

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)	Can propensity score matching be applied to cross-sectional data to evaluate Community-based Rehabilitation? Results of a survey implementing the World Health Organization's Community-based Rehabilitation Indicators in Vietnam
AUTHORS	Mason, Catherine; Sabariego, Carla; Thắng, Đoàn Mạnh; Weber, Jörg

VERSION 1 – REVIEW

REVIEWER	Anestis Divanoglou University of Iceland, Iceland
REVIEW RETURNED	26-Mar-2018

GENERAL COMMENTS	<p>This is a well-written manuscript and the introduction provides a good overview of the complexity in assessing the effectiveness of Community-based Rehabilitation (CBR) programs. The study used data from two Vietnamese provinces: one where CBR had been implemented and the other where CBR had not been implemented. The authors used data from these provinces to illustrate how propensity score matching (PSM) may be used to reduce bias between groups in cross-sectional studies evaluating the effectiveness of CBR.</p> <p>The following issues are not clarified:</p> <ol style="list-style-type: none">1. The manuscript does not provide sufficient details about the type and duration of the CBR intervention.2. According to Table 2, similar number of cases in both provinces had access to rehabilitation services. Does this mean that the main difference was that one was as CBR and the other was not?3. The manuscript does not mention the number of potential cases identified in each province through the governmental records. <p>Furthermore, the authors do not state the response rate, or whether they used some form of randomisation to identify the sample, or if they tried to include as many as possible. It is likely that the study had selection bias.</p> <p>According to the study findings, individuals from the CBR province had significantly worse community scores than those in the non-CBR province. The authors stated that two factors may have contributed to that: (1) cross-sectional data not being a suitable method to assess such outcomes, and (2) their outcome measures not being aligned with the purpose of the intervention (Page 8; lines 18-31). In my opinion, both these factors are likely and negatively impact the methodological quality of the manuscript.</p>
-------------------------	--

	<p>Because the original data seem not straight forward, with the control group scoring better as compared to the intervention group, it becomes a bit complex and confusing when using these “exemplar” data to explain the effect of the PSM model. PSM seems a complex concept, and if data used are also complex, then the illustrated analysis and interpretation become unnecessary convoluted. This may have negative implications on reader’s ability to understand the key message of the manuscript.</p> <p>Last, the importance and necessity for this manuscript is not established clearly. Why is it important to present such data and how is this useful?</p> <p>Overall, I have strong doubts whether cross-sectional design is suitable to evaluate CBR programs. The use of PSM as a statistical method may be able to control for specific bias, but probably cannot control for the strong limitations associated with using a cross-sectional design to evaluate CBR programs. Furthermore, I cannot see how the PMR as a technique will support researchers to use cross-sectional data to show effects of CBR.</p>
--	---

REVIEWER	Fred Stephen Sarfo Kwame Nkrumah University of Science & Technology, Ghana
REVIEW RETURNED	31-Mar-2018

GENERAL COMMENTS	<p>Well written manuscript detailing the use of Propensity-based Methods for cross-sectional data applied to Community-based rehabilitation.</p> <p>Minor edits 1. Change alinement to alignment on line 7 of page 4</p>
-------------------------	--

REVIEWER	Kelsey Chalmers Menzies Centre for Health Policy, School of Public Health, University of Sydney, Australia
REVIEW RETURNED	05-Jun-2018

GENERAL COMMENTS	<p>This is an interesting paper, and is potentially a useful resource for researchers in similar fields aiming to investigate complex interventions with limited data.</p> <p>The stated aim of the study is to examine whether PSM is an effective method to investigate CBR using cross-sectional data. Further discussions of the limitations and justification of the matching variables is probably required to convince readers that PSM is effective and useful for testing the CBR effect on the community inclusion score.</p> <p>Abstract results should include confidence intervals and the level of statistical significance (particularly for CBR comparison between groups). Also, given the aim of the study is to assess whether PSM is an effective method here, it seems necessary to clarify what covariates were included and the method ended up controlling for.</p> <p>Further information should be included in the methods in order to replicate and justify the matched sample, namely: which model (logit/probit regression?) were used for the propensity scores, and</p>
-------------------------	--

