Author’s response to reviews

Title: Diabetes screening intervals based on risk stratification

Authors:

Sachiko Ohde (saohde@luke.ac.jp)
Emily McFadden (ecm33@cantab.net)
Gautam Deshpande (drdeshpande@gmail.com)
Hiroshi Yokomichi (hyokomichi@yamanashi.ac.jp)
Osamu Takahashi (otakahas@gmail.com)
Tsuguya Fukui (fktts@luke.ac.jp)
Rafael Perera (rafael.perera@phc.ox.ac.uk)
Zentaro Yamagata (zenymgt@yamanashi.ac.jp)

Version: 1 Date: 27 Jul 2016

Reviewer #1: Sachiko Ohde and colleagues assessed appropriate HbA1c screening intervals in presence of T2D risk stratification. They employed retrospective design in two urban and rural setting to obtain repeated measurements data of HbA1c in individuals who attended for lab measurements. The paper is well written and is about a very important issue. However, there are some major considerations that need further elaboration as well as some minor comments:

1. The authors state "screening intervals may differ substantially when considering individual risk for diabetes" and seemingly used BMI as a surrogate for diabetes risk for stratification. However, they used international BMI cutoffs that were mostly defined in Caucasian populations. The cited reference for this classification recommended different definition of at risk populations for public health action that may have impact on screening intervals as well.

Thank you for your comment. According to the report by the Japan Society for the Study of Obesity (JASSO), BMI ≥25 is the suggested definition for obesity in not only Japanese, but most of the peoples of Asia-Oceania. In both women and men with BMI ≥25, a steep increase of incidence of obesity-associated chronic diseases (lifestyle-related diseases) including diabetes, hypertension and hyperlipidemia has been shown. We added this reference, “Kanazawa M,

2. The use of Framingham risk score as another stratification factor is a good idea, but mostly as a secondary aim, but needs to be justified in the text as it predicts CVD risk not diabetes risk which itself is a component of this risk score. To fulfil the aim of this study, use of risk scores that were developed to directly predict T2D risk should be a priority and may add more insight through employing more precise estimate of T2D risk compared to BMI alone.

Thank you for your thoughtful comment. We utilized the Framingham Score for several reasons. First, we considered a CVD-related tool as a particularly good stratifier since the end stage of T2D is typically CVD, and arguably its most frequent and dreaded complication. Second, many clinicians now consider T2D as a CVD “risk equivalent.” While we considered the Cambridge diabetes risk and Leicester practice scores as suggested in NICE guidelines, these score were, perhaps unfortunately, seldomly referred to in the literature and not frequently used in the clinical settings familiar to our clinicians, when compared to Framingham or ASCVD scoring tools. As there are multiple studies which indicate that diabetes and CVD are risk-equivalent, we have added the following references to the manuscript:


For clarification in the manuscript text, we added the following sentence: “We chose BMI as a basis of stratification as previous reports show different incidence rate of DM based on BMI classification; we chose FRS as for stratification to reflect that DM is considered a clinical CVD-risk equivalent” on Page 2, Line 14.
3. Page 2, line 14: A verb is missed; please revise the sentence as follow: "age (Range) and mean HbA1c (SD) was 48 (30-74) years old and 5.4 (0.4) %, respectively"

Thank you for catching this error. We have corrected this as suggested.

4. The statistical methods generally deserve further elaboration. For example, how "minimal informative screening interval" is determined based on signal to-noise ratio exceeding one?

Thank you for the valuable comment. In brief, we have added additional information in Figure 2 footnotes, including some useful equations in calculating both signal and noise, as well as more narrative explanation in the manuscript (Page 4, Line 15).

To elaborate a bit further, the Signal-to-Noise Ratio method is a relatively new way of assessing test characteristics, providing valuable supplemental information to the better-known sensitivity and specificity test characteristics. Because it is quite a new methodology, there is an ongoing discussion whether or not a ratio >1 can be a new “gold standard” for evaluating test efficacy. However, in line with previous studies by Perera R et al., we believe that when noise exceeds signal, there is a high probability that HbA1c may only be indicating measurement error, rather than reflecting true HbA1c change. When screening is performed at intervals which fail to account for measurement error, clinicians run the risk of labeling and treating disease when treatment is unwarranted (as well as de-escalating potentially appropriate regimens). The definition of minimal informative screening intervals are particularly well-described in “Perera R, McFadden E, McLellan J, Lung T, Clarke P, Perez T, Fanshawe T, Dalton A, Farmer A, Glasziou P et al: Optimal strategies for monitoring lipid levels in patients at risk or with cardiovascular disease: a systematic review with statistical and cost-effectiveness modelling. Health technology assessment 2015, 19(100):1-402.”

