BMJ Open

External validation of three diabetes prediction scores in a Spanish cohort: does adding high risk for depression improve the validation of the FINDRISC score (FINDRISC-MOOD)?

Miguel Salinero-Fort [ID] ,[1,2] Jose M Mostaza-Prieto,[3] Carlos Lahoz-Rallo,[3] Juan Cárdenas-Valladolid,[4,5] Victor Iriarte-Campo,[1] Eva Estirado-Decabo,[3] Francisca Garcia-Iglesias,[3] Teresa Gonzalez-Alegre,[3] Belen Fernandez-Puntero,[6] Victor M Cornejo-Del Rio,[7] Vanesa Sanchez-Arroyo,[3] Concesa Sabín-Rodríguez,[8] Silvia López-López,[8] Paloma Gómez-Campelo,[9] Belen Taulero-Escalera,[10] Fernando Rodriguez-Artalejo,[11,12,13] Francisco Javier San Andrés-Rebollo,[14] Carmen De Burgos-Lunar,[15] SPREDIA-2 Group

Correspondence to
Dr Miguel Salinero-Fort;
miguel.salinero@salud.madrid.org

ABSTRACT

Objectives To evaluate the external validity of the FINDRISC, DESIR and ADA risk scores for the prediction of diabetes in a Spanish population aged >45 years and to test the possible improvement of FINDRISC by adding a new variable of high risk of depression when Patient Health Questionnaire-9 (PHQ-9) questionnaire score ≥10 (FINDRISC-MOOD).

Design Prospective population-based cohort study.

Setting 10 primary healthcare centres in the north of the city of Madrid (Spain).

Participants A total of 1242 participants without a history of diabetes and with 2-hour oral glucose tolerance test (OGTT) plasma glucose <200 mg/dL (<11.1 mmol/L) were followed up for 7.3 years (median) using their electronic health records (EHRs) and telephone contact.

Primary and secondary outcome measures Diabetes risk scores (FINDRISC, DESIR, ADA), PHQ-9 questionnaire and 2-hour-OGTT were measured at baseline. Incident diabetes was defined as treatment for diabetes, fasting plasma glucose ≥126 mg/dL (≥7.0 mmol/L), new EHR diagnosis or self-reported diagnosis. External validation was performed according to optimal cut-off, sensitivity, specificity and Youden Index. Comparison between diabetes risk scores, including FINDRISC-MOOD (original FINDRISC score plus five points if PHQ-9 ≥10), was measured by area under the receiver operating characteristic curve (AUROC).

Results During follow-up, 104 (8.4%; 95% CI, 6.8 to 9.9) participants developed diabetes and 185 had a PHQ-9 score ≥10. The AUROC values were 0.70 (95% CI, 0.67 to 0.72) for FINDRISC-MOOD and 0.68 (95% CI, 0.65 to 0.71) for the original FINDRISC. The AUROCs for DESIR and ADA were 0.66 (95% CI, 0.63 to 0.68) and 0.66 (95% CI, 0.63 to 0.69), respectively. There were no significant differences in AUROC between FINDRISC-MOOD and the other scores.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The SPREDIA-2 study provides a well-characterised, longitudinal, prospective cohort of adults free of diabetes mellitus at baseline in the north of the city of Madrid (Spain).

⇒ Participants were followed up using electronic medical records from general practitioners and telephone calls to obtain robust and complete information on their health status.

⇒ The diabetes risk scores studied, and the Patient Health Questionnaire-9 are accurate tools widely used in primary care to measure the high risk of having or developing diabetes and depression, respectively.

⇒ We did not have an exact date for the self-reported cases of diabetes mellitus, which limits our ability to perform event-free analysis.

⇒ The higher socioeconomic level of the north of Madrid may limit the applicability of the results to areas in the south.

Conclusions The results of FINDRISC-MOOD were like those of the other risk scores and do not allow it to be recommended for clinical use.

INTRODUCTION

Diabetes risk scores such as FINDRISC,[1] DESIR[2] and ADA[3] can be used to identify individuals who may require laboratory tests such as fasting plasma glucose, HbA1c and oral glucose tolerance test (OGTT). These risk scores can help stratify populations according to prognosis and enable the implementation of interventions to prevent cardiovascular

complications and delay or halt the onset of diabetes mellitus (DM).

To ensure the accuracy of these risk scores, they are usually validated in the settings where they will be used (external validation), and the most appropriate cut-off point in terms of sensitivity and specificity is determined. In some cases, the FINDRISC score has been adapted to make it easier to use without compromising accuracy by removing some of the original variables[4 5] or adding new ones.[6]

Other risk scores, such as the German Diabetes Risk Score,[7] have been developed that include psychosocial variables such as perceived chronic stress. However, no diabetes risk score has included a history or high risk of depressive disorders, except the study of Atlantis et al.[8] A recent meta-analysis[9] showed that people with major depressive disorder have a higher risk of developing type 2 diabetes mellitus (T2DM). The Lifelines cohort study showed a relationship between high risk of depression and diabetes incidence in individuals with pre-diabetes at baseline after 9 years of follow-up.[10] Therefore, we think that the incorporation of depression risk, defined as a Patient Health Questionnaire-9 (PHQ-9) score of 10 or higher, into the original FINDRISC score could improve the external validation of FINDRISC to identify the risk of T2DM.

The present study aims to validate three diabetes risk scores (FINDRISC, DESIR, ADA) in the Spanish population to predict the incidence of T2DM after a long follow-up period (median 7.3 years) and to test the value of adding high risk of depression to the FINDRISC score in terms of area under the receiver operating characteristic curve (AUROC).

MATERIAL AND METHODS
Design
This study was conducted as part of a broader project funded by the Spanish Instituto de Salud Carlos III (PI 1500259). It included the Screening PRE-diabetes and type 2 DIAbetes (SPREDIA-2) study, which has been described in detail elsewhere.[11] SPREDIA-2 is a population-based prospective cohort study, in which baseline visits were scheduled from July 2010 to March 2014.

Population
Baseline visit
The study population comprised a random sample of 2553 subjects living in the north of the city of Madrid (Spain) in an area served by 10 primary healthcare centres. Of these, 1592 (62.4%) agreed to participate, and 1426 had not been previously diagnosed with DM.

Recruitment was divided into three stages:
1. Potential participants were sent a letter signed by their general practitioner explaining the aims of the study and inviting them to participate.
2. Subjects were contacted by telephone to clarify any doubts and, if interested, were given an appointment to be assessed.