	<p>was there any iteration in the variable selection for the propensity scores?</p> <p>The authors should also justify/discuss the included variables used for PSM. In particular, if the treatment (CBR) effects the variable then including the variable in the PSM can obscure part of the treatment effect that you are trying to assess. This seems like it might be a possibility for some of the matching variables (access to rehabilitation services, self-help group, financial awareness), especially if those receiving CBR are more likely to score higher in these variables because of the CBR itself. See the caution from: Garrido, Melissa M., et al. "Methods for constructing and assessing propensity scores." <i>Health services research</i> 49.5 (2014): 1701-1720.</p> <p>The authors should also reconsider using an independent t-test to compare the outcome between the matched samples. Matched pairs of individuals are more similar than randomly selected pairs of individuals between the CBR and non-CBR groups, and so the two groups in the matched samples are not independent. See the second section of the following for a guide on this: Austin, Peter C. "A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003." <i>Statistics in medicine</i> 27.12 (2008): 2037-2049.</p> <p>In the discussion section, paragraph starting line 18: would be interesting to know how long the CBR program has been in place (and whether this compares to the timing of the effect found in the longitudinal studies referenced).</p> <p>Limitations should also discuss the impact of the small sample size and the information (reduced sample size) lost because of PSM and how this affects the comparison of the treatment effect on the outcome.</p> <p>Also a typo on page 4, line 7: alignment not alinement.</p> <p>Citations for R packages (including their authors) can usually be found using the command 'citation(package="<code><package name here></code>")' and should probably be included.</p>
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Anestis Divanoglou

Institution and Country: University of Iceland, Iceland

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

Please leave your comments for the authors below

This is a well-written manuscript and the introduction provides a good overview of the complexity in assessing the effectiveness of Community-based Rehabilitation (CBR) programs. The study used

data from two Vietnamese provinces: one where CBR had been implemented and the other where CBR had not been implemented. The authors used data from these provinces to illustrate how propensity score matching (PSM) may be used to reduce bias between groups in cross-sectional studies evaluating the effectiveness of CBR.

The authors thank the reviewer for the thorough and valuable comments!

The following issues are not clarified:

1. The manuscript does not provide sufficient details about the type and duration of the CBR intervention.

The authors thank the reviewer for pointing out the need for this clarification. The following information has been added to the text in the *Methods – Data Collection* section:

The CBR program began in 2009 in cooperation with the Huế Rehabilitation Hospital. The program focused mainly on activities to increase capacity building for CBR workers, not only in terms of rehabilitation skills, but also working to improve their counselling and networking skills. The other focus of the program was to strengthen referral pathways for people with disabilities so that they could be connected with other existing services in the province, such as schools with teachers who were trained to support students with disabilities and vocational training centers.

Information regarding the duration of the program in comparison to a previous longitudinal CBR study has also been added to the discussion as follows:

PSM was used to reduce the bias between the CBR and non-CBR groups, with results showing that CBR participants had better health and livelihood outcomes, and that these differences generally increased over time at both four years and seven years. In our study, data was collected seven years after the program began, which would make the timing comparable and it is therefore plausible that the effect of CBR in our study could already be quantifiable.

2. According to Table 2, similar number of cases in both provinces had access to rehabilitation services. Does this mean that the main difference was that one was as CBR and the other was not?

Thank you for this question. The “access to rehabilitation services” variable was derived from the WHO CBR survey question that measures access to rehabilitation services when needed, using the question “In the last 12 months, has there been a time when you needed rehabilitation services, such as physical, occupational, or speech therapy, but did not get those services?”, with the dichotomization of the response options “no need for rehab” and “received services” both counting as a positive response, and the response option “did not receive needed services” coded as a negative response. This means that it is correct to say that in Hòa Bình (the province with no CBR) receiving rehab meant non-CBR rehab, while in Huế (the province with CBR) the rehab services received could have been CBR or possibly another form of rehab. The authors have now included a supplementary table where the survey questions and categorizations of the matching variables can be seen.

3. The manuscript does not mention the number of potential cases identified in each province through the governmental records. Furthermore, the authors do not state the response rate, or whether they used some form of randomisation to identify the sample, or if they tried to include as many as possible. It is likely that the study had selection bias.

The authors thank the reviewer for pointing out this important missing information. The authors have added the following text to the *Results* section:

In Hué, 575 people with disabilities were identified by government records and 147 were included, while in Hòa Bình 375 people were identified by government records and 151 were included (sample size calculated using an alpha significance level of 0.05 and power of 90%).

