

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. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

| | |
|----------------------------|--|
| TITLE (PROVISIONAL) | Estimating treatment effects in randomized controlled trials with non-compliance: a simulation study |
| AUTHORS | Ye, Chenglin; Beyene, Joseph; Browne, Gina; Thabane, Lehana |

VERSION 1 - REVIEW

| | |
|------------------------|---|
| REVIEWER | Conrad Kabali Health Quality Ontario, Canada |
| REVIEW RETURNED | 25-Apr-2014 |

| | |
|-------------------------|---|
| GENERAL COMMENTS | <p>Pg 2, line 12. It is never guaranteed that prognostic factors will balance in a single RCT. So the statement "...prognostic factors are balanced by randomization", should be rephrased into something like "....prognostic factors are expected to balance by randomization"</p> <p>Papers proposing solutions/remedial measures on broken randomization are vast. What does this paper add to existing methodological literature? This should explained.</p> <p>Pg 3, line 28. The phrase "....Randomized controlled trials (RCTs) have been the gold standard to assess new health interventions" can be modified to read "Randomized controlled trials (RCTs) are often considered as gold standard for assessing new health interventions" to avoid a misconception that RCTs are always least biased compared to other types of studies.</p> <p>Pg 5, line 10. The statement "For example, previous study did not consider the case of no crossover between treatment groups" suggest that this analysis is a special case of Bang et al when cross over is absent. The claim that a special case scenario is more insightful than a general case goes against the conversational wisdom. Can the authors defend their claim (e.g. by giving a specific example where the approach by Bang et al was less insightful compared to theirs)?</p> <p>In some sections authors would state that partial compliers scenario was not considered, but along the way the partial compliers appear. Can the authors make sure that the message is consistent throughout?</p> <p>Page 7, line 10-11. The sentence "The degree of non-compliance referred to the proportion of intervention that non- compliers still received even they did not fully comply" has grammatical errors and should be re-phrased. Similar grammatical errors can also be found on page 7, line 16-18.</p> <p>Page 8. The sentence "we chose $\mu_0 = 59$ for the usual care group; and $\mu_1 = 89, 74$ and 59 for the intervention group" is confusing. The authors assume that the counterfactual outcomes are stochastic with normal errors, which is fine. If that's the case I would expect each individual to have only one value for μ_0 and only value for</p> |
|-------------------------|---|

| | |
|-----------------|---|
| | <p>mu1. Instead 3 different values for mu1 are provided, what does this mean? I don't get the logic.</p> <p>Similarly, the statement "The mean score of 59 for the usual care group and the standard deviation of 10 were estimated from the CTN trial" needs further elaboration to be counterfactually meaningful. Just to give an example, in RCTs we can classify individuals into four causal (latent) types: 1) those who will never experience the outcome regardless of the treatment given, 2) those who will experience an outcome regardless of the treatment given, 3) those who will only experience an outcome if the treatment is given, and 4) those who will always experience an outcome regardless of the treatment is given. Assigning a single individual a counterfactual value from the mean of other individuals who can potentially be very different from him/her with regard to treatment response doesn't make sense in the counterfactual sense.</p> <p>Page 8, line 25. What does this statement mean "Assuming that the outcome score was proportional to the actual level of treatment received..."? The treatment is dichotomous but the outcome is continuous –I find it hard to follow this. Did they mean the treatment dose?</p> <p>Page 8, line 32. What does the proportion of intervention mean? Did they want to talk about incomplete dose? Does the pattern of non-compliance make a difference (e.g. if a subject comply in the first two occasions and not comply in the next two occasions will her outcome be different from the situation where she alternates treatment in each subsequent occasion?) This needs clarification.</p> <p>I can't distinguish counterfactual/causal notations (e.g. in the ITT formula) from purely statistical ones (e.g. in the AT formula). In ITT, the expression mu1 reads that: the mean outcome will be mu if subjects are ASSIGNED treatment. In AT mu1 means that the mean outcome is mu1 CONDITIONAL on the fact that were observed to be receiving treatment. These two statements do not mean the same thing, and different notations should be used so as not to confuse readers.</p> <p>Page 10. The statement "The theory and assumptions of IV analysis were thoroughly discussed in the literature" should re-worded to "The theory and assumptions of IV analysis are thoroughly discussed in the literature"</p> <p>Page 10. The CACE/principal stratum formula represents counterfactual outcomes. Again there is a mix between causal and purely statistical notations.</p> <p>It is unclear why the authors decided to impute counterfactual outcomes using the EM algorithm.</p> <p>In the limitation section of the discussion it should be mentioned that there exist other methods to account for non-compliance that were not considered in their paper e.g. Robins g-methods. Another oversight is the failure to mention that the instrumental variable formula applied in this study only works for linear models. For non linear models it is impossible to obtain an unbiased point estimate of the treatment effect without making strong (often unrealistic) assumptions.</p> |
| REVIEWER | Dr Freedom Gumedze Department of Statistical Sciences |