3. The patient attended the assessment at the Carlos III Hospital outpatient clinic after an overnight fast.

A fasting blood sample was taken on arrival at the outpatient clinic to determine levels of glucose, creatinine, uric acid, HbA1c, serum insulin, lipids and lipoproteins. Immediately after blood sampling, all subjects not previously diagnosed with DM underwent an OGTT with 75 g of anhydrous glucose in a total fluid volume of 300 mL. A second blood sample was taken 2 hours later. The measurement questionnaires were as follows: diabetes risk scores (FINDRISC, DESIR and ADA); the PHQ-9[12]; the 14-item Questionnaire to assess adherence to the Mediterranean diet (PREDIMED)[13] and the 12-item Short-Form Health Survey.[14] A full clinical history was taken. Alcohol consumption was measured as the number of units of alcohol per week.

Follow-up
Participants were followed up for a median of 7.3 years between the baseline visit and 31 December 2019 using their general practitioners' electronic health records (EHRs). The EHRs had previously been validated[15] and used in epidemiological studies.[16] The participants were also contacted by telephone during the last year of follow-up to ascertain whether they were alive and, if so, to record health status, including the incidence of T2DM or cardiovascular events. The interview was conducted by a researcher trained in obtaining medical data by telephone. The study flow chart is shown in figure 1.

Measurement tools and definitions of criteria
FINDRISC risk score
The Finnish Diabetes Risk Score is one of the most widely used.[1] It includes eight variables (anthropometric and lifestyle), namely, age, body mass index (BMI), waist circumference, family history of diabetes, use of blood pressure medication, history of high blood glucose levels, daily physical activity and daily intake of vegetables, fruits and berries. FINDRISC assesses the likelihood of developing T2DM over the next 10 years. The score ranges from 0 to 26 points, and the usual cut-off is 15. A risk score of 0–14 points indicates a low-moderate risk of diabetes (1%–17% risk of diabetes over 10 years), and 15–20 points indicates a high risk of diabetes (33% risk of diabetes over 10 years). A score of 20–26 points indicates a very high risk of diabetes (50% risk of diabetes over 10 years).[17]

DESIR risk score
DESIR was designed by Balkau et al[18] in the French population. The component variables differ by sex. In women, the variables include waist circumference (cm), family history of diabetes and arterial hypertension, while for men, they include waist circumference (cm), current smoking status and arterial hypertension. The waist circumference categories differ by sex (for women, <70, 70–79, 80–89, ≥90; for men, <80, 80–89, 90–99, ≥100). The score ranges from 0 to 5 points (a higher score means a higher risk).
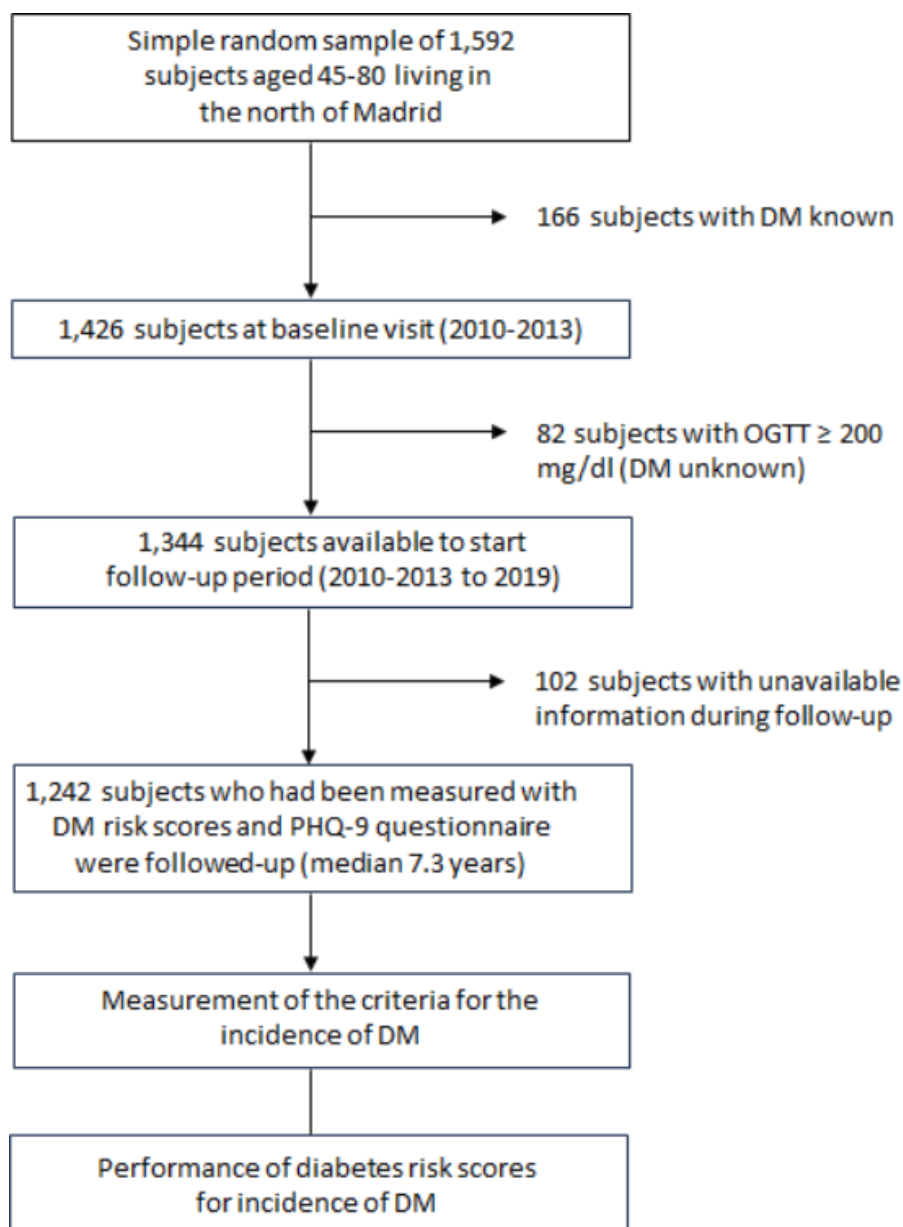
Figure 1 Flow chart of people included in the study. DM, diabetes mellitus; OGTT, oral glucose tolerance test; PHQ-9, Patient Health Questionnaire-9.

ADA risk score

The American Diabetes Association risk score[3] was developed based on the US population older than 20 years without DM to identify individuals at high risk for DM or pre-diabetes. It includes the following variables: age, sex, race, weight, height, family history of DM, history of gestational DM, history of arterial hypertension and physical activity. The total score ranges from 0 to 11. A score of 5 or higher indicates a high risk of DM.

