Author's response to reviews

Title: Shifting from glucose diagnosis to the new HbA1c diagnosis would reduce the ability of the Finnish Diabetes Risk Score -FINDRISC- in screening glucose abnormalities within a real-life primary healthcare preventive strategy.

Authors:

   Bernardo Costa (costaber@gmail.com)
   Francisco Barrio (ciscobarrio@gmail.com)
   Josep L Piñol (jpinyol.tarte.ics@gencat.cat)
   Juan J Cabre (jcabre.tarte.ics@gencat.cat)
   Xavier Mundet (xmundet.bcn.ics@gencat.cat)
   Jordi Salas (jordi.salas@urv.cat)
   Oriol Solà-Morales (osola@hittinnova.com)
   Research Group DE-PLAN-CAT/PREDICE (juanjocabre@gmail.com)

Version: 2 Date: 6 November 2012

Author's response to reviews: see over
Dear Dr. D’Souza

Many thanks for the consideration of the first version of this article on behalf of BMC-Medicine. We have revised the manuscript according to your comments and we enclose the amendments.

First we note that the contents of the previous version were altogether well valued by both reviewers, which has greatly encouraged us to undertake this revision. Above all, we wish to thank both of them for their contributions, which have undoubtedly increased the quality of the present version. We sincerely appreciate all these comments because they also reveal interest and a great deal of patience with this article. Obviously, the additional remarks have been taken into account carefully, as you can see throughout the entire text, which has been widely amended, with the main changes highlighted in yellow.

At this stage more specific information on the changes made in reply to the comments of the reviewers is given in the following sections by copying their specific comments first and then the manuscript paragraphs that reflect our response.

Reviewer #1

1. The authors state, that the recruited sample should be a representative one. Is the sample really representative for the persons using primary care services in Catalonia? For example the number of included men in comparison to women is lower and the age-distribution in men and women is not really the same. The proportion of included men # 65 years was higher than the proportion of included women of the same age. Contrary, more younger women than men participated in the study. Could this fact have influenced the results and if yes how?

Page 10, paragraphs 2/3

The general profile of the participants was similar to that of the primary care-attended population. In fact, women are the ones who most often use these services in Spain [7] and such predominance is similar to both widespread Finnish and American trials published concerning diabetes prevention [5,6]. Indeed, the number of included men in comparison to women is lower and the proportion of men over 65 years is higher than the proportion of included women of the same age. This distribution could perhaps explain why the risk of diabetes assessed by the score was higher in women and the risk assessed by the blood tests was higher in men.

In this regard, the age and sex distribution could be perceived as another bias at work, particularly when compared with larger population-based studies. Undoubtedly, the main reason is that the protocol was conducted under real working conditions in primary care which, in opposite, could be considered as an advantageous approach for this study. Moreover, it seems that individuals identified as at high risk at screening all benefit similarly from lifestyle
intervention, regardless of age, sex and different socioeconomic groups [22]. In previous controlled trials, older people seemed to benefit somewhat more than younger ones, but men and women both had similar outcomes. Accordingly, in specifying the target participant profile for diabetes prevention in primary care, it seems not necessary to pay too much attention to population subgroups; however, it is more important to plan properly for consistent preventive measures [22-24].

2. Persons with severe psychiatric disease, liver disease or blood disorders were excluded. It should be more specified what diseases were excluded.

Page 4, paragraph 1

Regarding the associated lifestyle intervention study (at least 5-year follow-up of the screened subjects) all individuals with severe psychiatric disease such as bipolar disorders or psychosis, chronic kidney (severe chronic renal failure) and serious chronic liver disease or blood disorders, such as severe iron-deficiency anaemia or others that may interfere with the HbA1c measurement, were excluded from the study.

3. The study is cross-sectional in design. This is a limitation of the study, which is already mentioned by the authors.

Page 9, last paragraph; Page 10, first paragraph

In fact, the diagnosis of diabetes and prediabetes were based on only one OGTT value, not two, but this is a commonly accepted procedure for screening in large samples. We tried first to evidence the FINDRISC performance in predicting current glucose disorders and then to compare the results depending on the set of diagnostic criteria applied. Obviously, we cannot rule out some selection bias. Nevertheless, available data on the four-year incidence of diabetes in the DE-PLAN-CAT cohort based on repeat testing have been contributed together with those from the derivation PREDIMED cohort to develop a new questionnaire tailored to the needs of our own Spanish-Mediterranean setting [21].