According to the study findings, individuals from the CBR province had significantly worse community scores than those in the non-CBR province. The authors stated that two factors may have contributed to that: (1) cross-sectional data not being a suitable method to assess such outcomes, and (2) their outcome measures not being aligned with the purpose of the intervention (Page 8; lines 18-31). In my opinion, both these factors are likely and negatively impact the methodological quality of the manuscript.

Thank you for emphasizing these highly relevant points, though the authors believe that the methodological quality of the manuscript is sound, and hope to clarify this as follows:

To address the first point, the authors believe the inherent limitations of cross-sectional data need to be kept in mind when looking to assess causality, but the importance of presenting data in the field of CBR cannot be understated. For example, in Vietnam where this study was carried out, the survey was a welcome initiative as it was the only attempt to quantitatively collect and compare data. Since there have been limited efforts in the field of CBR to evaluate results, the authors believe that publishing analyses using cross-sectional data, despite its shortcomings, represents progress in the field and may prompt future research. Furthermore, cross-sectional data is essential to disclose associations and differences between groups in order to generate hypotheses and formulate further studies. In this study, the original research question aimed to test whether PSM can be used to analyze cross-sectional data and reduce bias. Though cross-sectional data presents limitations, not being able to infer causality does not imply unsuitability of the method. The following text has been added to the discussion:

An explanation for our results could be that cross-sectional data do not allow causal inferences: results could simply point out that the province with highest problems has been selected for receiving CBR interventions. While the cross-sectional data collected in this study represent the first quantitative data from the region and therefore an important foundation for future work, the results emphasizes the general need for further collection and publication of CBR data, especially longitudinal data.

Regarding the second point, the authors apologize for not phrasing this point correctly. The ultimate goal of CBR is the social inclusion of people with disabilities into community life, and this can be accomplished through an extensive range of activities. The authors did not mean to imply that the outcome measure of community inclusion is not aligned with the purpose of the intervention. Creating the outcome of community inclusion through a sum score of the relevant CBR indicators (social base and supplementary, and base empowerment) ensured its alignment with the ultimate goal of CBR, namely community inclusion. However, since the specific activities associated with the program were capacity building for CBR workers and increasing the strength of referral pathways, especially for the medical and education sectors, it could be the case that the effect of the program on community inclusion may not be as immediate as the other outcome(s). To attempt to clarify this, the authors have altered the text in the discussion to the following:

Though CBR aims to impact all aspects of the lives of people with disabilities to increase community inclusion, the program in Hué does not directly target community inclusion. The program focuses on increasing the capacity of CBR workers and on strengthening referral pathways with the medical and educational sectors. Through these activities the community inclusion of people with disabilities should improve over time, but since community inclusion was not the direct target of the program, the community inclusion effects might only appear after a longer period, which could be a reason for the counter-intuitive results. Therefore, when

assessing a program in its early stages, it may be more important to match the indicators used with the targets of programs.

Because the original data seem not straight forward, with the control group scoring better as compared to the intervention group, it becomes a bit complex and confusing when using these “exemplar” data to explain the effect of the PSM model. PSM seems a complex concept, and if data used are also complex, then the illustrated analysis and interpretation become unnecessary convoluted. This may have negative implications on reader’s ability to understand the key message of the manuscript.

Thank you for this highly relevant comment. Indeed, results show a significant difference in the opposite direction to what was hypothesized and we have discussed potential explanations for this. However, the authors’ original research question was to test PSM for cross-sectional CBR data. The authors have added a further comment to the discussion:

This study presents a starting point to encourage the generation quantitative CBR research and demonstrates one possible method for reducing bias when the only data available for CBR is cross-sectional, as is the case in our study.

Last, the importance and necessity for this manuscript is not established clearly. Why is it important to present such data and how is this useful?

Thank you for this comment and for the chance to state the importance of this manuscript. As the authors mention, the complexity of CBR and often limited resources available in the field lead to challenges in research attempting to quantify its effectiveness and to a heavy reliance on non-randomized cross-sectional data. This implies that statistical approaches, such as PSM, need to be applied to account for these limitations. The potential of using PSM for analyzing cross-sectional CBR data was demonstrated, as biases detected in the distribution of covariates between groups before matching were successfully eliminated. The authors hope that despite the results going against the hypothesis that CBR participation should be associated with better community inclusion, this method can be used in other studies to reduce bias when assessing cross-sectional data especially for international comparisons where differences between populations may be greater than the within country differences observed in this study.

