Author's response to reviews

Title: Spousal diabetes as a diabetes risk factor: A systematic review and meta-analysis

Authors:

Aaron Leong (aaron.leong@mail.mcgill.ca)
Elham Rahme (elham.rahme@mcgill.ca)
Kaberi Dasgupta (kaberi.dasgupta@mcgill.ca)

Version: 3 Date: 5 November 2013

Author's response to reviews: see over
Dear Lin,

We are very grateful for the helpful comments of the Reviewers with respect to our manuscript 1912818447107451 - *Spousal diabetes as a diabetes risk factor: A systematic review and meta-analysis*. Please find below our point-by-point responses to their comments, along with excerpts of revised text. Please note that we have uploaded the manuscript with highlighted changes as the main file and the unmarked (clean) version as an additional file. We have done the same for appendices 2 and 3 that were also changed. Many thanks and we look forward to hearing from you.

Kaberi

Kaberi Dasgupta, MD, MSc, FRCPC
Associate Professor of Medicine, McGill University

**Reviewer: Solveig Cunningham**

This report presents a systematic review and meta-analysis to understand the clustering of diabetes within couples. The study was carefully designed and described. The limitations of the study come largely from the limitations of the literature it is summarizing.

1. The point system is quite arbitrary and it’s unclear to me that this is useful. For example, a study controlling for age and SES has equal points as a study that used GTT. Is this reasonable? Hard to tell. May be better to focus on assessment on multiple dimensions, eg. 1) satisfactory control variables 2) satisfactory diabetes assessment, etc.

We agree that it is very difficult to reduce study quality into a straightforward scoring system. As the Reviewer suggests, when comparing two given studies, both are likely to have their own strengths and weaknesses, and whether one’s strength is equivalent to another’s is a matter of reasonable debate. Indeed, summary scores can be potentially misleading as readers may simply focus on the scores when evaluating study quality. We do believe however that systematic glucose testing is a study strength because it ensures that all spouses had an equal opportunity to be detected to have diabetes. This is a potentially important measurement issue. On the other hand, socioeconomic status reflects an agglomeration of the challenges and opportunities that
one may have in life which may impact diabetes incidence and thus is also a potentially-important consideration.

In the revised manuscript, we have removed the summary scores and have reverted to the original system of awarding stars (*) that was originally proposed in the Newcastle-Ottawa quality assessment scale. To reflect the importance of systematic glucose testing, we have modified the scale and awarded one additional star (*) if blood glucose testing was used to ascertain exposure under the section ‘Selection’ for question number 3 (i.e., ascertainment of exposure). Similarly, we awarded an additional star (*) if blood glucose testing was used to assess outcome, under the section ‘Outcome’ for question 1 (i.e., ascertainment of outcome). Further, we have detailed similarities and differences across studies in our text and conclude that all retained studies are of reasonable quality and were thus used in the pooled estimate. We have modified the Results (Quality Assessment, Paragraph 1) section of the revised manuscript to reflect the question that Reviewer 1 has raised as follows:

Two key study strengths were identified. The first was performing systematic glucose testing on all participants as it ensured that all spouses had an equal opportunity to be detected to have diabetes. All participants underwent oral glucose tolerance testing in the study by Khan and colleagues [38] while, in the study from Kim and colleagues [37], fasting glucose measurements were used to detect diabetes. The second important study strength was the ability to capture diabetes incidence over time. The longitudinal cohort study by Hemminki and colleagues [9] followed 157,549 subjects for an average of 14.8 years (157,549) and was thus able to assess the impact of spousal diabetes on incident diabetes (Appendix 3).

Appended is Additional File 3 (Appendix 3):

<table>
<thead>
<tr>
<th>Questions</th>
<th>Selection</th>
<th>Comparability</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Maximum no. of stars (*)</td>
<td>*</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>Stimpson, 2005</td>
<td>a</td>
<td>*</td>
<td>a</td>
</tr>
<tr>
<td>Jurj, 2006</td>
<td>a</td>
<td>*</td>
<td>a</td>
</tr>
<tr>
<td>Kim, 2006</td>
<td>a</td>
<td>*</td>
<td>a</td>
</tr>
<tr>
<td>Khan, 2003</td>
<td>b</td>
<td>*</td>
<td>a</td>
</tr>
<tr>
<td>Hippisley-Cox, 2002</td>
<td>a</td>
<td>*</td>
<td>a</td>
</tr>
<tr>
<td>Hemminki, 2010</td>
<td>c</td>
<td>-</td>
<td>a</td>
</tr>
</tbody>
</table>
Letters a through d correspond to question responses on the quality assessment scale (Modified Newcastle-Ottawa quality assessment scale for nonrandomized observational studies). For Question 3 of “Selection”, we awarded an additional star (*) if blood glucose testing was performed on all participants to ascertain exposure. For Question 1 of “Outcome”, we awarded an additional star (*) if blood glucose testing was performed on all participants to assess outcome. For “Comparability”, the first star (*) was awarded if effect estimates were adjusted for age. The second star (*) was awarded if effect estimates were adjusted for markers of socioeconomic status.