PHQ-9

This validated and reliable scale has been used in many research studies.[19] It is used to assist primary care practitioners in diagnosing depression and monitoring treatment. The question put is 'Over the last 2 weeks, how often have you been bothered by any of the following

problems?', with four response options: (0) Not at all; (1) Several days; (2) More than half the days and (3) Nearly every day. The total PHQ-9 score is assessed by adding together the scores for all nine items. Higher scores on this measure indicate greater depression. Scores are categorised into five levels of severity: minimal=0–4; mild=5–9; moderate=10–14; moderately severe=15–19 and severe=20–27. The optimal cut-off for major depression disorder is 10.[20] The IPD meta-analysis confirmed a cut-off of 10 as yielding the best balance of sensitivity (85%) and specificity (85%).[21] It takes approximately 2–5 min to administer. Additionally, it can be self-administered.[22]

FINDRISC-MOOD risk score

This is an adaptation of the original FINDRISC risk score in which five new points are added if participants scored a

positive PHQ-9 for depression (≥10 points). In contrast, no points are added if the PHQ-9 is negative. We decided that five points was appropriate based on the value of the beta coefficient obtained in the PHQ-9 after adjustment for the FINDRISC score in the logistic multivariate analysis for the prediction of diabetes (beta coefficient=0.632). This coefficient was multiplied by 9, which was the smallest common multiplication factor possible to obtain a sensitive score.

Anthropometric measurements were obtained using standard methods.[23]

Diagnostic criteria

Incidence of DM (gold standard): Incident cases of diabetes were identified by treatment for diabetes (n=16), fasting plasma glucose ≥126 mg/dL (n=21), new diagnosis in the EHR (International Classification of Primary Care, Second Edition T90 code) (n=6) or self-reported diagnosis in the telephone interview (n=1). In addition, 60 individuals met more than one criterion.

Metabolic syndrome: It was defined according to ATPIII diagnostic criteria.[24]

Statistical methods

The statistical analyses were conducted using IBM SPSS Statistics for Windows, V.26.0 (IBM Corp, Armonk, New York, USA) and MedCalc for Windows, V.15.8 (MedCalc Software, Ostend, Belgium). The sociodemographic and clinical characteristics of the study population at baseline are presented as frequencies and percentages for categorical variables and as means and SDs for continuous variables. Between-group comparisons were performed using a χ^2 or Fisher's exact test for categorical variables and a t-test or Kruskal-Wallis test for continuous variables.

To calculate the sample size, the following assumptions were accepted: an α error of 0.05, a precision rate of 9% in a bilateral contrast, for an estimated specificity rate of 80% and an estimated incidence of DM of 6%; the total sample size required was 1217 participants.

The performance of the diabetes risk scores was assessed using the following indicators: sensitivity, specificity, positive and negative predictive values, Youden Index defined as (sensitivity+specificity−1) and positive and negative likelihood ratios. All scores were calculated using incident T2DM as the gold standard. Statistical significance was set at p<0.05 for a two-tailed test.

The discriminative accuracy of the different risk scores was assessed and expressed as the AUROC and corresponding 95% CIs. AUROCs were compared between scores using MedCalc software.

The cut-off points for the risk scores to identify incidence of T2DM were determined by the point with the shortest distance to the upper left corner of the ROC, as calculated using the Youden Index.

Patient and public nvolvement

It was not appropriate or possible to involve patients or the public in the design, or conduct, or reporting, or dissemination plans of our research.

RESULTS

Characteristics of the study population

Of the original 1426 participants, 1344 (94.2%) met the criteria for follow-up. Of these, 1242 (92.4%) were finally contactable via EHRs or telephone interviews. The reasons for exclusion are summarised in figure 1. The main characteristics of the participants stratified by sex are shown in table 1. At baseline, the mean age of the study population was 62 years. A high percentage had a family history of DM (31.6%) and a low prevalence of cardiovascular disease (eg, coronary artery disease, stroke and peripheral artery disease (3.1%, 2% and 0.8%, respectively)). One-third of the population met the criteria for current smoking, arterial hypertension and metabolic syndrome, and approximately 50% met the criteria for consumption ≥1 units of alcohol per week. Approximately half of the patients had dyslipidaemia, and one in five had regular or poor self-perceived health. The percentage of participants with a high score on the PHQ-9 was 11.6%; this was significantly higher among women. In terms of current treatment, nearly one in four participants were taking statins, and one in five were taking renin-angiotensin system blockers.

Incidence of T2DM

During 7.3 years (median) of follow-up, 104 participants (8.4%; 95% CI, 6.8% to 9.9%) developed T2DM. Table 2 shows the differences between participants with and without incident T2DM for the main characteristics examined. The risk factors for which values were significantly higher in the group with incident T2DM were hypertension, metabolic syndrome, BMI, waist circumference, systolic and diastolic blood pressures, fasting plasma glucose, OGTT result, HbA1c, impaired glucose tolerance, self-administered PHQ-9 score and diabetes risk scores (FINDRISC, DESIR and ADA).

Patients treated with renin-angiotensin system blockers were statistically significantly more likely to be in the diabetes group.

Diabetes risk scores usually include questions about lifestyle, diet and medical history. Online supplemental table 1 shows the differences between the two groups. The group without incident T2DM was more likely than the group with incident T2DM to perform at least 30 min of physical activity, to eat vegetables, fruit or berries every day, to have never taken medication for high blood pressure, to have never been diagnosed with high blood sugar and to have never had gestational diabetes.

Performance of diabetes risk scores

The performance of the FINDRISC score is shown in table 3. The best cut-off point was >14, achieving a sensitivity of 47.12% (95% CI, 37.2% to 57.2%), specificity of 81.37% (95% CI, 79% to 83.6%) and positive likelihood ratio of 2.53 (95% CI, 2.0 to 3.21). The AUROC was 0.68 (95% CI, 0.65 to 0.71).

