PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

Title (Provisional)

Association between gestational diabetes mellitus diagnostic criteria and adverse pregnancy outcomes– a systematic review and meta-analysis of adjusted effect sizes from studies using current diagnostic criteria

Authors

Mahmoud, Elhassan; Elsayed, Abdalla Moustafa; Elsayed, Basant; Elsalakawi, Yasmin; Gopinath, Aswathy; Chivese, Tawanda

VERSION 1 - REVIEW

Reviewer	**1**
Name	**Glastras, Sarah**
Affiliation	**University of Sydney CAR**
Date	**06-Sep-2024**
COI	**n/a**

This is a well-written manuscript that contributes significantly to the ongoing debate regarding diagnostic thresholds for gestational diabetes mellitus (GDM) and has the benefit of pooled meta-analysis.

Abstract: In the abstract, it is recommended to briefly describe the comparative groups based on the GDM diagnostic criteria used (i.e., specify which criteria were applied for diagnosis). This would enhance clarity and understanding.

Methods: A key concern is why the systematic review only included studies from 2010 onwards. Historically, studies that have utilized non-IADPSG (International Association of Diabetes and Pregnancy Study Groups) cutoffs are relevant and should be considered, provided they meet other inclusion criteria. Including these studies would improve the fairness of the comparison, potentially increasing the number of non-IADPSG studies beyond the current count of four. Statistical methods are sound.

Results: Well written, adequate figures sand supplementary material for SR.

Discussion: The association between GDM and adverse pregnancy outcomes is well-articulated. However, the significant increase in the risk of neonatal hypoglycemia requires contextualization. It

is important to acknowledge that neonates of mothers with GDM are routinely tested for blood glucose levels antenatally, a procedure not applied in non-GDM pregnancies. This difference introduces potential allocation bias, which should be addressed.

Furthermore, the implications of the finding that adverse pregnancy outcomes are comparable between IADPSG and non-IADPSG studies should be expanded upon. Specifically, the authors should explore what this finding means for healthcare systems, clinicians, and, most importantly, women. Although this topic is briefly mentioned, it deserves further emphasis.

Additionally, a discussion of whether universal testing is the optimal approach could add depth to the manuscript. Consider moving this section earlier in the discussion, perhaps to the second paragraph, as the findings comparing IADPSG and non-IADPSG criteria are the most compelling and novel aspect of the study. In contrast, the other findings related to GDM are already well-documented in the literature.

Reviewer	**2**
Name	**Berntorp, Kerstin**
Affiliation	**Lund University**
Date	**23-Sep-2024**
COI	**None**

This systemic review and meta-analysis aimed to quantify the association between contemporary GDM diagnostic criteria and adverse pregnancy outcomes for the mother and child based on observational studies published between 2010 and 2023 (after the IADPSG recommendation was published). Subgroup analysis was performed to compare studies adopting the IADPSG recommendation with those “adopting criteria with higher cut-offs”. It is concluded that GDM showed consistent associations with maternal and fetal adverse outcomes with no major differences in effects when different contemporary criteria were used.

As pointed out by the authors the study is limited by using data from observational studies, but nevertheless strengthened by including adjusted effect size. A few more limitations should be addressed as outlined below.

A major limitation of this analysis is that a vast majority of the included studies were based on Asian populations, according to Table 1, 19/35 (54%) studies and only 5/35 (14%) from Europe (3 from Croatia, 2 from Italy, the latter rather small numbers), Middle East 3/35, North America (Canada) 2/35, South America (Brazil) 1/35, Australia 3/35 (2/3 using Non-IADPSG criteria), Africa 2/35 (small numbers). This limitation should be addressed in the discussion and makes the result less generalizable.

According to the introduction an additional aim of the current analysis was to compare the effect size between criteria conformed to the IADPSG recommendation against non-IADPSG criteria “generally using higher cut-offs”. However, as also mentioned in the introduction IADPSG

recommend universal one-step 2-hour 75 g OGTT. In Table 1 the study by Liu 2020 (n=522) used selective screening. Why wasn´t this study excluded? In fact, a major strength of this study was the use of IADPSG criteria, with reference to both lower cut-offs and universal one-step OGTT, and its association with adverse pregnancy outcomes, as most previous studies dealing with this issue (including RCTs) are based on selective screening. As pointed out in the discussion page 13, line 401–403, universal screening leads to an increase in GDM prevalence without a concurrent increase in benefit (diluting the outcomes).

