Title: Validation of Seven Type 2 Diabetes Mellitus Risk Scores in a Population-Based Cohort: The CoLaus Study.

Authors: Kraege V, Fabecic J, Marques-Vidal P, Waeber G, Méan M

Journal: The Journal of clinical endocrinology and metabolism

Year: 2020 Mar 1

Issue: 105

Volume: 3

DOI: 10.1210/clinem/dgz220

VALIDATION OF SEVEN TYPE 2 DIABETES MELLITUS RISK SCORES IN A POPULATION-BASED COHORT. THE COLAUS STUDY

Short title: 10-y validation of 7 type 2 diabetes risk scores

Vanessa Kraege[1]*, Janko Fabecic[2]*, Pedro Marques Vidal[1], Gérard Waeber[1] and Marie Méan[1]

[1]Department of Medicine, Internal medicine, Lausanne university hospital (CHUV), Lausanne, Switzerland

[2]Faculty of Biology and Medicine, University of Lausanne (UNIL), Lausanne, Switzerland

Authors' emails:

Vanessa Kraege: Vanessa.Kraege@chuv.ch

Janko Fabecic: Janko.Fabecic@unil.ch

Pedro Marques-Vidal: Pedro-Manuel.Marques-Vidal@chuv.ch

Gérard Waeber: Gerard.Waeber@chuv.ch

Marie Méan: Marie.Mean@chuv.ch

***Contributed equally; co-first authors; Vanessa Kraege is sole corresponding author.**

Address for correspondence and reprints

Vanessa Kraege
Office BH10-672
Department of Medicine, Internal Medicine
Lausanne University Hospital (CHUV)
46, rue du Bugnon
1011 Lausanne
Switzerland
Email: Vanessa.kraege@chuv.ch
Phone: +41 (0) 21 314 41 85
ORCID 0000-0002-6654-8154

Word count: abstract and text: 3146; abstract: 249

Number of tables: 4 **Figures**: 2 **References**: 18

Supplementary tables: 3

Keywords: type 2 diabetes mellitus; risk scores; prospective study; epidemiology

FUNDING

The CoLaus study was and is supported by research grants from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (grants 33CSCO-122661, 33CS30-139468 and 33CS30-148401). Gérard Waeber is supported by the Swiss National Science Foundation grant 32003B 173092. The funding sources had no involvement in the study design, data collection, analysis and interpretation, writing of the report, or decision to submit the article for publication.

DISCLOSURE SUMMARY

The authors report no conflict of interest

ABSTRACT

Aim: To assess the validity of seven type 2 diabetes mellitus (T2DM) risk scores in predicting the 10-year incidence of T2DM in a Swiss population-based study.

Methods: Prospective study including 5131 participants (55% women, age range 35 to 75 years) living in Lausanne, Switzerland. The baseline survey was conducted between 2003 and 2006 and average follow-up was 10.9 years. Five clinically-based (Balkau, Kahn clinical, Griffin, Swiss diabetes association and Findrisc) and two clinically and biologically based scores (Kahn CB and Wilson) were tested.

Results: 405 (7.9%) participants developed T2DM. The overall prevalence of participants at high risk ranged from 13.7% for the Griffin score to 43.3% for the Balkau score. Prevalence of participants at high risk among those who developed T2DM ranged from 34.6% for the Griffin score to 82.0% for the Kahn CB score. The Kahn CB score had the highest area under the ROC [value and 95% confidence interval: 0.866 (0.849-0.883)], followed by the Findrisc [0.818 (0.798-0.838)] while the Griffin score had the lowest [0.740 (0.718-0.762)]. Sensitivities and specificities were above 70%, except for the Griffin and the Kahn C scores (for sensitivity) and the Balkau score (for specificity). The numbers needed to screen ranged from 15.5 for the Kahn CB score to 36.7 for the Griffin score.

Conclusion: The Kahn (CB) and the Findrisc performed best of all scores. Findrisc could be used in an epidemiological setting, while the need of blood sampling for the Kahn (CB) score restricts its use to a more clinical setting.

