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)	The development and internal validation of a multimorbidity index that predicts healthcare utilization using the Canadian Longitudinal Study on Aging
AUTHORS	Wang, Zhuoyu; Boulanger, Laurence; Berger, David; Gaudreau, Pierrette; Marrie, Ruth Ann; Potter, Brian; Wister, Andrew; Wolfson, Christina; Lefebvre, Genevieve; Sylvestre, Marie-Pierre; Keezer, M

VERSION 1 - REVIEW

REVIEWER	James Stanley University of Otago, Wellington New Zealand
REVIEW RETURNED	23-Sep-2019

GENERAL COMMENTS	Thank you for the opportunity to review this paper. It is clearly written overall: the aims are clear, the methods well described, and results presentation is clear, and the discussion is well framed. The data source (Canadian Longitudinal Study on Aging, CLSA) is robust and well described. I have a few relatively large concerns, and a few additional minor points. My biggest concern is with the model specification for the validation analysis: currently I think these decisions impact on the study results (particularly arising from how interactions are specified). While I have some concerns about the implications of these methodological decisions, I think addressing them is reasonably straightforward and will make for a more robust paper with increased validity. _____ The first concern is with the linearity specification for the morbidity indices in each model. This is not an issue in the development analysis (since each condition is a binary indicator), but does affect the validation analysis, and will be contributing somewhat to the performance metrics reported. Using a strictly linear predictor value is not best practice when adjusting for a continuous confounder, and it would be better to consider a more flexible modelling step that allows for a non-linear function (for instance, restricted cubic splines; or by grouping index scores into several categories for adjustment). e.g. at present, the difference in outcome risk for someone with two conditions compared to zero is modelled the same as the difference in outcome risk for someone with 12 conditions compared to ten.

	The situation is more complex for a weighted index, but the core issue is the same. The same is true of age as well: a non-linear functional form would be a more appropriate modelling strategy (here the linearity assumption effectively translates into a fixed odds ratio for a one year difference in age, which is over-simplified). _____ My second, and bigger, concern follows on from this, and relates to using interaction terms between two linear predictors (for the given morbidity index and for age in years). The more serious concern here is that the two-way interaction specified between each multimorbidity index and age is treating each element as a linear predictor, and treating the interaction as linear as well. Taking into account the issues raised above (that for both index and age, considered independently, a non-linear association with risk of outcome might be preferable) this specification of the interaction itself is a further unrealistic constraint imposed on the association between predictors and outcomes: that is, the impact of a given one-point difference on a multimorbidity index is assumed to change linearly with age (on the model fit using log-odds of outcome, so changing by a fixed ratio on the odds ratio scale, as explained in Table 3). This translates to a uniform relative change by age in the impact of multimorbidity: this seems like an unrealistic constraint from a clinical point of view (one might expect the association of multimorbidity and outcome risk to change in a non-monotonic fashion across age) but is also a major constraint on the fit of the model, and in my opinion may well be leading to poorer performance in the validation step than a more appropriate functional form (and/or leaving out the interaction terms). A related concern is the extent to which this interaction element is necessary to address the concerns of the paper, which is to assess which multimorbidity index performs best in the CLSA analysis setting. I would argue that this question could be addressed without needing to allow for a differential impact of multimorbidity across age (even if the inclusion of interactions likely allows for better overall predictions). I would suggest this could be addressed by a relatively minor change to the validation step: this would require re-running the validation models, which I appreciate would be time consuming. The first is removing the interaction components from the modelling altogether, on the basis that they do not necessarily help with answering the question at hand, and that the current specification is not appropriate. This would be my suggestion. The second would be altering the specification for the modelling of the interaction. This is intrinsically tricky: one would want to allow for non-linearities in both dimensions (morbidity index and age) and also in the interaction itself. The only straightforward way I can think of is using splines in the morbidity index, and then using age group as a main effect and in the interaction terms, which would effectively allow for stratified estimates of multimorbidity risk profile in each age group. However this might end up stretching the data quite thinly, and I don't think is necessarily warranted in a general

