Diabetes risk scores and death: predictability and practicability in two different populations

David Faeh, Pedro Marques-Vidal, Michael Brändle, Julia Braun, Sabine Rohrmann

1 Unit of Demography and Health Statistics and Division of Cancer Epidemiology and Prevention, Institute of Social and Preventive Medicine, University of Zurich, Zurich, Switzerland
2 Institute of Social and Preventive Medicine (IUMSP), Lausanne University Hospital, Lausanne, Switzerland
3 Division of Endocrinology and Diabetes, Kantonsspital St. Gallen, St. Gallen, Switzerland

The aim was to examine the capacity of commonly used type 2 diabetes mellitus (T2DM) risk scores to predict overall mortality. The US-based NHANES III (n = 3138; 982 deaths) and the Swiss-based CoLaus study (n = 3946; 191 deaths) were used. The predictive value of eight T2DM risk scores regarding overall mortality was tested. The Griffin score, based on few self-reported parameters, presented the best (NHANES III) and second best (CoLaus) predictive capacity. Generally, the predictive capacity of scores based on clinical (anthropometrics, lifestyle, history) and biological (blood parameters) data was not better than of scores based solely on clinical self-reported data. T2DM scores can be validly used to predict mortality risk in general populations without diabetes. Comparison with other scores could further show whether such scores also suit as a screening tool for quick overall health risk assessment.

Methods

Data from two population-based prospective studies was used. The US NHANES III was conducted in 1988–94 and had mortality follow-up until December 2006; for this study, only data from non-Hispanic white participants with fasting blood glucose <7 mmol/l and without T2DM was used (n = 3138; 982 deaths). The Swiss CoLaus study was conducted in 2003–06 and had mortality follow-up until September 2012; for this study, only data from participants without T2DM was used (n = 3946; 191 deaths). Eight T2DM risk scores relying on self-reported information on anthropometrics, lifestyle or medical history (clinical) or on a combination of clinical factors and blood parameters (biological) were used: Balkau, Wilson, Griffin, Kahn (clinical and clinical + biological), FINDRISK, the Swiss and the German Diabetes Risk Score (GDRS). Participants with missing information in any of the scores were excluded. The GDRS could only be calculated using the NHANES III data because habits is available in CoLaus. For each score, a continuous (score) and a binary (risk yes/no) variable were computed. The predictive capacity regarding death from all causes was assessed using Harrell’s C and area under the curve (AUC) for the continuous variable. For the binary variable, we used sensitivity, specificity and positive and negative predictive values (PPV and NPV). No adjustment was made for age or sex, as these variables were part of some scores. The predictive capacity of three individual biological parameters was also assessed: fasting blood glucose, glycated hemoglobin (HbA1c) and insulin resistance (HOMA-IR).

Results

The populations used to develop the T2DM scores and the variables the scores are composed of are shown in Supplementary tables A1 and A2 in the supplementary file. In NHANES III, the percentage of participants for whom the score was not assessable owing to missing information varied from 0% (Griffin) to 17% (Kahn’s clinical + biological), whereas no participant from CoLaus had missing information. The prevalence of participants at risk of T2DM varied between scores and, to a smaller extent, between surveys. In NHANES III, the Griffin score presented the best predictive capacity, as shown by the largest Harrell’s C, AUC (continuous variable), PPV and NPV (binary variable). In CoLaus, Kahn’s clinical score had the best predictive capacity, but the Griffin score ranked second best. Wilson, Balkau, Findrisc and Swiss Diabetes Risk Score from the Swiss Diabetes Association had a high specificity but a poor sensitivity (binary), thus being inappropriate for the prediction of death. The Griffin score offered the best combination of a high specificity and an acceptable sensitivity. In both surveys, the predictive capacity of scores based on clinical and biological data was not better than of scores based solely on clinical data (table 1).
Table 1 Characteristics of selected type 2 diabetes risk scores and single biological markers predicting all-cause mortality in two different surveys

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Type</th>
<th>Items</th>
<th>Assessment</th>
<th>Continuous variable</th>
<th>Binary variable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(n)</td>
<td></td>
<td>Harrell’s C</td>
<td>AUC</td>
</tr>
<tr>
<td>Wilson CB</td>
<td>CB</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>Balkau C</td>
<td>C</td>
<td>4</td>
<td>8</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Kahn 1 CB</td>
<td>CB</td>
<td>10</td>
<td>28</td>
<td>25</td>
<td>17</td>
</tr>
<tr>
<td>Kahn 2 C</td>
<td>C</td>
<td>13</td>
<td>37</td>
<td>33</td>
<td>8</td>
</tr>
<tr>
<td>SDRS C</td>
<td>C</td>
<td>8</td>
<td>1</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>GDRS C</td>
<td>C</td>
<td>10</td>
<td>12</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>Findrisc C</td>
<td>C</td>
<td>7</td>
<td>9</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>Griffin C</td>
<td>C</td>
<td>7</td>
<td>23</td>
<td>18</td>
<td>0</td>
</tr>
</tbody>
</table>

