Non-invasive risk scores for prediction of type 2 diabetes (EPIC-InterAct): a validation of existing models

Andre Pascal Kengne, Joline W J Beulens, Linda M Peelen, Karel G M Moons, Yvonne T van der Schouw, Matthias B Schulze, Annemieke M W Spijkerman, Simon J Griffin, Diederick E Grobbee, Luigi Palla, Maria-Jose Tormo, Larraitz Arriola, Noël C Barengo, Aurelio Barricarte, Heiner Boeing, Catalina Bonet, Françoise Clavel-Chapelon, Laureen Dartois, Guy Fagherazzi, Paul W Franks, José María Huerta, Rudolf Kaaks, Timothy J Key, Kay Tee Khaw, Kuanrong Li, Kristin Mühlenbruch, Peter M Nilsson, Kim Overvad, Thure F Overvad, Domenico Palli, Salvatore Panico, J Ramón Quirós, Olov Rolandsson, Nina Roswall, Carlotta Sacerdote, María-José Sánchez, Nadia Slimani, Giovanna Tagliabue, Anne Tjønneland, Rosario Tumino, Daphne L van der A, Nita G Forouhi, Stephen J Sharp, Claudia Langenberg, Elio Riboli, Nicholas J Wareham

Summary

Background The comparative performance of existing models for prediction of type 2 diabetes across populations has not been investigated. We validated existing non-laboratory-based models and assessed variability in predictive performance in European populations.

Methods We selected non-invasive prediction models for incident diabetes developed in populations of European ancestry and validated them using data from the EPIC-InterAct case-cohort sample (27 779 individuals from eight European countries, of whom 12 403 had incident diabetes). We assessed model discrimination and calibration for the first 10 years of follow-up. The models were first adjusted to the country-specific diabetes incidence. We did the main analyses for each country and for subgroups defined by sex, age (<60 years vs ≥60 years), BMI (<25 kg/m² vs ≥25 kg/m²), and waist circumference (men <102 cm vs ≥102 cm; women <88 cm vs ≥88 cm).

Findings We validated 12 prediction models. Discrimination was acceptable to good: C statistics ranged from 0·76 (95% CI 0·72–0·80) to 0·81 (0·77–0·84) overall, from 0·73 (0·70–0·76) to 0·79 (0·74–0·83) in men, and from 0·78 (0·74–0·82) to 0·81 (0·80–0·82) in women. We noted significant heterogeneity in discrimination ($p_{heterogeneity}$<0·0001) in all but one model. Calibration was good for most models, and consistent across countries ($p_{heterogeneity}$>0·05) except for three models. However, two models overestimated risk, DPoRT by 34% (95% CI 29–39%) and Cambridge by 40% (28–52%). Discrimination was always better in individuals younger than 60 years or with a low waist circumference than in those aged at least 60 years or with a large waist circumference. Patterns were inconsistent for BMI. All models overestimated risks for individuals with a BMI of <25 kg/m². Calibration patterns were inconsistent for age and waist-circumference subgroups.

Interpretation Existing diabetes prediction models can be used to identify individuals at high risk of type 2 diabetes in the general population. However, the performance of each model varies with country, age, sex, and adiposity.

Funding The European Union.

Introduction

The number of individuals with type 2 diabetes is high and increasing rapidly: 366 million people worldwide were estimated to have type 2 diabetes in 2011, and prevalence is expected to rise by 51% by 2030.[1] Diabetes is associated with increased morbidity and mortality, and accounts for a substantial proportion of use of health-care resources worldwide.[1] Several studies have convincingly shown that early interventions can prevent or postpone type 2 diabetes.[2,3] The cost of these interventions and their constraints for individuals are arguments against their provision, particularly for people at low risk of diabetes. Therefore, appropriate identification of individuals at high risk is important.