	analytical setting when adjusting for the impact of multimorbidity on outcomes. _____ My other major concern (though I think it could be dealt with relatively easily) is around the framing of the paper and how useful it could be for other readers. The goal of the study is presented as being "to develop a set of new indices to measure the multimorbidity burden of individuals in the CLSA cohort and to compare their internal validity". Broadly speaking, the aim is to solve the issue that common multimorbidity measures cannot be calculated in the CLSA as not all the relevant conditions were measured at baseline. This feels like a limited goal for a restricted audience (given it is framed as being useful for those using the CLSA). What can the study results tell us about multimorbidity indices more broadly, and how these can be deployed in settings where there is insufficient data to calculate an existing multimorbidity index? I feel that reframing the paper slightly along these lines (which would be largely in the introduction and discussion) would help make the paper more relevant to more readers. For example, this could be reframed as how best to adjust for multimorbidity when one has a relatively ad hoc set of chronic conditions: this specific study is fortunate to have a robust enough dataset to examine the performance of different approaches, whereas most smaller-scale studies would not be able to even create a weighted index (since doing so appropriately would require development in a separate data set from that analysed from an applied point of view). _____ Finally, for the intended purpose of the indices for adjusting for multimorbidity in analytical models, I would suggest that discrimination (i.e. the c-statistics) is more important than calibration (the calibration plots and related statistics). This might mean drawing conclusions prioritised on this basis (though presently the c-statistics are rather flat across all potential indices). This would be different if absolute risk prediction was more important, for example if one was determining risk for a particular patient. _____ MINOR COMMENTS: The introduction states that one prior study (ref #14) reported that unweighted indices are less valid than weighted indices: there have been more studies than this that have compared indices vs. counts. I would say that the idea that weighted indices can perform better than counts of conditions is more settled in favour of weighted indices than the current phrasing suggests. There is some additional discussion of this in: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3315139/. While the methods are very nicely written, it would be handy to include an initial unifying statement to say that chronic conditions were measured at baseline, and outcomes measured at follow-up (e.g. in the Variables section on p.7). There were a few places where it wasn't entirely clear as to timing of outcome measurement (e.g. on page 9: for correlations of indices with other health

	constructs, the timing of measurement with these other constructs isn't clear – I think it's fine if they are measured at baseline, but it should be mentioned). Relatedly, the section describing the different measures used for convergent validity (bottom of page 9, top of page 10) should probably be moved to the Variables section (leaving the analysis components where they are in the Analysis section). For convergent validity: given the issues with potential non-linearity of the indices (noted in the major concerns section) it might be preferable to calculate and report non-parametric correlations between each index and the various other measures. These would allow for measurement of monotonic associations (i.e. tendency to increase or decrease) without worrying about whether the interval spacing of each index is reasonable. It would be useful if the results text could present the model performance in the reference model (age/sex adjustment only) which is currently only presented in a table in the supplementary material. Bottom of page 11 (line 52): seems to be missing a word or two e.g. "reducing a complex [construct] into a single measure" References: The reference to Altman and Royston (ref 22) is perhaps superseded by more recent work (e.g. Steyerberg et al. 2009, https://www.ncbi.nlm.nih.gov/pubmed/20010215; or Steyerberg's "Clinical Prediction Models" textbook.) This might be useful to cite as a resource for future readers who are interested as well. Table 1/2 and Supp Table 5: It might be nice to have the model details summarised very briefly in the header under the Index number (i.e. "Condition count", "Weighted risk index", etc) . Currently this is summarised as a footnote for Table 1, which is a little awkward to read; also needs to be summarised for the other two tables. Table 3 is quite complex: it illustrates some of the earlier major points (linearity of age and morbidity index, and linearity of the interaction term) but if the interactions were to stay in then I think the implications in the Table would need to be reproduced in the body of the text. Supplementary Tables 2 and 3 reproduce some of the questionnaires used to measure the related health outcomes: I don't really think these are generally necessary to include (given they are reported elsewhere, and are a minor component of the reported research) and there may also be some copyright issues about reproducing these without prior permission (e.g. Supp Table 3 suggests only available through the Duke University creators).

REVIEWER	Manfred Gogol Institute of Gerontology Heidelberg University Germany
REVIEW RETURNED	07-Oct-2019

GENERAL COMMENTS	Dear colleagues, Thank you for submitting your paper “The development and internal validation of a multimorbidity indeces in the CLSA” to BMJ Open. I´ve following remarks: 1. Abstract: Please include mean age and gender distribution into the abstract. 2. What is your definition of multimorbidity. Or, in which way you differ from multimorbidity definition usual use 2 chronic definition to qualify for? 3. Please explain shortly your interaction term throughout the text, not only in table 3. 4. Discuss the probably bias by participant selection for CLSA. 5. Discuss the probably bias that you not access for acute conditions (not exacerbations of chronic conditions) for your proxy outcome. 6. Discuss the probably bias due to not adjust for disease severity, e.g. that chronic heart failure NYHA I counts the same as NYHA IV. 7. Giving the high number of probably over- and underdiagnosis for chronic conditions, e.g. COPD, what does this mean for your study? 8. Correct the reference section for using upper and lower cases in paper titles and use of appropriate abbreviations of journal titles. Sincerely

REVIEWER	Beatrice U Mueller, Prof. Dr. med. University Hospital Zürich Department for Internal Medicine Rämistrasse 100 8091 Zürich
REVIEW RETURNED	27-Oct-2019

GENERAL COMMENTS	1.The study was using data from the Canadian Longitudinal Study on Aging (CLSA). The cohort included 51338 participants between 2010-2015 and 40264 were analysed. There are several aspects that need to be addressed by the authors. Especially the introduction part, the results and the references: 2. Page 5, lane 7: Salive ME 's publication is not correctly cited. The authors should correct this. 3. Page 5, lane 10: "study conducted in Ontario, Canada the prevalence of multimorbidity increased by an average of 40% across all age groups should be cited in more details with the prevalence of multimorbidity in 2003 and with the prevalence of multimorbidity in 2009.

