Reviewer's report

Title: No consensus on gestational diabetes mellitus screening regimes in Sweden. Pregnancy outcomes in relation to different screening regimes 2011 to 2012: a cross-sectional study

Version: 1 Date: 29 November 2013

Reviewer: Sara Meltzer

Reviewer's report:

Summary:

The authors have used a well-established data base representing a large proportion (apparently 90%) of pregnant women in Sweden to explore the diagnostic criteria used for GDM in Sweden as well as the outcomes based on the presence or absence of GDM. They describe 43 Maternal Health Care Areas where 4 different screening and diagnostic practices are used varying from universal screening with a 2 hour cut-off of 10.0 (based on what criteria for these values?) to various levels of risk factor screening followed by a 75g diagnostic load and again, various values of 8.9, 10.0, and 12.2 as the diagnostic point where treatment would occur. They then use a population attributable risk for analysis, except they do not seem to have a ‘gold standard’ of unaffected women, since the criteria for diagnosis varies so considerably. They do mention that the incidence of GDM is low in Sweden, but the diagnostic criteria are very variable. What is the incidence of DM in Sweden (presumably using the same diagnostic criteria as much of the world now uses for the 75g OGTT or criteria for diagnosis)? Has this value remained unchanged in Sweden over the last 30 years? In much of the world, it has not and almost invariably, the rate of GDM mirrors the rate of DM in the population.

The authors discuss the risks increases for adverse outcomes and show that they were not dissimilar to the HAPO values. They do point out that the country of origin and socio-economic factors justify being included in the selective screening risk criteria – in Sweden this apparently represents 58.4% of women. Only 9.9% of women underwent universal screening. They discuss the fact that elevated random plasma glucose between 8 – 9 mmol/L was the most prevalent indicator for OGTT… but do not indicate how or when this would be done in the population.

Their objectives are broad and appear to be un-directed. They have shown that in Sweden, when GDM is diagnosed, it is associated with poorer outcomes. In their conclusion, they indicate some risk factors have been very strongly associated with GDM development. Unfortunately, they have not really
addressed the elephant in the room. Due to NO consistency in screening and diagnostic criteria, the importance of the findings is difficult to assess. Their findings on some of the screening methods in use might be enough to suggest they be dropped, but they do not go that far.

1. Is the question posed by the authors well defined?
   Yes and no… the authors address a number of issues but do not seem to have a clear question which they are attempting to answer. They assess the diagnostic criteria in use in Sweden and discuss the outcomes of the pregnancy in relation to the screening / diagnostic methods used.

2. Are the methods appropriate and well described?
   What is the background population prevalence of DM, not GDM in Sweden? Why was a random glucose the most frequent reason for an OGTT when no discussion of when or how it would be done is mentioned with the statement?

3. Are the data sound?
   The data appears to be from a very well-planned data base, although no discussion of potential for errors is made (e.g. is ht and wt always indicated, so do all women have a BMI even if self-reported) It would be nice if additional evaluation of the data would be possible to increase the value of the paper to a more general audience and not simply Sweden.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?
   It appears to.

5. Are the discussion and conclusions well balanced and adequately supported by the data?
   Please see comments below. I believe that more could be obtained by re-assessing some of the data slightly differently. In addition, possibly because the authors submitted this paper before the WHO came out with new recommendations which incorporate the IADPSG values but leave the methods for using them up to the individual countries (selective screen based on risk factors, glucose challenge test or FPG as initial screen universally, or universal testing of all women, the WHO recommendations are not addressed and likely should be.

6. Are limitations of the work clearly stated?
   I believe that they should discuss this more in depth – as mentioned below.

7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?
   It appears so.

8. Do the title and abstract accurately convey what has been found?
   Unfortunately, I think it does, however I wonder if they have not got valuable data to argue not simply for adding obesity, ethnicity and socio-economic factors to the considerations to make for selective screening, but they could argue that the
very high values carry a significant increase in outcome complications and likely should be changed to a different method.

9. Is the writing acceptable?
Overall, the writing is well-done with minimal grammatical errors or phrasing difficulties. The discussion could be more structured or organized to save space and to clarify the importance of their findings.

Particular areas where the manuscript could be strengthened:

1. Table 1 is very data heavy and difficult to follow – separated by 2011 vs 2012 cohort and statistical difference mentioned… why? Can the categories be moved together and fit all on one page with only the key variables of interest included (eg. only one smoking category, fewer age categories, either employed or on leave etc.) What does the p-value compare to for the OGTT group? What does the p-value refer to for the GDM group, especially since the sites had such variability? Table 2 adds to the understanding of the findings of the article and would be useful in the article but Table 1 likely should only be an on-line appendix or pared down considerably.

2. Would it be possible for them to identify a subpopulation of this database with NO risk factors and use those values to develop the baseline for population attributable risk, since there is no way to know if women un-screened definitely had not risk. If they did this already, it is not stated.

3. In table 4, could they not use already identifiable cut-offs for evaluation of the outcomes? Eg. under 7.8 (WHO 1999 values), 7.9 – 8.5 (old WHO to new IADPSG and WHO recommended criteria), 8.6 – 10.0 (values pertinent to Sweden historically) and greater than 12.2 (also pertinent to Swedish population historically).

4. In Table 6, where the odds ratios are developed and adjusted, did they adjust for the diagnostic criteria method used ie. For the centre were the data came from?

It seems to me that re-assessing the data, especially table 4 in a different manner would help to put the outcomes found in Sweden into a perspective which could be useful for other parts of the world. It would be helpful for the Swedish database and overall assessment of needs for care in Sweden, if some consistent values could be used. I am unaware of a Swedish guideline for GDM but no mention is made of it.

It seems plausible that the recommendation from this energetic and extensive study might be used to argue for a consistency in screening and diagnostic approach. Perhaps, based on their data, the use of 12.2mmol/L might be considered to be associated with too high risk to be justifiable at least? To have done all this work with the only hope that obesity be taken more seriously and ethnicity be considered may be a necessary step in Sweden, but seems to be a bar set too low. Re-calculating the outcomes with different cut-offs which fit
international criteria would possibly lead to a slightly different set of conclusions and may potentially justify one presently recommended diagnostic criteria over another for the Swedish population.

Of note, the prevalence of LGA is increased with higher values used for diagnosis but not so much in the immigrant populations. In Table 3, the GDM group had an increased incidence of SGA relative to the baseline – due to higher levels of obesity leading in some cases to poorer placental flow or due to an ethnic population picked up with probable type 2 DM? It would be nice if they could discuss this a bit.