4. In the methods section it is mentioned that data using the FINDRISC questionnaire was gathered? Were the questionnaires filled in by the patients or by the treating physician? Were there differences regarding the FINDRISC test scores and the distribution of the risk classes between the recruiting centers? Were the OGTTs carried out in the centers due to a standardized protocol and was the responsible personal in the centers trained and certified before recruitment? Were the different blood parameters measured in one central lab or locally?

Page 5, last paragraph

Given the stratification of the sample, a pooled analysis of all questionnaires was conducted.

Page 6, third paragraph

Most of the FINDRISC questionnaires were filled in by the professionals during the first interview (95%) but a small proportion was self-administered (5%).

Page 7, paragraph 2

No significant differences regarding the FINDRISC test scores and the distribution of the risk classes between the recruiting centres were found.

Page 4, last paragraph

The second screening was carried out using a 2-h 75-g OGTT according to the World Health Organization (WHO) standards in all recruiting centres, with measurements of FPG and 2hPG. All participants with FINDRISC scores >15 were asked to have a screening OGTT as part of the
Participants with FINDRISC scores <15 were offered an OGTT on a voluntary basis [7]. For the purposes of this part of the DE-PLAN-CAT (screening) the diagnostic of all glucose disorders were based on a single OGTT since a first FPG or 2hPG result suggestive of diabetes ruled out the participation in the subsequent part (lifestyle intervention). A second OGTT to confirm a diagnosis of diabetes was recommended in the study protocol but for those individuals who ultimately participated in the lifestyle intervention.

Page 5, first paragraph

Plasma glucose was determined by a uniform glucose oxidase–peroxidase method. The HbA1c determination was performed simultaneously by a standardized HPLC assay aligned to the Diabetes Control and Complications Trial in all laboratories [12]. Blood samples were analyzed at five laboratories, four of which corresponding to the same institution (Catalan Health Institute), using similar techniques. The intra- and interassay coefficients of variation for all assays ranged from 2 to 3%.

Page 3, last paragraph

All participating professionals were certified before recruitment after receiving several training meetings.

5. In the methods section it is further mentioned that a second OGTT was recommended in the study protocol to confirm diabetes. It is not clear for me, whether a second OGTT was carried out in the study and if yes, within what timeframe after the first OGTT. How was dealt with different results assessed through the two OGTTs, that means how was the participant categorized if the result of the first OGTT could not be confirmed?

Please see comment in page 4, last paragraph (page 2 of this cover letter). In addition:

Page 12, first paragraph

Contrary to the population based studies, the DE-PLAN-CAT survey was essentially focused on a representative sample of undiagnosed subjects in primary care where the likelihood of presenting glucose abnormalities obviously increases. Leaving aside the inconvenience of using the OGTT, both the WHO and the ADA criteria established that in the absence of unequivocal hyperglycaemia, results should be confirmed by repeat testing at least in clinical practice. Therefore, it is not surprising that about one third of the participants had any type of impaired glucose metabolism since screening was based on a single test.

6. How was dealt with missing data on any of the included variables?

Page 6, paragraph 4

A total of 1,746 participants (56%) also authorized the second screening by blood tests of which 1,712 (54.9%) cases were available for all requested data.

7. The risk of diabetes assessed by the FINDRISK score was higher in women, while the risk of diabetes assessed by either the glucose or the HbA1c measurements was higher in men. It would be good to discuss this finding.

Please see comment in page 10, paragraphs 2/3 (page 1 of this cover letter).

Minor Essential Revisions
Page 5, second line: the HbA1c criteria should be < 5.7, 5.7-6.4 and > 6.4%.

The paragraph has been edited as suggested by Reviewer #1

… (< 38, 38-48 and > 48 mmol/mol) or (< 5.7, 5.7-6.4 and > 6.4%).
**Reviewer #2**

Although we are honoured by the second reviewer’s comments, always based on solid scientific arguments, we think that their approach is that of a more stringent epidemiological survey when, in fact, the study is not. Nevertheless, the text has been corrected in accordance with their specific comments at all levels.