	4. References 3 is from 2003 and the data used in reference 3 were collected some years before 2003. They should have provided more recent references following 2009. 5. Page 10, lane 17-60: The results should be described in more detail. The result part should comment on the figures and tables in more detail in the text. 6. Page 10, lane: Online supplementary table 4 does not add important additional information. 7. Page 10, 11: The order of tables should be continuous as they are described in the manuscript. 8. Page 10, lane 45: "Index 1, but with an interaction term.... The authors should explain in this section index 1 in more detail. The authors should describe figure 1 in more detail in the manuscript.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: James Stanley

Institution and Country:

University of Otago, Wellington

New Zealand

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below

Thank you for the opportunity to review this paper. It is clearly written overall: the aims are clear, the methods well described, and results presentation is clear, and the discussion is well framed. The data source (Canadian Longitudinal Study on Aging, CLSA) is robust and well described. I have a few relatively large concerns, and a few additional minor points.

My biggest concern is with the model specification for the validation analysis: currently I think these decisions impact on the study results (particularly arising from how interactions are specified). While I have some concerns about the implications of these methodological decisions, I think addressing them is reasonably straightforward and will make for a more robust paper with increased validity.

———

The first concern is with the linearity specification for the morbidity indices in each model. This is not an issue in the development analysis (since each condition is a binary indicator), but does affect the validation analysis, and will be contributing somewhat to the performance metrics reported. Using a

strictly linear predictor value is not best practice when adjusting for a continuous confounder, and it would be better to consider a more flexible modelling step that allows for a non-linear function (for instance, restricted cubic splines; or by grouping index scores into several categories for adjustment).

e.g. at present, the difference in outcome risk for someone with two conditions compared to zero is modelled the same as the difference in outcome risk for someone with 12 conditions compared to ten. The situation is more complex for a weighted index, but the core issue is the same.

The same is true of age as well: a non-linear functional form would be a more appropriate modelling strategy (here the linearity assumption effectively translates into a fixed odds ratio for a one year difference in age, which is over-simplified).

AUTHORS: This is an excellent point. Confirming that there is no evidence of non-linearity is extremely important and we apologize that this was not emphasized in the submitted manuscript. Following the reviewer's suggestions we have introduced b-spline functions, 3 knots, to test for non-linearity in all of the validation regression models. We did not find evidence of non-linearity, however, and have left the final regression models as they previously were.

For example, for Index 1, we tested the linearity of age and the sum of chronic conditions and found the following. Of note, to be additionally thorough, we also examined whether the introduction of quadratic terms would provide evidence of non-linearity.

- AIC for the base model (predictors are sex, age, and the absolute sum of chronic conditions (sumCC)) = 15627
- Including an age quadratic term in the model, without interaction: p=0.3254 for the quadratic term, AIC=15597 for the model.
- Including a sumCC quadratic term in the model, without interaction: p=0.413 for the quadratic term, AIC=15628 for the model.
- Including a b-spine function for age: AIC=15630 for the model
- Including a b-spine function for sumCC: AIC=15630 for the model

We also assessed non-linearity in the interaction terms. For Index 1, our results were:

- AIC for the base model (predictors are sex, age and the absolute sum of chronic conditions (sumCC)) and an interaction between age and sumCC = 15594
- Including an age quadratic term in the model: p=0.5753 for the interaction term that includes the quadratic term, AIC=15597 for the model.
- Including a sumCC quadratic term in the model: p=0.4210 for the interaction term that includes the quadratic term, AIC=15597 for the model.

- Including a b-spine function for age: AIC=15600 for the model
- Including a b-spine function for sumCC: AIC=15630 for the model

We have added the following text to the 3rd to last paragraph of the Methods section: “We tested the linearity of the relationship between continuous predictors and healthcare utilization by introducing b-spline functions and observing whether this had any impact on model fit of the validation regression models (measured using AIC).” We have also added the following text to the 3rd paragraph of the Results section: “We did not find evidence of non-linearity between continuous predictors and healthcare utilization using b-spline functions and comparing AIC.”

———

My second, and bigger, concern follows on from this, and relates to using interaction terms between two linear predictors (for the given morbidity index and for age in years).

The more serious concern here is that the two-way interaction specified between each multimorbidity index and age is treating each element as a linear predictor, and treating the interaction as linear as well. Taking into account the issues raised above (that for both index and age, considered independently, a non-linear association with risk of outcome might be preferable) this specification of the interaction itself is a further unrealistic constraint imposed on the association between predictors and outcomes: that is, the impact of a given one-point difference on a multimorbidity index is assumed to change linearly with age (on the model fit using log-odds of outcome, so changing by a fixed ratio on the odds ratio scale, as explained in Table 3).

This translates to a uniform relative change by age in the impact of multimorbidity: this seems like an unrealistic constraint from a clinical point of view (one might expect the association of multimorbidity and outcome risk to change in a non-monotonic fashion across age) but is also a major constraint on the fit of the model, and in my opinion may well be leading to poorer performance in the validation step than a more appropriate functional form (and/or leaving out the interaction terms).

