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)	A simulation study to demonstrate biases created by diagnostic criteria of mental illnesses: major depressive episodes, dysthymia, and manic episodes
AUTHORS	Chao, Yi-Sheng; Lin, Kuan-Fu; Wu, Chao-Jung; Wu, Hsing-Chien; Hsu, Hui-Ting; Tsao, Lien-Cheng; Cheng, Yen-Po; Lai, Yi-Chun; Chen, Wei-Chih

VERSION 1 – REVIEW

REVIEWER	Roger Ho National University of Singapore Singapore
REVIEW RETURNED	30-Mar-2020

GENERAL COMMENTS	I have read this paper with great interest. This paper requires major revision.
	1) The PDF file has 85 pages and takes long time to download. It is way beyond page limit of a paper. Please shorten the paper to 20 -- 25 pages.
	2) Please offer more explanation that is understandable to doctors or psychiatrists. After reading this paper, I have difficulty understanding how the authors create bias variable.
	3) It is understandable that the current diagnostic criteria contain bias. As a doctor made a psychiatric diagnosis based on patient's report of symptoms (patient may hide psychiatric symptoms), the doctor's experience and effect of medication in alleviating some psychiatric symptoms. I have difficulty to understand how the mathematics formula able to calculate and address the above causes of biases.
	4) In line 376, the authors state " Amid these issues, we think the diagnostic criteria for mental illnesses should be reviewed and improved in a way that they can be easier to understand and use without introducing biases and can be closely linked to clinical decisions." This statement is not supported by any evidence. As long as the diagnosis is based on clinical criteria, it requires a doctor's interpretation and patient's willingness to report data. There will be bias. In order to avoid bias, psychiatry needs to move to use biological parameters for diagnosis. Please the latest diagnostic scan in psychiatry, functional near infrared spectroscopy or optical topography under future direction. Please amend the above statement as follows:

Amid these issues, we think the diagnostic criteria for mental illnesses should be reviewed and improved by incorporating biological parameters to reduce bias. Recent research found that functional near-infrared spectroscopy (fNIRS) or optical topography that use oxyhaemoglobin levels to classify different psychiatric disorders (e.g. major depressive disorder and borderline personality disorder) (Hussain et al 2020, Ho et al 2016).

References:

Husain SF, Tang TB, Yu R, et al. Cortical haemodynamic response measured by functional near infrared spectroscopy during a verbal fluency task in patients with major depression and borderline personality disorder. EBioMedicine. 2020;51:102586. doi:10.1016/j.ebiom.2019.11.047

Ho CS, Zhang MW, Ho RC. Optical Topography in Psychiatry: A Chip Off the Old Block or a New Look Beyond the Mind-Brain Frontiers?. Front Psychiatry. 2016;7:74. Published 2016 Apr 26. doi:10.3389/fpsyt.2016.00074

REVIEWER	Jan Christopher Cwik University of Cologne, Germany
REVIEW RETURNED	26-May-2020

GENERAL COMMENTS	The present study aimed to investigate the magnitudes of biases introduced to the diagnoses of major depressive episodes, dysthymic disorder, and manic episodes. Additionally, and the relationships between the diagnoses and the input symptoms should be investigated. Therefore, the authors used simulations of 100,000 subjects and assumed prevalence rates (0.05, 0.1, 0.3, 0.5, and 0.7) and correlations between symptoms (0, 0.1, 0.4, 0.7, and 0.9) were entered in the simulation. As outcomes, biases due to data censoring or categorization introduced to the intermediate variables, and the three diagnoses were measured. Additionally, the relationships between the input symptoms and diagnoses were interpreted using forward stepwise linear regressions. As a result, the authors report that the prevalence rates of the diagnoses were lower than those of the input symptoms and proportional to the assumed prevalence rates and the correlations between the input symptoms. The input variables could not fully explain the diagnoses except for zero assumed correlations and 0.7 prevalence rates of the input symptoms for the diagnosis of dysthymic disorder. The study focuses on a scientific topic that could potentially be of interest to readers of BMJ Open. However, the manuscript requires an extensive copy editorial process because there are several language issues and typos in the manuscript. However, as I am also not a native speaker, this should be evaluated elsewhere. The introduction extremely short, and a comprehensive introduction to the issue of this study is missing. Several important aspects are missing in the introduction (e.g., how often do biases during the diagnostic process occur? What are the consequences of these biases?). Additionally, several terms are named without and good introduction (e.g., the Biological Syndrome Model or the frailty syndrome). The formatting of the manuscript is inconsistent, and

	cross-references are missing ("Error! Reference source not found"). Also, the methods need some more enlightenment, and in the current version, from my point of view, the study could not be replicated, due to imprecise descriptions and missing information. The discussion focuses on criticism of the current diagnostic system and approaches, and from my perspective, this criticism is not indicated by the simulation of data without the basis of real-world data. However, for me, this manuscript was very hard to read, and admittedly, I was not able to follow the authors, and thus, I could be that my decision is biased by these circumstances.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1
I have read this paper with great interest. This paper requires major revision.
Authors' reply:
Thank you for your detailed comments. We appreciated reviewers' efforts to improve this manuscript and revised the manuscripts based on reviewers' comments.

