Author’s response to reviews

Title: Case Identification of Mental Health and Related Problems in Children and Young People using the New Zealand Integrated Data Infrastructure

Authors:

Nicholas Bowden (Nick.Bowden@otago.ac.nz)
Sheree Gibb (sheree.gibb@otago.ac.nz)
Hiran Thabrew (h.thabrew@auckland.ac.nz)
Jesse Kokaua (jesse.kokaua@otago.ac.nz)
Richard Audas (rick.audas@otago.ac.nz)
Sally Merry (s.merry@auckland.ac.nz)
Barry Taylor (barry.taylor@otago.ac.nz)
Sarah Hetrick (s.hetrick@auckland.ac.nz)

Version: 2 Date: 18 Jan 2020

Author’s response to reviews:

17 January 2020

Dear Dr. Alison Cuff,

Thank you for your consideration of our paper “Case Identification of Mental Health and Related Problems in Children and Young People using the New Zealand Integrated Data Infrastructure” for publication in BMC Medical Informatics and Decision Making. We were happy you feel the paper is potentially acceptable for publication, subject to minor essential revisions. As such we have made revisions to the paper based on reviewer’s feedback, which, again were valuable and we feel our manuscript is improved as a result. Please also see below our point-by-point responses below to all four reviewers.

Thank you for considering our revised manuscript for publication.
Yours sincerely,

Nick Bowden

Hugo Leroux, PhD (Reviewer 1):

I thank the authors for taking into consideration my suggestions in a previous review. I congratulate the authors on a much improved paper discussing a really important topic using really good evidence.

I am happy with the current manuscript and would recommend it for publication.

Thank you very much for your feedback. Your comments and suggestions in the first stage of review were extremely helpful in improving the manuscript to the level it is now at and we appreciate your recommendation for publication.

Elizabeth Ford, PhD (Reviewer 2):

Thank you for the opportunity to review once again this interesting manuscript, and congratulations to the authors for incorporating many of the previous reviewers' comments.

The manuscript is much improved. Just two comments before I can recommend acceptance:

Thank you for your feedback and for reviewing our revised manuscript. Your comments and suggestions are again extremely helpful and we have made changes as a result to address your concerns. We feel that again our manuscript is improved as a result.

1) Table 1 is still blank - what information should this contain?

We are very sorry about this; it has now been fixed. The table indicates the combinations of data sources case identifications for each mental health problem category draw from.
2a) I am still concerned about the issue of Pharms data over capturing potential cases (false positives). For example amitriptyline and diazepam are both included in the list which in the UK may be prescribed for neuropathic or muscle pain respectively. This could lead to over identification of cases, reflected by the high rate captured by this data. This is still nothing discussing this risk in the discussion - it should be directly alluded to in the discussion of the future method refinement - potential validation exercises might result in an adjustment of the list of included medications or result in a combination of codes being needed to identify a case, e.g. a pharm prescription as well as a code from another source.

We agree that these particular medications may be used for neuropathic or muscle pain respectively. We have tried to minimise this with the age restrictions we have imposed for case identifications using these medications. We believe that with these age restrictions they most likely reflect use for depression (Amitriptyline, 20 years and over) and for mental health not defined (Diazepam, 15 and over). More broadly, we acknowledge there will be a number of medications that may be used for things other than the mental health indication we have assigned them to. That said, we have tried to minimise overidentification by excluding medications we feel more more likely used for non-mental health reasons and as above by imposing age restrictions. We have now amended the Discussion section so that the risk of over identification in relation to medication-based case identifications is made clear to the reader (see second paragraph of page 19). Furthermore, as explained in the first paragraph of the Limitations and Strengths section of the Discussion and the paper’s Conclusion, we also remain committed to more formal validation of the IDI algorithm using other data sources.

2b) This brings me to another point which could be made in the discussion, one very early method of validating, might be to look at the rates of conditions identified by each source. It's tricky to work this out from Table 2, but could be proposed for future work. For example if one source identifies a prevalence of e.g. substance use of 2% and another of 10% you would have to conclude that one data source is likely identifying many false positives or the other is under-identifying cases.
We had considered analysing face validity, both in the development of the paper, and in response to reviewers concerns about validity. However, the challenge with what you propose is that because each of the datasets relate to different parts of the health system, we would expect to see large differences in identification rates between datasets. For example, in specialist secondary service data (PRIMHD) we would expect to see relatively low rates of case identification because this service is only tailored to those with the highest need (3% of the population), while case identifications from pharmaceutical data will predominantly reflect care in the primary setting, where we would expect much higher rates of case identifications. Depression is a good example where we would expect case identifications from pharmaceutical data to be many times higher than those from PRIMHD data, even without any false positives. This is because the vast majority of depression is treated in primary care in NZ. We have discussed how different datasets reflect different service provision and hence different sections of the mental health population in the methods section, pages 8-11. Therefore, in our opinion, inferring anything about false positives or under identification from rates comparisons by data source would not be very informative. For this reason, we have decided not to incorporate this into the paper, either as part of the analysis, or as suggested future work. We still believe the best course of action is a formal validation process as outlined on pages 18-19.

If the authors could address these issues in the discussion, I would be satisfied with the other changes made to the paper.

Thank you. We believe we have now addressed these issues as best we can and hope they are to your satisfaction. Again, we appreciate all your feedback.

Hamido Fujita (Reviewer 3):

The authors has responded to my question in satisfactory fashion.

Thank you for your feedback. Your comments and suggestions in the first stage of review were very helpful in improving the manuscript to the level it is now at and we appreciate your recommendation for publication.
Hung-Wen Chiu (Reviewer 4):

The authors have attempt to response the review comments and revise the manuscript according to these comment. Unfortunately, the authors cannot address my concerns.

Thank you for your feedback. We appreciate the concerns you have and are sorry we have been unable to address them to your satisfaction. That said, we do not believe we can do anything further given available data and information (please see below for more details).

1. Though the IDI is already linked together, the authors still have to provide the data schema to readers.

We have been in contact with Statistics New Zealand (the organisation who manages the IDI) and they are unable to provide us with a database schema. The best they can suggest is information accessible via the links we have included as footnotes on page 7. Statistics New Zealand do have their own internal information such as a data schema and additional technical details about the database, but they do not make this available to researchers or the general public. We believe that the descriptions provided in the methods, along with the additional links, provide enough information for readers to understand the structure of the data used.

2. I expect that the method should be a flow diagram or information algorithm with equations. But the method provided in the revised manuscript is just a coding table in Appendix. I cannot find the novelty in method for case identification.

As discussed previously, we do accept that the 'method' is not particularly complex. After the first review stage we changed the terminology from 'algorithm' to 'method' so as not to give the impression that the method is any more complex than it really is. However, we firmly believe the method is novel and does add to the body of mental health case identification methods in the literature, in particular moving beyond single source service use data. Furthermore, as per the discussion, once a data source is available to validate against, there is scope to further refine the method into a more complex algorithm, for example by utilising medication dose information, additional socio-demographic restrictions, and/or imposing conditions that case identifications require multiple sources of identification as discussed in the Discussion section, page 21.
3. This study still lack validation procedure in this revised version, thus the reader could not evaluate the accuracy of the proposed method.

Yes, we accept this. As discussed in the paper (Discussion section pages 18-21), this is something that is planned as soon as a data source becomes available to validate against. Until then, we simply cannot validate the method and we have noted this as a limitation, along with discussion of how it could be validated in the future. This limitation notwithstanding, we believe the method provides a better means for identifying mental health problems than existing methods using single source service use data and has research value.