A related concern is the extent to which this interaction element is necessary to address the concerns of the paper, which is to assess which multimorbidity index performs best in the CLSA analysis setting. I would argue that this question could be addressed without needing to allow for a differential impact of multimorbidity across age (even if the inclusion of interactions likely allows for better overall predictions).

I would suggest this could be addressed by a relatively minor change to the validation step: this would require re-running the validation models, which I appreciate would be time consuming.

The first is removing the interaction components from the modelling altogether, on the basis that they do not necessarily help with answering the question at hand, and that the current specification is not appropriate. This would be my suggestion.

The second would be altering the specification for the modelling of the interaction. This is intrinsically tricky: one would want to allow for non-linearities in both dimensions (morbidity index and age) and also in the interaction itself. The only straightforward way I can think of is using splines in the morbidity index, and then using age group as a main effect and in the interaction terms, which would effectively allow for stratified estimates of multimorbidity risk profile in each age group. However this might end up stretching the data quite thinly, and I don't think is necessarily warranted in a general analytical setting when adjusting for the impact of multimorbidity on outcomes.

AUTHORS: These are excellent points. That said, as outlined in our response to the preceding comment, we did not find evidence of non-linearity; for age, the multimorbidity indices, or the associated interaction terms. As a result, we have chosen to leave our final models unchanged.

———

My other major concern (though I think it could be dealt with relatively easily) is around the framing of the paper and how useful it could be for other readers.

The goal of the study is presented as being "to develop a set of new indices to measure the multimorbidity burden of individuals in the CLSA cohort and to compare their internal validity". Broadly speaking, the aim is to solve the issue that common multimorbidity measures cannot be calculated in the CLSA as not all the relevant conditions were measured at baseline.

This feels like a limited goal for a restricted audience (given it is framed as being useful for those using the CLSA). What can the study results tell us about multimorbidity indices more broadly, and how these can be deployed in settings where there is insufficient data to calculate an existing multimorbidity index? I feel that reframing the paper slightly along these lines (which would be largely in the introduction and discussion) would help make the paper more relevant to more readers.

For example, this could be reframed as how best to adjust for multimorbidity when one has a relatively ad hoc set of chronic conditions: this specific study is fortunate to have a robust enough dataset to examine the performance of different approaches, whereas most smaller-scale studies would not be able to even create a weighted index (since doing so appropriately would require development in a separate data set from that analysed from an applied point of view).

AUTHORS: Thank you. We have made modifications to the title (as stated above), the conclusion of the Abstract ("The utility of an age interaction term in measuring multimorbidity burden may be applicable to the study of chronic disease in cohorts other than the CLSA."), the end of the Introduction ("Therefore, a study of the performance of different approaches to measuring multimorbidity, applicable to in the CLSA and potentially other cohorts, is required."), and the conclusion of the Discussion ("The utility of an age interaction term in measuring multimorbidity burden may be applicable to the study of chronic disease in cohorts other than the CLSA.") to emphasise the potential transportability of our findings to other cohorts.

———

Finally, for the intended purpose of the indices for adjusting for multimorbidity in analytical models, I would suggest that discrimination (i.e. the c-statistics) is more important than calibration (the calibration plots and related statistics). This might mean drawing conclusions prioritised on this basis (though presently the c-statistics are rather flat across all potential indices). This would be different if

absolute risk prediction was more important, for example if one was determining risk for a particular patient.

AUTHORS: We agree that calibration, with its emphasis on averages predictions across groups rather than predictions for specific individual, is less relevant to regression modeling, That said, calibration is particularly relevant when evaluating the effect of healthcare policies, comparing the performance of different healthcare institutions, as described in the 2nd paragraph of the Introduction. Our calculations in Table 3 are particularly relevant to such situations.

———

MINOR COMMENTS:

The introduction states that one prior study (ref #14) reported that unweighted indices are less valid than weighted indices: there have been more studies than this that have compared indices vs. counts. I would say that the idea that weighted indices can perform better than counts of conditions is more settled in favour of weighted indices than the current phrasing suggests. There is some additional discussion of this in:
https://eur03.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.ncbi.nlm.nih.gov%2Fpmc%2Farticles%2FPMC3315139%2F&data=02%7C01%7C%7C0049db6a0dcf4381bfac08d75c460d74%7C84df9e7fe9f640afb435aaaaaaaaaaaa%7C1%7C0%7C637079328252625322&sdata=fDMPkkbzzAVTV3A7s4NZWv2EE6QYYJ5bTwqiMKYq%2F0Y%3D&reserved=0.

AUTHORS: Our understanding is that the conclusion of Huntley et al 2012 is that the evidence in favour of complex weighted indices as opposed to simple counts is not clear, in spite of what one would otherwise expect. This and related issues are discussed in the 2nd and 3rd paragraphs of the Discussion. We have modified the referenced sentence in the 3rd paragraph of the Introduction to: “Prior studies, however, have reported only modest benefits of an unweighted count of chronic diseases is a less valid measure of multimorbidity as compared to weighted indices.”