Several risk scores to predict type 2 diabetes have been developed.[4–7] They vary in many ways, such as the time horizon for prediction and number and nature of predictors. Some models are based on non-laboratory clinical variables (non-invasive risk scores); others have also incorporated biological variables (invasive risk scores). Non-invasive risk scores have been shown to identify a high risk of type 2 diabetes (C statistics ≥0·8), although invasive risk scores are more successful (C statistics ~0·9).[8] The use of non-invasive risk scores is more likely to be cost effective and feasible for large-scale screening than is use of invasive risk scores. Generally, the use of risk scores has been widely incorporated into strategies for diabetes prevention.[9,10] However, risk scores might not be generalisable from one population to another, and the validity of existing diabetes risk scores in different settings and subgroups on the basis of sex or BMI has not been established.[10]

We aimed to validate and compare existing non-invasive prediction models for type 2 diabetes in European populations. We assessed variability in predictive performance between countries and by sex, BMI, waist circumference, and age.

University Medical Center Utrecht, Utrecht, Netherlands (A P Kengne PhD, J W J Beulens PhD, L M Peelen PhD, Prof K G M Moons PhD, Prof Y T van der Schouw PhD, Prof D E Grobee PhD); University of Cape Town and South African Medical Research Council, Cape Town, South Africa (A P Kengne); The George Institute for Global Health, Sydney, NSW, Australia (A P Kengne); German Institute of Nutrition, Potsdam-Rehbruecke, Germany (Prof M B Schulze DrPH, Prof H Boeing PhD, K Mühlenbruch MSc); National Institute for Public Health and the Environment, Bilthoven, Netherlands (A M W Spijkerman PhD, D L van der A PhD); Medical Research Council Epidemiology Unit, Cambridge, UK (Prof S J Griffin MD, L Palla PhD, N G Forouhi PhD, S J Sharp MSc, C Langenberg PhD, Prof N J Wareham PhD); Murcia Regional Health Council, Murcia, Spain (M-J Tormo PhD, J M Huerta PhD); Public Health Division of Gipuzkoa, San Sebastian, Spain (L Arriola MD); Hjelt Institute, University of Helsinki, Helsinki, Finland (N C Barengo PhD); Navarre Public Health Institute, Pamplona, Spain (A Barricarte PhD); Catalan Institute of Oncology,

Barcelona, Spain
(C Bonet MStat); **Inserm, Centre
for Research in Epidemiology
and Population Health, U1018,
Villejuif, France**
(F Clavel-Chapelon PhD,
L Dartois MSc, G Fagherazzi PhD);
**Lund University, Malmö,
Sweden** (Prof P W Franks PhD,
Prof P M Nilsson PhD); **German
Cancer Research Centre,
Heidelberg, Germany**
(Prof R Kaaks PhD, K Li PhD);
**University of Oxford, Oxford,
UK** (Prof T J Key DPhil); **University
of Cambridge, Cambridge, UK**
(Prof K T Khaw FRCP);
**Department of Public Health,
Aarhus University, Aarhus,
Denmark** (Prof K Overvad PhD);
**Aalborg University Hospital,
Aalborg, Denmark**
(T F Overvad BSc); **Cancer
Research and Prevention
Institute, Florence, Italy**
(D Palli MD); **Federico II
University, Naples, Italy**
(S Panico MD); **Public Health
Directorate, Asturias, Spain**
(J R Quirós MD); **Umeå
University, Umeå, Sweden**
(O Rolandsson PhD); **Danish
Cancer Society Research Center,
Danish Cancer Society,
Copenhagen, Denmark**
(N Roswall PhD,
A Tjønneland Dr Med Sci); **Center
for Cancer Prevention, Turin,
Italy** (C Sacerdote PhD);
**Andalusian School of Public
Health, Granada, Spain**
(M-J Sánchez PhD); **International
Agency for Research on Cancer,
Lyon, France** (N Slimani PhD);
**Fondazione IRCCS, Istituto
Nazionale dei Tumori, Milan,
Italy** (G Tagliabue MD); **Cancer
Registry and Histopathology
Unit, Azienda Sanitaria
Provinciale 7, Ragusa, Italy**
(R Tumino MD); **and Imperial
College London, London, UK**
(Prof E Riboli MD)

Correspondence to:
Dr Joline W J Beulens, Julius
Center for Health Sciences and
Primary Care, University Medical
Center Utrecht, PO Box 85500,
3508 GA Utrecht, Netherlands
j.beulens@umcutrecht.nl

See **Online** for appendix

Methods

Study design and population

We used the EPIC-InterAct case-cohort sample (27 779 participants, of whom 12 403 had incident diabetes), which was selected from the European Investigation into Cancer and Nutrition (EPIC) cohort study (455 569 participants; appendix p 4).[11] The design and methods of the InterAct study have been described elsewhere.[11] Briefly, the InterAct study was designed to investigate how genes and lifestyle factors interact to influence risk of type 2 diabetes. The InterAct consortium partners ascertained and verified incident cases of type 2 diabetes in EPIC cohorts between 1991 and 2007, from eight of the ten EPIC countries (26 centres).

Participants gave written informed consent for participation in the EPIC study. The study was approved by the local ethics committees in the participating countries and the internal review board of the International Agency for Research on Cancer (Lyon, France).

Procedures

Self-administered questionnaires provided baseline information about lifestyle and medical history in the case-cohort sample. Validated instruments were used to assess nutritional variables[12] and physical activity.[13] Blood pressure and anthropometric measurements followed standard approaches (appendix p 5). We handled missing data in three ways: proxy variables (appendix p 6) or exclusion of the predictor from the model; exclusion of countries from specific analyses (appendix p 11); and imputation of missing values by country with R's AregImpute Function, which takes all aspects of uncertainty in the imputations into account by using the bootstrap to estimate the process of drawing predicted values from a full Bayesian predictive distribution.

Existing non-invasive prediction models for incident diabetes had been identified through a systematic review, as previously described.[8] We included in this study only models developed in general populations of European ancestry. We first validated the original models by computing the predicted probability of diabetes for each participant in the EPIC-InterAct case-cohort with baseline values (appendix p 7). We expressed models' performance in terms of discrimination (whether the model can distinguish between people who do and do not develop diabetes) and calibration (to what extent the predicted probabilities agree with the reported risk across groups of individuals). We assessed discrimination with the C statistic (which is comparable to the area under the receiver operator characteristic curve) adapted for time-to-event data.[14] C statistics vary from 0·5 (no discrimination) to 1 (perfect discrimination), with values of 0·7–0·8 deemed acceptable and 0·8–0·9 good.[15] We assessed calibration graphically with calibration plots, and by computing the ratio of

expected to recorded probabilities and the accompanying 95% CIs by assuming a Poisson variance.[16] We also calculated the Yates slope (difference between mean predicted probability of type 2 diabetes for participants with and without incident diabetes, with higher values indicating better performance), and Brier score (squared difference between predicted probability and actual outcome for each participant; these values vary from 0 for a perfect prediction model to 0·25 for a non-informative model with 50% incident outcome).[17] We did not use tests of calibration such as the Hosmer-Lemeshow test or U statistic,[18] because they are sensitive to study sample size.

Because we were comparing the performance of several models in a case-cohort design across many countries, we took some extra steps before model validation. First, the models have different so-called time horizons. In the main analysis, we compared all models for prediction of type 2 diabetes within 10 years. Therefore, participants who developed diabetes after 10 years of follow-up were included as non-cases. Second, calibration is strongly affected by the outcome's incidence in the population; we therefore recalibrated models to the country-specific incidence through intercept adjustment (appendix pp 9, 12).[18,19] These approaches eliminate differences between the models due to differences in incidence between the populations for which the scores were developed.[19] Third, the incidence of diabetes is artificially increased in the case-cohort design. To arrive at the true incidences from the original cohort, we reconstituted the EPIC cohort within countries by applying a so-called blow-up approach to extrapolate the case-cohort data to a full cohort (appendix p 10).[8] Finally, to account for any heterogeneity between countries, we estimated the C statistic and ratio of expected to recorded probabilities for the entire cohort by pooling country-level estimates with random effects models. We then used the Cochran's Q and I^2 statistics to assess heterogeneity across countries.