Overall, I have strong doubts whether cross-sectional design is suitable to evaluate CBR programs. The use of PSM as a statistical method may be able to control for specific bias, but probably cannot control for the strong limitations associated with using a cross-sectional design to evaluate CBR programs. Furthermore, I cannot see how the PSM as a technique will support researchers to use cross-sectional data to show effects of CBR.

Thank you for this comment. The authors agree that cross-sectional data is far from the ideal option for the evaluation of CBR programs. However, given the lack of longitudinal data in the field of CBR, and quantitative data in general, using and publishing CBR data can help researchers better explore the available data, which is very frequently cross-sectional data. Using PSM reduced bias for comparison of groups in observational studies, and going beyond descriptive statistics is already progress for the field of CBR. To emphasize the importance of this point, the following text has been added to the *Discussion*:

In accompaniment to increased data collection, we recommend PSM as a method to reduce bias in cross-sectional CBR data analyses, especially for international comparisons where differences between populations may be greater than the within country differences observed in this study. Since using cross-sectional data presents limitations even after adjusting for

bias, we also emphasize the need for future longitudinal data collection in order to assess effectiveness in the field of CBR.

Additionally, the following was added to the *Conclusion*:

While randomized and longitudinal data are ideal for evaluations, cross-sectional data presents the advantage of being more feasible to collect and thereby providing an essential basis to generate hypotheses and perform further studies. Therefore, it is essential that appropriate statistical methods are applied to capitalize on available data.

Reviewer: 2

Reviewer Name: Fred Stephen Sarfo

Institution and Country: Kwame Nkrumah University of Science & Technology, Ghana

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

Please leave your comments for the authors below

Well written manuscript detailing the use of Propensity-based Methods for cross-sectional data applied to Community-based rehabilitation.

Minor edits

1. Change alignment to alignment on line 7 of page 4

Thank you, the correction has been made.

Reviewer: 3

Reviewer Name: Kelsey Chalmers

Institution and Country: Menzies Centre for Health Policy, School of Public Health, University of Sydney, Australia

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

Please leave your comments for the authors below

This is an interesting paper, and is potentially a useful resource for researchers in similar fields aiming to investigate complex interventions with limited data.

The stated aim of the study is to examine whether PSM is an effective method to investigate CBR using cross-sectional data. Further discussions of the limitations and justification of the matching variables is probably required to convince readers that PSM is effective and useful for testing the CBR effect on the community inclusion score.

Abstract results should include confidence intervals and the level of statistical significance (particularly for CBR comparison between groups). Also, given the aim of the study is to assess whether PSM is an effective method here, it seems necessary to clarify what covariates were included and the method ended up controlling for.

The authors thank the reviewer for their detailed and pertinent comments. The abstract has been edited accordingly to include confidence intervals, however due to word limitations the list of covariates could not be included in the abstract. The authors thank the reviewer for requesting more information on the justification of matching variables which is discussed below.

Further information should be included in the methods in order to replicate and justify the matched sample, namely: which model (logit/probit regression?) were used for the propensity scores, and was there any iteration in the variable selection for the propensity scores?

Thanks for requesting this further information. It is specified in the original manuscript that logit regression was used in *Methods – Analysis – Matching on the Propensity Score* and in *Results* section. No iteration was used in the variable selection, but since MatchIt uses random sorting when matching it is possible that different results could be obtained in replication studies.

The specification about the iteration has now been added to the text and reads as follows:

Participants were matched using one-to-one nearest neighbour technique, which matched each treated unit to one control that was closest using calipers of width equal to 0.25 of the standard deviation (SD) of the logit of the estimated propensity score without iteration.

The authors should also justify/discuss the included variables used for PSM. In particular, if the treatment (CBR) effects the variable then including the variable in the PSM can obscure part of the treatment effect that you are trying to assess. This seems like it might be a possibility for some of the matching variables (access to rehabilitation services, self-help group, financial awareness), especially if those receiving CBR are more likely to score higher in these variables because of the CBR itself. See the caution from: Garrido, Melissa M., et al. "Methods for constructing and assessing propensity scores." *Health services research* 49.5 (2014): 1701-1720.