While the methods are very nicely written, it would be handy to include an initial unifying statement to say that chronic conditions were measured at baseline, and outcomes measured at follow-up (e.g. in the Variables section on p.7). There were a few places where it wasn’t entirely clear as to timing of outcome measurement (e.g. on page 9: for correlations of indices with other health constructs, the timing of measurement with these other constructs isn’t clear – I think it’s fine if they are measured at baseline, but it should be mentioned).

AUTHORS: We have added the following to the 4rd paragraph of the Statistical Analyses section: “All variables were measured at baseline with the exception of hospitalisation which was queried at approximately 18 months later (minimum 12 months).” We also added the following text to the 2nd paragraph of this same section: “Additional variables measured at the baseline assessment were life satisfaction, functional disability, as well as self-rated general health and mental health.”

Relatedly, the section describing the different measures used for convergent validity (bottom of page 9, top of page 10) should probably be moved to the Variables section (leaving the analysis components where they are in the Analysis section).

AUTHORS: Thank you for the suggestion and we have carried it out.

For convergent validity: given the issues with potential non-linearity of the indices (noted in the major concerns section) it might be preferable to calculate and report non-parametric correlations between each index and the various other measures. These would allow for measurement of monotonic associations (i.e. tendency to increase or decrease) without worrying about whether the interval spacing of each index is reasonable.

AUTHORS: We have found no evidence of non-linearity.

It would be useful if the results text could present the model performance in the reference model (age/sex adjustment only) which is currently only presented in a table in the supplementary material.

AUTHORS:. We appreciate this suggestion but to ensure the maximum readability of the article, we have respectfully chosen to keep this as an online document.

Bottom of page 11 (line 52): seems to be missing a word or two e.g. “reducing a complex [construct] into a single measure”

AUTHORS: Thank you. We have made this correction.

References: The reference to Altman and Royston (ref 22) is perhaps superseded by more recent work (e.g. Steyerberg et al. 2009, https://eur03.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.ncbi.nlm.nih.gov%2Fpubmed%2F20010215&data=02%7C01%7C%7C0049db6a0dcf4381bfac08d75c460d74%7C84df9e7fe9f640afb435aaaaaaaaaaaa%7C1%7C0%7C637079328252625322&sdata=vURCWghJ7HnvmcBwhLvoj%2FIN6sKSmHwSvxF9nNy%2F%2F4c%3D&reserved=0; or Steyerberg’s “Clinical Prediction Models” textbook.) This might be useful to cite as a resource for future readers who are interested as well.

AUTHORS: Thank you for this suggestion. We have replaced the Altman references with Steyerberg as suggested.

Table 1/2 and Supp Table 5: It might be nice to have the model details summarised very briefly in the header under the Index number (i.e. “Condition count”, “Weighted risk index”, etc) . Currently this is summarised as a footnote for Table 1, which is a little awkward to read; also needs to be summarised for the other two tables.

AUTHORS: We have made these suggested changes to the relevant tables.

Table 3 is quite complex: it illustrates some of the earlier major points (linearity of age and morbidity index, and linearity of the interaction term) but if the interactions were to stay in then I think the implications in the Table would need to be reproduced in the body of the text.

AUTHORS: Thank you. We have added the following sentence to the last paragraph of the Results: “This allows for the calculation of an “age-adjusted” sum of chronic diseases, as opposed to a simple absolute count.”

Supplementary Tables 2 and 3 reproduce some of the questionnaires used to measure the related health outcomes: I don't really think these are generally necessary to include (given they are reported elsewhere, and are a minor component of the reported research) and there may also be some copyright issues about reproducing these without prior permission (e.g. Supp Table 3 suggests only available through the Duke University creators).

AUTHORS:. Thank you for this suggestion. We have removed these tables from the online supplement.

Reviewer: 2

Reviewer Name: Manfred Gogol

Institution and Country:

Institute of Gerontology

Heidelberg University

Germany

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

Please leave your comments for the authors below

Dear colleagues,

Thank you for submitting your paper "The development and internal validation of a multimorbidity indeces in the CLSA" to BMJ Open.

I´ve following remarks:

1. Abstract: Please include mean age and gender distribution into the abstract.

AUTHORS: We have added the following to the Abstract: "Data from 40,264 CLSA participants (52% men) aged 45 to 85 years (a mean of 63 years) were analysed."

2. What is your definition of multimorbidity. Or, in which way you differ from multimorbidity definition usual use 2 chronic definition to qualify for?

AUTHORS: We apologize that this was not clear. Our definition includes the occurrence of ≥ 2 chronic diseases. We have modified the first sentence of the Introduction to the following: "Multimorbidity is defined as the co-occurrence of multiple chronic medical conditions in the same individual and affects at least 50% of individuals in the general population over age 65 years."