We did the main analyses for each country and for subgroups defined by sex, age (<60 years vs ≥60 years), BMI (<25 kg/m² vs ≥25 kg/m²), and waist circumference (men <102 cm vs ≥102 cm; women <88 cm vs ≥88 cm). In four of the eight countries (Denmark, Italy, Spain, and Sweden), not all models had been validated (appendix p 11). We did a sensitivity analysis by restricting the assessment of the heterogeneity between countries to the four countries in which all models had been validated (France, Germany, the Netherlands, and the UK) to investigate if assessments of heterogeneity of different models on the basis of a different number of countries affected the comparisons. We also compared all models for prediction of diabetes within 5 years, which corresponds to the time horizon for predicted probability for six models.

We used the Hmisc, Design, Survival, and meta packages of R (version 2.13.0) for data analysis.

Role of the funding source

The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. APK and JWJB had full access to all the data in the study; JWJB, YTvdS, and NJW had final responsibility for the decision to submit for publication.

Results

We validated 12 prediction models (table 1, appendix p 7). Age was a predictor of type 2 diabetes in all models except DESIR.[24] Most models were based on logistic regressions (table 1). Varying definitions for incident diabetes were applied (table 1). Data for baseline age and sex were available for all participants, but data were missing for some participants for some predictors (table 1).

More than half the overall EPIC-InterAct study population were women, but the proportion varied by country (table 2). French participants were all women, whereas the proportion of women among participants in other countries ranged from 44% to 81%. Mean baseline age was 54·1 years (SD 8·4) in men and 53·5 years (9·0) in women.

During the first 10 years of follow-up, almost 10 000 cases of incident diabetes were recorded in the EPIC-InterAct population (table 2). Discrimination of the 12 prediction models was acceptable to good, with C statistics ranging from 0·76 to 0·81 (figure 1, table 3). For nine models, the recalibrated calibration curves mostly followed the ideal calibration line (figure 2; see appendix p 21 for calibration curves before recalibration). For the ARIC 2009 model,[21] the curves were mostly steeper than the ideal calibration line, with increasing risk (figure 2), indicating risk underestimation in high-

risk participants. For the Cambridge[23] and DPoRT[25] models, calibration curves were fairly flat, with increasing risk (figure 2), indicating risk overestimation in high-risk participants. Four models accurately estimated the overall rate of incident diabetes, six marginally overestimated or underestimated, and two largely overestimated (figure 2, table 3).

We noted significant heterogeneity in discrimination across countries for all models except for the ARIC 2009 model (figure 1). The highest C statistics were almost always recorded in France, and the lowest mostly in Denmark (figure 1). The worst and best discriminating models varied by country (figure 1). Patterns of calibration were largely similar in each country to those recorded for the overall cohort (appendix pp 22–33). As expected, because of recalibration by country, we recorded almost no heterogeneity in calibration.

Discrimination was generally higher in women than men (table 3). C statistics in men ranged from 0·73 to 0·79 in the overall cohort (table 3), and from 0·68 to 0·86 across countries (appendix p 13). In men, the best discrimination was recorded in the Netherlands (appendix p 13). In women, C statistics ranged from 0·78 to 0·81 in the overall cohort (table 3), and from 0·70 to 0·85 across countries (appendix p 13). The highest C statistics in women were mostly recorded in Germany, and the lowest in Denmark (appendix p 13). Calibration differed greatly between men and women per country, which could be masked by the overall ratio of expected to recorded probabilities (appendix p 13).