| | |
|------------------------|--|
| | P D Hahn Building, Room 6.63 University of Cape Town Private Bag |
| REVIEW RETURNED | 05-May-2014 |

| | |
|-------------------------|---|
| GENERAL COMMENTS | <p>Major comments:</p> <ol style="list-style-type: none"> 1. Major shortcomings of the paper include non-incorporation of prognostic factors in their analyses and the assumption that outcomes for non-compliers are collectable. These have been acknowledged by the authors. 2. In their conclusions the authors seem to focus on bias in comparing the methods for the various scenarios. How important do they view MSE and coverage, especially since some of the methods gave 0% coverages for some of the scenarios? <p>Minor corrections:</p> <ol style="list-style-type: none"> 1. Page 7: Line 5: Define/explain the phrase "counterfactual outcomes" as a causal inference concept. 2. Page 8: Line 52: Rephrase end of sentence as "...that we performed." 3. Page 9: Line 3: Define the criteria: bias, MSE and 95% coverage. 4. Page 11: Line 43: Rephrase "...would approach infinity..." |
|-------------------------|---|

VERSION 1 – AUTHOR RESPONSE

Reviewer Name Conrad Kabali

Institution and Country Health Quality Ontario, Canada

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

Pg 2, line 12. It is never guaranteed that prognostic factors will balance in a single RCT. So the statement "...prognostic factors are balanced by randomization", should be rephrased into something like "...prognostic factors are expected to balance by randomization"

Response: The sentence has been revised and now reads as "..., whose prognostic factors are expected to balance by randomization" on page 2.

Papers proposing solutions/remedial measures on broken randomization are vast. What does this paper add to existing methodological literature? This should be explained.

Response: To strengthen the background, we have revised the Introduction section on pages 4-5 by adding more descriptions (with references) on the analyses to deal with non-compliance and highlighted how our simulation study is different from or add to previous studies.

Pg 3, line 28. The phrase "...Randomized controlled trials (RCTs) have been the gold standard to assess new health interventions" can be modified to read "Randomized controlled trials (RCTs) are often considered as gold standard for assessing new health interventions" to avoid a misconception that RCTs are always least biased compared to other types of studies.

Response: The sentence has been revised as suggested.

Pg 5, line 10. The statement “For example, previous study did not consider the case of no crossover between treatment groups” suggest that this analysis is a special case of Bang et al when cross over is absent. The claim that a special case scenario is more insightful than a general case goes against the conversational wisdom. Can the authors defend their claim (e.g. by giving a specific example where the approach by Bang et al was less insightful compared to theirs)?

Response: We apologize for the confusion. What we meant was that our simulation study considered additional non-compliant scenarios that were not considered by previous studies—another important contribution of our simulation study to the existing literature. We have revised the sentences accordingly to clarify the message.

In some sections authors would state that partial compliers scenario was not considered, but along the way the partial compliers appear. Can the authors make sure that the message is consistent throughout?