1) The PDF file has 85 pages and takes long time to download. It is way beyond page limit of a paper. Please shorten the paper to 20 -- 25 pages.
Authors' reply:
We agreed and assessed the necessity of all sections. We kept the main sections, Background, Methods, Results, and Discussion, to 9 pages and moved a figure to appendices. We worried further reduction in the texts might prevent readers using and understanding the R codes to reproduce the results.

2) Please offer more explanation that is understandable to doctors or psychiatrists. After reading this paper, I have difficulty understanding how the authors create bias variable.
Authors' reply:
Thank you for your comment. We agreed that this is a key part of this study. Bias variables were created by several mechanisms we mentioned in the texts: especially categorization of continuous variables, censoring of the sum of multiple variables, and multiplicative products of input variables. We revised the description of these mechanisms in the texts and added more examples. A visual presentation in one publication is very helpful and cited.

This is very easy to understand if seeing this figure:
https://journals.plos.org/plosone/article/figure/image?size=medium&id=info:doi/10.1371/journal.pone.0197859.g002

In our previous publication:
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0197859#pone-0197859-g002

3) It is understandable that the current diagnostic criteria contain bias. As a doctor made a psychiatric diagnosis based on patient's report of symptoms (patient may hide psychiatric symptoms), the doctor's experience and effect of medication in alleviating some psychiatric symptoms. I have difficulty to understand how the mathematics formula able to calculate and address the above causes of

biases.

Authors' reply:
Thank you for the reminder. Biases come from many sources. To be clear, this study focused on the biases created solely by the diagnostic criteria in simulated populations, while assuming perfect diagnostic processes and accurate symptom reporting (Table 1). We valued the comment a lot and added texts in all sections to emphasize our focus on the biases created by the diagnostic criteria. In the Discussion, we mentioned other sources of bias, for example diagnosis by non-psychiatrist professionals and inaccurate reporting. In the Limitations, we considered this a limitation to this study.

We also thank you for your attention on the math equations. They are accurate depiction of the complex diagnostic criteria used in the DSM-IV. For example, diagnosis based on whether subjects meeting both the major and minor criteria of dysthymic disorder is the same as identifying those with a multiplicative product of 1 of two binomial variables (0 and 1 for absence and presence of the major or minor criteria). We added this example in the Methods to help readers understand the complexity of the diagnostic criteria and the accuracy of the equations (and subsequent biases).

4) In line 376, the authors state " Amid these issues, we think the diagnostic criteria for mental illnesses should be reviewed and improved in a way that they can be easier to understand and use without introducing biases and can be closely linked to clinical decisions." This statement is not supported by any evidence. As long as the diagnosis is based on clinical criteria, it requires a doctor's interpretation and patient's willingness to report data. There will be bias. In order to avoid bias, psychiatry needs to move to use biological parameters for diagnosis. Please the latest diagnostic scan in psychiatry, functional near infrared spectroscopy or optical topography under future direction. Please amend the above statement as follows:
Amid these issues, we think the diagnostic criteria for mental illnesses should be reviewed and improved by incorporating biological parameters to reduce bias. Recent research found that functional near-infrared spectroscopy (fNIRS) or optical topography that use oxyhaemoglobin levels to classify different psychiatric disorders (e.g. major depressive disorder and borderline personality disorder) (Hussain et al 2020, Ho et al 2016).
References:
Husain SF, Tang TB, Yu R, et al. Cortical haemodynamic response measured by functional near infrared spectroscopy during a verbal fluency task in patients with major depression and borderline personality disorder. EBioMedicine. 2020;51:102586. doi:10.1016/j.ebiom.2019.11.047
Ho CS, Zhang MW, Ho RC. Optical Topography in Psychiatry: A Chip Off the Old Block or a New Look Beyond the Mind-Brain Frontiers?. Front Psychiatry. 2016;7:74. Published 2016 Apr 26. doi:10.3389/fpsyt.2016.00074

Authors' reply:
Thank you for the comment. They are very important. The main findings in this study include 1) there are biases created solely by the diagnostic criteria of three mental illnesses, 2) the design of the diagnostic criteria give more weight on one or two input symptoms much more than the others, 3) the three diagnoses could not be fully explained by input symptoms, except for one condition (>70% prevalence of input symptoms of dysthymic disorder), and 4) the design of diagnostic criteria determines the prevalence of the three mental illnesses (dysthymic disorder more prevalent, given the same prevalence for all input symptoms of three illnesses). In the equations, we identified the sources of biases in the diagnostic criteria of the three illness. We are confident that the revised texts can help readers understand the biases embedded in the diagnostic criteria and motivate researchers to review the problems we identified.

