96 Australian hospitals
AUTHORS	Mumford, Virginia; Greenfield, David; Hogden, Anne; Debono, Deborah; Gospodarevskaya, Elena; Forde, Kevin; Westbrook, Johanna; Braithwaite, Jeffrey

VERSION 1 - REVIEW

REVIEWER	Christopher Fuller UCL Research Dept of Infection and Population Health
REVIEW RETURNED	14-Apr-2014

GENERAL COMMENTS	<p>As a non-statistician I found this paper quite difficult to read, but I was glad that I made the effort. I think that I understood it eventually, but if I hadn't been reviewing it I would have given up in the methods section. If your target audience is readers like me you'll need to make it a bit more digestible.</p> <p>page 5 lines 38+. Might be useful to have a list of the standards or an overall summary if there are too many.</p> <p>page 6 lines 20-48. Would be easier to read if variables were bulleted or tabulated. List in the same order as table 2.</p> <p>What is a "cycle" and why is it necessary to have it in the model?</p> <p>What are "adjusted separations"?</p> <p>How large are "large", "medium" and "small"?</p> <p>page 7 lines 18-26. Does this belong in the methods section?</p> <p>Table 1. I don't think that it is very useful. the text is enough.</p> <p>Page 8 lines 19-25. Cycles again!</p> <p>lines 30-49. Would be easier if the results were reported in the same order as they are presented in the table.</p> <p>Table 2. Give a full title for the table. Might be useful to have the comparison group and numbers on the table as well. Didn't understand the survey models columns.</p> <p>page 9 lines 42-49. Don't understand this.</p> <p>page 9 line 54-page 10 line 15. Don't understand this either. I'm not</p>
-------------------------	--

	<p>even sure if this is results or methods.</p> <p>DISCUSSION section. This needs more structure. Would be easier if you summarised your main results first ie hand-hygiene increased over time, there was a relationship between compliance and hospital size, there was no relationship between IC/accreditation score and HHC. Then move on to the comparisons with other literature.</p>
--	--

REVIEWER	Susan Burnett CPSSQ, Imperial College London UK
REVIEW RETURNED	07-May-2014

GENERAL COMMENTS	<p>The acronyms need to be spelled out in full in the abstract</p> <p>There are a lot of acronyms used in the paper and whilst they are set out early in the text the paper became harder to read as the number of acronyms increased. I would suggest these are put in a box in the text for reference. Also where important points are being made, acronyms should be avoided so the reader can digest the point being made without having to go back to look at what each acronym means.</p> <p>First line of the introduction: should read 'assess compliance' not 'ensure compliance'?</p> <p>Results: page 7 lines 52-56: the fact that the large hospitals showed the biggest increase may be more likely due to the fact that they had further to go (lots of low hanging fruit) rather than that they were early adopters?</p> <p>Discussion: Page 10 line 26: What year are the results from the US study in 40 hospitals - how long after the guidelines were introduced in 2002? Page 10 line 50: suggest re-word from 'is not practical' to 'may not be practical' Page 11 lines 26-33: there are other limitations of accreditation, such as how leadership and other criteria are measured - as with HH some of this is open to bias Page 11 line 31 - I didn't understand the sentence about the model with higher compliance rates leading to higher IC scores in the next survey - needs more explaining.</p> <p>Policy Implications Page 11 lines 47-53: suggest other implications include a review of the accreditation standards with regard to research and leadership - to what extent do these really impact on IC if they are never achieved in small hospitals. Also to investigate what the impact is on small hospitals of never being able to achieve a high accreditation rating?</p> <p>Conclusions I thought the conclusions in the abstract summarised things better than the conclusions in the paper Page 12 line 9: the word 'complementary' - suggest this is re-worded to spell out that accreditation outcomes and HH audit data should be considered together since they report different aspects of IC.</p>
-------------------------	---

REVIEWER	Dr. M. Lindsay Grayson Hand Hygiene Australia Austin Health Australia
REVIEW RETURNED	19-Jun-2014