PRÉCIS

We tested 7 risk scores regarding their ability to predict incident type 2 diabetes after a 10-year follow-up in a Swiss population cohort. Kahn clinical-biological and Findrisc scores performed best.

INTRODUCTION

The prevalence of type 2 diabetes mellitus (T2DM) is increasing worldwide (1). In Switzerland, one out of sixteen persons aged between 35 and 75 years has diabetes mellitus, and almost one third of diabetic subjects is unaware of their status (2) despite easy access to screening in local pharmacies and general practitioners. T2DM carries a considerable economic burden (3) as patients with T2DM are at higher risk of developing cardiovascular, neurological, renal and ophthalmic complications. Hence, early diagnosis of T2DM is of major importance as the outcome of the disease can be modified through medical care and lifestyle changes (4). The identification of subjects at high risk of developing T2DM might also be cost-effective by reducing the incidence of T2DM (5). Therefore, multiple predictive risk scores have been developed to detect patients at high risk of developing T2DM(6). A review conducted in 2011 (7) identified as many as 145 diabetes risk models or scores and suggested a monthly increase of this number. Such scores rely mainly on anamnestic and clinical information such as personal or family history and on simple measurements such as blood pressure, weight or waist. Some scores use additional blood markers such as fasting glucose, cholesterol and triglycerides. Among the 94 risk prediction models studied by Noble et al, 40 were based on biological variables. (7) While scores including blood markers tend to perform better, their cost is higher (6). Further, most scores have been validated in selected populations, and their application in other settings or populations is not warranted. Health professionals need to be able to rely on robust scores to easily identify the people at risk of diabetes and prevent exaggerated and costly screening in those less at risk.

In a previous study, we assessed the 5.5-year predictive capacity of seven T2DM risk scores in a prospective, population-based sample (8). As many scores were originally developed using longer follow-up times, a further validation was deemed necessary. Hence, in this study, we aimed to validate the seven above-mentioned T2DM risk scores over a 10-year follow-up. Our initial hypothesis was that the predictive capacity of each score would not change significantly in a longer follow-up.

METHODS

The Colaus study

The sampling procedure of the CoLaus cohort has been described previously (9) and further details can be obtained under www.colaus-psycolaus.ch. Briefly, the source population was defined as all subjects aged

between 35 and 75 years registered in the population register of the city of Lausanne. The register includes all subjects living in this city for more than 90 days. A simple, non-stratified random sample of 19'830 subjects (corresponding to 35% of the source population) was drawn and the selected subjects were invited to participate by letter. If no answer was obtained, a second letter was sent, and if still no answer was obtained, the subjects were contacted by phone. Recruitment began in June 2003 and ended in May 2006, enrolling 6733 total participants who underwent an interview, a physical exam, and a blood analysis. The first follow-up was performed between April 2009 and September 2012, 5.6 years on average (median 5.4 years, range 4.5-8.8) after the collection of baseline data; the second follow-up was performed between May 2014 and April 2017, 10.9 years on average (median 10.7, range 8.8-13.6) after the collection of baseline data. The information collected was similar to that collected in the baseline examination.

Data collection

Participants were asked to attend an outpatient clinic at the Centre Hospitalier Universitaire Vaudois in the morning, after an overnight fast. History of disease (personal and familiar), socio-demographic and lifestyle data were collected by questionnaire. Smoking was categorized as never, former and current; alcohol consumption was assessed by the number of alcoholic drinks (i.e. glasses of wine, cans of beer or shots of spirit) consumed over the last seven days and categorized into none, moderate (1-13 units/week), high (14-27 units/week) and very high (28+ units/week). Educational level was categorized as low (primary), middle (apprenticeship), upper middle (high school), and high (university) for highest completed level of education. Physical activity was defined by exercising at least twice a week for at least 20 minutes per session. Prescribed and over-the-counter medicines were collected by questionnaire.