Thank you for this very important point. The authors were aware of this and took careful consideration when deciding what variables to include as the matching variables. The available matching variables were all collected using the WHO CBR survey, which presents advantages and disadvantages. The main disadvantage is that certain variables, such as ethnicity and disability type/severity were not collected (as mentioned in the discussion). The main advantage is that all the variables were derived from the WHO CBR Guidelines, the internationally accepted conceptual framework for CBR. The WHO CBR Indicators cover the aspects of health, education, livelihood, social life and empowerment. Since the CBR program in Hué focused on increasing referral pathways within the medical and education sectors the questions derived from the education component and many from the medical component were not included as matching variables.

Community inclusion is a highly debated and often controversial subject in the literature, and for this reason the variables to be included as matching variables can also be debated; however, it was through theory, personal experience and literature from the field (e.g. the WHO CBR Guidelines, OECD Promoting financial inclusion through financial education) that the variables were selected that affect community inclusion. Though the variables are most likely strongly related to the exposure, they are all strongly related to the community inclusion for people living with disability, as evident from their inclusion as indicators in the WHO consensus-based indicator development process. The authors apologize for not stating this point in the manuscript, and have added the following:

Matching variables were those available from the WHO CBR Indicators, and were selected based on their theoretical association with community inclusion, primarily using CBR Guidelines [2]... Since the CBR program in Hué focused on increasing referral pathways within the medical and education sectors, the questions derived from the education component and many from the medical component were not included as matching variables, since inclusion of covariates associated with CBR participation but not with community inclusion decrease model precision[24].

Despite the specific aims of the CBR program focusing on capacity building and increasing referral pathways, the authors agree that “access to rehabilitation” could be more strongly related to the exposure than the outcome, so to test if its inclusion is in fact detrimental to model we ran the full analysis excluding this variable for comparison. It resulted in 75 matched pairs (vs the original 74), the standardized differences ranged from 0 to 0.209 (with financial awareness having the highest value, though still falling below the threshold value of 0.25), and the paired t-test finding CBR participants to have worse community inclusion scores (mean=18.11, SD=5.981, 95%CI 16.72-19.47) than non-CBR participants (mean=21.17, SD=6.381, 95%CI 19.67-22.60); $t(74)=3.3098$, $p=0.0014$.

This has been briefly added to the results as follows:

To ensure that the covariate of “access to rehabilitation” did not bias the model by being strongly associated with receiving CBR rather than with the outcome of community inclusion, the model was run excluding this variable. The new model resulted in 75 matched pairs with all standardized differences falling below the threshold. The results of the t-test did not differ from the model including access to rehabilitation; CBR participants had worse community inclusion scores (mean=18.11, SD=5.981, 95%CI 16.72-19.47) than non-CBR participants (mean=21.17, SD=6.381, 95%CI 19.67-22.60); $t(74)=3.3098$, $p=0.0014$.

The authors should also reconsider using an independent t-test to compare the outcome between the matched samples. Matched pairs of individuals are more similar than randomly selected pairs of individuals between the CBR and non-CBR groups, and so the two groups in the matched samples are not independent. See the second section of the following for a guide on this: Austin, Peter C. "A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003." *Statistics in medicine* 27.12 (2008): 2037-2049.

The authors thank the reviewer for this comment and for providing the relevant reference. This reference is indeed more applicable than the initial citation supporting the use of the independent t-test. The authors have re-done the analysis of the matched pairs using the paired t-test and the results have been reported. The results remain similar to those from the independent t-test.

In the discussion section, paragraph starting line 18: would be interesting to know how long the CBR program has been in place (and whether this compares to the timing of the effect found in the longitudinal studies referenced).

Thank you for this suggestion. The authors agree that this would be an interesting comparison. The following has been added to the discussion:

PSM was used to reduce the bias between the CBR and non-CBR groups, with results showing that CBR participants had better health and livelihood outcomes, and that these differences generally increased over time at both four years and seven years. In our study, data was collected seven years after the program began, so in relation to timing it would be plausible that the effect of CBR could already be quantifiable.