The FINDRISC-MOOD diabetes risk score, that is, the original FINDRISC plus five points if PHQ-9 >10, showed

Table 1 Baseline characteristics of the population studied stratified by sex

	All (N=1242)	Men (N=506)	Women (N=736)	P value
Age, mean (SD)	61.8 (6.1)	61.6 (6.3)	62 (5.9)	0.214
University studies, n (%)	395 (31.9)	222 (44)	173 (23.5)	**<0.001**
Current smoking, n (%)	438 (35.3)	238 (47)	200 (27.2)	**<0.001**
Alcohol consumption, n (%)	598 (48.2)	349 (69.1)	249 (33.9)	**<0.001**
Sedentarism, n (%)	494 (40.0)	204 (40.6)	290 (39.7)	0.750
Family history of DM, n (%)	392 (31.6)	148 (29.4)	244 (33.2)	0.154
Family history of arterial hypertension, n (%)	616 (49.8)	204 (40.5)	412 (56.3)	**<0.001**
Hypertension, n (%)	400 (32.2)	185 (36.6)	215 (29.3)	**0.007**
Treated hypertension, n (%)	360 (90.0)	169 (91.4)	191 (88.9)	0.499
Coronary artery disease, n (%)	38 (3.1)	32 (6.3)	6 (0.8)	**<0.001**
Peripheral artery disease, n (%)	10 (0.8)	8 (1.6)	2 (0.3)	**0.011**
Stroke, n (%)	25 (2.0)	13 (2.6)	12 (1.6)	0.248
Chronic atrial fibrillation, n (%)	30 (2.4)	17 (3.4)	13 (1.8)	0.071
Dyslipidaemia, n (%)	562 (45.3)	226 (44.8)	336 (45.7)	0.754
GFRe ≤60 mL/min/1.73 m^2, n (%)	14 (1.1)	10 (2.0)	4 (0.5)	**0.019**
Metabolic syndrome, n (%)	396 (31.9)	176 (34.8)	220 (29.9)	0.069
Renin-angiotensin system blockers, n (%)	268 (21.6)	140 (27.7)	128 (17.4)	<0.001
Statins, n (%)	300 (24.2)	125 (24.7)	175 (23.8)	0.708
Aspirin, n (%)	76 (6.1)	50 (9.9)	26 (3.5)	**0.004**
Anticoagulants. n (%)	24 (1.9)	13 (2.6)	11 (1.5)	0.176
Waist circumference, cm, mean (SD)	93.9 (11.9)	99.9 (9.7)	89.7 (11.5)	**<0.001**
BMI, kg/m^2, mean (SD)	28 (4.5)	28.4 (3.7)	27.8 (5)	**0.010**
Systolic blood pressure, mm Hg, mean (SD)	123.4 (16.8)	121 (17.3)	126.8 (15.5)	**<0.001**
Diastolic blood pressure, mm Hg, mean (SD)	76.8 (9.8)	78.5 (9.7)	75.7 (9.7)	**<0.001**
Plasma glucose 0 hour, mg/dL, mean (SD)	100.6 (10.4)	103.6 (10.5)	98.5 (9.8)	**<0.001**
Plasma glucose 2 hours, mg/dL, mean (SD)	115.2 (30.6)	120.3 (31.2)	111.6 (29.6)	**<0.001**
HbA1c, %, mean (SD)	5.7 (0.3)	5.6 (0.4)	5.7 (0.3)	**0.006**
Psychological status				
PHQ-9 score, mean (SD)	4.2 (4.4)	3.2 (4)	4.9 (4.6)	**<0.001**
PHQ-9 ≥10 (high risk), n (%)	144 (11.6)	39 (7.7)	105 (14.3)	**<0.001**
Health perception (self-reported)				
Excellent, n (%)	36 (2.9)	23 (4.6)	13 (1.8)	**<0.001**
Very good, n (%)	248 (20.1)	121 (24.0)	17.5 (127)	
Good, n (%)	702 (57)	287 (56.8)	415 (57.2)	
Regular, n (%)	235 (19.1)	71 (14.1)	164 (22.6)	
Bad, n (%)	10 (0.8)	3 (0.6)	7 (1.0)	
Diabetes risk scores				
FINDRISC, mean (SD)	11.1 (4.3)	10.8 (4.2)	11.2 (4.3)	0.117
DESIR, mean (SD)	11.8 (1.2)	11.9 (1.2)	11.8 (1.2)	0.777
ADA, mean (SD)	5.1 (1.5)	5.8 (1.3)	4.6 (1.5)	**<0.001**

Bold means p <0.05
ADA, American Diabetes Association risk score; BMI, body mass index; DESIR, Data from an Epidemiological Study on the Insulin Resistance Syndrome; DM, diabetes mellitus; FINDRISC, Finish Diabetes Risk Score; PHQ-9, Patient Health Questionnaire-9.

Table 2 Baseline characteristics of the study population stratified by incidence of diabetes/no diabetes (median 7.3 years of follow-up)

	Diabetes, incidence (N=104)	No diabetes, incidence (N=1138)	P value
Sex, male, n (%)	51 (49.0)	455 (40.0)	0.072
Age, mean (SD)	62.5 (6.0)	61.7 (6.1)	0.223
University studies, n (%)	24 (23.1)	371 (32.7)	**0.018**
Current smoking, n (%)	22 (21.2)	179 (15.7)	0.211
Alcohol consumption, n (%)	51 (49.0)	547 (48.2)	0.862
Sedentarism, n (%)	46 (44.2)	448 (39.6)	0.363
Family history of DM, n (%)	36 (35.0)	356 (31.3)	0.450
Family history of arterial hypertension, n (%)	51 (49.5)	565 (49.9)	0.945
Hypertension, n (%)	49 (47.1)	351 (30.9)	**0.001**
Coronary artery disease, n (%)	2 (1.9)	36 (3.2)	0.482
Peripheral artery disease, n (%)	0 (0.0)	10 (0.9)	0.338
Stroke, n (%)	2 (1.9)	23 (2.0)	0.943
Chronic atrial fibrillation, n (%)	3.8 (4)	2.3 (26)	0.325
Dyslipidaemia, n (%)	47 (45.2)	515 (45.3)	0.984
GFRe ≤60 mL/min/1.73 m^2, n (%)	0 (0.0)	14 (1.2)	0.255
Metabolic syndrome, n (%)	64 (61.5)	332 (29.2)	**<0.001**
Renin-angiotensin system blockers, n (%)	37 (35.6)	231 (20.3)	**<0.001**
Statins, n (%)	25 (24.0)	275 (24.2)	0.977
Aspirin, n (%)	3 (2.9)	73 (6.4)	0.151
Anticoagulants, n (%)	1 (1.0)	13 (1.1)	0.867
Waist circumference, cm, mean (SD)	101.3 (11.1)	93.2 (11.8)	**<0.001**
BMI, kg/m^2, mean (SD)	30.6 (5.0)	27.8 (4.4)	**<0.001**
Systolic blood pressure, mm Hg, mean (SD)	129.3 (18.1)	122.8 (16.6)	**<0.001**
Diastolic blood pressure, mm Hg, mean (SD)	79.4 (9.4)	76.6 (9.8)	**0.005**
Fasting plasma glucose, mg/dL, mean (SD)	112.2 (14.3)	99.5 (9.3)	**<0.001**
Plasma glucose 2 hours, mg/dL, mean (SD)	136.5 (34.4)	113.2 (29.5)	**<0.001**
Impaired glucose tolerance, n (%)	52 (50.0)	209 (18.4)	**<0.001**
HbA1c, %, mean (SD)	5.9 (0.4)	5.6 (0.3)	**<0.001**
Mediterranean diet score (PREDIMED), mean (SD)	14.0 (0.1)	13.9 (0.9)	0.429
Psychological status			
PHQ-9 ≥10 (high risk), n (%)	22 (21.2)	122 (10.7)	**0.001**
Health perception (self-reported)			
Excellent, n (%)	4 (3.9)	32 (2.8)	**0.001**
Very good, n (%)	12 (11.5)	236 (20.9)	
Good, n (%)	59 (56.7)	642 (57)	
Regular, n (%)	24 (23.1)	211 (18.7)	
Bad, n (%)	5 (4.8)	6 (0.5)	
Diabetes risk scores			
FINDRISC, mean (SD)	13.9 (4.7)	10.8 (4.1)	**<0.001**
DESIR, mean (SD)	12.4 (1.4)	11.8 (1.2)	**<0.001**
ADA, mean (SD)	5.9 (1.4)	5.0 (1.5)	**<0.001**

