Reviewer’s report

Title: Reasons for participation and non-participation in a diabetes prevention trial among women with prior gestational diabetes mellitus (GDM)

Version: 1 Date: 17 October 2013

Reviewer: Matthew White

Reviewer’s report:

The authors are currently conducting a randomized controlled trial in a group of women who previously had gestational diabetes mellitus in order to compare a lifestyle intervention program vs. standard of care in the prevention of progression to type II diabetes in this population. The purpose of this manuscript is to describe differences between subjects who consented and who did not consent to the trial as well as to describe barriers to participation among non-consenters.

Major Compulsory Revisions

1) Methods, paragraph 1: IRB approval was given for the trial, but did non-consenters explicitly grant permission for their data to be used for research? It is not clear to me that this is the case.
2) Methods, paragraph 1: Why were so many consenters (46/89) and non-consenters (83/156) excluded from the analysis? Did you exclude a subject if they did not have complete data? If there is just intermittent missingness, consider using multiple imputation.
3) Background, paragraph 4: Did you attempt to evaluate if consenters’ and non-consenters’ perceptions of their personal risk of developing type II diabetes?
4) Methods, paragraph 3: Use Fisher’s exact test with categorical variables. This is especially needed when there are small cell counts (e.g. there is a 0 in the table for fruit and vegetable intake).
5) Methods, paragraph 3: Did you evaluate the distributions of the continuous variables in each group? If non-normal, transform the variable so that it is more normally distributed or use the Wilcoxon rank-sum test (non-parametric t-test). If normally distributed, use the Satterthwaite t-test which allows non-equal variances between the two groups.
6) Methods, paragraph 4: The classification tree analysis seems a bit excessive when also considering that you used logistic regression as well. How robust are the results from the classification analysis? I recommend performing a bootstrap analysis to see if the model you obtained holds up. In what fraction of the bootstrap samples were the same two covariates selected? In what fraction was the same age cutpoint is selected?
7) Methods, paragraph 4: How does LASSO compare to a forward model-selection procedure? Were the “meets daily fruit and vegetable intake”
and “IGT at follow-up” variables included in the logistic regression analysis? I do not think they should be due to the small cell counts, and this should be explained in the analysis.

8) Results, paragraph 5: Interpret the odds ratios from the logistic regression model. Also, please report the c-statistic from the logistic regression model.

9) Results, barriers to participation subsection: This description is quite long and could be shortened dramatically if in the “Description” column of Table 3 you include the number of subjects with each response in parentheses.

10) Discussion, paragraph 4: I do not like that you “speculate” here without this being explored at all. Was any data collected or analyzed that could speak to this point? If not, you should remove this speculation.

11) Discussion, paragraph 6: Since all subjects are female in this study, remove “female” from the patient-related factors. Also, were psychological barriers assessed? Please define “illness beliefs”.

12) Figures 2 and 3: I do not think these add to the paper and should be removed.

13) Additional table: I think it would be nice to present the logistic regression results in a table. On the left hand side, present the OR, 95% CI and p-value for each covariate from a univariate regression and then on the right present the OR, 95% CI and p-value for each covariate in the final multivariable model.

Minor Essential Revisions

1) Background, paragraph 8: I would make this the last paragraph of the Background section.

2) Methods, paragraph 2: Remove “This is a binomial response.” It’s a binary response, and I think this is understood from the definition of the primary response variable. Also, unless age at delivery is a risk factor for type II diabetes, I would move things around: “The explanatory variables included age at delivery as well as known type 2…”

3) Discussion paragraph 3. You have 36 listed as the age cutpoint, but I believe your analysis identified 34 as the cutpoint, correct?

4) Discussion, paragraph 5: 410 subjects were invited to participate in the trial, so there was not “a majority” that did not consider the intervention accessible. The majority would be for those who responded to the initial invitation.

5) Discussion, paragraph 11: The first sentence is not clear.

6) Endnote 1: The first sentence defines type I error while the second sentence describes a cause of type II error.

7) Table 1: Under “All” use “N=116” instead of “n=116”.

8) Table 1: Make the percents out of the column n and not the total N.

9) Table 2: If you find that a variable is not normally distributed and use the non-parametric test, report the median and interquartile range (Q1-Q3) for the variable instead of the mean and standard deviation.
Level of interest: An article of limited interest

Quality of written English: Needs some language corrections before being published

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests.