PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Application of the Ipswich Touch Test for Diabetic Peripheral Neuropathy Screening: A Systematic Review and Meta-Analysis
AUTHORS	Zhao, Nan; Xu, Jing-can; Zhou, Qiuhong; Li, Xinyi; Chen, Jiarui; Zhou, Jing; Zhou, Feng; Liang, Jinghong

VERSION 1 – REVIEW

REVIEWER	Doi, Suhail Australian National University, Department of Population Medicine We created the models we discuss
REVIEW RETURNED	11-Jan-2021

GENERAL COMMENTS	The authors present a diagnostic meta-analysis of the performance of the simple touch test for the diagnosis of sensory neuropathy and present the operating characteristics of the touch test as compared to several gold standards (mainly the 10g-MF) I have several methodology related comments a) The authors used the bivariate method for the synthesis of data. There is ample evidence that such models that make use of random effects approaches have unfavourable properties[1,2] and switching to the split component synthesis method is therefore advisable using either the IVhet (default) or quality effects model (*diagma* in Stata). [3] A quick analysis shows these results for the six datasets that are distinct from the authors results b) The subgroup analysis and sensitivity analyses are not adequate. Doing a leave one out sensitivity analysis is not useful and should be dropped. The secondary sensitivity analysis could be the five datasets using other gold standards (other than 10g-MF) to demonstrate the wide variability expected. There is not enough data to do a subgroup analysis and this should be dropped – there should be no analysis with less than five datasets. Table 2 should be dropped. c) The authors examined study quality using the QUADAS scale but then did not relate this to the results. This seems a waste of effort [4]. I recommend using the count of safeguards present (count of green dots in figure 3) as input into the quality effects model[5] in the diagnostic

meta-analysis. This can be done using *diagma* in Stata. Since there is a study with a zero count of safeguards, the authors should add 1 to all counts before entry to Stata to avoid that study dropping out of the analysis

d) The QUADAS is a methodological quality assessment scale – the authors should not mix up quality and risk of bias assessment in the paper [6]

e) "Publishing bias" should be "Publication bias". You cannot use Egger's test if there are less than 10 studies [7]. Please revert to the Doi plot and LFK index[8] based on the DOR

f) When you talk of post-test probabilities please first state the expected pre-test probabilities in various clinical settings and then factor in post-test probability based on the test.

References

[1] Doi SAR, Furuya-Kanamori L. Selecting the best meta-analytic estimator for evidence-based practice: a simulation study. *Int J Evid Based Healthc* 2020; **18**(1)**:**86-94.

[2] Doi SAR, Furuya-Kanamori L, Thalib L, Barendregt JJ. Meta-analysis in evidence-based healthcare: a paradigm shift away from random effects is overdue. *Int J Evid Based Healthc* 2017; **15**(4)**:**152-160.

[3] Furuya-Kanamori L, Kostoulas P, Doi SAR. A new method for synthesizing test accuracy data outperformed the bivariate method. *J Clin Epidemiol* 2020; **1 32:**51-58.

[4] Stone JC, Glass K, Munn Z, Tugwell P, Doi SAR. Comparison of bias adjustment methods in meta-analysis suggests that quality effects modeling may have less limitations than other approaches. *J Clin Epidemiol* 2019; **117:**36-45.

[5] Doi SA, Barendregt JJ, Khan S, Thalib L, Williams GM. Advances in the meta-analysis of heterogeneous clinical trials II: The quality effects model. *Contemp Clin Trials* 2015; **45**(Pt A)**:**123-9.

[6] Furuya-Kanamori L, Xu C, Hasan SS, Doi SA. Quality versus Risk-of-Bias assessment in clinical research. *J Clin* Epidemiol 2021; **129:**172-175.

[7] Sterne JA, Egger M, Smith GD. Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001; **323**(7304)**:**101-5.

