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 association between insulin therapy and depression in patients with type 2 diabetes mellitus: a meta-analysis
AUTHORS	Bai, Xiaosu; Liu, Zhiming; Li, Zhisen; Yan, Dewen

VERSION 1 – REVIEW

REVIEWER	J Tibaldi New York Presbyterian Queens USA Speakers Bureau Novo Nordisk Consultant Novo Nordisk
REVIEW RETURNED	04-Nov-2017

GENERAL COMMENTS	I am a clinician and always felt what you elegantly demonstrated that insulin use and depression are related. I particularly found your use of studies from all continents useful. I believe this is a useful addition to clinicians

REVIEWER	George Papandonatos Brown University, USA
REVIEW RETURNED	12-Dec-2017

GENERAL COMMENTS	The authors properly note that reliance on cross-sectional data does not allow the causal and temporal relationship between insulin use and depression to be established. Therefore, their results section uses measured language to note that insulin use is associated with depression, without inferring causality. In contrast, the abstract does away with such niceties, and claims that insulin therapy significantly increases the risk of depression, a much stronger conclusion not warranted by the data. As many more readers will focus on the abstract, than the body of the paper, it is strongly recommended that associational language be used in both sections. Similarly, "associations" in the results section became "correlations" in the abstract. As correlation is a very specific measure of association not actually employed by the authors, they are asked to switch back to "associations" when referring to the relationships conveyed by odds ratios. Similarly, relative risk may be close to odds ratios when the prevalence is low, but that does not mean that odds ratios of 1.41 imply an exact 41% increase in prevalence of depression; they measure a change in the odds of depression instead. Minor issues with the use of the English language remain in the rest of the paper, but they are merely annoying, rather than incorrect or misleading. Regarding the meta-analysis itself there are 3 possible p-values that should be included when conducting moderation analyses: i) a p-value for between-group heterogeneity (overall moderation test), ii) a

	p-value for within-group heterogeneity (homogeneity test) and iii) a p-value for the significance of the within-group odds ratio to be interpreted only in the absence of within-group heterogeneity. Assuming sufficient power for moderation testing, a significant p-value for moderation by, e.g., geographical region would lead one to look at stratified odds ratios, whose significance would be of interest only if they are themselves homogeneous within region. The authors only provide p-values for the stratified odds ratios themselves, with no indication of whether the interaction tests are significant or whether the stratified odds ratios are themselves homogeneous across studies. The reported I-square statistics seem to suggest significant residual heterogeneity even within moderator strata. If so, the stratum-specific odds ratios are poor summary measures of the insulin-therapy association with depression. For this very reason, it is recommended that the authors considerably augment their supplementary material, providing forest plots for each moderation analysis, with study names and characteristics clearly labelled in each plot. Funnel plots should also be added to the online supplements. Still, this is an interesting study in a matter of rising public health significance and the authors should be given a chance to improve its presentation.

REVIEWER	Marjolein M Iversen Western Norway University of Applied Sciences
REVIEW RETURNED	23-Feb-2018

GENERAL COMMENTS	In the abstract the aim is stated as: to evaluate the impact of insulin therapy on the development of depression. In the introduction: to clarify the association between insulin therapy and the development of depression in T2DM patients. As the most studies are cross sectional, it is not possible to address the aim. Furthermore, it seems little meaningful to use much focus on the unadjusted OR as people are more likely to start with insulin when diabetes is progressing with among others more complications.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1
General comments
I am a clinician and always felt what you elegantly demonstrated that insulin use and depression are related. I particularly found your use of studies from all continents useful. I believe this is a useful addition to clinicians
Response: On behalf of all co-authors, I would like to appreciate this kindly comment proposed during the peer review.

Reviewer #2
Question 1: The authors properly note that reliance on cross-sectional data does not allow the causal and temporal relationship between insulin use and depression to be established. Therefore, their results section uses measured language to note that insulin use is associated with depression, without inferring causality. In contrast, the abstract does away with such niceties, and claims that insulin therapy significantly increases the risk of depression, a much stronger conclusion not warranted by the data. As many more readers will focus on the abstract, than the body of the paper, it is strongly recommended that associational language be used in both sections. Similarly, "associations" in the results section became "correlations" in the abstract. As correlation is a very

specific measure of association not actually employed by the authors, they are asked to switch back to "associations" when referring to the relationships conveyed by odds ratios.
Similarly, relative risk may be close to odds ratios when the prevalence is low, but that does not mean that odds ratios of 1.41 imply an exact 41% increase in prevalence of depression; they measure a change in the odds of depression instead. Minor issues with the use of the English language remain in the rest of the paper, but they are merely annoying, rather than incorrect or misleading.
Response: On behalf of all co-authors, I would like to appreciate the thoughtful comments proposed during the peer review. We have already revised the causal and temporal relationship between insulin use and depression in the revised manuscript. All of changes have already marked “RED”.