Body weight and height were measured with participants barefoot and in light indoor clothes. Body weight was measured in kilograms to the nearest 100 g using a Seca® scale (Hamburg, Germany). Height was measured to the nearest 5 mm using a Seca® (Hamburg, Germany) height gauge. Waist circumference was measured mid-way between the lowest rib and the iliac crest using a non-stretchable tape and the average of two measurements was taken. Blood pressure (BP) and resting heart rate were measured thrice using an Omron® HEM-907 automated oscillometric sphygmomanometer after at least a 10-minute rest in a seated position. Different sized cuffs were available to take into account arm circumference and the average of the last two measurements was used.

Venous blood samples (50 mL) were drawn in the fasting state. Biological assays were performed at the clinical laboratory of the Lausanne university hospital within 2 hours of blood collection. Glucose was assessed by glucose dehydrogenase with a maximum inter- and intra-assay CV of 2.1% and 1.0%, respectively; high-density lipoprotein (HDL) cholesterol by CHOD-PAP + PEG + cyclodextrin (3.6%-0.9%); triglycerides by GPO-PAP (2.9%-1.5%), and uric acid by uricase-PAP (1.0%-0.5%). Glycated hemoglobin was measured by high performance liquid chromatography (HPCL) using Bio-Rad, D-10^{TM} system, with measurement range 3.8% (at 18 mmol/mol) to 18.5% (at 179 mmol/mol).

Diabetes risk scores

Seven T2DM risk scores were considered: 1) the FINDRISC (10) ; 2) the Swiss Diabetes Association (SDAS) (11); 3) the clinical and clinico-biological scores by Kahn et al., respectively (12) ; 4) the clinico-biological risk score by Wilson et al. (13); 5) the clinical risk score by Balkau et al. (14), and 6) the clinical risk score by Griffin et al. (15). The details of each score are summarized in **supplementary table 1** (16).

The FINDRISC score was derived from the 10-year follow-up FINRISK study consisting of 4435 participants (10) and consists of seven variables. The SDAS risk score is adapted from the FINDRISC, using familial history of diabetes as an additional variable. The risk scores by Kahn et al. (12) were derived in a cohort of 15'792 adults followed up during 10 years. The clinical risk score (C) consists of nine variables, while the clinic-biological risk score (CB) has four additional biological markers (glucose, triglycerides, high density lipoprotein, uric acid). The clinic-biological risk score of Wilson et al. (13) was derived from the Framingham Offspring Study, where 3140 participants were followed up for 8 years; the score consists of six variables among which three biological ones: glucose, triglycerides, and high density lipoprotein. The clinical risk score of Balkau et al. (14) was derived from the DESIR cohort, where 3817 participants were followed for nine years; it consists of four variables. Finally, the score of Griffin et al. (15) was derived from a cross-sectional study consisting of 1077 participants, and is composed of five clinical variables.

The FINDRISC, SDAS, Kahn (C and CB), Wilson and Balkau scores are based on a sum of allocated number of points per variable. For the FINDRISC and SDAS, nutritional variables and familial history of diabetes for second-degree parents were not available in our cohort at baseline; thus, the threshold was reduced by 1 point. The Griffin score uses a regression equation to calculate the probability of developing T2DM. As no threshold

had been proposed in the original study, a 37% probability was used to identify high-risk individuals, as proposed elsewhere (12).

Regarding ethnicity, the Wilson risk score was developed in a sample of 99% white and non-Hispanic subjects, and the Kahn scores were developed in a sample comprising 22% of black people, while no information regarding ethnicity was provided in Findrisc, Griffin and Balkau.

Outcome

The primary outcome was T2DM, defined as fasting blood glucose ≥7.0 mmol/l or taking insulin or oral antidiabetic medication as suggested by the American Diabetes association (17). HbA1C was not used in the definition of diabetes, as this variable was only available in the second follow-up.

Exclusion criteria

The original inclusion criteria were: 1) written informed consent; 2) willingness to take part in the examination and to provide blood samples; 3) French language ability. For this study, we added the following exclusion criteria: 1) diabetes (type 1 or 2) at baseline; 2) no follow-up (first or second); 3) missing variables to compute the scores and 4) no outcome data.

Statistical analysis

Statistical analyses were conducted using Stata version 15.1 for Windows (Stata Corp, College Station, Texas, USA). Participants' characteristics were expressed as number (percentage) for categorical variables or as average±standard deviation for continuous variables. Between-group comparisons were performed using chi-square or Fisher's exact test for categorical variables and student's t-test or Kruskal-Wallis test for continuous variables.