3. Please explain shortly your interaction term throughout the text, not only in table 3.

AUTHORS: We do not fully understand your comment. Table 3, which is mentioned at the end of the Results section, describes how to calculate a sum of chronic diseases that adjusts for the interaction with age. The remainder of our results discuss interaction terms within our regression models and

what impact they have on model calibration. We have highlighted the term “age-adjusted sum of chronic diseases” at the end of the Results and in Table 3 to help distinguish between this work and the age interaction term in our regression models.

4. Discuss the probably bias by participant selection for CLSA.

AUTHORS: We have added the following statement to the 2nd to last paragraph of the Discussion: “It is possible that healthy people are over-sampled in the CLSA, although the sampling weights are meant to at least in part correct for this.”

5. Discuss the probably bias that you not access for acute conditions (not exacerbations of chronic conditions) for your proxy outcome.

AUTHORS: Our outcome was hospitalisation in the last 12 months. If an acute condition is sufficiently severe (e.g. pneumonia or hip fracture), we are confident that this would have required a hospitalisation in the same way as for an exacerbation of a person’s chronic cardiac disease.

6. Discuss the probably bias due to not adjust for disease severity, e.g. that chronic heart failure NYHA I counts the same as NYHA IV.

AUTHORS: This is an important point and a limitation to our work. We added the following to the 2nd to last paragraph of the Discussion: “Each chronic disease was treated as present or absent, with no measure of severity. As a result, we likely underestimated the impact of more severe cases given their relative rarity in this cohort of more healthy individuals.”

7. Giving the high number of probably over- and underdiagnosis for chronic conditions, e.g. COPD, what does this mean for your study?

AUTHORS: We agree that the CLSA, like all large databases, uses an imperfect method of case ascertainment. Administrative databases generally rely on ICD codes, research databases often rely on self-reported diagnoses, both of which may suffer from inaccuracies. We have added a comment on this to the 2nd to last paragraph of the Discussion: “Chronic diseases in the CLSA are self-reported by participants, therefore there may be inaccuracies in case ascertainment. CLSA participants cannot be contacted to confirm their self-reported diagnoses. Prior validation studies have shown that self-report questions can accurately identify certain conditions such as multiple sclerosis.”

8. Correct the reference section for using upper and lower cases in paper titles and use of appropriate abbreviations of journal titles.

AUTHORS: We have corrected all of our references so that they respect the BMJ Open requirements.

Reviewer: 3

Reviewer Name: Beatrice U Mueller, Prof. Dr. med.

Institution and Country:

University Hospital Zürich

Department for Internal Medicine

Rämistrasse 100

8091 Zürich

Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below

1.The study was using data from the Canadian Longitudinal Study on Aging (CLSA). The cohort included 51338 participants between 2010-2015 and 40264 were analysed.

There are several aspects that need to be addressed by the authors. Especially the introduction part, the results and the references:

2. Page 5, lane 7: Salive ME 's publication is not correctly cited. The authors should correct this.

AUTHORS: Thank you. We have corrected this reference.

3. Page 5, lane 10: "study conducted in Ontario, Canada the prevalence of multimorbidity increased by an average of 40% across all age groups should be cited in more details with the prevalence of multimorbidity in 2003 and with the prevalence of multimorbidity in 2009.

AUTHORS: We have added the prevalence point estimates from this study.

4. References 3 is from 2003 and the data used in reference 3 were collected some years before 2003. They should have provided more recent references following 2009.

AUTHORS: We have replaced this reference with one from 2018.

5. Page 10, lane 17-60: The results should be described in more detail. The result part should comment on the figures and tables in more detail in the text.

AUTHORS: We have done our best to summarize the most salient points from the tables and figures in the text. We have also kept in mind that the general BMJ guidelines insist that data presented in tables not be reproduced in the text. (https://authors.bmj.com/writing-and-formatting/formatting-your-paper/) As a result, we apologize but are not able to provide further details on the tables and figures in the body of the manuscript.

6. Page 10, lane: Online supplementary table 4 does not add important additional information.

AUTHORS: We agree that the information is not essential to our overall message. That said, when there are missing data in a dataset, it is common practice to compare those with and without missing data, at least briefly, to help the reader understand whether these missing data may have affected the results. Our online table (now redefined as online table 2) supports our assertion that missing data should not have greatly affected our results. Finally, the TRIPOD reporting guidelines for prediction models recommends that such tables be included.

7. Page 10, 11: The order of tables should be continuous as they are described in the manuscript.

AUTHORS: We have reviewed all tables, both in-print and online, and have confirmed that they are numbered in the order that they appear. Of note, the number for in-print tables and figures is independent from those for online tables and figures. As a result, Online supplementary table 3 appears in the text before Table 1.

8. Page 10, lane 45: "Index 1, but with an interaction term....

The authors should explain in this section index 1 in more detail. The authors should describe figure 1 in more detail in the manuscript.

AUTHORS: We have added the following text to this section of the Results: “For the remaining multimorbidity indices, calibration was poor, especially when above a predicted probability of 0.20, where the curves diverged from a slope of 1.0 (Figure 1).” We would point out that the interpretation of calibration curves is described in the 3rd to last paragraph of the Methods section.

