# Edinburgh Research Explorer

## Development of a screening tool using electronic health records for undiagnosed Type 2 diabetes mellitus and impaired fasting glucose detection in the Slovenian population

## Research: Care Delivery

# Development of a screening tool using electronic health records for undiagnosed Type 2 diabetes mellitus and impaired fasting glucose detection in the Slovenian population

G. Štiglic[1,2], P. Kocbek[1], L. Cilar[1], N. Fijačko[1], A. Stožer[3], J. Zaletel[4], A. Sheikh[5,6] and P. Povalej Bržan[1,2]

Faculties of [1]Health Sciences, [2]Electrical Engineering and Computer Science and [3]Medicine, University of Maribor, Maribor, Slovenia [4]Department of Endocrinology, Diabetes and Metabolic Diseases, University Medical Centre, Ljubljana, Slovenia, [5]Centre for Medical Informatics, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK and

⁶Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital/Harvard Medical School, Boston, MA, USA

*Correspondence to*: Gregor Štiglic. E-mail: gregor.stiglic@um.si

## What's new?

- The increasing prevalence of Type 2 diabetes mellitus calls for more effective screening tests. This study introduces the first predictive models for undiagnosed Type 2 diabetes mellitus and impaired fasting glucose for the Slovenian population.
- Waist measurement is preferred to BMI in both screening tools.
- Two simplified models based on data from electronic health records were obtained by removing two questions from the Finnish Diabetes Risk Score (FINDRISC) questionnaire.
- The proposed models achieved a significantly higher predictive performance in comparison with the routinely used FINDRISC questionnaire.

## Abstract

**Aim** To develop and validate a simplified screening test for undiagnosed Type 2 diabetes mellitus and impaired fasting glucose for the Slovenian population (SloRisk) to be used in the general population.

**Methods** Data on 11 391 people were collected from the electronic health records of comprehensive medical examinations in five Slovenian healthcare centres. Fasting plasma glucose as well as information related to the Finnish Diabetes Risk Score questionnaire, FINDRISC, were collected for 2073 people to build predictive models. Bootstrapping-based evaluation was used to estimate the area

under the receiver-operating characteristic curve performance metric of two proposed logistic regression models as well as the Finnish Diabetes Risk Score model both at recommended and at alternative cut-off values.

**Results** The final model contained five questions for undiagnosed Type 2 diabetes prediction and achieved an area under the receiver-operating characteristic curve of 0.851 (95% CI 0.850–0.853). The impaired fasting glucose prediction model included six questions and achieved an area under the receiver-operating characteristic curve of 0.840 (95% CI 0.839–0.840). There were four questions that were included in both models (age, sex, waist circumference and blood sugar history), with physical activity selected only for undiagnosed Type 2 diabetes and questions on family history and hypertension drug use selected only for the impaired fasting glucose prediction model.

**Conclusion** This study proposes two simplified models based on FINDRISC questions for screening of undiagnosed Type 2 diabetes and impaired fasting glucose in the Slovenian population. A significant improvement in performance was achieved compared with the original FINDRISC questionnaire. Both models include waist circumference instead of BMI.

## Introduction

Recent studies report an increasing prevalence of Type 2 diabetes mellitus and its negative influence on healthcare systems and the affected individuals [1–4]. There is therefore an urgent need to identify people who are at high risk of developing Type 2 diabetes, especially those with impaired fasting glucose (IFG) or even undiagnosed Type 2 diabetes. Early detection of both IFG and Type 2 diabetes can significantly reduce public and personal costs [5].

Researchers in many countries are developing tools for predicting and detecting Type 2 diabetes. Studies show that the most cost-effective methods for screening Type 2 diabetes in the general population are the use of non-invasive screening tools for risk assessment, followed by blood tests for glycaemia [fasting plasma glucose (FPG), oral glucose tolerance test or HbA$_{1c}$] [6]. These screening tools are used for early detection of risk factors for developing Type 2 diabetes and for preventing the occurrence of full-blown disease and complications [7]. One of the most widely used non-invasive screening tools is the Finnish Diabetes Risk Score (FINDRISC). Its aim is to predict the likelihood of individuals developing diabetes in the next 10 years [8]. In its original form, FINDRISC consists of eight questions. Because gender is not collected as a separate question, but can be collected from the question on waist circumference, where a separate set of values is available for men and women, FINDRISC questionnaires can be used to collect data on nine non-invasive variables: age; gender; BMI; waist circumference; physical activity; daily consumption of fruit and vegetables; history of antihypertensive drug treatment; history of high blood glucose; and family history of diabetes. It has been validated in many countries; for example, Sweden [9], the Netherlands [10], Greece [11], Spain [12], Hungary [13], Bulgaria [14] and Slovenia [15]. FINDRISC has shown good results, but it has often been modified because of cultural differences; for example, by adjusting the BMI and waist circumference threshold to meet Asia-Pacific standards [16] or by adjusting the threshold and excluding some variables from the original FINDRISC [6,12]. As demonstrated in our previous work, focusing on differences in optimal FINDRISC thresholds for the male and female working population in Slovenia, gender plays an important role in the predictive modelling of undiagnosed Type 2 diabetes (UT2D) and IFG [15].