- The main concern is related with the criterion for define diabetes based on the FPG. The ADA criterion established that in the absence of unequivocal hyperglycemia, results should be confirmed by repeat testing. In this study, as the FPG was not repeated, the prevalence of glucose metabolic disorders, based on this diagnostic criterion, is seriously flawed. On this regard, it will be expected that the prevalence of diabetes and prediabetes, based on one measure of FPG, reflect a significant number of false negative tests, with a high prevalence of diabetes and prediabetes; however, the prevalence, based on one measure of FPG, was the lower and the prevalence of prediabetes similar as compared with other diagnostic criteria. These issues should be appropriately discussed, or the diagnostic criteria based on one measure of FPG deleted from analysis; taking into account that is not the original diagnostic criteria proposed by the ADA.

*Page 4, last paragraph*

The second screening was carried out using a 2-h 75-g OGTT according to the World Health Organization (WHO) standards in all recruiting centres, with measurements of FPG and 2hPG. All participants with FINDRISC scores >15 were asked to have a screening OGTT as part of the protocol. Participants with FINDRISC scores <15 were offered an OGTT on a voluntary basis [7]. For the purposes of this part of the DE-PLAN-CAT (screening) the diagnostic of all glucose disorders were based on a single OGTT since a first FPG or 2hPG result suggestive of diabetes ruled out the participation in the subsequent part (lifestyle intervention). A second OGTT to confirm a diagnosis of diabetes was recommended in the study protocol but for those individuals who ultimately participated in the lifestyle intervention.

*Page 9, last paragraph; Page 10, first paragraph*

In fact, the diagnosis of diabetes and prediabetes were based on only one OGTT value, not two, but this is a commonly accepted procedure for screening in large samples. We tried first to evidence the FINDRISC performance in predicting current glucose disorders and then to compare the results depending on the set of diagnostic criteria applied. Obviously, we cannot rule out some selection bias. Nevertheless, available data on the four-year incidence of diabetes in the DE-PLAN-CAT cohort based on repeat testing have been contributed together with those from the derivation PREDMED cohort to develop a new questionnaire tailored to the needs of our own Spanish-Mediterranean setting [21].

*Page 12, first paragraph*

Contrary to the population based studies, the DE-PLAN-CAT survey was essentially focused on a representative sample of undiagnosed subjects in primary care where the likelihood of presenting glucose abnormalities obviously increases. Leaving aside the inconvenience of using the OGTT, both the WHO and the ADA criteria established that in the absence of unequivocal hyperglycaemia, results should be confirmed by repeat testing at least in clinical practice. Therefore, it is not surprising that about one third of the participants had any type of impaired glucose metabolism since screening was based on a single test. In a previous work conducted in the same population we have already shown that defining diabetes by FPG resulted in a significant decrease in prevalence with regard to diabetes defined by 2hPG, even by repeat testing in those participants who agreed the lifestyle intervention [25]. In addition, a shift from the glucose-based diagnosis to the HbA1c-based diagnosis also reduced diabetes prevalence with a low overall or single degree of overlap between diagnostic categories [25].
Analyzing data showed in the Table 4, it is clear that the cutoff point of the FINDRISC is 14 (both the highest sensitivity and specificity). Why authors selected the cutoff point of 13? Decreasing the cutoff point, unnecessarily increases the false negative tests and the costs of confirming diagnosis.

Most of the proposed changes referred to the need to clarify the best cut-off point. We had chosen 13 because of its higher sensitivity, which is a useful property in screening programmes. However, if one looks at the sum of the values of sensitivity and specificity, the cut-off point 14 is more appropriate. That's why we drafted the manuscript accordingly.

Page 2, abstract

The FINDRISC at a cut-off of 14 had a reasonable high ability in predicting diabetes (AUC=0.71) and glucose abnormalities (AUC=0.67-0.69) using either 2-h or fasting glucose-based diagnostic criteria.

Page 6, first paragraph

The optimal cut-off points used were the peaks of the curve, where the sum of sensitivity and specificity is at maximum.