VERSION 2 – REVIEW

REVIEWER	James Stanley University of Otago, Wellington New Zealand
REVIEW RETURNED	09-Dec-2019

GENERAL COMMENTS	Thank you for the opportunity to re-review this paper, and for the additional information provided in the response letter and the paper itself. I have a few points of clarification relating to my original notes (following the revisions) plus a couple more minor points: all of these can be addressed with some minor clarification in the text. ** *Interaction terms in models*: I appreciate the notes on the testing of interaction terms. My remaining point here (also partly indicated by the other reviewers) is that the purpose of the interaction terms in the paper isn’t presented clearly: is this to illustrate that the interplay of multimorbidity and age is important, or to provide a quantification for how multimorbidity changes over age (for use in future research)? This point itself could be clarified in the paper. If the intended aim is to note that allowance for interaction between age and multimorbidity is important, then the information on improved model performance with interaction is probably more

	useful (this is effectively the final sentence of the conclusion on p. 13). If the aim is to quantify an adjustment element that can be used in subsequent models adjusting for multimorbidity as it varies with age, then the text and Table 3 are still quite opaque: the use/form of the "age adjusted count" is implicit in Table 3, but requires readers to critically engage with the presented formulae, and there is no information on how this could be used in future research. A couple of sentences might suffice to clarify, e.g. along the lines of "In future research, this age-adjusted count can be calculated for a given individual and used in an adjustment model in place of the simple count". (If there is no intention that the estimated interaction term be used in future research then I would consider moving this material to the supplementary material). ** *Discrimination performance of the reference model in main text*: I would respectfully disagree with the response statement that the baseline model performance should be kept as supplementary material: if the c-statistics were comparable between the baseline model and all the condition-adjusted models, then that would be very important to note (as it would indicate no improvement in model predictions with adjustment for multimorbidity). Giving the baseline age/sex model c-statistic in the results text should thus not distract from the main argument, and in fact directly supports the main argument that including information on multimorbidity is important (even if exact formulation isn't so relevant here). ** *Discussion on utility of weighted/unweighted indices*: The discussion states: "Such challenges likely explain, at least in part, why an unweighted count of chronic diseases performs generally as well as indices based upon more complex regression models. [ref 23]". However reference 23, and the Huntley reference as discussed in my previous review and the author's response, both specifically note that analysis of mortality outcomes appears to work best when using (some specified) weighted indices. This could be noted in the text at this point in the discussion, especially since the following sentence highlights that pharmaceutical-based indices might be better at predicting hospitalisation outcomes. ** New points (minor!) ** *Focus of study on hospitalisation*: the second to last paragraph of the introduction talks about face validity of unweighted morbidity counts, noting "(e.g. migraine is presumably associated with a smaller risk of death than metastatic cancer)". While I agree with the statement, given that the rest of the paper deals with hospitalisation as an outcome, the statement and the example conditions could perhaps be revised to better suit the utility of the approach described in the paper for hospitalisation outcomes, where conditions might be considered more equal in terms of risk of hospitalisation. It would also be good to explicitly note the outcome (hospitalisation) in the final sentence of the introduction. ** *Distribution of multiple chronic conditions*: I think it is important to include contextual information on multimorbidity in the results, e.g. a frequency breakdown on the number of identified conditions in

	the sample (e.g. add to Supp Table 1, which already includes the single-condition prevalences across the study sample) with something like the proportion of people with 0, 1, 2, 3 ... conditions (with some sensible capped upper limit e.g. 3+ or 4+ conditions).

REVIEWER	Manfred Gogol Institute of Gerontology, Heidelberg, Germany and Geriatric Trauma Center, Trauma Department, Hannover Medical School, Hannover, Germany
REVIEW RETURNED	02-Feb-2020

GENERAL COMMENTS	It´s an interisting approach to work on the problem of multimorbidity and will foster the discussion.

VERSION 2 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name

James Stanley

Institution and Country

University of Otago, Wellington New Zealand

Please state any competing interests or state ‘None declared’:

None declared

Please leave your comments for the authors below

Thank you for the opportunity to re-review this paper, and for the additional information provided in the response letter and the paper itself.

I have a few points of clarification relating to my original notes (following the revisions) plus a couple more minor points: all of these can be addressed with some minor clarification in the text.

**

Interaction terms in models: I appreciate the notes on the testing of interaction terms. My remaining point here (also partly indicated by the other reviewers) is that the purpose of the interaction terms in the paper isn’t presented clearly: is this to illustrate that the interplay of multimorbidity and age is important, or to provide a quantification for how multimorbidity changes over age (for use in future research)? This point itself could be clarified in the paper.

If the intended aim is to note that allowance for interaction between age and multimorbidity is important, then the information on improved model performance with interaction is probably more useful (this is effectively the final sentence of the conclusion on p. 13).