Risk scores were expressed as median and [interquartile range]. The diagnostic capacity of the different risk scores was assessed by the AUC [area under the ROC (receiver operating characteristic) curve] and corresponding 95% confidence intervals (CI). Comparisons of the AUC between scores were performed using the **roccomp** command of Stata. Sensitivity, specificity, positive and negative predictive values and their corresponding 95% CIs were computed using incident T2DM as gold standard. The number needed to screen (NNS) to detect one case of T2DM was computed as the total number of participants screened divided by the

number of detected T2DM cases (i.e. true positives). Statistical significance was assessed for a two-sided test with $p<0.05$.

RESULTS

Characteristics of participants

Of the initial 6733 participants, 5131 (76.2%) were retained for analysis. The reasons for exclusion are summarized in **figure 1** and the characteristics of the included and excluded participants are summarized in **supplementary table 2** (16). Included participants were younger, had lower waist, body mass index (BMI), prevalence of hypertension and family history of diabetes, and levels of fasting plasma glucose and uric acid than excluded ones. Included participants also had higher caffeine and alcohol consumption and higher levels of physical activity and HDL than excluded ones.

Figure 1: Exclusion criteria. Results expressed as number of participants and (percentage) using baseline number as denominator.

Incidence of T2DM

Upon second follow-up, 405 (7.9%) participants had developed T2DM. The baseline characteristics of the participants who developed T2DM or not are summarized in **table 1**. Participants who developed T2DM were more frequently male, older and former or current smokers, and had higher waist, BMI, prevalence of hypertension and of family history of diabetes, alcohol consumption and levels of fasting plasma glucose, triglycerides, HDL and uric acid, and more frequently received statins.

Performance of risk scores

For each risk score, the median score result and the prevalence of participants at high risk of developing T2DM, overall and according to development or not of T2DM, are provided in **table 2**. The overall prevalence of participants at high risk ranged from 13.7% for the Griffin score to 43.3% for the Balkau score. Prevalence of participants at high risk among those who developed T2DM ranged from 34.6% for the Griffin score to 82.0% for the Kahn CB score (**table 2**).

The AUC, sensitivity, specificity, positive and negative predictive values and the number needed to screen to detect one case of T2DM are summarized in **table 3** for each risk score. The AUCs for each score are also provided in **Figure 2** and the results of the bivariate comparisons of the AUCs are provided in **table 4**. The Kahn CB score had the highest AUC while the Griffin score had the lowest. Sensitivities and specificities were above 70%, except for the Griffin and the Kahn C scores (for sensitivity) and the Balkau score (for specificity). Positive predictive values were below 25%, while negative predictive values were above 90%. The numbers needed to screen ranged between 15.5 for the Kahn CB score to 36.7 for the Griffin score (**table 3**).

Figure 2: ROC curves of the seven diabetes risk scores.

DISCUSSION

Out of the seven diabetes risk scores evaluated, the two with the highest AUC were the Kahn et al. (CB), which includes biological variables, and the Findrisc, which is based on clinical data only. This finding is comparable to what was reported previously using a shorter follow-up period (5.5 vs. 10.9 years) (7). Importantly, our results confirm our hypothesis that the predictive value of a diabetes risk score does not change markedly when shorter follow-up times than the ones in the original validation are used.

Diabetes risk scores

The Kahn (CB) score showed the best metrics, a finding already reported previously (8). Several reasons might explain this performance: first, it was developed using a large sample size (12'729) and included four blood markers. Although the inclusion of blood markers might improve the predictive capacity of the score, it also makes it more expensive to use either for mass screening or for everyday clinical use. Hence, its applicability in settings with limited health resources is reduced. The Wilson score also includes biological markers but, contrary to the Kahn CB, its predictive capacity was rather low and its AUC was comparable to the clinical version of the Kahn score. One explanation could be that Kahn adapts the intervals of seven of its parameters according to sex, whereas Wilson only adapts HDL-C levels. The weighting of each parameter is also different and Wilson was developed on a cohort four times smaller than Kahn.