In the present study, we aimed to develop and validate a simplified screening test for the detection of UT2DM and IFG in the Slovenian population with a higher predictive performance and fewer questions in comparison with FINDRISC. Both models developed in the present study are primarily aimed for use in an online environment targeting a general adult population.

## Methods

### Study design and data source

A cross-sectional population-based study was performed on electronic healthcare records data collected from five primary care institutions located in different parts of Slovenia. The data were anonymized at the healthcare institution site and later centrally collected at a software vendor data warehouse. All variables collected in a dataset were routinely collected in healthcare centres during preventive health examinations. The study was approved by the institutional ethics committee. The Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) [17] and the Reporting of Studies Conducted Using Observational Routinely Collected Health Data (RECORD) [18] statements were followed.

### Study setting and sample

The initial dataset consisted of 11 391 electronic healthcare records of healthy adults with no prior Type 2 diabetes diagnosis. The data were collected during the period January 2015 to September 2016, with only the first visit after January 2015 used for model derivation. The International Classification of Diseases 10 code E11 was used to exclude people with prior Type 2 diabetes diagnosis. After removal of incomplete cases that were missing outcome or predictor variable values, the final sample consisted of 2073 records, summarized in Table 1. A test for possible bias attributable to missing data was conducted by comparing the difference between the group of complete records and the group of incomplete records, which were removed in the further steps of analysis (Tables S4 and S5 in the Supporting Information). This test was performed for all variables, except those with an extremely high number of missing values (i.e. physical activity, high blood pressure history, high blood glucose history, daily fruit and vegetables consumption, and the question on diabetes in family). The only significant difference could be found in the age variable, with a difference of up to 2.5 years; however, as the distribution of the difference was equal over all three groups [UT2DM, IFG and normal fasting glucose (NFG)], this difference should not have affected the

model significantly. The high number of missing values can be attributed to the fact that entry of detailed FINDRISC questionnaire data is not mandatory at the primary healthcare level in Slovenia [19].

**Key outcomes**

Two key outcomes were defined to allow derivation of a model predicting UT2DM and IFG. A laboratory measurement of FPG level was obtained from the electronic healthcare records for all 2073 participants to define an outcome for each individual as UT2DM (FPG ≥ 7.0 mmol/L), IFG (FPG 6.1 – 7.0 mmol/L), or normal fasting glucose (FPG < 6.1 mmol/L). Individuals classified in the IFG group ($n$ = 435, 21.0%) or UT2DM group ($n$ = 146, 7.0%) were both used as positive cases ($n$ = 581, 28.0%) in the IFG predictive model derivation. This approach, as opposed to treating the IFG group as a separate group, also allows a more favourable positive vs negative case balance and improves the fit of the derived predictive models. A higher variance in FPG measurements was found in the UT2D group than in the IFG or normal glucose groups (Table 1).

**Predictor variables**

Predictor variables included all FINDRISC questions, with a few exceptions where continuous values were used instead of discretized interval values. Continuous values included information on age (years), BMI (kg/m$^2$) and waist circumference (cm). Additionally, nominal variables representing gender, daily physical activity (>30 min), history of high blood pressure, history of high blood glucose, daily fruit and vegetable consumption, and family history of diabetes were used as binary variables to build the model. All nominal variables were dichotomous, except for family history of diabetes, for which three different answers were possible: 'no diabetes in family', 'diabetes present in grandparent, aunt, uncle or first cousin', and 'diabetes present in parent, brother, sister or own child'. This variable was dichotomized into two new binary variables representing presence of Type 2 diabetes in two different groups of relatives.

As in the original FINDRISC questionnaire, we conducted an additional *post hoc* analysis with introduction of the interaction term between gender and waist measurement. The interaction term allows modelling of the effect of waist measurement on the outcome that can be different for men and women. In our previous study on a smaller sample, we demonstrated that optimal cut-off values could differ significantly for men and women [15].

**Statistical analysis and model validation**

The models were derived using binomial logistic regression modelling. In each bootstrapping [20,21] iteration that was used to build and test the predictive model, we built two models, i.e. a full model and a simplified model that contained only features with statistically significant Wald test results from the full model. To allow an unbiased evaluation of the model performance [22], a full and a reduced model were built inside the same bootstrapping iteration. Because we were interested in detecting UT2DM and IFG, two full models and two simplified models were built for each bootstrapping iteration. The bootstrapping was performed by randomly sampling *n* samples with replacement from the initial set of samples, where *n* represents the number of all available samples. The samples that were not selected in the first step were used to evaluate the derived models, enabling testing of the derived models on independent samples. Additional experiments were conducted on two subsamples based on gender, where the predictive models were built and validated on data from men and women separately.