Page 9 line 252–260 gives the criteria for the studies using Non-IADPSG criteria. It would be helpful if the fasting, 1-hour and 2-hour ranges of cut-offs are given in the text along with the number of participants in the respective studies using the same cut-offs. One study used WHO 1999 cut-offs which probably stands out from the others using cut-offs more like the IADPSG recommendation. Furthermore, it could be understood that these studies used universal OGTT. Is this correct also for the two studies using Carpenter-Coustan criteria? The small numbers of studies using Non-IADPSG criteria, most of which obviously are close to IADPSG cut-offs makes such a comparison doubtful. This limitation should be further addressed and speculated on in the discussion.

A minor comment regarding “Characteristics of included studies” page 10, line 275–284. Numbers given for 31 included studies, Asia (18), Europe (4) and Australia (2), North America (1) are not in line with the numbers given I Table 1, referring to 35 included studies, Asia (19), Europe (5) and Australia (3), North America (2). What is the explanation for this discrepancy?

VERSION 1 - AUTHOR RESPONSE

Reviewer 1 Comments:

Comment 1*: This is a well-written manuscript that contributes significantly to the ongoing debate regarding diagnostic thresholds for gestational diabetes mellitus (GDM) and has the benefit of pooled meta-analysis.*

Response 1: Thank you.

Comment 2: *Abstract: In the abstract, it is recommended to briefly describe the comparative groups based on the GDM diagnostic criteria used (i.e., specify which criteria were applied for diagnosis). This would enhance clarity and understanding.*

Response 2: Thank you for this suggestion. The abstract has been revised and edited accordingly.

“The same meta-analytic models were used to synthesise the overall odds ratios (OR) and their 95% confidence intervals for comparisons of the criteria which followed the IADPSG recommendations to other criteria, mostly with higher blood glucose cut-offs.”

Comment 3*: Methods: A key concern is why the systematic review only included studies from 2010*

onwards. Historically, studies that have utilized non-IADPSG (International Association of Diabetes and Pregnancy Study Groups) cutoffs are relevant and should be considered, provided they meet other inclusion criteria. Including these studies would improve the fairness of the comparison, potentially increasing the number of non-IADPSG studies beyond the current count of four.

Response 3: Thank you. While we appreciate this valid concern, we restricted the inclusion to studies conducted from 2010 onwards to ensure a fair and consistent comparison between diagnostic criteria. The IADPSG recommendations were introduced in 2010, marking a significant shift in the diagnostic approach to GDM. Including studies conducted before 2010 may introduce bias due to factors beyond the criteria, such as a lower GDM detection rate and misclassification of women with overt diabetes in pregnancy – these women were classified as GDM in many guidelines prior to the HAPO and IADPSG. Further, some criteria that are pre-2010 are now redundant and including them in the analysis will not to the rigor of the current analysis.

***Comment 4**: Methods: Statistical methods are sound.*

Response 4: Thank you for the comment.

***Comment 5:** Results: Well written, adequate figures and supplementary material for SR.*

Response 5: Thank you for the encouraging comment.

***Comment 6**: Discussion: The association between GDM and adverse pregnancy outcomes is well-articulated. However, the significant increase in the risk of neonatal hypoglycemia requires contextualization. It is important to acknowledge that neonates of mothers with GDM are routinely*

tested for blood glucose levels antenatally, a procedure not applied in non-GDM pregnancies. This difference introduces potential allocation bias, which should be addressed.

Response 6: Thank you for the comment. This part of the discussion has been expanded upon to address this issue.

"Notably, the highest odds ratio was observed for neonatal hypoglycemia, with 3-fold higher odds for GDM exposed neonates compared to the non-GDM exposed neonates. However, it is important to consider that this risk could be exaggerated due to the possibility of allocation bias for this outcome. Neonates born to mothers with GDM are more likely to be routinely tested for blood glucose levels shortly after birth due to the known risks of hypoglycemia, whereas neonates of non-GDM pregnancies do not typically undergo such testing unless clinically indicated. This difference in clinical practice likely increases the detection rate of hypoglycemia in the GDM group, which could lead to an overestimation of the association between GDM and neonatal hypoglycemia."

***Comment 7**: Discussion: The implications of the finding that adverse pregnancy outcomes are comparable between IADPSG and non-IADPSG studies should be expanded upon. Specifically, the authors should explore what this finding means for healthcare systems, clinicians, and, most importantly, women. Although this topic is briefly mentioned, it deserves further emphasis.*

Response 7: Thank you for the comment. We have expanded the discussion to further emphasize on the implications of the findings for healthcare systems, clinicians, and women.