Discrimination was acceptable to good across age strata, and was always better in the younger age group

	Incident diabetes risk models												Variable or proxy in InterAct	Participants with missing data for InterAct
	ARIC 2005[20]	ARIC 2009[21]	AUS-DRISK[22]	Cam-bridge[23]	DESIR[24]	DPoRT[25]	FINDRISK concise[6]	FINDRISK full[6]	Framingham personal[26]	KORA S4/F4[27]	EPIC-Potsdam[28]	QD Score[29]		
Year	2005	2009	2010	2008	2008	2010	2003	2003	2007	2010	2007	2009	..	..
Cohort size	7915	12729	6060	24495	3817	19861	4595	4435	3140	873	25167	2540753	..	..
Follow-up (years)	9	10	5	4·6	9	9	5	5	7	5	5	10	..	..
Definition of incident diabetes	Any incident	Any incident	Treated with drugs, diag-nosed with OGTT	Self-report, registries	Treated with drugs, diagnosed with fasting glucose measure-ments	Any incident	Treated with drugs	Treated with drugs	Treated with drugs, diagnosed with fasting glucose measure-ments	Diag-nosed with OGTT	Diag-nosed with OGTT	Any incident	..	..
Number of cases of incident diabetes	1292	2407	362	323	203	1410	194	182	160	91	849	78081	..	..
Country	USA	USA	Australia	UK	France	Canada	Finland	Finland	USA	Germany	Germany	UK	..	..
Age group included (years)	≥25	45–64	≥25	40–79	30–64	>20	35–64	35–64	54*	55–74	35–65	25–79	20–79	..
Statistical model	Logistic	Weibull	Logistic	Logistic	Logistic	Weibull	Logistic	Logistic	Logistic	Logistic	Cox	Cox	NA	NA
Validation	External	External	External	External	External	External	External	External	External	External	External	External	NA	NA

(Continues on next page)

	Incident diabetes risk models												Variable or proxy in InterAct	Participants with missing data for InterAct
	ARIC 2005[20]	ARIC 2009[21]	AUS-DRISK[22]	Cam-bridge[23]	DESIR[24]	DPoRT[25]	FINDRISK concise[6]	FINDRISK full[6]	Framingham personal[26]	KORA S4/F4[27]	EPIC-Potsdam[28]	QD Score[29]		
(Continued from previous page)														
Predictors														
Age	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	0
Sex	No	No	Yes	Yes	Sex specific†	Sex specific†	No	No	Yes	Yes	No	Sex specific†	Yes	0
Smoking	No	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	339
Family history	No	No	No	Yes	Yes	No	No	No	No	No	No	Yes	Yes	13795‡
Parental diabetes	Yes	Yes	Yes	No	Sex specific†	No	No	No	Yes	Yes	No	No	Yes	15761‡
Maternal diabetes	No	Yes	No	No	No	No	No	No	No	No	No	No	Yes	15761‡
Hypertension	No	Yes	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	556
Pulse	No	Yes	No	No	No	No	No	No	No	No	No	No	Yes	16311§
Ethnic origin	Yes	Yes	Yes	No	No	Yes	No	No	No	No	No	Yes	Yes	0
BMI	No	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	197
Body weight	No	Yes	No	No	No	No	No	No	No	No	No	No	Yes	177
Waist circumference	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	No	Yes	No	Yes	2018
Height	Yes	Yes	No	No	No	No	No	No	No	No	Yes	No	Yes	113
Steroid use	No	No	No	Yes	No	No	No	No	No	No	No	Yes	NA	All¶
High glucose concentration	No	No	Yes	No	No	No	Yes	Yes	No	No	No	No	NA	All¶
Physical activity	No	No	Yes	No	No	No	No	Yes	No	No	Yes	No	Yes	2217
Townsend score‖	No	No	No	No	No	No	No	No	No	No	No	Yes	NA	All¶
Education	No	Yes	No	No	No	Yes	No	No	No	No	No	No	Yes	330
Consumption of red meat	No	No	No	No	No	No	No	No	No	No	Yes	No	Yes	736
Consumption of whole-grain bread	No	No	No	No	No	No	No	No	No	No	Yes	No	Yes	736
Coffee intake	No	No	No	No	No	No	No	No	No	No	Yes	No	Yes	736
Alcohol intake	No	No	No	No	No	No	No	No	No	No	Yes	No	Yes	117
Immigration into country where study done	No	No	Yes	No	No	No	No	No	No	No	No	No	NA	All¶
Systolic blood pressure	Yes	No	No	No	No	No	No	No	No	No	No	No	Yes	6343
Heart disease	No	No	No	No	No	Yes	No	No	No	No	No	No	Yes	55