The following performance metrics were calculated for each derived prediction model in each bootstrapping iteration: area under the receiver-operating characteristic curve (AUC), accuracy, sensitivity (true positive rate), specificity (true negative rate), positive predictive performance, negative predictive performance, and percentage of positively classified samples. For details regarding all predictive performance metrics used in this study see Harrell [23].

## Results

This section presents results of both UT2DM and IFG prediction models that were tested using 1000 bootstrapping iterations for both the full model and the simplified model. The bootstrapping-based evaluations were followed by derivation of the final UT2DM simplified model and IFG simplified model that can be applied to the Slovenian population.

### Bootstrap validation

Figure 1 shows a comparison of classification performance (measured in AUC) between FINDRISC tests with different thresholds ranging from 10 to 20 (models F10–F20), including the FINDRISC test threshold value of 15 (F15) recommended in the Slovenian Type 2 diabetes prevention guidelines [24]. Additionally, both full model and simplified model were compared with the FINDRISC screening test for predicting UT2DM and IFG.

A difference between variabilities of the derived models can be observed when comparing UT2DM to IFG prediction models (Fig. 1). As expected, UT2DM prediction models showed more variability compared with IFG models. For the best-performing model predicting UT2DM an AUC of 0.853 (95% CI 0.811–0.898) was obtained for the full model and an AUC of 0.851 (95% CI 0.807–0.895) for the simplified model. Similarly, the best-performing full model predicting IFG achieved an AUC of 0.841 (95% CI 0.816–0.866) and an AUC of 0.840 (95% CI 0.814–0.865) for the simplified model. Detailed results on sensitivity, specificity, PPV, NPV and percentage of positively classified individuals can be found in the Supporting Information (Table S1).

Next, we observed the frequency of variables selected in 1000 bootstrap iterations to build the simplified model (Fig. 2). There were three variables that were selected in >90% of bootstrapping iterations in UT2DM as well as IFG prediction: history of high blood sugar; age; and gender. For UT2DM prediction, there were two more variables that were selected in >50% bootstrap iterations when building the simplified predictive model. Both physical activity and waist circumference are variables that can be influenced through lifestyle-based interventions aimed at risk reduction.

Waist circumference was also part of the simplified model for IFG prediction in the majority of bootstrapping iterations. An interesting variable selection pattern could be observed in both simplified models, with waist circumference being kept in the model at the cost of BMI, which was excluded from both UT2DM and IFG models for most of the bootstrap iterations. A significant difference in the frequency of selection for the family history-related variables can be observed between the IFG and UT2DM models (Fig. 2). Two binary variables were used in the models as a result of dichotomization of the variables related to the question on family history of diabetes mellitus, for which participants were offered three answers.

Additionally, we introduced an interaction term between waist measurement and gender as in the original FINDRISC questionnaire. The same experimental setup as used in the initial validation was used for the interaction term validation. In comparison with the original results, we did not observe any significant improvement in AUC. More specifically, the full prediction models achieved AUC values of 0.854 (95% CI 0.812–0.898) for UT2DM and 0.841 (95% CI 0.815–0.865) for IFG. In the case of the simplified models, we even observed a small but insignificant deterioration of performance, with AUC values of 0.849 (95% CI 0.798–0.895) for UT2DM and 0.829 (95% CI 0.800–0.856) for IFG. More detailed results can be found in the Supporting Information.

Finally, we built and validated separate models for male and female samples (Tables S2 and S3 in the Supporting Information). Figure 3 shows the results for UT2DM and IFG prediction models for both gender-based groups. Observing the predictive performance of the FINDRISC model with different thresholds, we noted that the optimal AUC-based threshold for men lay at 14 and 12 for UT2DM and IFG, respectively. In women, the threshold of 14 provides the best AUC results for both UT2DM and IFG prediction. It should be noted that there were very small differences for thresholds between 12 and 15 for both groups. The predictive performance of the proposed full model and simplified model did not differ significantly between men and women.

**Model derivation**

After assessing the predictive performance and variable importance using bootstrap-based evaluation, we applied the same predictive model derivation procedure to all available data ($n$ = 2073) to obtain the final simplified models for predicting UT2DM and IFG. Tables 2 and 3 show the binomial logistic regression model characteristics for the UT2DM and IFG models, respectively.

The selection of variables in the final two simplified models strongly correlates with the results of the bootstrap-based variable selection evaluation. In the UT2DM model, we observed five variables (age, sex, waist circumference, blood sugar history and physical activity), of which two, physical activity and waist circumference, are modifiable and can be targeted by lifestyle change interventions (Table 2). In the case of IFG model (Table 3), six variables were selected (age, sex, waist circumference, blood sugar history, family history of diabetes, and hypertension drug use), with four of them overlapping with the five UT2DM model variables. The correlation between numerical variables and log odds of the model response variable was significant for all variables ($P$<0.001). The Spearman correlation coefficient of age and waist circumference for the UT2DM (IFG) model was 0.428 (0.514) for age and 0.607 (0.561). The optimal threshold in the UT2DM and IFG predictive model was set to 0.126 and 0.224, respectively.

## Discussion

Many risk-scoring models have been developed for detecting individuals at high risk of developing Type 2 diabetes and for detecting undiagnosed Type 2 diabetes, to reduce healthcare and personal costs [25,26]. Kengne *et al.* [27] analysed 12 prediction models for Type 2 diabetes in eight European countries. They concluded that every predictive model needs to be evaluated and adjusted to the needs of the population in which it will be implemented. In the present study, we proposed two predictive risk models based on the FINDRISC model with simplifications (i.e. reduction of the number of questions) that reflect the characteristics of the data collected in Slovenia.

Our results indicate that the performance of both simplified risk models in screening for UT2DM and IFG was very high, as indicated by the AUC and other predictive performance measures. Significant differences were observed when we compared the performance of both simplified models to the original FINDRISC models with different thresholds. The models developed in the present study included all eight non-invasive variables of FINDRISC, but we used continuous values (age, weight, height, BMI) instead of the intervals that are used in the original FINDRISC questionnaire. Additional differences in the performance gap between the proposed binomial logistic regression models and the original FINDRISC model are assumed to originate from the rounding of the regression coefficients in the original FINDRISC. We believe that in the era of digital screening tests there is no need to simplify regression functions to simple scoring systems leading to huge differences in the predictive performance of such models.

In the present study, the optimal threshold for detection of UT2DM using the FINDRISC was 14, with a sensitivity of 76%, specificity of 73% and an AUC of 0.75 (95% CI 0.70–0.79). For predicting IFG, an optimal threshold of 13 resulted in a sensitivity of 68% and a specificity of 76%, with an AUC of 0.72 (95% CI 0.70–0.75). A related study from Finland [28] showed similar results with an UT2DM threshold of 11, sensitivity of 66% in men and 70% in women and an AUC of 0.72 in men and 0.73 in women. In other similar studies where the FINDRISC test was used to detect UT2DM, thresholds ranged from 9 to 15, with AUC results very similar to the results obtained in the present study. More specifically, the AUC of the proposed models achieved 0.74 (threshold of 9) [16], 0.72 (threshold of 13) [12], 0.72 (threshold of 15) [11] and 0.70 (threshold of 10) [14]. A study from Germany proposed a simplified FINDRISC-based model to predict UT2DM using only age, BMI, waist circumference, use of blood pressure medication, and history of high blood glucose to obtain AUC values of 0.88 [29]. Different thresholds were also demonstrated when separate models were built for men and women. In the present study, we did not find any significant difference between the predictive performances of both proposed models in different gender groups. These results confirm the findings from our previous study, in which a large gap in predictive performance was shown when comparing classic paper and pencil questionnaires with fully developed regression models [30]. Nevertheless,

using logistic regression models instead of simplified tests does limit the spectrum of usage for such a test, mainly to information technology-supported environments. We believe, however, that in modern society the online environment is the most appropriate one for the deployment of screening tests intended to reach a wide target population. Additionally, we believe that both models can be used in the clinical environment to support the preventive work of healthcare professionals by visual demonstration of risk and influence of lifestyle changes at different ages. It should be noted that at least an external validation of the proposed models should be performed, before they can be used in the clinical environment.

Studies conducted in different European countries show a higher probability of developing overweight, obesity, and central obesity in people with higher levels of blood glucose [25]. Physical activity, diet, weight, well-being, alcohol consumption and smoking are the most commonly mentioned risk factors related to the development of Type 2 diabetes that one can target using lifestyle-related interventions. The Diabetes Prevention Program shows that the intensive lifestyle interventions could reduce the incidence of Type 2 diabetes over 3 years by up to 58% [31]. Such interventions include weight loss and moderate physical activity [32].

Our results from binomial logistic regression (Tables 2 and 3) show that the incidence of Type 2 diabetes was higher among men, older people, physically inactive people, people with a large waist circumference, and those with a history of high levels of blood glucose. Similar results were obtained in a study conducted among eight European countries. The incidence of Type 2 diabetes was higher in men, people aged >60 years, and people with a large waist circumference [27].

Several studies show that we can prevent the development of Type 2 diabetes with specific interventions. To apply these interventions, we need to identify individuals who are at high risk of developing Type 2 diabetes at the right time. The proposed models represent a simple non-invasive tool for predicting UT2DM and IFG. Using the proposed models, participants need to answer seven questions, the answers to which would be used to simultaneously estimate the risk for UT2DM and IFG. Alternatively, it is possible to ask five Type 2 diabetes-related questions first and then ask two

additional IFG-related questions to estimate the risk of pre-diabetes (IFG) in case of a negative outcome for UT2DM.

Two variables out of nine derived from the original FINDRISC questionnaire were omitted altogether in our models. Unsurprisingly, our results showed that BMI and waist circumference were closely related. Consequently, our models included waist circumference instead of BMI. The question arises as to whether replacing BMI by waist circumference makes sense from the end-user perspective. One could assume that for most people it is easier to remember their height and weight than their waist circumference, which is rarely measured. However, inclusion of only BMI did not significantly change the results. The only question, therefore, should be whether waist circumference or BMI should be used; asking both questions in the same questionnaire should be avoided to save time in filling out the questionnaire. There is also a wide body of literature that confirms problems with BMI as it is not able to distinguish fat from fat-free mass, such as muscle and bone [33].

The second omitted variable was daily fruit and vegetable consumption. Table 1 shows that the majority of the population answered this question positively (89% in the UT2DM group, 90% in the FPG ≥ 6.1 mmol/L group, and 92% in the normal FPG group). It should be noted that eating habits differ between Slovenia and Finland, for example, in different sources of flavonoid-rich plant products that are frequently found in Mediterranean and Scandinavian diets [34]. Perhaps even more importantly, eating habits do change over time. More precisely, the original FINDRISC test [8], that actually points out the insignificant contribution of this question was developed using data from a cohort sampled in 1987. The same study mentioned a significant difference between the answers to this question in the development and validation data collected in 1992 that were used to validate the predictive model. Perhaps these questions should be set out more explicitly nowadays, possibly broken into two or more separate questions, especially given the fact that there are many studies confirming new associations between different food consumption patterns and Type 2 diabetes [35].

The present study has some limitations that should be mentioned. First, an even higher predictive performance of the models could possibly be achieved by using state-of-the-art machine learning

approaches, such as boosting approaches [36] or deep neural networks [37], to name just two. However, one should be aware that introduction of advanced predictive models often leads to different levels of complexity that can strongly reduce the interpretability of prediction models. As inclusion of the variable representing history of high blood glucose in most bootstrap iterations suggests, this variable represents a strong predictor of UT2DM or IFG. This may represent a problem as the mere fact that blood sugar measurement was carried out in the past may suggest that this person already has Type 2 diabetes or IFG. The only way to check for this was exclusion of records with a reported diagnosis of Type 2 diabetes. A further limitation of this study is the potential for bias attributable to missing data as only the complete samples were used to develop prediction models. As reported in the study by Stiglic *et al.* [19] on a smaller subset (*n*=952) of data from three healthcare centres in Slovenia, the imputation of missing data has a significant impact on the performance of a predictive model, but only for smaller samples. As demonstrated in Stiglic *et al.* [19] where an UT2DM predictive model was built on increasing fraction of complete samples, the predictive performance stabilizes at ~600 complete records used to build the model.

Although the bias towards a too optimistic predictive performance was reduced by introduction of bootstrap-based evaluation of the models, other external validations should be performed to further evaluate the predictive performance of the proposed models.

In conclusion, in the present study we present two simplified prediction models for the screening of undiagnosed Type 2 diabetes and IFG for the Slovenian population. The results show that a significant improvement in performance can be achieved compared with the original FINDRISC questionnaire. Two variables used in the original FINDRISC questionnaire were excluded in the simplified predictive models derived in the present study. Waist circumference was used instead of the highly correlated variable BMI and the variable consumption of fruit and vegetables was excluded because of its non-significant contribution to prediction performance. The proposed models were also

applied in the web-based application for the general population and are available at http://www.ri.fzv.um.si/modest2/slorisk/en/.

As already mentioned, the external validation of the proposed prediction model was not part of this study; however, the arrangements have already been made for evaluation of the proposed prediction models in five additional healthcare centres in Slovenia. Additionally, a temporal external validation is planned for all five healthcare centres included in this study. By further validation, the developed models might become ready for clinical use. The clinical use of such models should not be limited exclusively to screening purposes as there is the potential to use both risk tests presented in this study also for educational purposes, mainly to demonstrate the consequences of lifestyle changes in terms of changes in the estimated risk.

**Competing interests**

None declared.

# References

1 Choi SB, Kim WJ, Yoo TK, Park JS, Chung JW, Lee Y *et al*. Screening for prediabetes using machine learning models. *Conf Proc IEEE Eng Med Biol Soc* 2014; **2014:** 2472–2475.

2 Zhuo X, Zhang P, Barker L, Albright A, Thompson TJ, Gregg E. The lifetime cost of diabetes and its implications for diabetes prevention. *Diabetes care* 2014; **37**: 2557–2564.

3 Barengo NC, Tuomilehto JO. How can we identify candidates at highest risk–to screen or not to screen? *Herz* 2016; **41:** 175–183.

4 Ogurtsova K, da Rocha Fernandes JD, Huang Y, Linnenkamp U, Guariguata L, Cho NH *et al*. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Res Clin Pract* 2017; **128:** 40–50.

5 Cavan D. Why screen for type 2 diabetes? *Diabetes Res Clin Pract* 2016; **121:** 215–217.

6 Barengo NC, Tamayo DC, Tono T, Tuomilehto J. A Colombian diabetes risk score for detecting undiagnosed diabetes and impaired glucose regulation. *Prim Care Diabetes* 2017; **11**: 86–93.