If the aim is to quantify an adjustment element that can be used in subsequent models adjusting for multimorbidity as it varies with age, then the text and Table 3 are still quite opaque: the use/form of the "age adjusted count" is implicit in Table 3, but requires readers to critically engage with the presented formulae, and there is no information on how this could be used in future research. A couple of sentences might suffice to clarify, e.g. along the lines of "In future research, this age-adjusted count can be calculated for a given individual and used in an adjustment model in place of the simple count". (If there is no intention that the estimated interaction term be used in future research then I would consider moving this material to the supplementary material).

AUTHORS: Our aim is to highlight the importance of allowing for an interaction between age and multimorbidity in any model that studies multimorbidity. We adhere to the assumption that any precise quantification derived from our dataset is not likely transportable to other datasets. Table 3, on the other hand, serves a dual purpose. On the one hand it illustrates the concept and approach to deriving an age-adjusted sum of chronic conditions. At the same time, it presents specific adjustment factors that could reasonably be used by other CLSA researchers. We have added the following to the 1st paragraph of the Discussion: "We also present a method of calculating an absolute sum of chronic conditions that adjusts for the interaction with age, as opposed to a simple count, which could be used in future research. The precise age adjustment factors presented are specific to the CLSA database and may not be transportable to other databases although the mathematical approach is applicable to any database."

**

Discrimination performance of the reference model in main text: I would respectfully disagree with the response statement that the baseline model performance should be kept as supplementary material: if the c-statistics were comparable between the baseline model and all the condition-adjusted models, then that would be very important to note (as it would indicate no improvement in model predictions with adjustment for multimorbidity). Giving the baseline age/sex model c-statistic in the results text should thus not distract from the main argument, and in fact directly supports the main argument that including information on multimorbidity is important (even if exact formulation isn't so relevant here).

AUTHORS: We have added the following to the end of the 4th paragraph of the Results

"although consistently higher than a model including only sex and age, with the exception of Index 5 (Table 1, Online supplementary table 3)." We have also added a footnote to Table 1 that specifies the R-squared, correlation coefficient, and C-statistic for the reference model.

**

Discussion on utility of weighted/unweighted indices: The discussion states: "Such challenges likely explain, at least in part, why an unweighted count of chronic diseases performs generally as well as indices based upon more complex regression models. [ref 23]". However reference 23, and the Huntley reference as discussed in my previous review and the author's response, both specifically note that analysis of mortality outcomes appears to work best when using (some specified) weighted indices. This could be noted in the text at this point in the discussion, especially since the following sentence highlights that pharmaceutical-based indices might be better at predicting hospitalisation outcomes.

AUTHORS: Thank you for pointing this out. In fact, this was the wrong reference. It has been changed to Huntley. In Huntley's review, they cite two comparative studies that conclude there was little difference in performance between weighted indices and simple counts. We have clarified the end of the 1st paragraph of the Discussion with: "Such challenges likely explain, at least in part, why some comparative studies have reported that an unweighted count of chronic diseases performs almost as well as an index based upon a more complex regression model.16"

**

New points (minor!)

**

Focus of study on hospitalisation: the second to last paragraph of the introduction talks about face validity of unweighted morbidity counts, noting "(e.g. migraine is presumably associated with a smaller risk of death than metastatic cancer)". While I agree with the statement, given that the rest of the paper deals with hospitalisation as an outcome, the statement and the example conditions could perhaps be revised to better suit the utility of the approach described in the paper for hospitalisation outcomes, where conditions might be considered more equal in terms of risk of hospitalisation.

It would also be good to explicitly note the outcome (hospitalisation) in the final sentence of the introduction.

AUTHORS: We have modified the example to discuss the risk of hospitalisation rather than death. We have also added hospitalisation to the end of the Introduction.

**

Distribution of multiple chronic conditions: I think it is important to include contextual information on multimorbidity in the results, e.g. a frequency breakdown on the number of identified conditions in the sample (e.g. add to Supp Table 1, which already includes the singlecondition prevalences across the study sample) with something like the proportion of people with 0, 1, 2, 3 ... conditions (with some sensible capped upper limit e.g. 3+ or 4+ conditions).

AUTHORS: We have added the following sentence to the end of the 1st paragraph of the results:

"We found 1.8% of participants with 0 chronic conditions, 10.6% with one, 16.5% with two,

17.1% with three, 14.4% with four, 11.1% with five, and 20.0% with six or more conditions

(examining only individuals with complete data)."

Reviewer: 2

Reviewer Name

Manfred Gogol

Institution and Country

Institute of Gerontology, Heidelberg, Germany and

Geriatric Trauma Center, Trauma Department, Hannover Medical School, Hannover, Germany

Please state any competing interests or state ‘None declared’:

None declared

Please leave your comments for the authors below

It´s an interisting approach to work on the problem of multimorbidity and will foster the discussion.

AUTHORS: Thank you for your comments.

VERSION 3 - REVIEW

REVIEWER	James Stanley University of Otago, Wellington
REVIEW RETURNED	14-Feb-2020

GENERAL COMMENTS	Thank you for the amended manuscript -- these read well and I have no further comments.