5. According to authors descriptions, this method is supposed to distinguish true change from apparent change due to noise, but how about clinical significance of these true changes? If we can detect true changes in 2-year intervals, does it mean it is appropriate to detect a change in disease status?

Thank you for the interesting comment. To address this relevant question, we have added a table (Appendix 3) in the appendix in which we show true change (“signal”) per year, as well noise, among each group. We used variance for both signal and noise, and time at which signal exceeds
noise follows a quadratic curve. The table shows clearly that compare true change per year, which is quite small, HbA1c evidences quite a lot of noise—suggesting that clinicians ought to be wary about the very real possibility that they may be interpreting noise, rather than signal, when evaluating HbA1c values for their screening patients.

<table>
<thead>
<tr>
<th>Age Category</th>
<th>Stratification</th>
<th>HbA1c signal change per year (%)</th>
<th>HbA1c noise (%)</th>
<th>Time when Signal &gt; Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>30-44</td>
<td>BMI underweight</td>
<td>0.02</td>
<td>0.18</td>
<td>10.5</td>
</tr>
<tr>
<td></td>
<td>normal</td>
<td>0.03</td>
<td>0.18</td>
<td>6.7</td>
</tr>
<tr>
<td></td>
<td>overweight</td>
<td>0.05</td>
<td>0.22</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>obese</td>
<td>0.13</td>
<td>0.30</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>FRS 0-10%</td>
<td>0.03</td>
<td>0.19</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td>10-20%</td>
<td>0.11</td>
<td>0.23</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>over20%</td>
<td>0.10</td>
<td>0.20</td>
<td>2.0</td>
</tr>
<tr>
<td>45-59</td>
<td>BMI underweight</td>
<td>0.02</td>
<td>0.18</td>
<td>10.4</td>
</tr>
<tr>
<td></td>
<td>normal</td>
<td>0.03</td>
<td>0.18</td>
<td>6.3</td>
</tr>
<tr>
<td></td>
<td>overweight</td>
<td>0.05</td>
<td>0.22</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>obese</td>
<td>0.08</td>
<td>0.31</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>FRS 0-10%</td>
<td>0.03</td>
<td>0.18</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>10-20%</td>
<td>0.04</td>
<td>0.21</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>over20%</td>
<td>0.06</td>
<td>0.25</td>
<td>4.1</td>
</tr>
<tr>
<td>60-74</td>
<td>BMI underweight</td>
<td>0.03</td>
<td>0.17</td>
<td>6.6</td>
</tr>
</tbody>
</table>
normal 0.03 0.18 7.2
overweight 0.04 0.21 5.0
obese 0.06 0.23 3.9

FRS
0-10% 0.02 0.17 8.0
10-20% 0.03 0.19 6.7
over20% 0.04 0.21 5.3

*Signal defined as variance of random slope; noise defined as variance of residual between model-generated and observed HbA1c. Values represent square root of above measures to obtain HbA1c as %.

6. As this project aimed to define appropriate screening intervals, the performance of obtained intervals with regard to T2D diagnosis could be assessed too. For example, how many more new T2D cases would be diagnosed if these intervals are implemented compared to conventional recommendations? How sooner?

Thank you for your comment, which is both relevant to diabetes research and clinical practice. While your questions are both interesting and relevant, a comparison between a conventional screening strategy and a tailored screening interval is probably more appropriate for a future validation study of our work and, therefore, beyond the scope of the present study. Our group is currently planning a cost-effectiveness analysis comparing a conventional and risk-tailored strategy, which will include a statistical simulation addressing your interesting questions.

7. Authors adjusted their model describing HbA1c progression for gender, baseline age and BMI by year; first, what was the main independent variable in this model? Second, these adjusting factors appear indicators of HbA1c; but how they were selected and why other important factors like fasting plasma glucose were not included?

Thanks for your valuable questions. The independent variable in the random effect model is HbA1c. We used longitudinal data for the random effect model with random intercept and random slope to calculate signal and noise of HbA1c. Per stipulations of the Japanese Enforcement of Industrial Safety and Health Act, participants in our study were diagnosed with diabetes based on either HbA1c or fasting plasma glucose (FPG). We chose not to include fasting plasma glucose because of a high number of missing values; noncompliance with FPG
protocols continue to plague clinical practice and have been a driver of HbA1c adoption. In addition, we tried several combination of covariates in the model, confirming that results remained the same.