The Findrisc score ranked second highest among all scores. Contrary to the Kahn (CB) score, it is based solely on clinical data and can thus be applied in screening campaigns or in communities with limited health resources. Importantly, although the complete version of the Findrisc score could not be used in this study, still,

the reduced version performed well, suggesting that the performance of the complete version, if computable, could even be better. Further, the SDAS, which is based on the Findrisc, also showed an adequate performance, albeit with a lower AUC than the Findrisc. The likely reason is an arbitrary addition of 5 points for family history of diabetes on the SDAS, which does not seem to improve its performance. The Balkau score had the lowest number of components. Although this small number of items might facilitate its applicability in public health or in clinical practice, its predictive capacity was modest, and it led to a very high number of participants classified as being "at risk". Finally, the Griffin score had the lowest prediction capacity. A probable explanation is that it was developed in a cross-sectional setting, whereas the other scores were developed in a prospective setting. Overall, our results indicate that a prospective setting is paramount to adequately derive and validate a risk prediction score. Indeed, most scores perform well in the populations they were developed in, but their predictive value drops when applied to a different cohort. Hence, externally validating predictive scores on different populations is essential to assess their generalizability and performance.

Comparison with previous findings

Our initial hypothesis was that the predictive capacity of each score would not change significantly in a longer follow-up. The ranking of the scores according to performance is the same for 5- or 10-year follow-ups for the best 3 performers (Kahn CB, Findrisc, Swiss Diabetes association) **(Supplementary Table 3)**(16)**.** Overall, the AUC are slightly lower using 10-year follow-up. Follow-up period transportability is a well-known challenge, as it demands that scores maintain accuracy when predictions are tested over longer versus shorter follow-up periods.(18) In our study, AUC changes are probably due to time-sensitive variables such as anthropometric data.

Strengths and limitations

The main strength of our study is that we used a follow-up time similar to the one the scores were developed for. The second strength is the relatively large sample size, which provided an adequate number of incident events.

We also acknowledge several limitations. No nutritional data was collected at baseline; hence, we had to adapt the threshold of the Findrisc and SDAS scores by reducing the threshold by one unit, a procedure also

performed in our previous study (8). Still, despite this limitation, both the Findrisc and the SDAS scores performed better than other scores. The CoLaus study includes a majority (90%) of Caucasian subjects living in a high-income country and urban setting, who volunteered to participate in the study; hence, full representativeness and generalizability to other ethnicities or settings might not be warranted. Still, no differences were found between the CoLaus cohort and the target population regarding gender and zip code distribution, while subjects aged <65 were underrepresented. Also, it would be important to replicate our study in other cohorts to validate the robustness of our findings. Approximately 15% of our baseline cohort was lost during the first and second follow-up, which might have reduced the number of incident T2DM events. Still, this affected all scores equally and would not change the conclusions of the study. Finally, statins are diabetogenic (19) and participants with incident DM were more frequently prescribed statins at baseline. In this study, we did not assess the predictive value of statin therapy, as the objective was to validate existing DM risk scores. Nevertheless, statins should be considered as a possible T2DM determinant when developing T2DM risk scores.

CONCLUSION

The Kahn (CB) and the Findrisc performed best of all scores. Findrisc could be used in an epidemiological setting, while the need of blood sampling for the Kahn (CB) score restricts its use to a more clinical setting.

LIST OF ABBREVIATIONS

AUC	Area under the ROC (receiver operating characteristic) curve
BMI	Body mass index
BP	Blood pressure
HDL cholesterol	High-density lipoprotein cholesterol
HPCL	High performance liquid chromatography
Kahn C	Kahn clinical
Kahn CB	Kahn clinic-biological
ROC	Receiver operating characteristic
SDAS	Swiss diabetes association score
T2DM	Type 2 diabetes mellitus

DECLARATIONS

Ethics approval and consent to participate

The institutional Ethics Committee of the University of Lausanne, which afterwards became the Ethics Commission of Canton Vaud (www.cer-vd.ch), approved the baseline CoLaus study (reference 16/03, decisions of 13th January and 10th February 2003); the approval was renewed for the first (reference 33/09, decision of 23rd February 2009) and the second (reference 26/14, decision of 11th March 2014) follow-up. The study was performed in agreement with the Helsinki declaration and its former amendments, and in accordance with the applicable Swiss legislation. All participants gave their signed informed consent before entering the study.