"Our findings have several implications. For healthcare systems, adopting the IADSPG criteria, i.e. universal screening and lower glycemic thresholds compared to targeted screening and generally higher glycemic diagnostic thresholds, may strain resources, as more women would require screening, monitoring, and interventions. This could lead to an increase in healthcare costs [59, 60], which will lead to an increased burden, especially in settings where resources are already constrained. On the other hand, selective or targeted screening may result in some proportions of women progressing with undiagnosed hyperglycemia in pregnancy, and the consequent higher risk of adverse pregnancy outcomes. The NICE, for example, has opted to keep their guidelines which use risk factor based screening and higher glycemic thresholds. It is crucial to balance the costs and benefits of adopting either the IADPSG recommendations or selective screening, higher glycemic threshold approaches such as that used by the NICE. These considerations may be different for different health systems, depending on affordability and healthcare system capacity. For clinicians, these findings highlight the need for careful consideration when diagnosing and managing GDM, as they should be mindful of the potential for over-diagnosis and over-treatment, and they should tailor management strategies based on each patient's individual risk factors, ensuring that interventions are justified and beneficial. For women, the increased likelihood of a GDM diagnosis that comes with universal screening and lower glycaemic thresholds may result in increased anxiety and an increased likelihood of medical interventions, without a clear improvement of outcomes. GDM diagnosis has been associated with a higher occurrence of mental health problems in pregnant women [61, 62]. It is therefore critical to provide women with clear and balanced information along with the implications, and to promote shared decision-making.

More research is needed to identify appropriate blood glucose cut-offs where the benefit of GDM diagnosis outweighs the unintended negative consequences."

***Comment 8**: Discussion: A discussion of whether universal testing is the optimal approach could add depth to the manuscript. Consider moving this section earlier in the discussion, perhaps to the second paragraph, as the findings comparing IADPSG and non-IADPSG criteria are the most compelling and novel aspect of the study. In contrast, the other findings related to GDM are already well-documented in the literature.*

Response 8: Thank you for your comment. We agree and this has been edited accordingly. Please note that this point overlaps with the previous point and therefore the discussion has been combined for the two points.

"Our findings have several implications. For healthcare systems, adopting the IADSPG criteria, i.e. universal screening and lower glycemic thresholds compared to targeted screening and generally higher glycemic diagnostic thresholds, may strain resources, as more women would require screening, monitoring, and interventions. This could lead to an increase in healthcare costs [59, 60], which will lead to an increased burden, especially in settings where resources are already constrained. On the other hand, selective or targeted screening may result in some proportions of women progressing with undiagnosed hyperglycemia in pregnancy, and the consequent higher risk of adverse pregnancy outcomes. The NICE, for example, has opted to keep their guidelines which use risk factor based screening and higher glycemic thresholds. It is crucial to balance the costs and benefits of adopting either the IADPSG recommendations or selective screening, higher glycemic threshold approaches such as that used by the NICE. These considerations may be different for different health systems, depending on affordability and healthcare system capacity. For clinicians, these findings highlight the need for careful consideration when diagnosing and managing GDM, as they should be mindful of the potential for over-diagnosis and over-treatment, and they should tailor management strategies based on each patient's individual risk factors, ensuring that interventions are justified and beneficial. For women, the increased likelihood of a GDM diagnosis that comes with universal screening and lower glycaemic thresholds may result in increased anxiety and an increased likelihood of medical interventions, without a clear improvement of outcomes. GDM diagnosis has been associated with a higher occurrence of mental health problems in pregnant women [61, 62]. It is therefore critical to provide women with clear and balanced information along with the implications, and to promote shared decision-making. More research is needed to identify appropriate blood glucose cut-offs where the benefit of GDM diagnosis outweighs the unintended negative consequences."

Reviewer 2 Comments:

***Comment 1**: A major limitation of this analysis is that a vast majority of the included studies were based on Asian populations, according to Table 1, 19/35 (54%) studies and only 5/35 (14%) from Europe (3 from Croatia, 2 from Italy, the latter rather small numbers), Middle East 3/35, North America (Canada) 2/35, South America (Brazil) 1/35, Australia 3/35 (2/3 using Non-IADPSG criteria), Africa 2/35 (small numbers). This limitation should be addressed in the discussion and makes the result less generalizable.*

Response 1: Thank you for your comment. This has been added to the discussion.

"Additionally, most of the included studies were conducted in Asia, and relatively fewer studies from the other regions. This limits the generalizability of our findings to non-Asian populations."