8. "Age at baseline" as an adjusting factor may mean at recruitment or at first visit, but it may be more rational to use age at first measurement of HbA1c in each period like what usually happens in practice. Please clarify which one was the case. For example, if there are 10 measurement points, age at point 9 should be the adjusting factor for changes between point 9 and 10 not age at point 1.

Age at baseline refers to at first visit, which was also at first measurement of HbA1c. This has been clarified in the manuscript in Page4.Line 10, “adjusted for gender, age and BMI as continuous values at first HbA1c measurement”.

9. Page 5, line 15: In the phrase: "Mean (SD) FRS was slightly higher in rural than urban cohort, at 6.0 (0.07)% and 9.0 (0.08)%, respectively"; please follow a consistent order as 6 is less than 9 but you have phrased it as higher. Moreover, if it is "mean score" and its corresponding SD, what is the meaning of % symbol here?

Thank you for your comment. We have corrected the manuscript text as suggested.

10. Page 6, line 3: Please change "patients" to participants.

Thank you for your comment. We have corrected the manuscript text as suggested.

11. Page 6, line 3: The term "similarly" appears wrong as there is substantial difference between 97.7% and 36.7.

We apologize for the confusion and have corrected the manuscript text as suggested, deleting “similarly.”

12. Page 6, line 3: "In those 60-74" should be revised to "In those aged 60-74 years"

We believe the reviewer is referring to Page 6, line 10 (rather than line 3). We have corrected the text to reflect the reviewer`s comments as suggested.
13. Page 9, line 10: The following sentence is hard to understand and needs revision: "More specifically, while the previously reported intervals of approximately 3-5 years may represent reasonable minimal thresholds of re-screening per economic and behavioural pressures placed on physicians by both patients and society, it inadequately addresses ceilings of re-screening, which may be substantially longer in low-risk patients."

Thank you for your comment. We meant to say that current physician practice, driven by both their own training, economic pressures, as well as patient demands for screening, may allow 3-5 year intervals between screenings. These might be considered “shortest tolerable durations”. As difficult as practitioners may find allowing this length of time between tests, however, our data suggests that for lower risk groups, practitioners should steel themselves to go even longer between screening. Admittedly, the sentence was convoluted, though, and has been simplified to: “Based on actual physician practice, the previously reported intervals of approximately 3-5 years may represent reasonable minimal thresholds of re-screening; however, they appear to be inadequate representations of re-screening ceilings, which may be substantially longer in low-risk patients.”

14. Please move study limitation to the end of discussion before conclusion

Yes, we moved the study limitation section as you have suggested. Thanks for the valuable suggestion.

Reviewer #2: This is an interesting paper. It aims to demonstrate different screening interval of T2D should be applied according to the baseline status of people. I have some comments for the authors to consider.

1. The authors classified participants according to their Framingham risk score, which include several components, such as age, cholesterol level, and smoking. Have they tried the analysis after classifying participants using individual risk factors. It is possible that all data are available in practice. Besides, it will be of interesting to know which factors could be most informative when estimating screening intervals on an individual level.

We used Framingham risk score for CVD risk stratification, as it accounts for several important individual risk factors in a validated model. The role of individual risk factors is interesting, but beyond the scope of this study; however, are considering this approach in a future study currently under development.
2. It has been argued that benefits of screening for diabetes could be less than screening for pre-diabetes [PMID:26881373]. Have the authors considered an analysis for pre-diabetes defined by HbA1c.

This is very interesting point and one that we feel it is important enough to warrant its own manuscript. We are considering applying our SNR methodology to the issue of pre-DM screening in a future study.

3. Details of statistical method is not clear to me even after referring to the citations. It will be informative if the authors could provide more details in the modelling so that eligible readers could follow their strategy

Thanks for your comment. We have added some more information in Figure 2 with footnotes and some equations that will be helpful for readers to more fully understand the methodology. We have also added supplemental explanation in manuscript on Page 4, Line15.

4. The authors mentioned HbA1c measurement, but how were Framingham risk score components measured?

Framingham Risk Score parameters were collected during the routine health check up at the same time and in the same facility in which HbA1c was measured. In brief, main parameters in Framingham Risk Score are age, total cholesterol, HDL, systolic blood pressure, smoking status, and diabetes treatment status. Smoking status, age, and diabetes treatment status were based on a questionnaire which all patients are required to complete, and reviewed for accuracy by an interview nurse. Please note that we only included those with no baseline diabetes, so this component of the FRS was nil.