7 Muchira JM, Stuart-Shor EM. Using Noninvasive Tools to Screen for Prediabetes and Undiagnosed Diabetes: Value Added to Cardiovascular Disease Prevention. *J Cardiovasc Nurs* 2016; **31**: 488–491.

8 Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 2003; **26**: 725–731.

9 Hellgren MI, Petzold M, Björkelund C, Wedel H, Jansson PA, Lindblad U. Feasibility of the FINDRISC questionnaire to identify individuals with impaired glucose tolerance in Swedish primary care. A cross  sectional population  based study. *Diabet Med* 2012; **29**: 1501–1505.

10 Alssema M, Feskens EJ, Bakker SJ, Gansevoort RT, Boer JM, Heine RJ *et al*. Finnish questionnaire reasonably good predictor of the incidence of diabetes in The Netherlands. *Ned Tijdschr Geneeskd* 2008; **152:** 2418–2424.

11 Makrilakis K, Liatis S, Grammatikou S, Perrea D, Stathi C, Tsiligros P *et al*. Validation of the Finnish diabetes risk score (FINDRISC) questionnaire for screening for undiagnosed type 2 diabetes, dysglycaemia and the metabolic syndrome in Greece. *Diabetes Metab* 2011; **37**:144–151.

12 Salinero-Fort MA, Burgos-Lunar C, Lahoz C, Mostaza JM, Abánades-Herranz JC, Laguna-Cuesta F *et al*. Performance of the Finnish Diabetes Risk Score and a Simplified Finnish Diabetes Risk Score in a Community-Based, Cross-Sectional Programme for Screening of Undiagnosed Type 2 Diabetes Mellitus and Dysglycaemia in Madrid, Spain: The SPREDIA-2 Study. *PloS One* 2016; **11**: e0158489.

13 Winkler G, Hidvégi T, Vándorfi G, Balogh S, Jermendy G. Prevalence of undiagnosed abnormal glucose tolerance in adult patients cared for by general practitioners in Hungary. Results of a risk-stratified screening based on FINDRISC questionnaire. *Med Sci Monit* 2013; **19:** 67–72.

14 Tankova T, Chakarova N, Atanassova I, Dakovska L. Evaluation of the Finnish Diabetes Risk Score as a screening tool for impaired fasting glucose, impaired glucose tolerance and undetected diabetes. *Diabetes Res Clin Pract* 2011; **92:** 46–52.

15 Štiglic G, Fijačko N, Stožer A, Sheikh A, Pajnkihar M. Validation of the Finnish Diabetes Risk Score (FINDRISC) questionnaire for undiagnosed type 2 diabetes screening in the Slovenian working population. *Diabetes Res Clin Pract* 2016; **120:** 194–197.

16 Ku GM, Kegels G. The performance of the Finnish Diabetes Risk Score, a modified Finnish Diabetes Risk Score and a simplified Finnish Diabetes Risk Score in community-based cross-sectional screening of undiagnosed type 2 diabetes in the Philippines. *Prim Care Diabetes* 2013; **7:** 249–259.

17 Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Ann Intern Med* 2015; **162:** 55–63.

18 Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I *et al*. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 2015; **12:** e1001885.

19 Stiglic G, Kocbek P, Fijacko N, Sheikh A, Pajnkihar M. Challenges associated with missing data in electronic health records: A case study of a risk prediction model for diabetes using data from Slovenian primary care. *Health Informatics J* 2017; DOI: 10.1177/1460458217733288 [Epub ahead of print]

20 Mooney CZ, Duval RD. Bootstrapping: A nonparametric approach to statistical inference. London: 1993.

21 Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. *J Clin Epidemiol* 2016; **69:** 245–247.

22 Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol* 2003; **56:** 441–447.

23 Harrell F. Regression modeling strategies. New York, 2001.

24 Zaletel J, Ravnik Oblak M. Slovenske smernice za klinično obravnavo sladkorne bolezni tipa 2. Klinični oddelek za endokrinologijo, diabetes in presnovne bolezni, Interna klinika, Univerzitetni klinični center Ljubljana and Klinični oddelek za endokrinologijo, diabetes in bolezni presnove. Ljubljana, 2016.

25 Hippisley-Cox J, Coupland C, Robson J, Sheikh A, Brindle P. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. *BMJ* 2009; **38:** b880.

26 Kahn R. Risk prediction models in diabetes prevention. *Lancet Diabetes Endocrinol* 2014; **2:** 2–3.

27 Kengne AP, Beulens JW, Peelen LM, Moons KG, van der Schouw YT, Schulze MB, *et al*. Non-invasive risk scores for prediction of type 2 diabetes (EPIC-InterAct): a validation of existing models. *Lancet Diabetes Endocrinol* 2014; **2**: 19–29.

28 Saaristo T, Peltonen M, Lindström J, Saarikoski L, Sundvall J, Eriksson JG, *et al*. Cross-sectional evaluation of the Finnish Diabetes Risk Score: a tool to identify undetected type 2 diabetes, abnormal glucose tolerance and metabolic syndrome. *Diab Vasc Dis Res* 2005; **2:** 67–72.