[8] Furuya-Kanamori L, Barendregt JJ, Doi SAR. A new improved graphical and quantitative method for detecting bias in meta-analysis. *Int J Evid Based Healthc* 2018; **16**(4)**:**195-203.

	g)

REVIEWER	Duhamel, Todd A. University of Manitoba Faculty of Kinesiology and Recreation Management
REVIEW RETURNED	11-Jan-2021

GENERAL COMMENTS	This systematic review seeks to determine the sensitivity and specificity of the Ipswich Touch Test (IPTT) compared to reference methods typically used to diagnose Diabetic peripheral neuropathy. The review appears to have followed standardized reporting approaches and is, generally, well written. A few comments to enhance the clarity of the manuscript are recommended. Page number references below will refer to the page number of the pdf file reviewed, which was 37 pages in length. Page 6 - Lines 51-55: state “Currently, IPTT has been applied in some countries, and previous studies have reported differences in the results of the screening value of DPN. However, neither a meta-analysis nor a systematic review has been conducted on the screening value of IPTT.” However, a reference is not provided. Such a reference is required to support this statement. Additionally, please identify the countries that have approved the IPTT method for clinical use. Page 9 – Lines 51-56: “10g-MF, VPT, NDS, pin prick, tuning fork 128Hz, and ankle reflex were used as the reference standard to explore the accuracy of IPTT in DPN.” Table 1 report that VPT was used as a reference for 3 studies. However, even though page 11 lines 35-37 state “In general, when 10g-MF and VPT were used as reference standards, the sensitivity and specificity of IPTT were relatively high”, this outcome is not fully described in the methods and is not described in the abstract or conclusion. The authors must also describe the outcomes of for the reference comparison of the IPTT to the NDS, pin prick, tuning fork 128Hz, and ankle reflex in the abstract and conclusion. This information is required because the specificity and the sensitivity of the IPTT to these outcomes were examined, but not summarized. For example, the NDS as a reference standard had a specificity with the IPTT of 0.53. The authors must disclose this and discuss the implications that this outcome has within the context of the overall conclusion identified in the research. Page 11 – Line 3: It is stated that “Six datasets were included to evaluate the overall effect of IPTT in the screening of DPN.” Please identify which 6 datasets were used in the meta-analysis. Also, please identify why the other studies were not included in the meta-analysis. This is important information to include in the manuscript, as the decision must be justified.

	Table 2 must be revised to include additional information to better support the reader to understand the information listed on the table. For example, what does “No.” represent on the table legend? What studies are being included in each subgroup analysis approach? That information is not clear and should be articulated. Page 13 – Lines 49-54: It is written “The results of the meta-analysis found the combined sensitivity and specificity of IPTT to be 0.78 (95%CI 0.65–0.87) and 0.95 (95%CI 0.89–0.98), respectively, and AUC to be 0.93 (95%CI 0.90-0.95)”. However, it is unclear what reference measure this statement is referring to, as a number of comparator tests (10g-MF, VPT, NDS, pin prick, tuning fork 128Hz, and ankle reflex were described as the reference standard in the methods). If this data is referring to the comparison of IPTT to 10g-MF only, please clearly articulate this. Page 14 – Lines 8-10: What is meant by “has a certain potential to improve”. Please report the specific improvement that can be achieved with data or a relative description (is it a moderate improvement?).

REVIEWER	Jensen, Troels Staehelin Aarhus University, Neurology
REVIEW RETURNED	25-Mar-2021

GENERAL COMMENTS	In a metaanalysis Zhao et al., have examined the specificitity and sesnitivity of the Ipswich touch test in detecting diabetic neuropathy. In a careful search of 441 records 7 studies were found to meet the inclusion criteria i.e. all had diabetes and the Ipswich touch test was an index test (1510 pts).. Five of these studies were eventually included in a meta-analysis, involving altogether 1162 patients. The combined sensitivity and specificity was 0.78 and 0.96, respectively. In comparison to the 10 g MF the touch test had a reasonable sensitivity but a high specificity. A few additional limiting points needs consideration. The study is cross sectional and without a Gold standard for DPN neither the touch test or for that matter other of the reference test can predict the presence of DPN. The included studies in the metaanalysis used different reference standards for DPN (MF, NDS, VPT, tuning fork), so we can conclude that there is a high degree of agreement between results from the touch test on one hand and the MF and VPT on the other . The studies examined only patients with proven diabetes but no documentation such as the Toronto concensus panel definition for probable or definite diabetic neuropathy or any neurophysiological measure of DN. It would be interesting to get the sensitivity and specificity figures against these more hard core reference points. A high degree of heterogeneity was present between the studies which apparently was not related to number of patients in the study, age or ethnicity, but other factors such as methods, study performance was not analyzed

	A series of statistical analysis were carried out that needs specific statistical assessment

VERSION 1 – AUTHOR RESPONSE

To Reviewer 1:
Dear Dr. Suhail Doi, thank you very much for reviewing my manuscript during your busy schedule and providing professional and in-depth review comments on it, We are deeply inspired and follow your comments to supplement some basic experiments and conduct the revised manuscript. Next, I will give a point-to-point response to your suggestions.