GENERAL COMMENTS	<p>Overall, this is an interesting (albeit rather statistically complex) study that assesses the correlation between two measures of hospital quality performance. The results are notable - although whether they will be of specific interest to readers of the BMJ Open, is less clear.</p> <p>However, I have a number of queries that should be addressed:</p> <ol style="list-style-type: none"> 1. Australia has recently changed its approach to hospital accreditation and now uses the Australian Commission for Safety and Quality in Health Care (ACSQHC) Standards rather than those previously used by ACHS. Can the authors confirm that all the accreditation scoring data reported here is using the previous ACHS Standards, rather than the new ACSQHC Standards? 2. Was all the hospital accreditation standardised across sites and time periods? One presumes, yes, but even within the ACHS Accreditation approach there was often some variability between assessors – so, perhaps this should be acknowledged as a potential (unavoidable) limitation. 3. The authors compare the rates of Hand Hygiene Compliance (HHC) between institutions and across various time periods. However, a key additional factor that needs to be considered is whether each site submitted the required minimum number of HH Moments for each audit period, since a good rate of HHC on a small (inadequate) number of Moments is not as valid as if the required amount of data was submitted. One option would be to only analyse HHC data on sites and time periods where the minimum number of audited Moments required by Hand Hygiene Australia was reported and exclude the other data. What do the authors think about this? 4. The findings that small and medium-sized hospitals regularly reported higher rates of HHC than larger hospitals is consistent with national data on the Hand Hygiene Australia website. As the authors suggest, it is likely to be related to the manageable organisational structure in these sites that helps embed quality culture-change activities such as improving HHC. 5. Page 10/lines 25-27. This statement regarding HHC rates in the USA is misleading since the US authors did not use the WHO 5 Moments audit tool – thus the quoted rates of HHC may not be comparable with those identified in Australia or other sites to use the WHO tool. <p>I think an opinion on whether the statistical methods used were appropriate, would be worthwhile -- my sense is that the statistical methods are OK, but I am not a statistician and the study's conclusions are heavily based on the statistical approach used.</p>
-------------------------	--

REVIEWER	Jennifer Kuntz, PhD KP Center for Health Research, USA
REVIEW RETURNED	10-Jul-2014

- The reviewer completed the checklist but made no further comments.

REVIEWER	Eric Lau The University of Hong Kong, Hong Kong
REVIEW RETURNED	15-Jul-2014

GENERAL COMMENTS	<p>The author attempted to disentangle quality and safety indicator data, using accreditation outcomes and hand hygiene compliance rates as an example. While the mutual dependence of the two main outcomes brings complexity to the problem, the authors also identified confounding by the effect of hospital size on the accreditation program. There are some concerns in the modeling approach which needs to be clarified. Overall, this study draws attention to the limitation of using a single indicator versus multidimensional assessment.</p> <p>Specific comments</p> <p>Abstract</p> <ul style="list-style-type: none"> • Results: Please clarify if you mean “high IC accreditation scores in one survey”. It seems to me that “higher IC accreditation scores in one survey” means to have different IC accreditation scores in the two surveys. <p>Analysis</p> <ul style="list-style-type: none"> • Please briefly explain why a restrictive model is fitted (audits 3 to 8). Was it fitted to exclude early adopters of the accreditation program? • Please describe the rationale and detail of the “Survey models” • A multilevel linear model was fitted for the binary outcome, with hand hygiene compliance rates up to 80-100% in many hospitals plus a “ceiling effect” as described in the results. Have you considered using a logistic model? • Please describe that the models will be tested using likelihood ratio test. <p>Results</p> <ul style="list-style-type: none"> • Line 23: data “from” some of the smaller hospitals... • Table 2 title: please add the outcome of the multilevel model. • Table 2: “higher” should be “high” for the first two rows • Hospital group was accounted for in the main and restricted models. Is there any reason that it is not included in the survey models? • Figure 2: add axis label for no of hospitals • Page 9, line 47. The p values were not found in Table 2. • The section “testing the model” is better placed before the result of the final multilevel model. • Page 9, line 56. Please change $p=0.000$ to $p < 0.001$
-------------------------	---

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

As a non-statistician I found this paper quite difficult to read, but I was glad that I made the effort. I think that I understood it eventually, but if I hadn't been reviewing it I would have given up in the methods section. If your target audience is readers like me you'll need to make it a bit more digestible. Thank you, this is useful feedback. We have re-written sections of the methods and we believe this

section is now easier to read.

page 5 lines 38+. Might be useful to have a list of the standards or an overall summary if there are too many.