2. More clarity is needed on the exclusion of studies with only ethnic groups. Provide explanations for why we could expect the associations to be different between groups.

We would like to emphasize that we did not exclude studies with particular ethnocultural groups in the main analysis. We apologize if this was not clear. As indicated under Methods, Study Selection, Paragraph 1, our criteria were as follows:

We used the following inclusion criteria: 1) cross-sectional, case-control or cohort design; 2) study population with married couples selected from public health records, or administrative, hospital or clinic databases; 3) outcomes were diabetes and/or prediabetes, defined as impaired fasting glucose (IFG) or impaired glucose tolerance (IGT)[25]; 4) effect measures reported as adjusted or unadjusted odds ratios (OR), risk ratios, hazard ratios or rate ratios. We excluded studies that did not specifically address spousal concordance and those that reported simple linear correlations of metabolic syndrome criteria only. We excluded studies that examined between-spouse correlations for absolute glucose levels rather than diabetes given that there is high intra-individual variability within both the abnormal and normal absolute glucose range[26, 27]. In contrast, a diabetes diagnosis generally requires a clinical assessment that includes more than one glucose measurement and/or glycated hemoglobin or glucose tolerance testing[7, 28].

In our original Data Synthesis and Analysis, we explored the effects of excluding studies that focused on specific ethnocultural groups (i.e., homogenous Asian population) in a sensitivity analysis. Our rationale for this investigation was to explore whether differences in ethnocultural background contributed to the heterogeneity observed across studies as norms of interaction between spouses (e.g., shared meals, shared decision-making, shared perspectives), which may potentially differ across groups and thus potentially modify the impact of a spousal diabetes history on diabetes risk. However, as highlighted by Reviewer 2, performing subgroup comparisons on a small number of
studies could result in chance findings. Moreover, our subgroup analysis on non-Asian studies [1.31 (95%CI 1.27, 1.36)] demonstrated that while the point estimate was slightly higher, the confidence intervals greatly overlapped with the overall effect estimate [1.26 (95%CI 1.08, 1.45)]. Thus, in the revised manuscript, we have elected to remove the subgroup analysis that stratifies on ethnocultural groups. We have added the underlined section noted below in the Discussion section (Strengths and Limitations, Paragraph 2):

Given the small number of studies, we were unable to perform meta-regression or subgroup analyses to describe the effect of other study characteristics on outcome measures or statistically explore the possibility of publication bias[55]. Results from individual studies should also be interpreted with caution as differences observed may be merely chance findings [56]; for example, although studies differed in ethnocultural composition, there were not sufficient numbers of studies within individual ethnocultural groups for conclusions about any ethnocultural variations in spousal concordance.

3. Provide a bit more interpretation of some of the considerations in shared risks – what variables should future studies consider that may increase the shared risk? Along the same lines, explain the implications of duration of partnership and age – how are these factors likely to shape shared risks?

The ideal study would include systematic assessments of health behaviours, sociodemographic profiles, living arrangements and glucose handling (oral glucose tolerance testing) of the spouses of individuals with or without type 2 diabetes. Couples would be recruited from similar geographic areas at similar points in time and studied longitudinally with periodic assessments of glucose tolerance. Such an observational study would permit investigators to definitively link shared habits with shared diabetes risk and better clarify the importance of assortative mating. Longitudinal studies would ideally be powered to determine whether longer duration of partnership increases both shared behaviours and diabetes risk, whether the age of marriage modifies these associations, and if concordance persist in couples who have separated.

We have added the following paragraph to the Discussion at the end of the section Strengths and Limitations:

Stronger evidence for causal associations, mediated through shared behavioral habits within families, could be derived from longitudinal studies assessing the concordance of incident rather than prevalent disease. These cohort studies could include systematic assessments of health behaviours, sociodemographic profiles, living arrangements, and glucose handling (oral glucose tolerance testing) of the spouses of
individuals with or without type 2 diabetes. Couples would be recruited from similar geographic areas at similar points in time and studied longitudinally with periodic assessments of glucose tolerance. Such an investigation could definitively demonstrate a link between shared habits with shared diabetes risk and better clarify the importance of assortative mating. Longitudinal studies can also be powered to determine whether longer duration of partnership increases both shared behaviours and diabetes risk, whether the age of marriage modifies these associations and if concordance persist in couples who have separated.

