**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:** 1  **Date:** 28 Oct 2019

**Author’s response to reviews:**

29 October 2019

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 found the feedback from the four reviewers and yourself to be particularly useful, have discussed this in detail among the authorship team, and have responded point-by-point below in blue. As a result, we have made major revisions to the manuscript (see attached) and feel we have a much-improved paper to re-submit.

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

Nick Bowden

Editor Comments:

Thank you for submitting your paper, which has now been reviewed. The general conclusion is that the paper is relevant, well written, and appropriate for publication after it has been revised. The main problem identified by the reviewers, and with which I also concur, is that the methods are not properly described. This applies to both how the coding schemes were created and the algorithm that assigns people to mental-health categories.

The first method needs more details on the consensus approach, who was involved in it, how they were selected, what validation took place, and so on. I would suggest that you might also look at whether coding ontologies such as Snomed CT have a role at this point. Would these ontologies help validate the manual approach or even be a more tractable substitute for them if the approach was to be extended?

Thank you for your feedback and comments. Along with the reviewer's they have been very useful. As a result, we have made a number of revisions and now feel we have a much-improved manuscript to re-submit. Specifically, we have made some major revisions to the methods section that makes it clearer and more transparent. In addition, we have included tables in the appendix that were left out in error in the original submission that detail each code that is used in the method to identify the mental health problem groups. Please see our responses to the relevant reviewers’ comments for more details.
Regarding Snomed CT, at present New Zealand hospital and outpatient events are coded using ICD clinical coding systems. Snomed CT is not widely used in New Zealand. Therefore, we are unable to incorporate Snomed CT codes into our algorithms or validation at this time. New Zealand health coding systems are gradually moving from Read and ICD coding to greater uptake of Snomed CT. As the use of Snomed CT becomes more widespread it may be possible to incorporate Snomed CT codes into future revisions of this method.

The second case-identification algorithm is a mathematical one that should be easy to specify in detail. However, its performance does also need validation. You mention this as a weakness in the paper but the reviewers see it as more than future work but, rather, something that needs doing before this paper can be published. You need to show some evidence of the algorithm's accuracy.

As per our letter, the authorship team have discussed this issue in depth and have tried to think of strategies or approaches that would satisfy the reviewers. The main issue is that with the currently available data in the Integrated Data Infrastructure (IDI), a formal empirical validation is not possible. As a result, while validation for this method is viewed by the authorship group as very important, we see it as a whole new study, one that would require new ethics approval, applications to access new data, the creation of a new bespoke dataset, and therefore a study likely conducted over the next 12-18 months. With this in mind, we feel that presently, our only feasible response to the concerns raised over validation would be in line with the suggestion from reviewer #2 (Elizabeth Ford) who, after posing several questions about validation to us says: “Even if you can't actually answer these questions, a discussion of the limitation and much more detailed suggestions for future validation should be included in the discussion. You could include here literature on validation methods, and literature on influences on recording of conditions by medical practitioners (e.g. stigma, uncertainty), which may inform our understanding of what might be missing.” For details of how we have addressed this in the manuscript please see the response to that suggestion under reviewer 2.

There are some minor typos that need addressing and Table 1 is incomplete: the sources for each problem are not showing on my PDF of the submission.

This has now been fixed and the paper has now been proofread.
Hugo Leroux, PhD (Reviewer 1):

This paper describes research into case identification method for research into the mental health of children and young people in New Zealand. This is a very well-written paper that is substantiated with really good evidence.

My main concern is that the authors present the data sets used and mention the "algorithm" used to link and reason over the data. However, the algorithm is rather implicit in nature and is not presented clearly to the reader. For example, on page 16, lines 25-32, the authors state: "The algorithm identified 84,000 individuals …. This algorithm is not designed to estimate prevalence …". This may well be true (and I don't dispute this fact), however, the content of the paper provides no guidance as to how this is achieved.

Thank you for your feedback, we found it very useful in improving the paper. In the first instance, we have added a table to an appendix that includes all of the codes used to identify cases of each of the 13 mental health problem groups. We feel this adds much needed detail to the method. This table was left out in error in the original submission. We have also made significant modifications to the manuscript to add clarity to other aspects of the method (see our response to your second point below).

Furthermore, on page 17, line 20, the authors state: "… used a careful and transparent process to assign codes …". It would be useful to the reader if they would indeed explain the process to make it more transparent.