Q: a) The authors used the bivariate method for the synthesis of data. There is ample evidence that such models that make use of random effects approaches have unfavourable properties[1,2] and switching to the split component synthesis method is therefore advisable using either the IVhet (default) or quality effects model (diagma in Stata). [3] A: Dear Dr. Suhail Doi, thank you very much for your comments on statistical analysis. We have benefited a lot from this. We have used the quality effects model to analyze the data, it's more appropriate for our study. Obviously, the problem of high heterogeneity can be well solved by using this model.

Q: b) The subgroup analysis and sensitivity analyses are not adequate. Doing a leave one out sensitivity analysis is not useful and should be dropped. The secondary sensitivity analysis could be the five datasets using other gold standards (other than 10g-MF) to demonstrate the wide variability expected. There is not enough data to do a subgroup analysis and this should be dropped – there should be no analysis with less than five datasets. Table 2 should be dropped.
A: Thanks for your seriousness and professionalism. Under your advice, we have dropped the Table 2(subgroup analysis) in the manuscript. Due to the low heterogeneity in the revised draft, we did not further conduct sensitivity analysis to explore the source of heterogeneity.

Q:c) The authors examined study quality using the QUADAS scale but then did not relate this to the results. This seems a waste of effort [4]. I recommend using the count of safeguards present (count of green dots in figure 3) as input into the quality effects model[5] in the diagnostic meta-analysis. This can be done using diagma in Stata. Since there is a study with a zero count of safeguards, the authors should add 1 to all counts before entry to Stata to avoid that study dropping out of the analysis.
A: We followed your suggestion and input the quality score(Qj) into the quality effects model. We are very sincere to your professional guidance and related literature sharing, under your advice, the quality of our articles has been greatly improved.

Q: d) The QUADAS is a methodological quality assessment scale – the authors should not mix up quality and risk of bias assessment in the paper [6]
A: After we have read the literature you shared with us, we find the confused quality evaluation and risk bias in the first draft. We have input the results of quality evaluation into the quality effects model in the revised manuscript.(Table 1 and Finger 2)

Q: e) “Publishing bias” should be “Publication bias”. You cannot use Egger’s test if there are less than 10 studies [7]. Please revert to the Doi plot and LFK index[8] based on the DOR

A: We sincerely apologize for the trouble caused to you by our negligence. After that, we have corrected it in the manuscript. And we also use Doi plot and LFK index based on the DOR to explain the publication bias. we have completed the description in the “Publication Bias” (Figure 5).

Q: f) When you talk of post-test probabilities please first state the expected pre-test probabilities in various clinical settings and then factor in post-test probability based on the test.
A: Thank you for your suggestion, we have elaborated the expected pre-test probabilities in various clinical settings and then factor in post-test probability based on the test in paragraph 3 of the “DISCUSSION” section. All the latest changes are highlighted in yellow in the revised manuscript.

To Reviewer 2:
Dear Dr. Todd A. Duhamel, thank you very much for reviewing my manuscript during your busy schedule and providing professional and in-depth review comments on it, We are deeply inspired and follow your comments to supplement some basic experiments and conduct the revised manuscript. Next, I will give a point-to-point response to your suggestions.

Q1: Page 6 - Lines 51-55: state “Currently, IPTT has been applied in some countries, and previous studies have reported differences in the results of the screening value of DPN. However, neither a meta-analysis nor a systematic review has been conducted on the screening value of IPTT.” However, a reference is not provided. Such a reference is required to support this statement. Additionally, please identify the countries that have approved the IPTT method for clinical use.
A1: Thank you for your very helpful suggestion. We have added the references and clarify the countries used IPTT in clinical setting, we have completed the description in the part of “INTRODUCTION” all the latest changes are highlighted in yellow in the revised manuscript.