29 Li J, Bergmann A, Reimann M, Bornstein SR, Schwarz PE. A more simplified Finnish diabetes risk score for opportunistic screening of undiagnosed type 2 diabetes in a German population with a family history of the metabolic syndrome. *Horm Metab Res* 2009; **41**: 98–103.

30 Stiglic G, Pajnkihar M. Evaluation of major online diabetes risk calculators and computerized predictive models. *PloS One* 2015; **10**: e0142827.

31 American Diabetes Association. 5. Prevention or Delay of Type 2 Diabetes. *Diabetes Care* 2017; **40**: S44–S47.

32 Steven S, Hollingsworth KG, Al-Mrabeh A, Avery L, Aribisala B, Caslake M *et al*. Very low-calorie diet and 6 months of weight stability in type 2 diabetes: pathophysiological changes in responders and nonresponders. *Diabetes Care* 2016; **39**: 808–815.

33 Burkhauser RV, Cawley J. Beyond BMI: the value of more accurate measures of fatness and obesity in social science research. *J Health Econ* 2008; **27**: 519–529.

34 Zamora-Ros R, Knaze V, Luján-Barroso L, Romieu I, Scalbert A, Slimani N *et al*. Differences in dietary intakes, food sources and determinants of total flavonoids between Mediterranean and non-Mediterranean countries participating in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Br J Nutr* 2013; **109**: 1498–1507.

35 Lajous M, Bijon A, Fagherazzi G, Balkau B, Boutron-Ruault MC, Clavel-Chapelon F. Egg and cholesterol intake and incident type 2 diabetes among French women. *Br J Nutr* 2015; **114**: 1667–1673.

36 Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 785–794.

37 Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform* 2017; DOI: 10.1093/bib/bbx044. [Epub ahead of print]

## Supporting information

Additional Supporting Information may be found in the online version of this article:

**Table S1.** Predictive performance for FINDRISC at different cut-off values ranging from 10 to 20, for the full model (FM) and the simplified model (SM) with and without the interaction term (IT).

**Table S2.** Predictive performance for FINDRISC at different cut-off values ranging from 10 to 20, for the full model (FM), and the simplified model (SM) for the female population.

**Table S3.** Predictive performance for FINDRISC at different cut-off values ranging from 10 to 20, for the full model (FM), and the simplified model (SM) for the male population.

**Table S4.** Study participant demographics, FINDRISC related characteristics and percentage of missing values for incomplete cases belonging to the undiagnosed UT2DM, IFG, and NFG groups, defined by the corresponding FPG threshold values.

**Table S5.** Difference between mean values for complete and incomplete cases belonging to the undiagnosed UT2DM, IFG, and NFG groups, defined by the corresponding FPG threshold values.

**FIGURE 1** Boxplot visualization of predictive performance (AUC) based on 1000 bootstrap iterations for FINDRISC at different cut-off values ranging from 10 to 20, the full model (FM) and the simplified model (SM).

**FIGURE 2** Relative variable importance measured as the frequency of variable inclusion in the simplified impaired fasting glucose and undiagnosed Type 2 diabetes predictive models.

**FIGURE 3** Boxplot visualization of predictive performance [area under the receiver-operating characteristic curve (AUC)] by gender based on 1000 bootstrap iterations for FINDRISC at different cut-off values ranging from 10 to 20, the full model and the simplified model.

**Table 1** Study participant demographics and FINDRISC-related characteristics for individuals belonging to the undiagnosed Type 2 diabetes (UT2DM), impaired fasting glucose and normal fasting glucose groups, defined by the corresponding fasting plasma glucose thresholds