There were 14 standards during our study period. We have incorporated a summary of the standard topics into the supplemental material.

page 6 lines 20-48. Would be easier to read if variables were bulleted or tabulated. List in the same order as table 2

We have restructured the description of the variables in the methods section in line with the findings presented in table 2.

What is a "cycle" and why is it necessary to have it in the model?

We have clarified this in the methods and results section. We used the cycle variable to test for a timing effect between hospitals having accreditation surveys in 2009 or 2011 (cycle =0) and surveys in 2010 and 2012 (cycle =1).

What are "adjusted separations"?

Separations are episodes of care (i.e. in essence an admission to hospital) and in this case are adjusted for diagnostic complexity and resources needed to treat that episode. We have clarified this in the text.

How large are "large", "medium" and "small"?

The size is determined by size and activity according to the Australian Institute of Health and Welfare. We have added a copy of the table to the supplemental materials and clarified this in the text.

page 7 lines 18-26. Does this belong in the methods section?

Thank you - we have moved the discussion of missing data to the methods section.

Table 1. I don't think that it is very useful. the text is enough.

We have replaced this table with a table describing the basic characteristics of the variables.

Page 8 lines 19-25. Cycles again!

We have reworded the text to clarify the discussion.

lines 30-49. Would be easier if the results were reported in the same order as they are presented in the table

Thank you - we have amended the text.

Table 2. Give a full title for the table. Might be useful to have the comparison group and numbers on the table as well. Didn't understand the survey models columns

Thank you, we have amended the title and deleted the survey columns in the model.

page 9 lines 42-49. Don't understand this

We have removed the discussion on surveys as the results were not significant and detracted from our main argument.

page 9 line 54-page 10 line 15. Don't understand this either. I'm not even sure if this is results or methods

We have moved this section to above the results of the multilevel model on the advice of another reviewer. We believe an appraisal of the model is required to give readers an understanding of our modelling approach and its robustness.

Reviewer 2

The acronyms need to be spelled out in full in the abstract
We have amended the text in the abstract to reflect this.

There are a lot of acronyms used in the paper and whilst they are set out early in the text the paper became harder to read as the number of acronyms increased. I would suggest these are put in a box in the text for reference. Also where important points are being made, acronyms should be avoided so the reader can digest the point being made without having to go back to look at what each acronym means

We have reduced the number of acronyms and spelled these out in the text.

First line of the introduction: should read 'assess compliance' not 'ensure compliance'?
Agreed - thank you for picking this up – we have amended the text.

Results: page 7 lines 52-56: the fact that the large hospitals showed the biggest increase may be more likely due to the fact that they had further to go (lots of low hanging fruit) rather than that they were early adopters

We have tried to clarify this further. The large hospitals went from having the lowest percentage meeting target to the highest, but we also wanted to point out that the two largest groups (principal referral and large hospitals) were also over-represented in the first two surveys which would impact the results. We have replaced figure 2 with two graphs (figures 2 and 3) which show the mean compliance rates by hospital and also the percentage of hospitals meeting the target to help illustrate this point.

Page 10 line 26: What year are the results from the US study in 40 hospitals - how long after the guidelines were introduced in 2002?

We have amended this section in line with comments from one of the other reviewers in order to clarify the study data.

Page 10 line 50: suggest re-word from 'is not practical' to 'may not be practical'
Agreed - thank you for picking this up – we have amended the text.