Q2:.Page 9 – Lines 51-56: “10g-MF, VPT, NDS, pin prick, tuning fork 128Hz, and ankle reflex were used as the reference standard to explore the accuracy of IPTT in DPN.” Table 1 report that VPT was used as a reference for 3 studies. However, even though page 11 lines 35-37 state “In general, when 10g-MF and VPT were used as reference standards, the sensitivity and specificity of IPTT were relatively high”, this outcome is not fully described in the methods and is not described in the abstract or conclusion. The authors must also describe the outcomes of for the reference comparison of the IPTT to the NDS, pin prick, tuning fork 128Hz, and ankle reflex in the abstract and conclusion. This information is required because the specificity and the sensitivity of the IPTT to these outcomes were examined, but not summarized. For example, the NDS as a reference standard had a specificity with the IPTT of 0.53. The authors must disclose this and discuss the implications that this outcome has within the context of the overall conclusion identified in the research.
A2: Thanks for your seriousness and professionalism. We fully agree with your comments. We have made supplements in the revised manuscript. it can be seen in paragraph 2 of the “DISCUSSION” and “CONCLUSIONS” section.

Q3: Page 11 – Line 3: It is stated that “Six datasets were included to evaluate the overall effect of IPTT in the screening of DPN.” Please identify which 6 datasets were used in the meta-analysis. Also, please identify why the other studies were not included in the meta-analysis. This is important information to include in the manuscript, as the decision must be justified.
A3: Thank you for your suggestion. We have included 5 studies [22-26]. Because Sharma2012 [22] was conducted in two places (patient’s home and the clinic), it has two datasets. Therefore, a total of 6 datasets were used in the meta-analysis. We have also made supplements in the revised manuscript, we

have completed the description in the part of Meta-analysis Results Using 10g-MF as the Reference Standard (Screening Accuracy). In addition, it can also be seen in Table 1

Q4: Table 2 must be revised to include additional information to better support the reader to understand the information listed on the table. For example, what does “No.” represent on the table legend? What studies are being included in each subgroup analysis approach? That information is not clear and should be articulated.
A4: Thank you for your suggestion. “No” refers to the number of studies. Considering the small number of studies we included and the suggestions of other reviewers, so we deleted Table 2 in the revised manuscript.

Q5: Page 13 – Lines 49-54: It is written “The results of the meta-analysis found the combined sensitivity and specificity of IPTT to be 0.78 (95%CI 0.65–0.87) and 0.95 (95%CI 0.89–0.98), respectively, and AUC to be 0.93 (95%CI 0.90-0.95)”. However, it is unclear what reference measure this statement is referring to, as a number of comparator tests (10g-MF, VPT, NDS, pin prick, tuning fork 128Hz, and ankle reflex were described as the reference standard in the methods). If this data is referring to the comparison of IPTT to 10g-MF only, please clearly articulate this.
A5: We are sorry for our negligence to not make the manuscript fully described. this data is only referring to the comparison of IPTT to 10g-MF. We have added a description in the revised manuscript, it can be seen in paragraph 2 of the “DISCUSSION” section.

Q6: Page 14 – Lines 8-10: What is meant by “has a certain potential to improve”. Please report the specific improvement that can be achieved with data or a relative description (is it a moderate improvement?).
A6: Thanks for your advice, the potential value here refers to the potential screening value of IPTT in DPN. We have made relevant explanations and descriptions in the revised manuscript, the description in paragraph 3 of the “DISCUSSION”. All the latest changes are highlighted in yellow in the revised manuscript.

To Reviewer 3:
Dear Dr. Troels Staehelin Jensen, thank you very much for reviewing my manuscript during your busy schedule and providing professional and in-depth review comments on it, We are deeply inspired and follow your comments to supplement some basic experiments and conduct the revised manuscript. Next, I will give a point-to-point response to your suggestions.

Q1: The study is cross sectional and without a Gold standard for DPN neither the touch test or for that matter other of the reference test can predict the presence of DPN. The included studies in the meta-analysis used different reference standards for DPN (MF, NDS, VPT, tuning fork), so we can conclude that there is a high degree of agreement between results from the touch test on one hand and the MF and VPT on the other
A1: We completely agree with your opinions. After discussing with our research team, we followed your suggestion to make our conclusion more clearly. The description in “CONCLUSIONS” is “IPTT has a high degree of agreement in DPN screening with commonly used screening tool for DPN." All the latest changes are highlighted in yellow in the revised manuscript.