| | UT2DM<br>FPG ≥ 7.0 mmol/L<br>(*n* = 146, 7.0%) | IFG<br>6.1 ≤ FPG < 7.0<br>(*n* = 435, 21.0%) | NFG<br>FPG < 6.1 mmol/L<br>(*n* = 1,492, 72.0%) | *P** |
|---|---|---|---|---|
| Mean (SD) age | 59.8 (9.4) | 58.8 (9.4) | 53.5 (11.4) | < 0.001 |
| Age, *n* (%) | | | | |
| < 45 years | 11 (7.5) | 33 (7.6) | 351 (23.5) | |
| 45–54 years | 30 (20.5) | 107 (24.6) | 417 (27.9) | < 0.001 |
| 55–64 year | 60 (41.1) | 169 (38.9) | 457 (30.6) | |
| > 64 years | 45 (30.8) | 126 (29.0) | 267 (17.9) | |
| Weight, kg | 91.8 (20.8) | 86.2 (14.8) | 79.2 (15.6) | < 0.001 |
| Height, m | 168.6 (11.2) | 168.6 (9.1) | 167.8 (34.7) | < 0.001 |
| Mean (SD) BMI, kg/m$^2$ | 33.0 (16.2) | 30.3 (4.8) | 28.5 (8.0) | < 0.001 |
| BMI, *n* (%) | | | | |
| <25 kg/m$^2$ | 15 (10.3) | 49 (11.3) | 376 (25.2) | |
| 25–30 kg/m$^2$ | 42 (28.8) | 175 (40.2) | 633 (42.4) | < 0.001 |
| >30 kg/m$^2$ | 89 (61.0) | 211 (48.5) | 483 (32.4) | |
| Waist circumference measured below the ribs at the level of the navel | | | | |
| Mean (SD), cm | 104.2 (13.0) | 100.4 (12.7) | 94.2 (12.1) | < 0.001 |
| Men < 94 cm, *n* (%)<br>Women < 80 cm, *n* (%) | 16 (11.0) | 65 (14.9) | 357 (23.9) | |
| Men 94–102 cm, *n* (%)<br>Women 80–88 cm, *n* (%) | 30 (20.5) | 103 (23.7) | 377 (25.3) | < 0.001 |
| Men >102 cm<br>Women >88 cm | 100 (68.5) | 267 (61.4) | 758 (50.8) | |
| Women, *n* (%) | 49 (34) | 186 (43) | 888 (60) | < 0.001 |
| Do you usually have daily at least 30 min of physical activity? Yes, *n* (%) | 78 (53) | 164 (38) | 459 (31) | < 0.001 |
| Have you ever taken medication for high blood pressure? Yes, *n* (%) | 81 (55) | 210 (48) | 352 (24) | <0.001 |
| Have you ever been found to have high blood glucose? Yes, *n* (%) | 117 (80) | 234 (54) | 117 (8) | < 0.001 |
| How often do you eat fruit or vegetables? (0, daily; 1, less frequently) Daily, *n* (%) | 16 (11) | 42 (10) | 122 (8) | 0.376 |

| Have any of the members of your immediate family or other relatives been diagnosed with diabetes? | | | | |
|---|---|---|---|---|
| No, n (%) | 87 (59.6) | 259 (59.5) | 1072 (71.8) | |
| Yes; grandparent, aunt, uncle, or first cousin (but not own parent, brother, sister or child), n (%) | 10 (6.8) | 32 (7.4) | 99 (6.6) | < 0.001 |
| Yes; parent, brother, sister or child, n (%) | 49 (33.6) | 144 (33.1) | 321 (21.5) | |
| FPG, mmol/l | 8.1 (1.9) | 6.4 (0.2) | 5.3 (0.4) | < 0.001 |

FPG, fasting plasma glucose; IFG, impaired fasting glucose; NFG, normal fasting glucose.

*P value of the statistical test performed between the UT2DM, IFG and NFG groups, where for continuous values either parametric one-way ANOVA or non-parametric Kruskal–Wallis test was performed and for binary values a chi-squared test was performed.

**Table 2** Binomial logistic regression model for predicting undiagnosed Type 2 diabetes

| | Estimate | SE | z value | Odds ratio (95% CI) | Pr(>|z|) |
|---|---|---|---|---|---|
| (Intercept) | –8.49 | 1.04 | –8.19 | 0.00 (0.00–0.00) | < 0.001 |
| Age (years) | 0.03 | 0.01 | 3.26 | 1.03 (1.01–1.05) | 0.001 |
| Waist circumference (cm) | 0.03 | 0.01 | 3.52 | 1.03 (1.01–1.045) | < 0.001 |
| Female | –0.75 | 0.20 | –3.68 | 0.47 (0.32–0.70) | < 0.001 |
| Physical inactivity | 0.62 | 0.20 | 3.20 | 1.87 (1.27–2.74) | 0.001 |
| Blood sugar history | 2.54 | 0.22 | 11.47 | 12.64 (8.31–19.82) | < 0.001 |

**Table 3** Binomial logistic regression model for predicting impaired fasting glucose

| | Estimate | SE | z value | Odds ratio (95% CI) | Pr(>|z|) |
|---|---|---|---|---|---|
| (Intercept) | –6.48 | 0.66 | –9.89 | 0.00 (0–0.01) | < 0.001 |
| Age (years) | 0.04 | 0.01 | 6.23 | 1.04 (1.03–1.05) | < 0.001 |
| Waist circumference (cm) | 0.03 | 0.01 | 5.05 | 1.03 (1.02–1.04) | < 0.001 |
| Female | –0.78 | 0.13 | –6.13 | 0.46 (0.36–0.59) | < 0.001 |
| Taking hypertension drugs | 0.32 | 0.14 | 2.31 | 1.37 (1.05–1.80) | 0.021 |
| Blood sugar history | 2.58 | 0.14 | 19.01 | 13.17 (10.13–17.24) | < 0.001 |
| Diabetes in family | | | | | |
| Grandparent, aunt, uncle or first cousin | 0.57 | 0.24 | 2.34 | 1.77 (1.09–2.84) | 0.019 |
| Parent, brother, sister or own Child | 0.52 | 0.14 | 3.68 | 1.69 (1.28–2.22) | < 0.001 |