Question 2: Regarding the meta-analysis itself there are 3 possible p-values that should be included when conducting moderation analyses: i) a p-value for between-group heterogeneity (overall moderation test), ii) a p-value for within-group heterogeneity (homogeneity test) and iii) a p-value for the significance of the within-group odds ratio to be interpreted only in the absence of within-group heterogeneity. Assuming sufficient power for moderation testing, a significant p-value for moderation by, e.g., geographical region would lead one to look at stratified odds ratios, whose significance would be of interest only if they are themselves homogeneous within region. The authors only provide p-values for the stratified odds ratios themselves, with no indication of whether the interaction tests are significant or whether the stratified odds ratios are themselves homogeneous across studies. The reported I-square statistics seem to suggest significant residual heterogeneity even within moderator strata. If so, the stratum-specific odds ratios are poor summary measures of the insulin-therapy association with depression. For this very reason, it is recommended that the authors considerably augment their supplementary material, providing forest plots for each moderation analysis, with study names and characteristics clearly labelled in each plot. Funnel plots should also be added to the online supplements.
Response: On behalf of all co-authors, I would like to appreciate this kindly comment proposed during the peer review. P value for pooled results, heterogeneity, and between subgroups have already calculated and provided in Tables 2 and 3. All of changes have already marked “RED”. Further, significant residual heterogeneity was observed, while these results were restricted by uncontrolled baseline characteristics of included studies. Therefore, we added one sentence in Limitation section and marked “RED”.

Question 3: Still, this is an interesting study in a matter of rising public health significance and the authors should be given a chance to improve its presentation.
Response: On behalf of all co-authors, I would like to appreciate this kindly comment proposed during the peer review.

Reviewer #3: In the abstract the aim is stated as: to evaluate the impact of insulin therapy on the development of depression. In the introduction: to clarify the association between insulin therapy and the development of depression in T2DM patients. As the most studies are cross sectional, it is not possible to address the aim. Furthermore, it seems little meaningful to use much focus on the unadjusted OR as people are more likely to start with insulin when diabetes is progressing with among others more complications.
Response: On behalf of all co-authors, I would like to appreciate this kindly comment proposed during the peer review.
First, we have already changed the causal and temporal sentences in the revised manuscript, and all of changes are marked “RED”.
Second, in the planning stages, the studies reported multivariable-adjusted OR was intended to pooled, whereas numerous studies reported raw data should be excluded, and the reliable of this study may be uncertain. Therefore, the summary results for crude and adjusted OR were analyzed separately, and the result of crude OR could provide a reference.

VERSION 2 – REVIEW

REVIEWER	George Papandonatos Brown University, U.S.A
REVIEW RETURNED	13-Aug-2018

GENERAL COMMENTS	The authors have made all substantive changes requested by the reviewers. My only request is that in Tables 3 and 4 they relabel the last 2 p-value columns as "P value for within-stratum heterogeneity" and "P value for between-stratum heterogeneity". The standard of written English is very poor and detracts from the presentation of the findings. However, I found that in a similar instance with a past submission, the use of a professional company to polish the paper led to factual errors, as the internal reviewers spoke English very well, but had no understanding of the statistical concepts the authors were trying to convey! Therefore, I am reluctant to suggest a rewrite by professinals.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2 (George Papandonatos)
Question 1. The authors have made all substantive changes requested by the reviewers. My only request is that in Tables 3 and 4 they relabel the last 2 p-value columns as "P value for within-stratum heterogeneity" and "P value for between-stratum heterogeneity".
Response: We thank the reviewer’s positive comments. Thanks for your kind comment, it has been modified.

Question 2. The standard of written English is very poor and detracts from the presentation of the findings. However, I found that in a similar instance with a past submission, the use of a professional company to polish the paper led to factual errors, as the internal reviewers spoke English very well, but had no understanding of the statistical concepts the authors were trying to convey! Therefore, I am reluctant to suggest a rewrite by professinals.
Response: We are sorry for this, the language has been proofread by a native English speaker.