Page 11, last paragraph

In any case, the ROC curves indicated that a lower cut-off of 14 for detecting diabetes or any glucose metabolic abnormality offered the best balance in this population irrespective to the set of diagnostic criteria used. This cut-off is one point lower than 15, the most commonly found [23], but even lower cut-off points have been considered suitable for screening in other community-based diabetes prevention programmes [24]. It is likely that if we had practiced blood tests to all participants who answered the FINDRISC questionnaire, the cut-off had also increased. However, this is a realistic strategy to identify individuals at high risk who might be offered a preventive intervention and not a stringent experimental study aimed at validating the scale. At this point the maximum sensitivity and specificity of the FINDRISC for detecting diabetes were about 76% and 52% being 68% and 56% for detecting all glucose abnormalities. When diabetes was defined by a single HbA1c measurement resulted in a small decrease in sensitivity ranging from 1.3% (in comparison with 2hPG-based diagnosis) to 1.7% (with respect to FPG-based diagnosis). As for detecting all glucose abnormalities the use of HbA1c-based criteria led to a greater reduction in sensitivity ranging from 11.4% (fasting plasma glucose diagnosis) to 13.6% (2-h plasma glucose diagnosis). Corresponding specificity findings also showed a reduction though much more moderate, which reached a maximum of 5.4%.

Minor Essential Revisions

• Data on text, to describe AUC, unnecessarily repeat data shown in Figure 3. It should be deleted the text, or delete the Figure 3.

We agree with the conceptual comments raised by the reviewer. We would not wish to delete the third figure because we consider it highly didactic to allow visual comparison of ROC curves. Data on text to describe AUC results has been edited considering the deliberation from reviewer #2. The AUC values by sex (not showed in Figure 3) were retained and new sensitivity, specificity and negative predictive values by applying the cut-off point of 14 were introduced.

Page 7, last paragraph; Page 8, first paragraph
Drawing on this cut-off greater than or equal to 14 on the FINDRISC scale and regarding the diagnostic classification by the WHO criteria (that includes 2hPG) the area under the ROC curve for detecting unknown diabetes (Fig. 3A) was 0.67 (95% CI: 0.59–0.72) for men and 0.76 (95% CI: 0.70–0.81) for women. These values for all glucose abnormalities (Fig. 3B) were 0.64 (95% CI: 0.60–0.69) for men and 0.70 (95% CI: 0.66–0.73) for women. The sensitivity and specificity were 75.9% and 52.3% for the detection of type 2 diabetes alone with 65.8% and 56.7% for detecting any degree of abnormal glucose metabolism. The cut-off of 14 had a negative predictive value of 95.5% for diabetes and 78.4% for glucose abnormalities (Table 4). 

Corresponding FPG-based findings (ADA criteria) for diabetes (Fig. 3C) were 0.72 (95% CI: 0.63–0.82) for men and 0.70 (95% CI: 0.61–0.79) for women. Equivalent values for all glucose metabolic abnormalities (Fig. 3D) were 0.64 (95% CI: 0.60–0.69) and 0.73 (95% CI: 0.69–0.76), respectively. For a FINDRISC greater or equal to 14 the sensitivity and specificity were 75.5% and 50.5% for the detection of diabetes with 68.0% and 56.6% for detecting glucose abnormalities. The negative predictive values were 98.5% and 81.8%, respectively (Table 4). 

Finally, parallel findings based on HbA1c for diabetes (Fig. 3E) were 0.62 (95% CI: 0.53–0.70) for men and 0.70 (95% CI: 0.61–0.80) for women. Corresponding values for all glucose metabolic abnormalities (Fig. 3F) were 0.51 (95% CI: 0.46–0.57) and 0.57 (95% CI: 0.53–0.61), respectively. For a cut-off point of 14 the sensitivity and specificity were 74.2% and 50.5% for the detection of diabetes with 54.4% and 51.3% for detecting glucose abnormalities. The corresponding negative predictive values were 98.1% and 73.6%, respectively (Table 4). 

We hope that this second version is found suitable for publication. If so, we would very much appreciate a letter of acceptance. Of course, we continue at your disposal if you require further corrections of this manuscript. Thank you again for your time and attention.

We look forward to hearing from you as soon as possible.

Sincerely yours

Dr. Bernardo Costa
Diabetes and Metabolism Research Area
Jordi Gol Primary Health Care Research Institute
Catalan Health Institute