According to D’Agostino et al., Framingham Score can be calculated as below. We should have referred to this article, so we added in Page 2 Line 17,

For clarification, the following sentence has been added: “BMI and Framingham Risk Score parameters were collected during the routine health check up at the same time and in the same facility in which HbA1c was measured.” was added Page 4, Line 4.

5. Page 5, line 9-10: parentheses did not seem to match.
We deleted the first “(“. Thank you for catching this error.

Reviewer #3: Reviewer comments: BEND-D-16-00101

This large retrospective cohort study sought to determine whether optimal diabetes screening intervals with HbA1c differ when other factors such as age category and BMI are taken into consideration. The question is a novel and interesting one.

The introduction is well written and referenced, concise and interesting. The rationale for the proposed work is clearly established.

Thank you for your kind comments.

In the methods, it is unclear why age was treated as a categorical rather than a continuous variable.

Thanks for your comment. To clarify, we only stratified participants based on age category (30-44, 45-59, 60-74), hypothesizing that informative HbA1c interval would differ by age category. We believe that this approach probably has more clinical meaning for practitioners. We then adjusted our age-stratified models using age as a continuous value, in order to fully account for the effect of the 15 year span within in each age category. This has been clarified in the manuscript on Page 4, Line 11: “…adjusted for gender, age and BMI as continuous values at first measurement of HbA1c”.

The description of the statistical methodological approach is good, but many readers will not be familiar with signal to noise ratio as it applies to diabetes epidemiology. Little is considered of the sensitivity and specificity of using different intervals of HbA1c for diagnosing diabetes based on differences in age, BMI or other factors.

Signal to Noise Ratio method is a new way of assessing test characteristics in addition to the better-known sensitivity and specificity. As is shown in the table on page 3, while HbA1c’s true change per year in those 30-44 years old at normal BMI is around 0.03, HbA1c has a much
larger noise (i.e., short-term variation) of 0.18. This suggests that prior to about 6 years (when signal begins to exceed noise), it may be difficult judge if HbA1c changes represent actual change or just measurement error. In order to better understand a test’s short-term variation, we believe the efficacy of laboratory tests should also be evaluated based on their signal and noise characteristics, in addition to sensitivity and specificity.

This would make the paper more relevant to clinicians and would allow people to make a determination of how well the (screening intervals for the) test might perform at different age and BMI thresholds, even if these were arbitrarily defined by the authors.

Thank you for your comment. To better address this with the reading audience, we have added a Table on Page 3 demonstrating HbA1c’s individual signal and noise by age stratum.

A risk "matrix" similar to ones used for determining low, intermediate and high cardiovascular risk might help to illustrate the clinical importance of the very interesting findings. This might have age category on one axis and BMI category on the other, with the cells within the matrix containing the optimal screening interval in years. However this additional work would not be essential to allow the paper to be published.

Thank you for the valuable and interesting suggestion. We will certainly consider this interesting visual summary in our next signal-to-noise study.

It would be worth stating explicitly that though ethics approval was obtained for the study, patients did not provide written informed consent and the data used was obtained through routine clinical practice.

As suggested, this has been clarified in the manuscript on page 5, line 11.

It would be helpful to clarify why approximately 30% of the cohort were not eligible for inclusion.

Among 149,191 patients presenting for health check up, 40,467 were excluded for having only a single visit, 5,285 were excluded for age <30 years old, 1237 were excluded for age >75 years old, 2700 were excluded for having DM at baseline, 1524 were excluded for a CVD history, 1276 had missing Framingham Risk Score parameters, and 241 were excluded for missing HbA1c data. The exclusion criteria flow chart is now provided as Figure 1.
Define FRS at its first use (or better still just give it its full name throughout).

“FRS” is now changed to Framingham Risk Score throughout as suggested.

In addition, we added the reference in Page 2 Line 12, “


Did you report the influence of smoking status on diabetes risk? If not, you should.

As one of the parameter of Framingham Risk Score is smoking status, we have accounted for this in the study’s models. As another reviewer pointed out, it will be interesting to see the individual role of smoking and other factors on screening intervals and we are considering this for a future study.

The limitations are very comprehensively addressed in the discussion.

Overall, this is a strong, important paper which ought to be highly cited by other scientists with an interest in screening methodology and diabetes epidemiology. It is a clever and succinct description of an important observation. It is an excellent paper.