Consent for publication

Not applicable.

Availability of data and material

The datasets analysed during the current study are not publicly available as cohort participants did not consent to this. Non-identifiable individual-level data are available for researchers who seek to answer questions related to health and disease in the context of research projects who meet the criteria for data sharing by research committees. Please follow the instructions at https://www.colaus-psycolaus.ch/ for information on how to submit an application for gaining access to CoLaus/PsyCoLaus data.

Competing interests

The authors report no conflict of interest.

Funding

The CoLaus study was and is supported by research grants from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (grants 33CSCO-122661, 33CS30-139468 and 33CS30-148401). Gérard Waeber is supported by the Swiss National Science Foundation grant 32003B 173092. The funding sources had no involvement in the study design, data collection, analysis and interpretation, writing of the report, or decision to submit the article for publication.

Authors' contributions

VK wrote part of the article and revised it for important intellectual content. JF wrote most of the article. GW revised the article for important intellectual content. PMV had full access to the data and is the guarantor of the study. MM was the initiator of the study and revised the article for important intellectual content.

ACKNOWLEDGEMENTS

Not applicable

REFERENCES

1. Ogurtsova K, da Rocha Fernandes JD, Huang Y, et al. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Res Clin Pract.* 2017;128:40-50.
2. Kaiser A, Vollenweider P, Waeber G, Marques-Vidal P. Prevalence, awareness and treatment of type 2 diabetes mellitus in Switzerland: the CoLaus study. *Diabet Med.* 2012;29(2):190-197.
3. Huber CA, Schwenkglenks M, Rapold R, Reich O. Epidemiology and costs of diabetes mellitus in Switzerland: an analysis of health care claims data, 2006 and 2011. *BMC Endocr Disord.* 2014;14:44.
4. American Diabetes A. Standards of medical care in diabetes--2014. *Diabetes Care.* 2014;37 Suppl 1:S14-80.
5. Chamnan P, Simmons RK, Khaw KT, Wareham NJ, Griffin SJ. Estimating the potential population impact of stepwise screening strategies for identifying and treating individuals at high risk of Type 2 diabetes: a modelling study. *Diabet Med.* 2012;29(7):893-904.
6. Abbasi A, Peelen LM, Corpeleijn E, et al. Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. *Bmj.* 2012;345:e5900.
7. Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. *Bmj.* 2011;343:d7163.
8. Schmid R, Vollenweider P, Bastardot F, Waeber G, Marques-Vidal P. Validation of 7 type 2 diabetes mellitus risk scores in a population-based cohort: CoLaus study. *Arch Intern Med.* 2012;172(2):188-189.
9. Firmann M, Mayor V, Vidal PM, et al. The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc Disord.* 2008;8:6.
10. Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care.* 2003;26(3):725-731.
11. Diabète Suisse. Test diabète. 2017; https://www.diabetesschweiz.ch/fr/le-diabete/test-diabete/. Accessed February 28,, 208.
12. Kahn HS, Cheng YJ, Thompson TJ, Imperatore G, Gregg EW. Two risk-scoring systems for predicting incident diabetes mellitus in U.S. adults age 45 to 64 years. *Ann Intern Med.* 2009;150(11):741-751.
13. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB, Sr. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern Med.* 2007;167(10):1068-1074.
14. Balkau B, Lange C, Fezeu L, et al. Predicting diabetes: clinical, biological, and genetic approaches: data from the Epidemiological Study on the Insulin Resistance Syndrome (DESIR). *Diabetes Care.* 2008;31(10):2056-2061.
15. Griffin SJ, Little PS, Hales CN, Kinmonth AL, Wareham NJ. Diabetes risk score: towards earlier detection of type 2 diabetes in general practice. *Diabetes Metab Res Rev.* 2000;16(3):164-171.
16. Kraege V, Fabecic J, Marques Vidal P, Waeber G, Méan M. Data from: Validation of Seven Type 2 Diabetes Mellitus Risk Scores in a Population-Based Cohort. The CoLaus Study. Zenodo Data Repository 2013. Deposited 28 October 2019. 10.5281/zenodo.3520776.
17. American Diabetes A. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2018. *Diabetes Care.* 2018;41(Suppl 1):S13-S27.
18. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med.* 1999;130(6):515-524.
19. Laakso M, Kuusisto J. Diabetes Secondary to Treatment with Statins. *Curr Diab Rep.* 2017;17(2):10.