4. Be sure to distinguish between risk of bias and risk of noise – it is bias that we should worry about more and it’s important to distinguish the two.

We agree that we should clearly distinguish between factors that may lead to bias towards an inflated estimate of association versus factors that may lead to bias towards the null (i.e., ‘noise.’)

Under Discussion, Strengths and Limitations, Paragraph 3, we have included the following:

*Individual studies may have potential limitations that impact the accuracy of our findings. For example, determination of diabetes or prediabetes status was more rigorous in some studies than others. Only two studies performed systematic glucose testing on all participants [37, 38]. Another study likely captured only more advanced diabetes cases as its diabetes definition required a hospital discharge diagnosis [9]; while the probable under detection is expected to be similar for individuals with or without a spousal diabetes history, it potentially reduces power to detect spousal associations or bias effects towards the null, although this may not have been a major concern given the large sample size.

Conversely, spouses of diabetes patients could have greater understanding of diabetes and seek medical assistance in the event of relevant symptoms. Similarly, physicians may enforce greater surveillance for these spouses; this detection bias could inflate estimates of association.

5. Were weights applied so that the big study (66,130) did not dominate the results?

The larger study did to some degree dominate the results precisely because weights were applied, as is standard for meta-analyses. Nevertheless, studies of all sample sizes were considered important in our qualitative synthesis and quantitative synthesis where we systematically compared the strengths, weaknesses and reported results of all included studies. We ensured that larger studies did not dominate the overall results in our conclusions. For example, we highlighted specific studies (i.e., that by Khan and
colleagues) that, while having a smaller sample size, had performed systematic glucose testing on all participants and thus identified spouses of diabetic participants with undiagnosed diabetes.

Reviewer: Ross Harris

Reviewer's report:
This study presents a systematic review of spousal diabetes as a risk factor. It makes for interesting reading, and the authors have conducted the study carefully. The conclusions are well-thought out and interesting, although I have some concerns that some of the final conclusions have little evidence basis. In general though, this is a very interesting paper and for the most part needs only minor improvements.

Major Compulsory Revision:

1. In the discussion, the authors mention that the true effect is likely to be higher than their pooled estimate, and more akin to the highest estimate of all studies. I think this is a dangerous conclusion: why is this study likely to be the most representative? The only cohort study available, which is well-recognised as being a preferable study design, could as well be taken as the “best estimate”. This study did not adjust for BMI, but given that adjusting for BMI would be more likely to attenuate effect estimates, the more modest association in Hemminki makes me wonder whether the effect is really as strong as the authors speculate.

We agree that we should be more balanced in our inference that the Khan and colleagues’ study provides the most accurate assessment. Our reasons for emphasizing this study relates to potential for misclassification. Diabetes may be present but remain undiagnosed if an individual does not undergo a clinical evaluation. Such individuals could report themselves as non-diabetic in surveys. Further, conventional screening has traditionally relied on a fasting glucose value, previously judged to be the most cost efficient approach. However, it is known that some individuals with diabetes may have a normal fasting glucose but have difficulty handling a glucose load, with an abnormal rise in glucose values after ingesting a glucose ‘drink.’ The oral glucose tolerance test permits detection of such individuals. The study with the highest estimate by Khan and colleagues performed systematic glucose tolerance testing on all participants, and was therefore able to detect undiagnosed diabetes in the spouses of diabetic patients.
On the other hand, cohort study by Hemminki and colleagues does have the preferred study design to not only make causal inferences but also determine the incidence risk increase reflecting shared risk within couples. This study, however, did not perform systematic glucose testing on all participants and was therefore unable to account for undiagnosed diabetes. We have reworked our Discussion to temper our conclusions, as suggested by Reviewer 2. The first paragraph of the Discussion now reads as follows:

Our analyses demonstrate spousal diabetes concordance. The degree of concordance estimated was lowest in a study that relied on women’s reports of diabetes in themselves and their spouses (effect estimate 1.1, 95% CI 1.0 to 1.30)[20] and highest in a study with systematic assessment of glucose tolerance (2.11, 95% CI 1.74 to 5.10)[38]. The random effects pooled estimate suggests that a spousal history of diabetes is associated with a 26% risk increase for diabetes overall without adjustments for BMI (effect estimate 1.26, 95%CI 1.08, 1.45) and 18% with BMI adjustment (effect estimate 1.18, 95%CI 0.97, 1.40). This effect size is similar to the incidence risk increase of approximately 30% attributed to spousal diabetes that is reported by the single longitudinal cohort study [Standardized incidence ratios 1.31 (95%CI 1.26-1.35) for men; 1.33 (95%CI 1.29-1.38) for women][9].