We strongly agreed with you that the diagnosis of mental illnesses should be improved and made with precise measures, such as biomarkers. The references you suggested were included. Thank you.

Reviewer 2
The present study aimed to investigate the magnitudes of biases introduced to the diagnoses of major depressive episodes, dysthymic disorder, and manic episodes. Additionally, and the relationships between the diagnoses and the input symptoms should be investigated. Therefore, the authors used simulations of 100,000 subjects and assumed prevalence rates (0.05, 0.1, 0.3, 0.5, and 0.7) and correlations between symptoms (0, 0.1, 0.4, 0.7, and 0.9) were entered in the simulation. As outcomes, biases due to data censoring or categorization introduced to the intermediate variables, and the three diagnoses were measured. Additionally, the relationships between the input symptoms and diagnoses were interpreted using forward stepwise linear regressions.

Authors' reply: thank you for your summary.

As a result, the authors report that the prevalence rates of the diagnoses were lower than those of the input symptoms and proportional to the assumed prevalence rates and the correlations between the input symptoms. The input variables could not fully explain the diagnoses except for zero assumed correlations and 0.7 prevalence rates of the input symptoms for the diagnosis of dysthymic disorder.

Authors' reply:
Thank you. This is an accurate summary.

The study focuses on a scientific topic that could potentially be of interest to readers of BMJ Open. However, the manuscript requires an extensive copy editorial process because there are several language issues and typos in the manuscript. However, as I am also not a native speaker, this should be evaluated elsewhere. The introduction extremely short, and a comprehensive introduction to the issue of this study is missing. Several important aspects are missing in the introduction (e.g., how often do biases during the diagnostic process occur? What are the consequences of these biases?).

Authors' reply:
Thank you for the comments and advices. In the Introduction, we added texts to better explain the issue (biases created by composite diagnostic criteria) and the lack of research (only selected diagnoses reviewed for such biases). The magnitudes of the biases (> 71% diagnosis variances explained by biases alone) and consequences of the biases (biases better explaining the diagnosis than input symptoms) were revised in the Introduction.

We had a native speaker proofread the revised manuscript. Hope this addresses your comments well.

Additionally, several terms are named without and good introduction (e.g., the Biological Syndrome Model or the frailty syndrome). The formatting of the manuscript is inconsistent, and cross-references are missing ("Error! Reference source not found"). Also, the methods need some more enlightenment, and in the current version, from my point of view, the study could not be replicated, due to imprecise descriptions and missing information. The discussion focuses on criticism of the current diagnostic system and approaches, and from my perspective, this criticism is not indicated by the simulation of data without the basis of real-world data.

Authors' reply:
We apologize for the formatting errors. This manuscript was written on a Mac and somehow some cross-references could not be processed correctly in PDFs. This mistake was not repeated in this revision.

In the Introduction, we made changes based on your comments. We had more description about the biases embedded in composite diagnostic criteria, frailty, and study objectives.

In the Methods, we provided more descriptions and examples about how biases were generated by composite diagnostic criteria. The Methods section described how we conceived the research questions, why we needed to simulate, what the mechanisms to create biases, and how to evaluate the role of the biases in the diagnosis. This study is reproducible by using the R codes we provided in the Supplemental materials. Thank you for your attention to these issues.

For the criticisms to current diagnostic approaches, this study focused on the biased created solely by the diagnostic criteria. We cited other studies using real-world data to mention other sources of biases to mental illness diagnoses. In the Discussion, we described the importance of our results that demonstrated the diagnostic criteria introduced information unrelated to input symptoms into the diagnoses and determined the prevalence. Certain input symptoms were weighted much more than others without being noticed.

However, for me, this manuscript was very hard to read, and admittedly, I was not able to follow the authors, and thus, I could be that my decision is biased by these circumstances.

Authors' reply:

Thank you for your comments. We appreciate all the advices you provided and revised accordingly. This study took months to finish the simulations and had a lot of details that might be intimidating to other researchers. However, the biases created by the diagnostic criteria are an important issue that have never been well investigated. Your comments are helping us to increase the impact of the results we proudly discovered. The revised manuscript was proofread by a native speaker specialized in education research. We believe the revised manuscript provides a better introduction to the research questions and related issues, an improved description of the methods, and improved discussion without cross-referencing errors. We appreciate your input to our draft.

VERSION 2 – REVIEW

REVIEWER	Roger Ho National University of Singapore Singapore
REVIEW RETURNED	24-Jun-2020

GENERAL COMMENTS	I recommend publication

REVIEWER	Dr. Jan Christopher Cwik Universität zu Köln Germany
REVIEW RETURNED	25-Jun-2020

GENERAL COMMENTS	The authors revised the manuscript very carefully and took all of my recommendations into consideration. From my point of view, the paper is of good quality and could be accepted for publication.