Q2: The studies examined only patients with proven diabetes but no documentation such as the Toronto concensus panel definition for probable or definite diabetic neuropathy or any neurophysiological measure of DN. It would be interesting to get the sensitivity and specificity figures against these more hard core reference points.

A2: Thank you very much for your comments, we deeply agree with your comments. The Toronto Consensus Group divides DPN into typical DPN (DSPN) and atypical DPNs, and DSPN is divided into Possible DSPN, Probable DSPN, Confirmed DSPN, and Subclinical DSPN. However, since we conducted a secondary analysis of previous published literatures, DN was not classified in more detail in the five original studies, meanwhile, the studies we included indicated that the subjects were diabetic patients, and there was no specific definition of diabetic neuropathy. The purpose of these studies is to carry out the preliminary screening of DPN in diabetic patients, IPTT is only a tool for primary screening, not a tool for diagnosing DPN, therefore, we cannot synthesize evidence based on a core reference point. We also think that the sensitivity and specificity of these reference points will be very interesting and more meaningful, At the same time, thank you very much for your guidance, we also regard it as our future research direction.

Q3: A high degree of heterogeneity was present between the studies which apparently was not related to number of patients in the study, age or ethnicity, but other factors such as methods, study performance was not analyzed.
A3: Thanks for your seriousness and professionalism, the research in our initial version is highly heterogeneous and has no connection with the number, age or ethnicity of the patients. We re-selected a new model based on the recommendation of the reviewer. By using this model, we found the heterogeneity decreased from 95.88% to 40.5%, so we did not further discuss the heterogeneity in the revised manuscript.

Q4: A series of statistical analysis were carried out that needs specific statistical assessment
A4: Thank you for your very helpful suggestion. We have made the analysis in more detail in the part of statistical method.

VERSION 2 – REVIEW

REVIEWER	Doi, Suhail Australian National University, Department of Population Medicine
REVIEW RETURNED	07-May-2021

GENERAL COMMENTS	I thank he authors for taking time to consider the comments and implement changes to the analysis which is much improved. However, the presentation and interpretation of the analysis requires more attention as follows: 1. Inclusion (1): “the study was designed as a diagnostic test and systematic reviews” – I am not sure what this means as reviews cannot really be an eligibility criterion and neither can “diagnostic test” 2. Under Data Synthesis: “The sensitivity, specificity, positive likelihood ratios (PLR), negative likelihood ratios (NLR), and corresponding 95% confidence intervals (95% CIs) were calculated using the TP, FP, FN, and TN values, which were extracted from each study prior to data pooling.” This whole statement does not make sense. Similarly “The quality effects model in meta-analysis was used to estimate variance between studies by using STATA, version15.1” is misleading. I assume the synthesis of DTA data was undertaken using a split component synthesis method - or was it not? This was not

	mentioned anywhere. The quality effects model is the synthesis model for diagnostic odds ratios within the split component synthesis method and it has nothing to do with estimating variance. Its not sufficient to use a tool to run an analysis – authors need to familiarize themselves with the research methods they use and report them accurately otherwise how can the results be interpreted? 3. Stata should be spelt Stata and not STATA 4. Likelihood ratios are important – granted. However other test accuracy indices have a role to play too and stating that "The likelihood ratio is more clinically significant than summary receiver operating characteristic curve (SROC) and diagnostic odds ratio (DOR) value" in my view is overkill. This statement needs to be more balanced and utility of all DTA metrics put in context. A table is needed highlighting the different DTA measures from meta-analysis and their interpretation discussed 5. The authors say that "In addition, Fagan nomograms were generated to evaluate the clinical utility of the two screening methods." This nomogram simply translates a pretest probability to a post-test probability and this is not an accurate representation 6. Heterogeneity and publication bias used which DTA measure and why? This is not stated 7. I would not do a leave-one-out sensitivity analysis when you have only five studies in a meta-analysis 8. The statement "Fagan's analysis showed that the pre-test probability was 50%, the probability of a positive result for DPN detected by IPTT was 94%, and the probability of a negative result for DPN detected was 23%. Further, the positive likelihood ratio was 15, and the negative likelihood ratio was 0.23. The above results demonstrate that there is a good diagnostic value of IPTT for DPN (Figure 4)." Is misleading because there is no such thing as "Fagan's analysis" and a pre-test probability comes from an external source and not from an analysis. The meta-analysis result tells us if there is good diagnostic value of IPTT – computing post-test probabilities (if done properly) can tell us how decision making can change with a test result but that has not been addressed. 9. Minor asymmetry does NOT mean there is slight publication bias – this needs a correct interpretation 10. Figure 4 (Fagan's nomogram) can be dropped as it does not add any useful information and should be replaced with the table suggested in comment 4