Page 11 line 31 - I didn't understand the sentence about the model with higher compliance rates leading to higher IC scores in the next survey - needs more explaining

We have expanded the section on inter-reliability in the limitations section. We have also included a discussion as to whether the criteria for higher scores need to include measures that are achievable by all groups of hospitals. We agree with the reviewer comments about the general structure of the accreditation process but felt this was beyond the scope of this article

Policy Implications

Page 11 lines 47-53: suggest other implications include a review of the accreditation standards with regard to research and leadership - to what extent do these really impact on IC if they are never achieved in small hospitals. Also to investigate what the impact is on small hospitals of never being able to achieve a high accreditation rating?

These are excellent suggestions and we have incorporated some of this discussion into the implications section as per the comments above.

Conclusions: I thought the conclusions in the abstract summarised things better than the conclusions in the paper

These are excellent suggestions and we have incorporated some of this discussion into the implications section as per the comments above

Page 12 line 9: the word 'complementary' - suggest this is re-worded to spell out that accreditation outcomes and HH audit data should be considered together since they report different aspects of IC. Agreed – we have amended the text.

Reviewer 3

Overall, this is an interesting (albeit rather statistically complex) study that assesses the correlation between two measures of hospital quality performance. The results are notable - although whether they will be of specific interest to readers of the BMJ Open, is less clear.

We thank the reviewer for the endorsement of the study.

Australia has recently changed its approach to hospital accreditation and now uses the Australian Commission for Safety and Quality in Health Care (ACSQHC) Standards rather than those previously used by ACHS. Can the authors confirm that all the accreditation scoring data reported here is using the previous ACHS Standards, rather than the new ACSQHC Standards?

This is a good distinction – we used accreditation data from hospitals surveyed against ACHS standards. Our study period is before the introduction of the NSQHS standards and we have made this clearer in the text.

Was all the hospital accreditation standardised across sites and time periods? One presumes, yes, but even within the ACHS Accreditation approach there was often some variability between assessors – so, perhaps this should be acknowledged as a potential (unavoidable) limitation.

Inter-surveyor reliability is definitely an issue – we have clarified the text to highlight this point better.

The authors compare the rates of Hand Hygiene Compliance (HHC) between institutions and across various time periods. However, a key additional factor that needs to be considered is whether each site submitted the required minimum number of HH Moments for each audit period, since a good rate of HHC on a small (inadequate) number of Moments is not as valid as if the required amount of data was submitted. One option would be to only analyse HHC data on sites and time periods where the minimum number of audited Moments required by Hand Hygiene Australia was reported and exclude the other data. What do the authors think about this?

This is an important issue, especially for the smaller facilities. We discuss using the minimum of 50 moments recommended by Hand Hygiene Australia at the time of the study. We understand this has been subsequently changed to a minimum of 100 moments.

The findings that small and medium-sized hospitals regularly reported higher rates of HHC than larger hospitals is consistent with national data on the Hand Hygiene Australia website. As the authors suggest, it is likely to be related to the manageable organisational structure in these sites that helps embed quality culture-change activities such as improving HHC.

It is positive to have our analysis and ideas confirmed; we thank the reviewer for this.

Page 10/lines 25-27. This statement regarding HHC rates in the USA is misleading since the US authors did not use the WHO 5 Moments audit tool – thus the quoted rates of HHC may not be comparable with those identified in Australia or other sites to use the WHO tool.

This is a helpful insight and we have clarified the text.

I think an opinion on whether the statistical methods used were appropriate, would be worthwhile -- my sense is that the statistical methods are OK, but I am not a statistician and the study's conclusions are heavily based on the statistical approach used.

We believe our responses to Reviewer 5 address this issue.

Reviewer 4

No Comments

We thank the reviewer for their effort and acceptance of the manuscript.

Reviewer 5

The author attempted to disentangle quality and safety indicator data, using accreditation outcomes and hand hygiene compliance rates as an example. While the mutual dependence of the two main outcomes brings complexity to the problem, the authors also identified confounding by the effect of hospital size on the accreditation program. There are some concerns in the modeling approach which needs to be clarified. Overall, this study draws attention to the limitation of using a single indicator versus multidimensional assessment.