Bold means p <0.05
ADA, American Diabetes Association risk score; BMI, body mass index; DESIR, Data from an Epidemiological Study on the Insulin Resistance Syndrome; DM, diabetes mellitus; FINDRISC, Finish Diabetes Risk Score.

Table 3 Performance of the FINDRISC diabetes risk score in predicting incident diabetes mellitus after 7.3 years (median) of follow-up

Criterion	Sensitivity	Specificity	Youden Index	LR+	LR−	PPV	NPV
≥2	100	0	0	1.00		8.40	
>2	100	1.49	0.01	1.02	0.00	8.52	100
>3	100	4.57	0.05	1.05	0.00	8.77	100
>4	98.08	7.03	0.05	1.05	0.27	8.82	97.56
>5	97.12	9.84	0.07	1.08	0.29	8.99	97.39
>6	96.15	14.76	0.11	1.13	0.26	9.37	97.66
>7	91.35	21.62	0.13	1.17	0.40	9.66	96.46
>8	87.50	30.14	0.18	1.25	0.41	10.30	96.34
>9	81.73	38.84	0.21	1.34	0.47	10.92	95.86
>10	74.04	47.80	0.22	1.42	0.54	11.51	95.26
>11	65.38	57.12	0.23	1.52	0.61	12.27	94.73
>12	58.65	65.73	0.24	1.71	0.63	13.57	94.55
>13	50.96	75.04	0.26	2.04	0.65	15.77	94.35
>14	**47.12**	**81.37**	**0.28**	**2.53**	**0.65**	**18.83**	**94.38**
>15	33.65	87.26	0.21	2.64	0.76	19.50	93.48
>16	26.92	91.12	0.18	3.03	0.80	21.75	93.15
>17	23.08	93.76	0.17	3.70	0.82	25.33	93.00
>18	17.31	96.92	0.14	5.63	0.85	34.01	92.74
>19	11.54	98.42	0.10	7.29	0.90	40.11	92.39
>20	6.73	98.77	0.05	5.47	0.94	33.41	92.03
>21	5.77	99.38	0.05	9.38	0.95	46.05	92.00
>22	4.81	99.56	0.04	10.94	0.96	50.06	91.94
>23	3.85	99.74	0.04	14.59	0.96	57.59	91.88
>24	3.85	100	0.04		0.96	100	91.90
>25	0.00	100	0		1.00		91.60

LR+, positive likelihood ratio; LR−, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.

the same cut-off point as FINDRISC, although the sensitivity and specificity were more balanced (56.7% and 76.7%, respectively) (table 4). The negative predictive value was 95.08%, slightly higher than that of the original FINDRISC score. The AUROC increased to 0.70 (95% CI, 0.67 to 0.72).

Finally, the performance of DESIR and ADA is shown in online supplemental table 2. The best cut-off points were >12 and >5, respectively. The AUC of the ROC was almost equal, with 0.66 (95% CI, 0.63 to 0.68) for DESIR and 0.661 (95% CI, 0.63 to 0.69) for ADA.

The results of the bivariate comparisons of the AUROCs are provided in online supplemental table 3. The differences between values were not statistically significant. The greatest difference was between the FINDRSC-MOOD and DESIR scores (z=1.841, p=0.0657).

Mortality
There were 24 deaths during follow-up, that is, a crude mortality rate of 2.57 (95% CI, 1.64 to 3.82) per 1000 person-years. When stratified by FINDRISC score, those with a FINDRISC ≤14 had a crude mortality rate of 2.44 (95% CI, 1.45 to 3.86) per 1000 person-years and those with a score >14 had a crude mortality rate of 3.03 (95% CI, 1.11 to 6.61) per 1000 person-years. The rate ratio was 1.24 (95% CI, 0.41 to 3.27), p=0.628.

The crude mortality rate among PHQ-9 negative participants was 2.42 (95% CI, 1.48 to 3.73), and the crude mortality rate among PHQ-9 positive participants was 3.70 (95% CI, 1.01 to 9.48). The rate ratio between the two groups was 1.53 (95% CI, 0.38 to 4.57), p=0.434.

Similar results were observed with FINDRISC-MOOD. Those with a score below 14 had a crude mortality rate of 2.32 (95% CI, 1.32 to 3.76) per 1000 person-years, and those with a score >14 had a crude mortality rate of 3.27 (95% CI, 1.41 to 6.45) per 1000 person-years. The rate ratio was 1.41 (95% CI, 0.52 to 3.50), p=0.427. In this sense, a FINDRISC-MOOD score >14 indicates a slight increase in crude mortality compared with the same score in the traditional FINDRISC questionnaire,

Table 4 Performance of the FINDRISC-MOOD questionnaire in predicting incident diabetes mellitus after 7.3 years (median) of follow-up