REVIEWER	Duhamel, Todd A. University of Manitoba Faculty of Kinesiology and Recreation Management
REVIEW RETURNED	22-May-2021

GENERAL COMMENTS	The authors have adequately addressed the comments raised in my initial review.

REVIEWER	Jensen, Troels Staehelin Aarhus University, Neurology
REVIEW RETURNED	26-May-2021

GENERAL COMMENTS	I am satisfied with the answers and the revised manuscript

VERSION 2 – AUTHOR RESPONSE

To Reviewer 1:

Dear Dr. Suhail Doi, thank you very much for reviewing my manuscript during your busy schedule and providing professional and in-depth review comments on it. We are deeply inspired and follow your comments to supplement some basic experiments and conduct the revised manuscript. Next, I will give a point-to-point response to your suggestions.

Q1. Inclusion (1): “the study was designed as a diagnostic test and systematic reviews” – I am not sure what this means as reviews cannot really be an eligibility criterion and neither can “diagnostic test”.

Response: Thank you for your comments. We are sorry that we were unlcear in the previous manuscript. In the revised manuscript, we described it with reference to previous published studies, and changed it to “(1) the study examined the screening accuracy of the IPTT test for detecting DPN”.

Reference:

[1] Wang F, Zhang J, Yu J, Liu S, Zhang R, Ma X, Yang Y, Wang P. Diagnostic Accuracy of Monofilament Tests for Detecting Diabetic Peripheral Neuropathy: A Systematic Review and Meta-Analysis. J Diabetes Res. 2017;2017:8787261.

Q2. Under Data Synthesis: “The sensitivity, specificity, positive likelihood ratios (PLR), negative likelihood ratios (NLR), and corresponding 95% confidence intervals (95% CIs) were calculated using the TP, FP, FN, and TN values, which were extracted from each study prior to data pooling.” This whole statement does not make sense. Similarly “The quality effects model in meta-analysis was used to estimate variance between studies by using STATA, version15.1” is misleading.

I assume the synthesis of DTA data was undertaken using a split component synthesis method - or was it not? This was not mentioned anywhere. The quality effects model is the synthesis model for diagnostic odds ratios within the split component synthesis method and it has nothing to do with estimating variance. Its not sufficient to use a tool to run an analysis – authors need to familiarize themselves with the research methods they use and report them accurately otherwise how can the results be interpreted?

Response: Thank you. We have benefited from your comments. We changed it to“The sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR) and corresponding 95% confidence interval (95% CI) were synthesized using the quality effects model, it is the synthesis model for diagnostic odds ratios within the split component synthesis method.”

Q3. Stata should be spelt Stata and not STATA.

Response: Thank you for the correction. We have corrected it in the revised manuscript.

Q4. Likelihood ratios are important – granted. However other test accuracy indices have a role to play too and stating that “The likelihood ratio is more clinically significant than summary receiver operating characteristic curve (SROC) and diagnostic odds ratio (DOR) value” in my view is overkill. This statement needs to be more balanced and utility of all DTA metrics put in context. A table is needed highlighting the different DTA measures from meta-analysis and their interpretation discussed.

Response: Thank you. Following your advice, we have carefully revised this part in revision, we added a table (Table 2) describing these DTA metrics in the “RESULTS” section, and the results are discussed in paragraph 2 of the “DISCUSSION” section. All the latest changes are highlighted in yellow in the revised manuscript.

Q5. The authors say that “In addition, Fagan nomograms were generated to evaluate the clinical utility of the two screening methods.” This nomogram simply translates a pretest probability to a post-test probability and this is not an accurate representation.

Response: We have benefited a lot from this. This will also help us to better use Fagan nomograms in the future. Based on the limitations of Fagan's Nomogram (Recommendation 5, Recommendation 8), and Recommendation 10, we have dropped the Fagan's Nomogram.