Response: Partial complier scenarios were considered in our simulation. We have checked the paper thoroughly to make sure that the message is consistent throughout.

Page 7, line 10-11. The sentence “The degree of non-compliance referred to the proportion of intervention that non- compliers still received even they did not fully comply” has grammatical errors and should be re-phrased. Similar grammatical errors can also be found on page 7, line 16-18.

Response: We have revised the sentences on pages 7 and 8. Now it reads as: the degree of non-compliance refers to the proportion of interventional components that a patient did not receive according to the protocol.

Page 8. The sentence “we chose $\mu_0 = 59$ for the usual care group; and $\mu_1 = 89, 74$ and 59 for the intervention group” is confusing. The authors assume that the counterfactual outcomes are stochastic with normal errors, which is fine. If that’s the case I would expect each individual to have only one value for μ_0 and only value for μ_1 . Instead 3 different values for μ_1 are provided, what does this mean? I don’t get the logic.

Response: We apologize for the confusion. What we meant was that we chose $\mu_0 = 59$ for the effect of usual care and three different cases for the effect of treatment μ_1 : $89, 74$ and 59 . Each case was considered in a separate simulation. We have revised the sentences accordingly to clarify the message.

Similarly, the statement “The mean score of 59 for the usual care group and the standard deviation of 10 were estimated from the CTN trial” needs further elaboration to be counterfactually meaningful. Just to give an example, in RCTs we can classify individuals into four causal (latent) types: 1) those who will never experience the outcome regardless of the treatment given, 2) those who will experience an outcome regardless of the treatment given, 3) those who will only experience an outcome if the treatment is given, and 4) those who will always experience an outcome regardless of the treatment is given. Assigning a single individual a counterfactual value from the mean of other individuals who can potentially be very different from him/her with regard to treatment response doesn’t make sense in the counterfactual sense.

Response: We realize that the wording we used is rather confusing. We actually defined the counterfactual outcomes from the population level instead of the patient level and then generated individual patients’ outcomes from a Normal distribution using the assumed population mean and

standard deviation. We have revised the paragraph on pages 8-9 to clarify the message.

Page 8, line 25. What does this statement mean “Assuming that the outcome score was proportional to the actual level of treatment received...”? The treatment is dichotomous but the outcome is continuous –I find it hard to follow this. Did they mean the treatment dose?

Response: This sentence has been removed to avoid confusion without loss of information.

Page 8, line 32. What does the proportion of intervention mean? Did they want to talk about incomplete dose? Does the pattern of non-compliance make a difference (e.g. if a subject comply in the first two occasions and not comply in the next two occasions will her outcome be different from the situation where she alternates treatment in each subsequent occasion?) This needs clarification.

Response: This sentence has been revised and now reads as “...was the degree of treatment compliance with the protocol for patient i”.

I can’t distinguish counterfactual/causal notations (e.g. in the ITT formula) from purely statistical ones (e.g. in the AT formula). In ITT, the expression μ_1 reads that: the mean outcome will be μ_1 if subjects are ASSIGNED treatment. In AT μ_1 means that the mean outcome is μ_1 CONDITIONAL on the fact that were observed to be receiving treatment. These two statements do not mean the same thing, and different notations should be used so as not to confuse readers.

Response: We have replaced the notations on pages 9-10 to avoid confusion. We also added a sentence in Method section on page 4 to indicate: “A conceptual difference among all those methods is that the ITT, IV, and CACE approaches estimate treatment effects by preserving randomization or accounting for potential confounding, but the AT and PP approaches do not.”.

Page 10. The statement “The theory and assumptions of IV analysis were thoroughly discussed in the literature” should re-worded to “The theory and assumptions of IV analysis are thoroughly discussed in the literature”

Response: This sentence has been revised as suggested.

Page 10. The CACE/principal stratum formula represents counterfactual outcomes. Again there is a mix between causal and purely statistical notations.

It is unclear why the authors decided to impute counterfactual outcomes using the EM algorithm. In the limitation section of the discussion it should be mentioned that there exist other methods to account for non-compliance that were not considered in their paper e.g. Robins g-methods.