Comment 2: *According to the introduction an additional aim of the current analysis was to compare the effect size between criteria conformed to the IADPSG recommendation against non- IADPSG criteria "generally using higher cut-offs". However, as also mentioned in the introduction IADPSG recommend universal one-step 2-hour 75 g OGTT. In Table 1 the study by Liu 2020 (n=522) used selective screening. Why wasn´t this study excluded? In fact, a major strength of this study was the use of IADPSG criteria, with reference to both lower cut-offs and universal one-step OGTT, and its association with adverse pregnancy outcomes, as most previous studies dealing with this issue (including RCTs) are based on selective screening. As pointed out in the discussion page 13, line 401–403, universal screening leads to an increase in GDM prevalence without a concurrent increase in benefit (diluting the outcomes).*

Response 2: Thank you for this important comment. The study by Liu did indeed use selective screening, as it excluded women at high risk for GDM. We understand that the selective approach does not fully align with the IADPSG recommendation for universal screening, and therefore, we decided to exclude the study and we have edited the results accordingly.

Comment 3*: Page 9 line 252–260 gives the criteria for the studies using Non-IADPSG criteria. It would be helpful if the fasting, 1-hour and 2-hour ranges of cut-offs are given in the text along with the number of participants in the respective studies using the same cut-offs. One study used WHO 1999 cut-offs which probably stands out from the others using cut-offs more like the IADPSG recommendation. Furthermore, it could be understood that these studies used universal OGTT. Is this correct also for the two studies using Carpenter-Coustan criteria? The small numbers of studies using Non-IADPSG criteria, most of which obviously are close to IADPSG cut-offs makes such a comparison doubtful. This limitation should be further addressed and speculated on in the discussion.*

Response 3: Thank you for the insightful comments. We have now revised the methods section to include the blood glucose cut-offs, along with the number of participants in studies using these cut-offs. The studies employing CC did indeed use universal screening, which we have now clarified in the results section. Additionally, we have expanded on the limitation of the few number of non-IADPSG studies.

"Non-IADPSG criteria in this study were Carpenter-Coustan (CC) (2 studies [25, 26]), 2008 Canadian Diabetes Association (CDA) (1 study [27]), ADA 2014 (1 study [28]), World Health Organization 1999 (1 study [29]) and the ADIPS (1 study [30]). The studies using CC criteria employed universal OGTT screening. The cut-offs used in these studies varied. For CC and ADA 2014 criteria, fasting glucose ≥ 5.3 mmol/L, 1-hour ≥ 10.0 mmol/L, and 2-hour ≥ 8.6 mmol/L were used (n = 7,612). The WHO 1999 cut-offs included fasting glucose ≥ 7.0 mmol/L or 2-hour glucose ≥ 7.8 mmol/L (n = 42,656). The 2008 CDA criteria used fasting glucose ≥ 5.3 mmol/L, 1-hour ≥ 10.6 mmol/L, and 2-hour ≥ 8.9 mmol/L (n = 270,843). The ADIPS cut-offs used included fasting glucose ≥ 5.5 mmol/L and 2-hour ≥ 8.0 mmol/L (n = 32,013)"

“Finally, the small number of studies using non-IADPSG criteria, most of which employed cut- offs relatively close to those recommended by IADPSG, limits the strength of the comparison between IADPSG and non-IADPSG criteria, as the non-IADPSG group may not fully represent the diversity of diagnostic approaches in use.”

Comment 4*: A minor comment regarding “Characteristics of included studies” page 10, line 275–*
284. Numbers given for 31 included studies, Asia (18), Europe (4) and Australia (2), North America (1) are not in line with the numbers given I Table 1, referring to 35 included studies, Asia (19), Europe (5) and Australia (3), North America (2). What is the explanation for this discrepancy?

Response 4: Thank you for pointing out this discrepancy. The explanation is that some studies contain two independent populations that were analyzed separately in the meta-analysis. Specifically, four studies each have two distinct populations, which are labeled in Table 1 as "Author, Year A" for the first population and "Author, Year B" for the second population. While the total number of included populations in the meta-analysis is 35, the total number of studies is
31. We have now clarified this in the text to avoid confusion.

VERSION 2 - REVIEW

Reviewer	**2**
Name	**Berntorp, Kerstin**
Affiliation	**Lund University**
Date	**26-Oct-2024**
COI	

I am pleased with the authors response to my comments and their revision of the text accordingly. I have no further comments.