Criterion	Sensitivity	Specificity	Youden Index	LR+	LR–	PPV	NPV
≥2	100	0	0	1		8.40	
>2	100	1.41	0.01	1.01	0	8.51	100
>3	100	4.22	0.04	1.04	0	8.74	100
>4	99.04	6.59	0.06	1.06	0.15	8.86	98.68
>5	98.08	9.31	0.07	1.08	0.21	9.02	98.14
>6	97.12	13.53	0.11	1.12	0.21	9.34	98.09
>7	93.27	19.68	0.13	1.16	0.34	9.62	96.96
>8	89.42	27.86	0.17	1.24	0.38	10.21	96.64
>9	84.62	35.68	0.20	1.32	0.43	10.77	96.20
>10	78.85	43.76	0.23	1.4	0.48	11.39	95.76
>11	72.12	52.99	0.25	1.53	0.53	12.33	95.40
>12	67.31	61.42	0.29	1.74	0.53	13.79	95.35
>13	61.54	70.47	0.32	2.08	0.55	16.05	95.23
>14	**56.73**	**76.71**	**0.33**	**2.44**	**0.56**	**18.26**	**95.08**
>15	42.31	83.04	0.25	2.49	0.69	18.62	94.01
>16	34.62	86.82	0.21	2.63	0.75	19.41	93.54
>17	27.88	90.07	0.18	2.81	0.8	20.48	93.16
>18	21.15	93.76	0.15	3.39	0.84	23.71	92.84
>19	18.27	95.96	0.14	4.52	0.85	29.31	92.76
>20	11.54	96.75	0.08	3.55	0.91	24.56	92.26
>21	11.54	98.07	0.10	5.97	0.9	35.41	92.36
>22	10.58	98.42	0.09	6.69	0.91	38.05	92.31
>23	8.65	99.21	0.08	10.94	0.92	50.10	92.21
>24	4.81	99.82	0.05	27.36	0.95	71.02	91.96
>25	1.92	99.82	0.02	10.94	0.98	49.40	91.73

Bold means the values of the better cutpoint
LR+, positive likelihood ratio; LR–, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.

probably due to the increased mortality risk in PHQ-9 positive individuals, as we found (online supplemental figure 1).

DISCUSSION

It is widely accepted that it is not possible to create a perfect prediction rule. In this sense, when using questionnaires to predict future diabetes, achieving a sensitivity of at least 80% (20% false negatives) is an adequate result. This approach is optimal because diabetes risk scores are cost-effective and can be administered in a clinical or community setting.

The original FINDRISC score of ≥9 showed this high sensitivity (80%). However, high sensitivity can also lead to low specificity, which means a higher rate of false positives. This misdiagnosis can lead to unnecessary pharmacological treatment and psychological harm to patients. Therefore, obtaining laboratory measurements to identify false positives is essential for questionnaires with low

specificity. The Youden Index showed an adequate cut-off of >14 for the original FINDRISC to maintain an appropriate balance between sensitivity and specificity.

To compare the discriminatory ability of each diabetes risk score, we calculated the AUROC. Several authors have attempted to improve the AUROC of predictive risk scores by adding laboratory measures such as fasting plasma glucose,[25 26] HbA1c alone[27] and HbA1c plus fasting plasma glucose.[28] This strategy improves discriminative ability but hinders the goal of rapid and noninvasive prediction using easily measured variables. The predictive diabetes risk scores could be considered as a pre-screening tool to identify individuals who would benefit from the measurement of fasting plasma glucose or HbA1c. This approach represents a relatively cost-effective screening programme, as highlighted by the International Diabetes Federation.[29]

Our strategy to improve the AUROC of original FIND-RISC was the addition of five points when the PHQ-9

questionnaire score is greater than 10 points.[21] However, the result was not statistically significantly different from the original FINDRISC score. Although, the AUROC of FINDRISC-MOOD presented an essential advantage given that it reaches a value of 0.70, the minimum to consider that a test provides sufficient discrimination. The p value for comparing the AUROCs was less than 0.10 between FINDRISC-MOOD and ADA (difference between areas=0.0374; p=0.0808), and very close to 0.05 between FINDRISC-MOOD and DESIR (difference between areas=0.0413; p=0.0657).

Our study was not designed to detect small differences between AUROCs. We would have required a larger sample size to achieve that. However, if there had been sufficient power to detect differences, the question that arises is whether the PHQ-9 questionnaire can be used in conjunction with the FINDRISC questionnaire in primary clinical practice. The PHQ-9 questionnaire is self-administered and does not require much time. This fact is particularly interesting because depression should be screened for regularly, as per the recommendations of the US Preventive Services Task Force.[30] Additionally, the ADA recommends screening for depression in patients with diabetes.[31] Depression is associated with up to a 65% increased risk of DM,[32 33] making the addition of the PHQ-9 questionnaire a useful tool for assessing DM risk. Moreover, the PHQ-9 can be considered a first-line tool for diagnosing depression in primary care settings due to its ease of administration, good acceptability and sensitivity in detecting depression.[30]

Multiple authors have created predictive models for diabetes using variables commonly included in most diabetes risk questionnaires. Atlantis et al developed one such model, which also accounted for the high risk of anxiety and affective disorders by including the Kessler Psychological Distress Scale (K10). Atlantis et al compared two models, one with the K10 and one without. However, they did not use hypothesis testing to compare the two AUROC values. They found that the differences were of low magnitude.[8] Also, the AUROC of their model was similar to that of FINDRISC-MOOD. Our study took a different approach. We tested the performance of three diabetes risk scores and found that adding five points to the original FINDRISC score when the PHQ-9 score was higher than 10 points did not statistically improve the AUROC.

Owing to differences in lifestyle and the prevalence of chronic diseases such as obesity, pre-diabetes and diabetes between communities and ethnicities, it is often necessary to calculate the optimal cut-off point for diabetes risk scores for each country, even state or region. From a statistical point of view, the best cut-off point is the one that achieves a higher Youden Index, although as previously mentioned, this strategy can be modified depending on the purpose of screening. According to the highest Youden Index, the best cut-off points for the FINDRISC and FINDRISC-MOOD questionnaires are the same (≥14).

The cutoffs used change over time and according to study site. In Europe, the original FINDRISC designed by Linström and Tuomilheto in Finland[1] showed an optimal cutoff ≥9 in two consecutive cohorts after 10 years of follow-up with the following performance parameters (Youden Index, sensitivity, specificity and AUC) 0.59, 78%, 81% and 0.85 in the 1987 cohort and 0.53, 77%, 76% and 0.87 in the 1992 cohort. In the Norwegian general population aged over 20 years, the best FINDRISC cut-off, according to the highest Youden Index, was ≥11 (sensitivity, 73%; specificity, 67%) after 10 years of follow-up.

The study by Alssema et al with subjects from the HOORN[34 35] (n=1434), PREVEND[36 37] (n=2713) and MORGEN[38 39] (n=863) cohorts in the Netherlands included six variables from FINDRISC (age, BMI, waist circumference, use of blood pressure medication, history of high blood glucose and family history of DM). The total score of the original FINDRISC ranged from 0 to 22, and it yielded two acceptable cutoffs for each cohort. Values (Youden Index, sensitivity and specificity) were as follows: HOORN cohort, 0.28, 52% and 76% for a cutoff ≥10 and 0.26, 84% and 42% for a cutoff ≥7; PREVEND cohort, 0.28, 43% and 85% for a cutoff ≥10 and 0.42, 78% and 64% for a cutoff ≥7; MORGEN cohort, 0.30, 47% and 83% for a cutoff ≥10 and 0.27, 75% and 52% for a cutoff ≥7.[40] The second study by Alssema et al[41] included the previous variables used in 2008, as well as sex and smoking, with an optimal cut-off of ≥7 (Youden Index, 0.39; sensitivity, 76%; specificity, 63%).