We have significantly revised the methods section on page 12-14 to make this clearer to the reader. This includes clarity on who was involved in developing the method, clarity on the rationale for code assignment decisions, additional references to reflect evidence used, and examples to illustrate.
Page 6, line 43: "… according to Statistics New Zealand protocols …". Please state the protocols or include a reference (I'm assuming that it's "Statistics New Zealand 2017c")

Reference added.

Page 16, line 46: please insert a comma after "compose" and after "far" - i.e. "compose, by far, most of …"

This is now fixed.

Elizabeth Ford, PhD (Reviewer 2):

I congratulate the authors on tackling an important topic and using a strong data source to do so. There are many strengths to this paper, but also some lack of clarity and depth of thought which I feel should be rectified before I could recommend publication, and which preclude the use of the method they present in further data sources. Detailed comments listed below:

1) The introduction makes a good case and rationale for the study and is concise and clearly written.

2) The methods sections on the data source are well written and clear.

3) The description of the development of the case definitions needs more work. For example:

   a) How were the 13 MH problems of interest established; by what process and using what justifications were these reached? What did each of the headings encompass, for example, were OCD, phobias, PTSD included under the anxiety heading?
Thank you very much for your detailed feedback and suggestions. We have addressed these point by point below and feel the paper is much-improved as a result. The 13 mental health categories were decided on by a team of experienced clinicians and researchers. The emphasis on this decision-making process was on creating categories of "common" mental health or related problems among this age group that presented to primary and secondary care, but also categories that were possible given available IDI data. For these reasons we argued that broader categories would be the most appropriate. With this in mind Phobias, OCD and PTSD were included in the anxiety category even though they are no longer in the DSM-V anxiety category. The groupings were therefore in some cases based on pragmatism as opposed to diagnostic purity. We have made a number of changes to manuscript to make all of this clearer (see pg. 12-14). A table has been added to an Appendix to make the specific constituents of each diagnostic explicit as well. This was left out in error in the original submission.

b) Table 1 is blank - please supply a new version

This is now fixed.

c) It appears that the second stage of work involved drawing up code lists of codes or drug names which were agreed by consensus to represent the MH problems of interest. Given that the authors recognise that different sources of data are more or less reliable, and more or less indicative of a definitive diagnosis, was any attempt made to weight the codes by how much it was likely to indicate a diagnosis? E.g. were medications given a lower weight than ICD diagnostic codes in specialist mental health care? If not, why was this not considered?

Codes were not weighted although this approach was discussed in detail among the authorship team. We felt accurate weighting would require detailed information about the relationship between observed diagnoses/medications and the 'true' diagnosis status (i.e. validation of the weighting). At present we do not have this information and validation was not part of the planned first stage of the project (see our response to your comment on validation below). For these reasons the authors decided against weighting and elected to take a simpler approach. Over time we feel that code weighting, or further refinement of age restrictions, or code allocations should be considered once a data source is available to validate against.
d) On the basis of what evidence were the age categories created and restrictions applied for age threshold of diagnosis?

These age bands were used as the epidemiology and recommended management strategies depend on developmental level. In addition, secondary mental health services are approximately organised around these age bands. Our method of classification drew on the clinical experience, took into account the prevalence of disorder and the likely treatment within these age bands (see changes on pg. 13/14).

e) In the section on data management page 13. Can you clarify that a single code from the code list in any of the categories was sufficient to get a 1 in the dichotomous mental health problem indicator variable? If there were more than one code per indicator, this was disregarded? Could you give a good justification for this approach?

Yes, a single code only was required (see above (c) for the rationale). Additional codes for the same mental health indicator did not result in any change. An individual could get a 1 in multiple mental health indicators if the appropriate codes were identified. We have now made this clearer in the data management section (pg. 15). We do see value in leveraging information on multiple codes for the same indicator, however in line with (c) above, we feel that this would be something we could include in future iterations of the method, once a data source is available to validate against.