Q6. Heterogeneity and publication bias used which DTA measure and why? This is not stated.

Response: Heterogeneity is measured by I2 test, since Q test is affected by the number of studies included, the value of I2 statistic will not change with the number of studies, and the results of heterogeneity test are more reliable, so we used I2 to evaluate the magnitude of heterogeneity.

Publication bias is measured by Doi plot and LFK index, because the LFK index can detect and quantify the asymmetry in the Doi plots, and related studies have shown that the methods can markedly improve the ability of researchers to detect bias in meta-analysis. we presented in the second paragraph of the “Data Synthesis” section.

Reference:

[2] Wang D, Mou ZY, Zhai JX, et al. Application of Stata software to test heterogeneity in meta-analysis method. Zhonghua Liu Xing Bing Xue Za Zhi. 2008 Jul;29(7):726-9.

[3] Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, et al. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods. 2006 Jun;11(2):193-206.

[4] Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21(11):1539-58.

Q7. I would not do a leave-one-out sensitivity analysis when you have only five studies in a meta-analysis.

Response: Thank you for your suggestion. We have dropped the sensitivity analysis.

Q8. The statement “Fagan’s analysis showed that the pre-test probability was 50%, the probability of a positive result for DPN detected by IPTT was 94%, and the probability of a negative result for DPN detected was 23%. Further, the positive likelihood ratio was 15, and the negative likelihood ratio was 0.23. The above results demonstrate that there is a good diagnostic value of IPTT for DPN (Figure 4).” Is misleading because there is no such thing as “Fagan’s analysis” and a pre-test probability comes from an external source and not from an analysis. The meta-analysis result tells us if there is good diagnostic value of IPTT – computing post-test probabilities (if done properly) can tell us how decision making can change with a test result but that has not been addressed.

Response: We fully agree with your comments. Based on the limitations of Fagan's Nomogram (Recommendation 5, Recommendation 8), and Recommendation 10, we have dropped the Fagan's Nomogram.

Q9. Minor asymmetry does NOT mean there is slight publication bias – this needs a correct interpretation.

Response: We have benefited a lot from this, we have made a remarkable revision in the manuscript. Minor asymmetry was present in the Doi plot and the results of the LFK index also suggested minor negative asymmetry of the Doi plot (LFK index =-1.68), indicating the publication bias existed between the studies, this was also one of the limitations of our study, we also added in paragraph 3 of the “DISCUSSION” section.

Q10. Figure 4 (Fagan’s nomogram) can be dropped as it does not add any useful information and should be replaced with the table suggested in comment 4.

Response: Thank you for your suggestion. We have dropped Fagan's Nomogram and added a table in the results section (Table 2).

VERSION 3 – REVIEW

REVIEWER	Doi, Suhail Australian National University, Department of Population Medicine
REVIEW RETURNED	21-Jul-2021

GENERAL COMMENTS	The reviewer provided a marked copy with additional comments. Please contact the publisher for full details.

VERSION 3 – AUTHOR RESPONSE

To Dear Dr. Suhail Doi:
Thank you very much for reviewing my manuscript during your busy schedule and providing professional and in-depth review comments on it, we are deeply inspired and follow your comments to supplement some basic experiments and conduct the revised manuscript. Next, I will give a point-to-point response to your suggestions.

Q: Commented [S1]: “The likelihood ratio is an independent indicator to assess authenticity, which can simultaneously reflect sensitivity and specificity” is a meaningless description, please re-write.
A: Thanks for your seriousness and professionalism. Under your advice, we have already rewritten according to your suggestion, we presented in the “Data Synthesis” section.

Q: Commented [S2]:
A: We sincerely apologize for the trouble caused to you by our negligence. The word is “manual searches”, we have corrected it in the manuscript.

Q: Commented [S3]: This is NOT the interpretation of observed asymmetry.
A: Thanks for your seriousness and professionalism. We have reinterpreted Doi plot and the LFK index in the part of “Publication Bias”.

Q: Commented [S4]: Incorrect interpretation of NLR and PLR
A: Thank you for your very helpful suggestion. PLR is the ratio of the true positive rate to the false positive rate of IPTT screening for DPN, and NLR is the ratio of false negative rate to true negative rate. We have reinterpreted in the part of “DISCUSSION”, all the latest changes are highlighted in yellow in the revised manuscript.