Limitations should also discuss the impact of the small sample size and the information (reduced sample size) lost because of PSM and how this affects the comparison of the treatment effect on the outcome.

Thank you for this comment. The authors had already included a short comment on this in the discussion, but agree that it is important to expand on this. The text in the discussion now reads:

Another limitation of PSM is that it leads to reduced sample size which could limit generalizability, though this is partly addressed through the provided sensitivity analysis. The reduced sample size also increases the risk of type II error (35) but the sample size of this study does meet the commonly recommended minimum sample size of $10(p + 1)$, where p is

the number of matching variables(36).

Also a typo on page 4, line 7: alignment not alinement.

Thank you. This has been corrected.

Citations for R packages (including their authors) can usually be found using the command 'citation(package="<package name here>")' and should probably be included.

Thank you. The recommended full citations have now been included.

VERSION 2 – REVIEW

REVIEWER	Anestis Divanoglou University of Iceland
REVIEW RETURNED	28-Aug-2018

GENERAL COMMENTS	<p>The revised manuscript is improved. Please see below my last two questions:</p> <ul style="list-style-type: none"> - Did the authors seek ethical approval from the National Human Research Ethics Committee in Vietnam? If not, why? - Were the objectives of this manuscript, i.e. “to examine whether PSM can be an effective method to facilitate evaluations of results in CBR when data are cross-sectional” stated in the ethics applications in Germany and in Vietnam?
-------------------------	--

REVIEWER	Kelsey Chalmers Menzies Centre for Health Policy, School of Public Health, University of Sydney, Australia
REVIEW RETURNED	24-Jul-2018

GENERAL COMMENTS	<p>Thank you to the authors for their response and resulting edits; they have greatly improved the original manuscript and the analysis is much clearer. The abstract is now much more informative. I have some additional comments on the discussion/conclusions, which has some inconsistencies with the conclusion of the manuscript.</p> <p>The ‘Matching variables’ section (line 7 – introduction). Matching variables were/should be selected based on their theoretical association with both the outcome (community inclusion) and treatment group (CBR or non-CBR) assignment. The authors have only stated they were associated with community inclusion, which may confuse readers not familiar with PSM trying to follow the methods used.</p> <p>Page 8; line 12 – 14. Is this a result of the CBR, or the treatment assignment group? (Basically – would we expect that this group would have a higher health status, etc, before CBR was implemented 7 years ago).</p> <p>Page 9; Line 26 – 28. This conclusion is unconvincing; in what scenario would the authors have shown that they cannot successfully apply PSM to the data? The authors have stated that their aim is to test whether PSM can be used in this scenario, but it</p>
-------------------------	--

	<p>is not clear how this is defined/achieved other than applying the method.</p> <p>This is especially confusing when the authors explain in line 12 (page 10) that a reason for their findings could be that the province with the highest problems has been selected for receiving CBR. This type of difference is the precise reason someone would apply PSM – to remove the treatment allocation bias – so we are able to test the effect of CBR after this bias has been accounted for. It's not clear, therefore, what using PSM has achieved and why readers should agree with the authors' conclusions.</p>
--	---

VERSION 2 – AUTHOR RESPONSE

The authors thank the reviewers for taking the time to provide further comments and for the chance to improve the manuscript through this revision. Please find our response to the second round of comments below.

Reviewer 3:

- 1) The 'Matching variables' section (line 7 – introduction). Matching variables were/should be selected based on their theoretical association with both the outcome (community inclusion) and treatment group (CBR or non-CBR) assignment. The authors have only stated they were associated with community inclusion, which may confuse readers not familiar with PSM trying to follow the methods used.

Thank you for pointing this out. This is of course true, but the authors originally did not explicitly mention the association with the treatment groups, as they thought this might be redundant since the CBR indicators compare the effect of CBR. We agree with the reviewer's suggestion and have added the recommended text as follows:

Matching variables were those available from the WHO CBR Indicators, and were selected based on their theoretical association with community inclusion and CBR group assignment, primarily using CBR Guidelines.

- 2) Page 8; line 12 – 14. Is this a result of the CBR, or the treatment assignment group? (Basically – would we expect that this group would have a higher health status, etc, before CBR was implemented 7 years ago).