Response: Dealing with unknown compliant status for some patients, e.g. the ones whose compliance status could not be directly observed, is a key challenge in estimating the CACE. The EM algorithm is one solution, which has been widely used in making CACE inferences [1]. We have revised the corresponding paragraph on page 11 to clarify this message.

In the Method section on pages 4-5, we also mention other proposed approaches to correct for non-compliance in RCTs and provide the key reference. However, we only included the ITT, AT, PP, IV, and CACE methods in our comparison since other approaches have not been widely adopted.

Another oversight is the failure to mention that the instrumental variable formula applied in this study only works for linear models. For non linear models it is impossible to obtain an unbiased point estimate of the treatment effect without making strong (often unrealistic) assumptions.

Response: We have revised the paragraph on page 10 and mentioned that the IV estimator used is

for linear models.

Reference

1. Yau LHY, Little RJ. Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association* 2001, 96(456):1232-1244.

Reviewer Name Dr Freedom Gumedze

Institution and Country Department of Statistical Sciences

P D Hahn Building, Room 6.63

University of Cape Town

Private Bag

Rondebosch 7701

South Africa

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

Major comments:

1. Major shortcomings of the paper include non-incorporation of prognostic factors in their analyses and the assumption that outcomes for non-compliers are collectable. These have been acknowledged by the authors.

Response: We thank the reviewer for commenting the limitations of our simulation. We have acknowledged those limitations in the Discussion. In addition, we would like to point out that we did consider different associations between patients' outcome and non-compliant behaviours in our simulation although specific prognostic factors were not considered. Also, imputation techniques can be applied to handle missing data. The Discussion on pages 13-14 has been revised to clarify the message.

2. In their conclusions the authors seem to focus on bias in comparing the methods for the various scenarios. How important do they view MSE and coverage, especially since some of the methods gave 0% coverages for some of the scenarios?

Response: The MSE and coverage were important when considering the methods of analysis. We have added summary of the MSE and 95% coverage results in the Results section on page 12. We did not focus on the discussion of MSE and 95% coverage because the results from comparing the MSE and 95% coverage were consistent with those from comparing the bias of estimates. We have revised the corresponding paragraph on page 12 to clarify the message.

Minor corrections:

1. Page 7: Line 5: Define/explain the phrase "counterfactual outcomes" as a causal inference concept.

Response: We have revised the corresponding paragraph on page 7 to describe 'counterfactual outcomes' in a causal inference concept. Please refer to our earlier responses to comments 8 and 11 made by reviewer 1.

2. Page 8: Line 52: Rephrase end of sentence as "...that we performed."

Response: The sentence has been rephrased and now reads as "...methods that we compared".

3. Page 9: Line 3: Define the criteria: bias, MSE and 95% coverage.

Response: We have added descriptions on how bias, MSE, and 95% coverage were defined on pages 9-10 with reference.

4. Page 11: Line 43: Rephrase "....would approach infinity..."

Response: The sentence has been rephrased and now reads as "when non-compliant rate was equal between groups, the denominator was zero and the estimate became undefined".

VERSION 2 – REVIEW

| | |
|------------------------|---|
| REVIEWER | Conrad Kabali Health Quality Ontario, Canada |
| REVIEW RETURNED | 01-Jun-2014 |

| | |
|-------------------------|---|
| GENERAL COMMENTS | Authors statement that counterfactual outcomes are fully observable at the population level is not accurate. Can that be re-edited? |
|-------------------------|---|

| | |
|------------------------|---|
| REVIEWER | Dr Freedom gumedze Dr Freedom N. Gumedze Department of Statistical Sciences P D Hahn Building, Room 6.63 University of Cape Town Private Bag |
| REVIEW RETURNED | 30-May-2014 |

| | |
|-------------------------|---|
| GENERAL COMMENTS | The authors have addressed my earlier concerns. i am happy for the paper to be published. |
|-------------------------|---|