2. Further, the authors are, in a way, discounting the results of their own meta-analysis – if they felt that the studies were not comparable, they should not have combined them in a pooled estimate.

While the studies were different in study design, study populations and case ascertainment methods, we felt they were comparable enough to be pooled bearing in mind that there was some heterogeneity. We thus elected to use a random-effects model that took between-study and within-study variability into account in generating overall effect estimates. The point is well-taken, however, and we have modified our Abstract, Discussion, and Conclusion and, as recommended by Reviewer 2 (Please also see response to Reviewer 2, Comment 1).

The concluding statement in the revised Abstract is as follows:

Our pooled estimate suggests that a spousal history of diabetes is associated with a 26% diabetes risk increase. Recognizing shared risk between spouses may improve diabetes detection and motivate couples to increase collaborative efforts to optimize eating and physical activity habits.

The first line of our Conclusions in the main body of the text is now:
In summary, spousal diabetes history confers an increased risk for diabetes that our pooled estimate suggests is 26%.

3. The study by Khan is given the highest quality score, but this is based on a somewhat subjective method, which the authors have further modified. The confidence interval for the study by Khan is also very wide, indicating substantial uncertainty in their results. I therefore struggle to see the reasoning behind this.

We agree that there is uncertainty in the reported results by Khan and colleagues as the study had a relatively small sample size; hence, effect estimates (i.e., odds ratios, incidence rate ratios, etc.) have wide confidence intervals. Studies of all sample sizes were considered important in our qualitative synthesis and quantitative synthesis where we systematically compared the strengths, weaknesses and reported results of all included studies. We have also modified the quality assessment and removed summary scores to provide a more balanced perspective on study quality. Please see response to Reviewer 1, Comment 1, for details.

4. The authors state that this may be due to the “systematic glucose testing” employed by the study – if this is the case, then further explanation of how not using this method would weaken the association should be presented.

Without systematic glucose testing on all subjects, the study would be unable to detect undiagnosed diabetes in spouses of diabetic individuals and miss cases when both partners have undiagnosed diabetes. This under detection likely weakens the ability to detect spousal concordance of diabetes. Please see our response to Reviewer 1, Comment 4, for details.

5. The authors state “It may be easier to capture concordance for diabetes/prediabetes than for diabetes alone because diabetes/prediabetes has a higher prevalence” – but odds ratios or other measures of association should not depend on background prevalence, so the reasoning here does not make sense to me. In any case, the authors should keep in mind the wide confidence intervals of the Kahn study, and the conclusions should be more cautious.

We agree that background prevalence should not affect measures of association. However, smaller numbers of cases could weaken the ability to detect an association. We
have clarified this point in the discussion section of the revised manuscript as follows (Discussion, second paragraph):

The between-spouse association was higher for the broader definition of ‘dysglycemia’ that encompassed pre-diabetes (IGT, IFG) and diabetes in the two studies that examined this issue, with an approximately two-fold risk increase for dysglycemia with spousal dysglycemia history (OR 1.92, 95%CI 1.55, 2.37 by Kim and colleagues [37]; OR 2.32, 95%CI 1.87, 3.98 by Khan and colleagues [38]). This broader definition potentially improves the power to detect spousal associations.

Minor Essential Revisions:

6. There is not much evidence to go on here: although a pooled estimate may of course be derived, performing subgroup analyses is going to be somewhat akin to randomly selecting subsets of studies. The authors should be careful to acknowledge the limitations in comparing small subsets of studies, or even the results of single studies, and the danger of chance findings (see, e.g., Higgins et al, Controlling the risk of spurious findings from meta regression, Statsmed 2004)

Thank you for highlighting this important point. We agree that subgroup analyses could have chance findings which are potentially misleading, and have now removed the subgroup comparisons between ethnocultural groups in the revised manuscript. Please see responses to Reviewer 1, Comment 2. We now acknowledge this limitation in our discussion section and referenced the Higgins and colleagues’ paper, as Reviewer 2 has suggested:

Given the small number of studies, we were unable to perform meta-regression or subgroup analyses to describe the effect of other study characteristics on outcome measures or statistically explore the possibility of publication bias[55]. Results from individual studies should also be interpreted with caution as differences observed may be merely chance findings [56]; for example, although studies differed in ethnocultural composition, there were not sufficient numbers of studies within individual ethnocultural groups for conclusions about any ethnocultural variations in spousal concordance.