ARIC=Atherosclerosis Risk in Communities. AUSDRISK=Australian Type 2 Diabetes Risk Assessment Tool. DESIR=Epidemiological Study on the Insulin Resistance Syndrome. DPoRT=Diabetes Population Risk Tool. FINDRISK=Finnish Diabetes Risk Score. KORA S4/F4=Cooperative Health Research in the Region of Augsburg (KORA), Survey 4. EPIC=European Investigation into Cancer and Nutrition. OGTT=oral glucose tolerance test. NA=not applicable. *Mean. †Model built separately for women and men. ‡Predictor completely missing in centres in Italy and Spain. §Predictor completely missing in centres in Denmark and Sweden. ¶All 27779 participants. ‖Index of social and economic deprivation of a location.

Table 1: Models of incident diabetes risk prediction included

than in the older (table 3). Discrimination was below the acceptable range (C statistic <0·70) in the lower stratum of waist circumference but was acceptable in the upper one. Discrimination was similar across BMI groups, although C statistics were higher in people with increased BMI than in those with lower BMIs in seven of the 12 models (table 3).

In terms of calibration, the ARIC 2005,[20] AUSDRISK,[22] FINDRISK concise,[6] and FINDRISK full models[6] acceptably predicted the overall rate of incident diabetes in age subgroups (table 3). The Cambridge model[23] overestimated risk in all participants (table 3). Most other models overestimated risk in younger participants and underestimated risk in the older people, but we recorded

	Overall	Denmark	France	Germany	Italy	Netherlands	Spain	Sweden	UK
Participating centres	26	2	6	2	5	2	5	2	2
Total eligible EPIC cohort	326 805	20 410	45 877	36 755	35 471	35 093	46 104	52 146	54 949
Total InterAct case-cohort	27 779	4037	867	3578	3393	2290	5889	5401	2324
Women	15 887 (57·2%)	1787 (44·3%)	867 (100%)	1795 (50·2%)	2177 (64·2%)	1845 (80·6%)	3323 (56·4%)	2824 (52·3%)	1269 (54·6%)
Age (years)	53·7 (8·7)	57·0 (4·4)	56·9 (6·5)	52·4 (8·3)	51·3 (7·7)	53·7 (10·2)	50·3 (7·8)	54·9 (9·7)	58·5 (10·2)
Smoking									
Never	12 335 (44·4%)	1264 (31·3%)	588 (67·8%)	1529 (42·7%)	1509 (44·5%)	908 (39·7%)	3149 (53·5%)	2309 (42·8%)	1079 (46·4%)
Former	8026 (28·9%)	1281 (31·7%)	200 (23·1%)	1286 (35·9%)	925 (27·3%)	762 (33·3%)	1028 (17·5%)	1618 (30·0%)	926 (39·8%)
Current	7418 (26·7%)	1492 (37·0%)	79 (9·1%)	763 (21·3%)	959 (28·3%)	620 (27·1%)	1712 (29·1%)	1474 (27·3%)	319 (13·7%)
Hypertension*	7461 (26·9%)	1020 (25·3%)	172 (19·8%)	1506 (42·1%)	912 (26·9%)	626 (27·3%)	1469 (24·9%)	1332 (24·7%)	424 (18·2%)
Systolic blood pressure (mm Hg)	136 (21)	144 (21)	130 (20)	134 (19)	134 (19)	133 (22)	132 (20)	140 (20)	139 (20)
Pulse rate (beats per min)	73 (12)	NA	70 (11)	74 (12)	71 (9)	74 (12)	74 (12)	NA	71 (12)
BMI (kg/m²)	27·6 (4·8)	27·4 (4·6)	24·6 (4·8)	27·7 (4·9)	27·5 (4·9)	26·6 (4·5)	29·4 (4·6)	26·8 (4·6)	27·1 (4·7)
Waist circumference (cm)	91 (14)	93 (14)	80 (12)	92 (14)	88 (13)	87 (13)	95 (12)	89 (14)	89 (14)
Hip circumference (cm)	103 (9)	103 (9)	101 (10)	103 (9)	102 (9)	104 (9)	107 (9)	101 (10)	103 (9)
Family history of diabetes	4907 (26·5%)†	955 (23·7%)	147 (17·0%)	1114 (31·1%)	NA	771 (33·7%)	NA	1534 (28·4%)	386 (16·6%)
Follow-up (days)	3968 (2603–4634)	3745 (2299–4242)	3391 (2603–3835)	3472 (1766–4099)	3967 (2480–4638)	4065 (2389–4655)	4557 (3262–4975)	4334 (3335–4952)	3797 (2294–4401)
Cases of incident diabetes during follow-up									
Overall	12 403	2055	288	1584	1437	828	2564	2622	1025
Within 5 years	3698	633	132	916	470	406	418	418	305
Within 10 years	9951	1845	278	1554	1292	778	1759	1520	925