FIGURE TITLES AND LEGENDS

Figure 1: exclusion criteria. Results expressed as number of participants and (percentage) using baseline number as denominator.

Figure 2: ROC curves of the seven diabetes risk scores.

356 **Tables**

357 **Table 1**: Factors associated with incident diabetes, 10.9-year follow-up, CoLaus study, Lausanne, Switzerland

	No diabetes	Incident diabetes	P-value
N	4726	405	
Gender (women)	2669 (56.5)	149 (36.8)	<0.001
Age (years)	51.3 ± 10.5	55.9 ± 9.8	<0.001
Clinical data			
Weight (kg)	71.3 ± 13.9	82.5 ± 14.2	<0.001
Body mass index (kg/m^2)	25.1 ± 4.0	28.7 ± 4.3	<0.001
Waist (cm)	86.7 ± 12.1	98.7 ± 11.5	<0.001
Hypertension (%)	2073 (43.9)	294 (72.6)	<0.001
Resting heart rate (bpm)	67 ± 9	69 ± 10	0.008
Family history of diabetes (%)	980 (20.7)	137 (33.8)	<0.001
High glucose (≥ 6.1 mmol/L)	1339 (28.4)	318 (78.7)	<0.001
Prescribed steroids (%)	20 (0.4)	3 (0.7)	
Prescribed statins (%)	411 (8.7)	70 (17.3)	<0.001
Lifestyle data			
Alcohol consumption (%)			<0.001
None	1254 (26.5)	107 (26.4)	
1-13 UA/week	2733 (57.8)	212 (52.4)	
14-27 UA/week	601 (12.7)	58 (14.3)	
≥28 UA/week	138 (2.9)	28 (6.9)	
Smoking categories (%)			0.001
Never	1987 (42.0)	132 (32.6)	
Former	1519 (32.1)	151 (37.3)	
Current	1220 (25.8)	122 (30.1)	
Caffeinated drinks consumption (%)			0.135
None	289 (6.1)	25 (6.2)	
1-3 u/day	3075 (65.1)	264 (65.2)	
4-6 u/day	1158 (24.5)	89 (22.0)	
>6 u/day	204 (4.3)	27 (6.7)	
Physical activity (%)	2687 (56.9)	181 (44.7)	<0.001
Blood markers			
Fasting plasma glucose (mmol/l)	5.3 ± 0.5	6.0 ± 0.6	<0.001
Triglycerides (mmol/l)	1.27 ± 0.96	1.99 ± 1.99	<0.001
HDL (mmol/l)	1.68 ± 0.44	1.46 ± 0.37	<0.001
Uric acid (µmol/l)	303 ± 81	353 ± 87	<0.001

358 UA, units of alcohol. Results expressed as average ± standard deviation or as number of participants and
359 (percentage). Between-group comparisons performed using student's t-test or Kruskal-Wallis test (§) for
360 continuous variables and chi-square or Fisher's exact test (†) for categorical variables.