7. The authors acknowledge that there are too few studies for meta-regression, which also precludes any other valid subgroup comparisons, and therefore need to be cautious in their conclusions about different ethnic groups, etc. Further, the authors should acknowledge the small number of studies, and heterogeneity in their results, in the strengths and limitations section. With
the evidence available there is no possibility for exploring differences between studies.

We agree that we need to be cautious in our conclusions on ethnic differences as they have only arisen from subgroup analyses of a small subset of studies. Replication of these findings are required before firm conclusions can be made. We have now removed this subgroup comparison from the revised manuscript and added the following:

Some of the heterogeneity observed in the meta-analysis could be attributed to differing ethnocultural composition of study populations, diabetes/prediabetes ascertainment methods, study design, reference groups and characteristics of participants used to adjust effect estimates. Unmeasured confounders/mediating variables such as dietary information, physical activity level, marriage duration and time of diagnosis were not uniformly obtained across all included studies. Therefore, in pooling effect estimates, we generated random-effects models that accounted for between-study and within-study variability. Given the small number of studies, we were unable to perform meta-regression or subgroup analyses to describe the effect of other study characteristics on outcome measures or statistically explore the possibility of publication bias[55]. Results from individual studies should also be interpreted with caution as differences observed may be merely chance findings.

8. Abstract – not clear what the “effect” is – if a mix of OR/IRR could just say “relative risks”

We agree that the term ‘effect’ in the Abstract is not clear. We have modified as follows:

“Effect estimates (i.e., odds ratios, incidence rate ratios, etc) with Body Mass Index (BMI) adjustment were pooled separately were pooled separately from those without BMI adjustment (random effects models) to distinguish BMI-dependent and independent concordance”.

Further, under Methods, Data Synthesis and Analysis, we have modified the first line as follows (additional words underlined):

“All data analyses were performed using STATA (version 11 StataCorp, Texas, USA). We extracted the reported effect estimates (i.e., odds ratios, incidence rate ratios, et cetera) and 95% confidence intervals from each study to generate forest plots and visually inspected for heterogeneity across studies”.

Royal Victoria Hospital, V Pavilion (V1.08), 687 Pine Avenue West, Montréal, QC H3A 1A1
Tel: 514-934-1934, ext. 44715   Fax: 514-934-8293   Email: Kaberi.Dasgupta@mcgill.ca
9. P12, meta-analysis “There was suggestion of heterogeneity (Higgin’s I-squared statistic = 65.4%, P-value = 0.03)” – I would say “some evidence” of heterogeneity, given the low p-value.

We have changed the phrasing of this sentence as Reviewer 2 recommends.

10. Table 1: marriage duration in Kim 2006 says “530” – not clear what measure this is, or what the “(16.9)” is (standard deviation?). Also, the second half of the table appears to be pre-diabetes, but this is not clear.

We apologize for this typo. The study by Kim and colleagues did not report information on marriage duration. The number of participants with prediabetes (Fasting glucose above or equal to 6 mmol/l) was 530 (16.9%) of the study sample. We have corrected this error in the revised Table 1.

Discretionary Revisions:

11. The Newcastle-Ottowa scale has been used to evaluate risk of bias. Although this provides an easy-to-use measure of study quality, its limitations should be considered by the authors: in general, Cochrane guidelines do not recommend the use of summary scores, particularly for analytical purposes, and generally prefer the use of checklists and so on. It must also be borne in mind that the reference for the NOS is not from a peer-reviewed publication, and aspects of it have been criticised (see, e.g., Stang A, European Journal of Epidemiology 2010). However, the authors have given good descriptions of the individual studies and their limitations; therefore I have no particular criticism apart from whether it has led to the specific part of the conclusion that I am concerned about above.

Thank you for raising this valid point. Indeed the Newcastle-Ottowa scale has limitations. A comprehensive qualitative and quantitative assessment of each included study is most appropriate to judge study quality and assess bias. In our revised quality assessment, we have removed the summary scores and focused on a more qualitative evaluation of the included studies. Please refer to our response to Comment 1 by Reviewer 1 for further details.