4) Results. As Pharms identifies the most individuals, it flags to me that there may be a problem with using Pharm data as a proxy for diagnosis. I understand it may be capturing primary care diagnoses, which the other datasets cannot do. But it may also be throwing up a lot of false positives.
We were not surprised to find that most case identifications came from Pharms data as it is the only measure of primary mental health care treatment available. New Zealand secondary mental health care services aim to treat the 3% of the population with the highest mental health needs, yet we know that around 20 to 30% of the population will experience a mental health problem. Therefore, the majority of mental health care in New Zealand happens in primary care, and we would expect to source most of our mental health population through primary care records. As the reviewer has pointed out, false positives are possible, especially with Pharms data, and these are a concern. We have attempted to minimise these by carefully reviewing the list of medications used and excluding medications known to be used mostly for non-mental health conditions. We have also refined the medication indicators using age restrictions that should further reduce false positives, as some medications have different indications for different age groups. Without validation (see our response to your comment of validation below) we feel this is the best approach at the present time. Importantly, we have now also included a section in the discussion (see pg. 20 and our response to your comment below) that summarises the implications of the various limitations (including the pharmaceutical related one you highlight) and suggests that end users of the method should tailor it to suit their needs. This might for example involve excluding pharmaceutical data from the method if they feel this best suit their research goals. We feel that at this stage, this is the most responsible and transparent way to present this method.

5) Discussion. Page 16. This seems to be the first time you have described your approach as an algorithm. I think this is overstating the approach and is misleading. It seems like case-finding was just matching single codes from a code list? It would be more likely to be called an algorithm if various codes had different weights in terms of determining whether a diagnosis was present, or if different combinations of codes indicated a diagnosis (e.g. secondary diagnosis code from one dataset + prescription from another). Suggest you change Algorithm to another term.

Yes, good point, we agree and have changed "algorithm" to "method" throughout the manuscript.

6) What is the likely quality of your case finding method in terms of sensitivity and specificity? This has not been addressed. How does your method work as a whole? What is the rate of false positives from this method? How many cases might you miss? Does having a code on the list from one data source make you more likely to be a true case than if the code came from another data source? Are the codes within the code lists all equally likely to be representative of a true diagnosis? Even if you can't actually answer these questions, a discussion of the limitation and much more detailed suggestions for future validation should be included in the discussion. You
could include here literature on validation methods, and literature on influences on recording of conditions by medical practitioners (e.g. stigma, uncertainty), which may inform our understanding of what might be missing.

The authorship team have discussed this issue in depth and have tried to think of strategies or approaches that would satisfy the reviewers. The main issue is that with the currently available data in the Integrated Data Infrastructure (IDI), a formal empirical validation is not possible. As a result, while validation for this method is viewed by the authorship group as very important, we see it as a whole new study, one that would require new ethics approval, applications to access new data, the creation of a new bespoke dataset, and therefore a study likely conducted over the next 12-18 months. With this in mind, we feel that presently, our only feasible response to the concerns raised over validation would be in line with your suggestion: “Even if you can't actually answer these questions, a discussion of the limitation and much more detailed suggestions for future validation should be included in the discussion. You could include here literature on validation methods, and literature on influences on recording of conditions by medical practitioners (e.g. stigma, uncertainty), which may inform our understanding of what might be missing.”

We have added a substantive paragraph at the beginning of the limitations subsection (see pg. 18/19) concerning validation. We discuss lack of validation as a limitation, why it cannot currently be conducted, and how validation would be useful. We discuss and reference existing validation methods in the literature which are relevant. In addition, we have added detail to the section on future research on validation. Finally, when discussing implications for researchers using this method (see below) we have now emphasised that we have presented a broad case identification approach that could be tailored to suit the needs of individual research projects. For example, if false negatives were a concern, researchers may wish to exclude pharmaceutical based case identifications from the method.

7) Page 17 you say "The assignment of diagnostic categories using medications should be considered an informed guess, rather than a definitive classification". It's great that you recognise this but what should be done about it? How will it affect the next stages of development and validation of your method?
We have added an additional paragraph at the end of the limitations subsection (pg. 20) that highlights the implications of the stated limitations of the method for researchers who might use this method. We agree that pharmaceutical indications are the main area where validation is required. As per our response above, we have also advised that users of this method should tailor it to their needs. For example, if false negatives were a concern, researchers may wish to exclude pharmaceutical based case identifications from the method.

8) Page 18 you say "administrative data lacks clinical detail and have known quality issues which affect accuracy of case identification" - so what should be done about this for your case definition? what implications does this have for using the method you present?

Ideally, we think that formal validation of the method against a dataset containing 'true' diagnoses would help to strengthen the case definitions and highlight any areas of weakness. However, as discussed above, this is not possible at the present time. We have expanded the discussion to include some discussion of this specific limitation (pg. 18). In terms of implications, please see comments above.

Hamido Fujita (Reviewer 3):

The article is well written, it main contribution is the data studies provided for mental health care, it is provide good review on mental health studies, based on provided data. I suggest to accept the article after minor change. I think the structure of the paper should be redone for better comprehension instrumenting the main contribution. simple comparison with othe data set for mental health care can be also highlighted to demonstrate the value of the data used in the study of this article.
Thank you very much for your feedback. We have highlighted that this is a novel approach and its contributions further in the conclusion. In terms of comparisons to other data, at this stage we have already used all relevant available data within the IDI. While it may be possible to compare our prevalence rates against published prevalence rates, we have avoided doing this for two reasons. First, it is difficult to find New Zealand-specific prevalence rates that have a similar definition to those in the current paper (which is more of a 'service use prevalence' than a true prevalence) and therefore our rates are always likely to be lower. We have added an explicit comment about this in the Limitations subsection. Second, comparison against published prevalence rates was not the original intent of the paper, and we feel that the amount of discussion that would be required to compare and contrast different prevalence definitions would detract from the main focus of the paper (which is to present a broad method for identifying a mental health service use population in the IDI).

Proofreading is required for to enhance the readability.

The paper has now been proofread.

In total i suggest minor change revision.

Hung-Wen Chiu (Reviewer 4):

This study presented a work about the development of an administrative data-based case identification of mental health using New Zealand Integrated Data Infrastructure (IDI). I have some concerns about this study:

1. The method about how to integrate these dataset is unclear. The authors should provide a data schema for linking theses data and describing the relation between datasets.

Thank you for your feedback, you have highlighted some concerns that we appreciate and have made a number of changes as a result. We have added some additional detail and a further link for details of how data linking is achieved in the IDI (see pg. 7). Please note however that IDI is already linked together, the researchers don’t do the linking themselves.
2. The algorithm for case identification was not described clearly. The only description in "The method built on a similar approach taken by the SIA (Social Investment Agency, 2017)" on line 28 in page 11. The authors should provide more detailed information about this algorithm.

Yes, we agree.

In the first instance, we have added a table to an appendix that includes all of the codes used to identify cases of each of the 13 mental health problem groups. We feel this adds much needed detail to the method. This table was left out in error in the original submission. We have also made significant modifications to the methods section of the manuscript (page 12-14), to add clarity to other aspects of the method:

We have made clear who was involved in developing the method (including clinical specialty), clarity on the rationale for code assignment decisions, additional references to reflect evidence, as well as examples to illustrate.

We have added detail regarding how the 13 mental health categories were decided. The emphasis on this decision-making process was on creating categories of "common" mental health or related problems among this age group that presented to primary and secondary care, but also categories that were possible given available IDI data.

We have also added detail to the discussion of how age restrictions were employed highlighting that epidemiology and recommended management strategies depend on developmental level. In addition, secondary mental health services are approximately organised around these age bands. Our method of classification drew on the clinical experience, took into account the prevalence of disorder and the likely treatment within these age bands.

We believe that these revisions now make the method and the process undertaken to develop it clearer and more transparent.
3. The results for case identification were presented in Table 2. I wonder how to verify this results? The authors should provide more information about the accuracy for case identification.

The authorship team have discussed this issue in depth and have tried to think of strategies or approaches that would satisfy the reviewers. The main issue is that with the currently available data in the Integrated Data Infrastructure (IDI), a formal empirical validation is not possible. As a result, while validation for this method is viewed by the authorship group as very important, we see it as a whole new study, one that would require new ethics approval, applications to access new data, the creation of a new bespoke dataset, and therefore a study likely conducted over the next 12-18 months. With this in mind, we feel that presently, our only feasible response to the concerns raised over validation would be in line with the suggestion from reviewer #2 (Elizabeth Ford) who, after posing several questions to us (about the validity of the method) says: “Even if you can't actually answer these questions, a discussion of the limitation and much more detailed suggestions for future validation should be included in the discussion. You could include here literature on validation methods, and literature on influences on recording of conditions by medical practitioners (e.g. stigma, uncertainty), which may inform our understanding of what might be missing.”

We have added a substantive paragraph at the beginning of the limitations subsection (see pg. 18/19) concerning validation. We discuss lack of validation as a limitation, why it cannot currently be conducted, and how validation would be useful. We discuss and reference existing validation methods in the literature which are relevant. In addition, we have added detail to the section on future research on validation. Finally, when discussing implications for researchers using this method (see below) we have now emphasised that we have presented a broad case identification approach that could be tailored to suit the needs of individual research projects. For example, if false negatives were a concern, researchers may wish to exclude pharmaceutical based case identifications from the method.