We thank the reviewer for their summary, useful suggestions for improvement and overall endorsement of the manuscript.

Results: Please clarify if you mean “high IC accreditation scores in one survey”. It seems to me that “higher IC accreditation scores in one survey” means to have different IC accreditation scores in the two surveys.

We have clarified the text to explain this categorical variable comprises three outcomes: zero (the reference case) where infection control scores met accreditation standards but higher scores were not achieved in either survey; hospitals achieving higher scores in one survey; and hospitals achieving higher scores in both surveys.

Analysis : Please briefly explain why a restrictive model is fitted (audits 3 to 8). Was it fitted to exclude early adopters of the accreditation program

We used a restrictive model as the programme was progressively rolled out across the state and we wanted to investigate any differences due to principal and large hospitals being over-represented in the first two audits.

Please describe the rationale and detail of the “Survey models”

We have deleted the section on the survey models as the results were not significant and we felt they detracted from the main argument.

A multilevel linear model was fitted for the binary outcome, with hand hygiene compliance rates up to 80-100% in many hospitals plus a “ceiling effect” as described in the results. Have you considered using a logistic model?

We have clarified the text to state that “To analyse the matched data we used hand hygiene compliance rates as our outcome of interest, expressed as a continuous outcome variable” on Page 6. Logistic regression requires a binary variable outcome, which makes it unsuitable for our purposes.

Please describe that the models will be tested using likelihood ratio test

We describe using a likelihood ratio to test a linear regression model against a random intercept model in the model selection section.

Line 23: data “from” some of the smaller hospitals...

Thank you the text has been amended.

Table 2 title: please add the outcome of the multilevel model.

Thank you the title has been amended.

Table 2: “higher” should be “high” for the first two rows

We assessed that higher scores more accurately described the relative nature of the accreditation scoring process.

Hospital group was accounted for in the main and restricted models. Is there any reason that it is not included in the survey models?

We have deleted the section on survey models.

Figure 2: add axis label for no of hospitals

We have replaced figure 2 and deleted the number of hospitals for clarity (the numbers are expressed in table 2).

Page 9, line 47. The p values were not found in Table 2.

Thank you - these have now been added.

The section "testing the model" is better placed before the result of the final multilevel model

Thank you the text has been amended.

Page 9, line 56. Please change $p=0.000$ to $p < 0.001$

Thank you the text has been amended.

VERSION 2 – REVIEW

REVIEWER	Christopher Fuller University College London
REVIEW RETURNED	28-Aug-2014

GENERAL COMMENTS	I still do not understand the accreditation process. In particular I don't understand the difference between survey 1 survey 2, cycle 0, cycle 1. Minor points line 13 p8 I presume that a higher infection control score is EA. Table 1 you have a row for higher IC scores in one survey and higher IC scores in either survey. This doesn't make sense.
-------------------------	---

REVIEWER	Eric Lau The University of Hong Kong, Hong Kong
REVIEW RETURNED	15-Aug-2014

GENERAL COMMENTS	The authors have answered all of my previous comments satisfactorily. I have only one more comment. In the main text and the tables, please use the term "high IC scores" instead of "higher IC scores" to avoid confusion.
-------------------------	--

VERSION 2 – AUTHOR RESPONSE

Comments from Reviewer 1

Minor points line 13 p8 I presume that a higher infection control score is EA. Table 1 you have a row for higher IC scores in one survey and higher IC scores in either survey. This doesn't make sense. The data for high IC scores in either survey indicate hospitals that either had a high score in one survey or in both, i.e. a combination of these two. We agree this might appear confusing and have deleted the line in Table 1 for higher scores for either survey.

Comments from Reviewer 5

In the main text and the tables, please use the term “high IC scores” instead of “higher IC scores” to avoid confusion.

We have made this change in the text and the tables