Thank you for the comment. When we say "before matching" we meant at the point of data collection, and due to the cross-sectional nature of the data, we cannot be sure if this is a result of the CBR implementation or of the treatment assignment group (i.e. province of residence). We might assume that the differences came as a result of CBR implementation, however, since the province is known to generally have better resources it might have been possible that this was at least partially the case before CBR. This is one limitation that comes with the use of cross-sectional data, which we addressed in the discussion (this is what the authors were trying to emphasize with the sentence mentioned in point 4 below). To clarify what was meant with before and after matching, we have changed the wording as follows:

In the unmatched sample ~~Before matching~~, CBR participants had higher health status,...

This wording change was implemented throughout the manuscript to replace "before matching" and "after matching" for consistency and clarity.

- 3) Page 9; Line 26 – 28. This conclusion is unconvincing; in what scenario would the authors have shown that they cannot successfully apply PSM to the data? The authors have stated

that their aim is to test whether PSM can be used in this scenario, but it is not clear how this is defined/achieved other than applying the method.

Thank you for requesting this clarification. The authors meant here that the application of the method was statistically successful to reduce bias (e.g. the standardized difference of the means falling below the threshold value for all variables). The authors agree that a small change in the text might reduce potential for confusion:

We conclude that PSM can be ~~successfully~~ applied to cross-sectional CBR data, though in this case the bias reduction provided by PSM did not affect the tested outcome.

- 4) This is especially confusing when the authors explain in line 12 (page 10) that a reason for their findings could be that the province with the highest problems has been selected for receiving CBR. This type of difference is the precise reason someone would apply PSM – to remove the treatment allocation bias – so we are able to test the effect of CBR after this bias has been accounted for. It's not clear, therefore, what using PSM has achieved and why readers should agree with the authors' conclusions.

Thank you for allowing us to further clarify our meaning. We completely agree that the wording choice here could be improved. As mentioned above in point 2, the cross-sectional nature of the data means that we cannot be sure if differences are a result of the CBR implementation or of the treatment assignment group (i.e. province of residence). We might assume that the differences came as a result of CBR implementation, however, since the province is known to generally have better resources it might have been possible that this was at least partially the case before CBR. What we can do is reduce the bias between groups and compare the groups at a single time point in order to make conclusions about the success of the programs and to encourage further data collection activities so causal conclusions can be made. The authors have edited the sentence as follows:

An explanation for our results could be that cross-sectional data allow for comparisons between groups at a single time-point, and even after PSM is applied to reduce bias the causal relationship between CBR implementation and social inclusion cannot be determined. ~~results could simply point out that the province with highest problems has been selected for receiving CBR interventions.~~

Reviewer: 1

- 1) Did the authors seek ethical approval from the National Human Research Ethics Committee in Vietnam? If not, why?

The authors received approval for the study from the local Ministries of Health and from the LMU Munich Ethics Commission. In Vietnam, the regulatory authority responsible for the approval of research studies is the Ministry of Health, and since the study was to be carried out locally and was approved by the respective Ministries of Health, the authors did not seek further approval for the study at the national level.

- 2) Were the objectives of this manuscript, i.e. "to examine whether PSM can be an effective method to facilitate evaluations of results in CBR when data are cross-sectional" stated in the ethics applications in Germany and in Vietnam?

The primary goal of the data collection was an evaluation of the situation of persons with disabilities in each province using the CBR indicators. This evaluation was carried out alongside a qualitative evaluation and was intended to compliment these activities. The evaluation was carried out by CBM to inform local authorities (including the Ministry and Health, Ministry of Education, and Ministry of Labour and Social Affairs) and the ethics approval therefore has other content, with the main product being a report for the local authorities. This methodological

study is a secondary data analyses using this data. To clarify this, the authors have changed the text as follows:

~~Data collection involved~~ This study presents a secondary analysis of data collected during a multi-site cross-sectional survey in 2016 in two Vietnamese provinces.

VERSION 3 – REVIEW

REVIEWER	Kelsey Chalmers Menzies Centre for Health Policy, School of Public Health, The University of Sydney, Australia
REVIEW RETURNED	18-Nov-2018
GENERAL COMMENTS	Thank you for your revisions and the improved manuscript. The change in wording that the authors have used for the unmatched and matched samples have really made it easier to interpret. There is a typo on page 10, line 30: "WHO CRB Indicators" should be "WHO CBR Indicators".