361

362

Table 2: Bivariate analysis of diabetic risk scores, 10-year follow-up (2003-2006 to 2014-2017) CoLaus study, Lausanne, Switzerland

	All	No diabetes	Incident diabetes
N	5131	4726	405
Griffin et al.			
Score	11 [3 - 35]	10 [3 - 32]	38 [17 - 66]
High risk (%)	702 (13.7)	562 (11.9)	140 (34.6)
Balkau et al.			
Score	2 [1 - 3]	2 [1 - 3]	3 [3 - 4]
High risk (%)	2222 (43.3)	1905 (40.3)	317 (78.3)
Kahn et al (C)			
Score	25 [12 - 40]	23 [12 - 38]	44 [32 - 57]
High risk (%)	1443 (28.1)	1184 (25.1)	259 (64.0)
Wilson et al			
Probability	3 [3 - 4]	3 [3 - 4]	7 [4 - 18]
High risk (%)	1552 (30.3)	1236 (26.2)	316 (78.0)
Swiss Diabetes association			
Score	7 [4 - 12]	7 [3 - 11]	14 [11 - 17]
High risk (%)	1586 (30.9)	1282 (27.1)	304 (75.1)
Findrisc			
Score	6 [3 - 10]	6 [3 - 9]	12 [9 - 14]
High risk (%)	1388 (27.1)	1096 (23.2)	292 (72.1)
Kahn et al (CB)			
Score	19 [9 - 33]	18 [9 - 30]	47 [35 - 56]
High risk (%)	1426 (27.8)	1094 (23.2)	332 (82.0)

Results expressed as median [interquartile range] or as number of participants and (percentage). Between-group (diabetes and non-diabetes) comparisons using Kruskal-Wallis test or chi-square test. All differences are significant at $p<0.001$

Table 3: Diagnostic performance of diabetic risk scores, 10-year follow-up (2003-2006 to 2014-2017) CoLaus study, Lausanne, Switzerland

	AUC	Sensitivity §	Specificity §	Positive predictive value §	Negative predictive value §	Number needed to screen §§
Griffin et al.	0.740 (0.718 - 0.762)	34.6 (29.9 - 39.4)	88.1 (87.2 - 89.0)	19.9 (17.0 - 23.1)	94.0 (93.3 - 94.7)	36.7
Balkau et al.	0.750 (0.728 - 0.771)	78.3 (73.9 - 82.2)	59.7 (58.3 - 61.1)	14.3 (12.8 - 15.8)	97.0 (96.3 - 97.6)	16.2
Kahn et al (C)	0.777 (0.755 - 0.798)	64.0 (59.1 - 68.6)	74.9 (73.7 - 76.2)	17.9 (16.0 - 20.0)	96.0 (95.4 - 96.6)	19.8
Wilson et al.	0.788 (0.765 - 0.811)	78.0 (73.7 - 82.0)	73.8 (72.6 - 75.1)	20.4 (18.4 - 22.5)	97.5 (96.9 - 98.0)	16.2
Swiss Diabetes association	0.807 (0.787 - 0.828)	75.1 (70.6 - 79.2)	72.9 (71.6 - 74.1)	19.2 (17.3 - 21.2)	97.2 (96.5 - 97.7)	16.9
Findrisc	0.818 (0.798 - 0.838)	72.1 (67.5 - 76.4)	76.8 (75.6 - 78.0)	21.0 (18.9 - 23.3)	97.0 (96.4 - 97.5)	17.6
Kahn et al (CB)	0.866 (0.849 - 0.883)	82.0 (77.9 - 85.6)	76.9 (75.6 - 78.0)	23.3 (21.1 - 25.6)	98.0 (97.5 - 98.5)	15.5

Results expressed as value (95% confidence interval). § Of high vs. low risk; §§ to detect one diabetic case.

Table 4: Results of the bivariate comparison of the AUCs between diabetes risk scores.

		Model 2					
		Griffin	Kahn C	Kahn CB	Balkau	SDAS	Findrisc
Model 1	Wilson	**<0.001**	0.358	*<0.001*	**0.004**	*0.047*	*0.002*
	Griffin		*<0.001*	*<0.001*	0.319	*<0.001*	*<0.001*
	Kahn C			*<0.001*	**<0.001**	*0.001*	*<0.001*
	Kahn CB				**<0.001**	**<0.001**	**<0.001**
	Balkau					*<0.001*	*<0.001*
	SDAS						*0.029*

Comparisons were performed using the **roccomp** command of Stata. SDAS, Swiss diabetes association score. In bold, model 1 performs better. In italics, model 2 performs better.