In Spain, a population-based prospective study performed in the town of Pizarra (Málaga) followed 824 individuals for 6 years to evaluate the performance of FINDRISC. The best prediction of the risk of incident T2DM was found in subjects with a FINDRISC cut-off of 9 with an AUROC of 0.75. No information was provided on sensitivity, specificity or the Youden Index.[42]

In contrast to most studies, the AUROC of the Pizarra study was closer to our results, probably owing to a common lifestyle pattern and a progressive decrease in discriminative ability, as is relatively common in contemporary studies.

Our study has several limitations. First, the incidence of DM was not measured by the same method as at baseline (OGTT). However, because we used four sources of data (self-reported diagnosis, diagnosis by general practitioner, baseline plasma glucose levels and use of hypoglycaemic medications), which are consistent with other authors,[43 44] information bias is unlikely. Second, the FINDRISC score estimates the risk of participants aged 35–64 years developing T2DM within 10 years; our study included some patients aged 65 years or older with a follow-up period of 7.3 years. This may have changed the accuracy of the results. Third, the diabetes risk scores in our study were performed on a representative population in the north of the city of Madrid. However, lifestyle, diet and prevalence of obesity may differ from the rest of the city, given the lower gross domestic product in the southern area. This may limit extrapolation to these areas or other Spanish cities. Finally, we did not have an exact

date for the self-referred cases, which limits our ability to perform event-free analysis.

Our study also has important strengths. The study population reflects the average risk of DM, without a large number of people at high risk of developing T2DM, as in other studies conducted in hospital samples,[27 45] patients with chronic infectious diseases[46] and regions with a high prevalence of sedentary lifestyle and obesity.[27] Another strength is that the extension of the original FINDRISC, FINDRISC-MOOD, is an easy-to-use tool that does not require additional investment in health professionals, as the PHQ-9 is typically self-administered.[47]

CONCLUSION

The results of FINDRISC-MOOD were similar to those of the other risk scores and do not allow it to be recommended for clinical use. Nevertheless, further studies can be conducted in different settings to address the observed limitations and improve its accuracy.

Author affiliations
[1]FIIBAP, Madrid, Spain
[2]Frailty, patterns of multimorbidity and mortality in the community-dwelling elderly population, IdiPAZ, Madrid, Spain
[3]Medicina Interna, Hospital Carlos III, Madrid, Spain
[4]Gerencia Asistencial de Atención Primaria, Comunidad de Madrid Servicio Madrileno de Salud, Madrid, Spain
[5]Enfermería, Universidad Alfonso X El Sabio, Villanueva de la Canada, Spain
[6]Servicio de Bioquímica, Hospital Carlos III, Madrid, Spain
[7]Supervisión de Enfermería, Hospital Carlos III, Madrid, Spain
[8]Unidad de Día, Hospital Carlos III, Madrid, Spain
[9]Fundación de Investigación, La Paz University Hospital Health Research Institute, Madrid, Spain
[10]Foundation for Research and Biomedical Innovation of Primary Care of the Community of Madrid (FIIBAP), Madrid, Spain
[11]Department of Preventive Medicine and Public Health, Universidad Autonoma de Madrid, Madrid, Spain
[12]CIBERESP, Madrid, Spain
[13]IMDEA-Food, CEI UAM+CSIC, Madrid, Spain
[14]Centro de Salud Las Calesas, SERMAS, Madrid, Spain
[15]Medicina Preventiva, Hospital Clinico San Carlos, Madrid, Spain