Single biological markers

| HOMA-IR ≥ 2.5 | B | 1 | 35 | 23 | 10 | 0.561 | 0.538 | 57 | 54 | 42 | 33 | 69 | 76 | 37 | 4 | 73 | 97 |
| FBG ≥ 6.1 mmol/l | B | 1 | 10 | 13 | 10 | 0.626 | 0.547 | 66 | 54 | 18 | 16 | 93 | 93 | 54 | 6 | 72 | 97 |
| HbA1c ≥ 5.9% | B | 1 | 10 | - | 4 | 0.679 | - | 70 | - | 22 | - | 90 | - | 38 | - | 81 | - |

\* not assessed in the CoLaus Study.

a: of persons at risk of type 2 diabetes.
b: only available for NHANES III.

The scores were calculated based on a sample excluding participants with missing information for any of the scores, i.e. 17%.

AUC, area under the receiver operating curve; PPV, positive predictive value; NP, negative predictive value; CoLaus, Cohorte Lausannoise; NHANES III, third National Health and Nutrition Examination Survey; C, clinical; CB, clinical + biological; B, biological; SDRS, Swiss Diabetes Risk Score from the Swiss Diabetes Association; GDRS, German Diabetes Risk Score from the EPIC-Potsdam cohort; HOMA-IR, Homeostasis Model Assessment of insulin resistance, defined as fasting insulin (mU/L) x fasting blood glucose (mmol/L) / 22.5; FBG, fasting blood glucose; HbA1c, glycated hemoglobin.
Among individual biological parameters, HbA1c had the best predictive capacity in NHANES III, while insulin resistance (HOMA-IR) had the lowest predictive capacity in both studies (table 1).

**Discussion**

We examined whether T2DM risk scores could be envisaged as screening tool for overall health risk assessment and used all-cause mortality as outcome proxy. The capacity of T2DM scores to predict mortality differed considerably within a study population. However, the scores also performed differently in the two studies. The scores also varied regarding the number and the type of required variables, which was reflected in a broad range of the percentage of participants excluded because of missing values.

Looking at both studies, the Griffin score generally had the best predictive capacity. Further, as it does not require blood sampling, the Griffin score is less costly and could be used for self-assessment by Internet or paper questionnaire. Moreover, the applicability of this score appeared universal, as it performed similarly well in a European and in a US population. Still, the number of participants classified as being at risk using the Griffin score was high. This may be problematic because monitoring, coaching and treating a potentially large number of individuals at risk could be expensive and logistically demanding. It would be preferable to have a score with similar sensitivity/specificity values but leading to a lower prevalence of individuals at risk. In all scores and both study population samples, the sensitivity was relatively low. One reason for this may be that these scores were not developed to predict death as outcome and/or that the cut-off point should be adapted. Low sensitivity is problematic because a substantial proportion of persons who will die in the future are missed. Interestingly, and in agreement with the literature, scores including biological data were not better than scores based solely on clinical data. This finding suggests that the inclusion of biological variables might not improve the predictive capacity of T2DM risk scores regarding overall mortality. Further studies are needed to better assess whether biological variables could be replaced by more easily obtainable parameters, e.g. also in scores predicting a cardiovascular outcome. HbA1c had a good predictive capacity and led to a relatively low prevalence of participants at risk. Hence, HbA1c could be used as a risk factor for overall mortality. In fact, HbA1c performed significantly better than the lipid parameters traditionally used in models predicting cardiovascular disease mortality. HbA1c can also be measured and interpreted in the non-fasting state, thus facilitating screening procedures.

We conclude that T2DM risk scores might be useful to predict the risk of death as a proxy for overall health risk. In both populations, scores based solely on clinical data were as efficient as scores additionally including biological variables. This could open doors for the development of simple self-administrable screening questionnaires, allowing us to assess and approach lifestyle factors.

**Supplementary data**

Supplementary data are available at EURPUB online.

**Funding**

This work was supported by the Swiss National Science Foundation (grants 3347CO-108806, 33CS30-134273, 32473B-125710). The Colaus study was supported by grants from the Swiss National Science Foundation [grant no: 33CSCO-122661 and 3CSCO-139468]; GlaxoSmithKline and the Faculty of Biology and Medicine of Lausanne.

**Conflict of interest:** None declared.

**Key points**

- Numerous scores exist helping to identify persons at risk for type 2 diabetes mellitus
- The capacity of the scores to predict all-cause mortality and the usefulness in clinical practice has not been examined previously
- There were large differences regarding the prevalence of persons defined at risk and concerning the predictive capacity of the scores
- Scores based on solely clinical data may perform better than scores relying on clinical plus biological data
- This could open doors for the development of simple self-administrable screening questionnaires, allowing us to assess and approach lifestyle factors.

**References**