ORCID iD
Miguel Salinero-Fort http://orcid.org/0000-0002-9450-4116

REFERENCES

1 Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 2003;26:725–31.
2 Balkau B, Lange C, Fezeu L, *et al*. Predicting diabetes: clinical, biological, and genetic approaches. *Diabetes Care* 2008;31:2056–61.
3 American Diabetes Association. Classification and diagnosis of diabetes: standards of medical care in diabetes-2020 . *Diabetes Care* 2020;43:S14–31.
4 Bernabe-Ortiz A, Perel P, Miranda JJ, *et al*. Diagnostic accuracy of the Finnish diabetes risk score (FINDRISC) for Undiagnosed T2Dm in Peruvian population. *Prim Care Diabetes* 2018;12:517–25.
5 Bergmann A, Li J, Wang L, *et al*. A simplified Finnish diabetes risk score to predict type 2 diabetes risk and disease evolution in a German population. *Horm Metab Res* 2007;39:677–82.
6 Robinson CA, Agarwal G, Nerenberg K. Validating the CANRISK Prognostic model for assessing diabetes risk in Canada's multi-ethnic population. *Chronic Dis Inj Can* 2011;32:19–31.
7 Nübel J, Du Y, Baumert J, *et al*. Perceived chronic stress is associated with the German diabetes risk score among adults without known diabetes in Germany. *Psychosom Med* 2023;85:332–40.
8 Atlantis E, Ghassem Pour S, Girosi F. Incremental predictive value of screening for anxiety and depression beyond current type 2 diabetes risk models: a prospective cohort study. *BMJ Open* 2018;8:e018255.
9 Vancampfort D, Mitchell AJ, De Hert M, *et al*. TYPE 2 diabetes in patients with major depressive disorder: a meta-analysis of prevalence estimates and predictors. *Depress Anxiety* 2015;32:763–73.
10 Deschênes SS, McInerney A, Nearchou F, *et al*. Prediabetes and the risk of type 2 diabetes: investigating the roles of depressive and anxiety symptoms in the Lifelines cohort study. *Diabet Med* 2023;40:e15061.
11 Salinero-Fort MÁ, de Burgos-Lunar C, Mostaza Prieto J, *et al*. Validating prediction scales of type 2 diabetes mellitus in Spain: the SPREDIA-2 population-based prospective cohort study protocol. *BMJ Open* 2015;5:e007195.
12 Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *Primary Care Evaluation of Mental Disorders Patient Health Questionnaire JAMA* 1999;282:1737–44.
13 Martínez-González MA, García-Arellano A, Toledo E, *et al*. A 14-item Mediterranean diet assessment tool and obesity indexes among high-risk subjects: the PREDIMED trial. *PLoS One* 2012;7:e43134.
14 Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34:220–33.
15 de Burgos-Lunar C, Salinero-Fort MA, Cárdenas-Valladolid J, *et al*. Validation of diabetes mellitus and hypertension diagnosis in computerized medical records in primary health care. *BMC Med Res Methodol* 2011;11:146.
16 Gil Montalbán E, Ortiz Marrón H, López-Gay Lucio-Villegas D, *et al*. Validity and concordance of electronic health records in primary care (AP-Madrid) for surveillance of diabetes mellitus [PREDIMERC study]. *Gac Sanit* 2014;28:393–6.
17 Riise HKR, Graue M, Igland J, *et al*. Prevalence of increased risk of type 2 diabetes in general practice: a cross-sectional study in Norway. *BMC Prim Care* 2023;24:151.
18 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;31:2056–61.
19 Martin A, Rief W, Klaiberg A, *et al*. Validity of the brief patient health questionnaire mood scale (PHQ-9) in the general population. *Gen Hosp Psychiatry* 2006;28:71–7.
20 Kroenke K, Spitzer RL, Williams JBW, *et al*. The patient health questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. *Gen Hosp Psychiatry* 2010;32:345–59.
21 Levis B, Benedetti A, Thombs BD, *et al*. Depression screening data (DEPRESSD) collaboration. accuracy of patient health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. *BMJ* 2019;365:l1476.
22 Costantini L, Pasquarella C, Odone A, *et al*. Screening for depression in primary care with patient health Questionnaire-9 (PHQ-9): A systematic review. *J Affect Disord* 2021;279:473–83.
23 Cameron N. Essential anthropometry: baseline anthropometric methods for human biologists in laboratory and field situations. *Am J Hum Biol* 2013;25:291–9.
24 Ford ES, Giles WH, Dietz WH. Prevalence of the metabolic syndrome among US adults: findings from the third national health and nutrition examination survey. *JAMA* 2002;287:356–9.
25 Bozorgmanesh M, Hadaegh F, Ghaffari S, *et al*. A simple risk score effectively predicted type 2 diabetes in Iranian adult population: population-based cohort study. *Eur J Public Health* 2011;21:554–9.
26 Guasch-Ferré M, Bulló M, Costa B, *et al*. A risk score to predict type 2 diabetes mellitus in an elderly Spanish mediterranean population at high cardiovascular risk. *PLoS ONE* 2012;7:e33437.
27 Meijnikman AS, De Block CEM, Verrijken A, *et al*. Predicting type 2 diabetes mellitus: a comparison between the FINDRISC score and the metabolic syndrome. *Diabetol Metab Syndr* 2018;10:12.
28 Hippisley-Cox J, Coupland C. Development and validation of Qdiabetes-2018 risk prediction algorithm to estimate future risk of type 2 diabetes: cohort study. *BMJ* 2017;359:j5019.
29 International Diabetes Federation Guideline Development Group. Global guideline for type 2 diabetes. *Diabetes Res Clin Pract* 2014;104:1–52.
30 Siu AL, Bibbins-Domingo K, Grossman DC, *et al*. Screening for depression in adults: US preventive services task force recommendation statement. *JAMA* 2016;315:380–7.
31 American Diabetes Association. Lifestyle management. *Diabetes Care* 2017;40:S33–43.
32 Mezuk B, Eaton WW, Albrecht S, *et al*. Depression and type 2 diabetes over the LifeSpan: a meta-analysis. *Diabetes Care* 2008;31:2383–90.
33 Campayo A, de Jonge P, Roy JF, *et al*. Depressive disorder and incident diabetes mellitus: the effect of characteristics of depression. *Am J Psychiatry* 2010;167:580–8.
34 Mooy JM, Grootenhuis PA, de Vries H, *et al*. Prevalence and determinants of glucose intolerance in a Dutch Caucasian population. *Diabetes Care* 1995;18:1270–3.
35 de Vegt F, Dekker JM, Jager A, *et al*. Relation of impaired fasting and postload glucose with incident type 2 diabetes in a Dutch population: the Hoorn study. *JAMA* 2001;285:2109.
36 Pinto-Sietsma SJ, Janssen WMT, Hillege HL, *et al*. Urinary albumin excretion is associated with renal functional abnormalities in a nondiabetic population. *J Am Soc Nephrol* 2000;11:1882–8.
37 Verhave JC, Hillege HL, Burgerhof JGM, *et al*. The association between atherosclerotic risk factors and renal function in the general population. *Kidney Int* 2005;67:1967–73.
38 Mensink M, Corpeleijn E, Feskens EJM, *et al*. Study on Lifestyle-intervention and impaired glucose tolerance Maastricht (SLIM): design and screening results. *Diabetes Res Clin Pract* 2003;61:49–58.
39 van Dam RM, Boer JM, Feskens EJ, *et al*. Parental history of diabetes modifies the association between abdominal Adiposity and hyperglycemia. *Diabetes Care* 2001;24:1454–9.
40 Alssema M, Feskens EJM, Bakker SJL, *et al*. Finnish questionnaire reasonably good Predictor of the incidence of diabetes in the Netherlands. *Ned Tijdschr Geneeskd* 2008;152:2418–24.
41 Alssema M, Vistisen D, Heymans MW, *et al*. The evaluation of screening and early detection strategies for type 2 diabetes and impaired glucose tolerance (DETECT-2) update of the Finnish diabetes risk score for prediction of incident type 2 diabetes. *Diabetologia* 2011;54:1004–12.
42 Soriguer F, Valdés S, Tapia MJ, *et al*. Validation of the FINDRISC (Finnish diabetes risk score) for prediction of the risk of type 2 diabetes in a population of Southern Spain. *Med Clin (Barc)* 2012;138:371–6.
43 NCD Risk Factor. Effects of diabetes definition on global surveillance of diabetes prevalence and diagnosis: a pooled analysis of 96 population-based studies with 331,288 participants. *Lancet Diabetes Endocrinol* 2015;3:624–37.
44 Kazemian P, Shebl FM, McCann N, *et al*. Evaluation of the Cascade of diabetes care in the United States, 2005-2016. *JAMA Intern Med* 2019;179:1376–85.
45 Jiménez-Lucena R, Rangel-Zúñiga OA, Alcalá-Díaz JF, *et al*. Circulating miRNAs as predictive biomarkers of type 2 diabetes mellitus development in coronary heart disease patients from the CORDIOPREV study. *Mol Ther Nucleic Acids* 2018;12:146–57.
46 Galaviz KI, Schneider MF, Tien PC, *et al*. Expanding the Finnish diabetes risk score for predicting diabetes incidence in people living with HIV. *AIDS Res Hum Retroviruses* 2021;37:373–9.
47 Kroenke K, Spitzer RL, Williams JBW. The patient health Questionnaire-2: validity of a two-item depression Screener. *Med Care* 2003;41:1284–92.