Data are n, n (%), mean (SD), or median (IQR). Includes imputed data for those with missing values. NA=not available. *Self-reported diagnosis or use of drugs, or both. †Percentage calculated with number of individuals for whom data available (ie, not Italy and Spain; n=18 497).

Table 2: Baseline characteristics of the InterAct case-cohort and cases of incident diabetes during follow-up, overall and by country

the opposite finding for KORA[27] and Potsdam[28] models. Absolute risk of diabetes was variably overestimated or underestimated by models across weight circumference strata, while all models systematically overestimated the risk within the lower BMI stratum (table 3).

The pattern of overall event rate prediction in the four countries with valid data for all models (France, Germany, the Netherlands, and the UK) was very similar to that across all countries (data not shown), again with significant heterogeneity for the Cambridge[23] and DPoRT[25] models (both p≤0·0002). Heterogeneity in discrimination across countries decreased when we examined data for countries with full data, with the ARIC 2009,[21] DPoRT,[25] and FINDRISK[6] models showing no heterogeneity (p>0·11).

During the first 5 years of follow-up, 3698 cases of incident diabetes were recorded (table 2). Discrimination was slightly better and less heterogeneous at 5 years than at 10 years; otherwise, the pattern of predictions was broadly similar to that recorded at 10 years (appendix pp 16–20, 34–36).

Discussion

We have shown that non-invasive models for prediction of incident type 2 diabetes have acceptable to good discrimination over 10 years, both overall and across countries. After recalibration, most models showed good calibration, which was consistent across countries, although discrimination varied significantly. We showed that the models' performance is worse in men than in women. Discrimination is better in people younger than 60 years, but risk can be overestimated in this age group. Discrimination is generally lower in participants with a BMI of less than 25 kg/m² than in those with a BMI of at least 25 kg/m². Risk is systematically overestimated in participants with a BMI of less than 25 kg/m². No model significantly outperforms others enough to be uniquely recommended for routine risk stratification.

A few previous validation studies of incident diabetes models have been done.[4,5,8,21,24,30–34] Most of those—mainly cross-sectional studies—compared several incident or prevalent risk scores in one population,[21,24,32–34] and another study compared three risk scores in a multiethnic population in the same country.[30] In one study,[8] the performance of both invasive and non-invasive prediction models for incident diabetes in a population was assessed. Overall, the previous studies showed that the models had modest to good discrimination and poor calibration. However, intercept adjustment to correct for differences in diabetes incidence between development and validation populations was done in only one study.

Models' performance (mostly discrimination) differed across countries in our study. This variation could be a result of differences in how predictors and outcomes
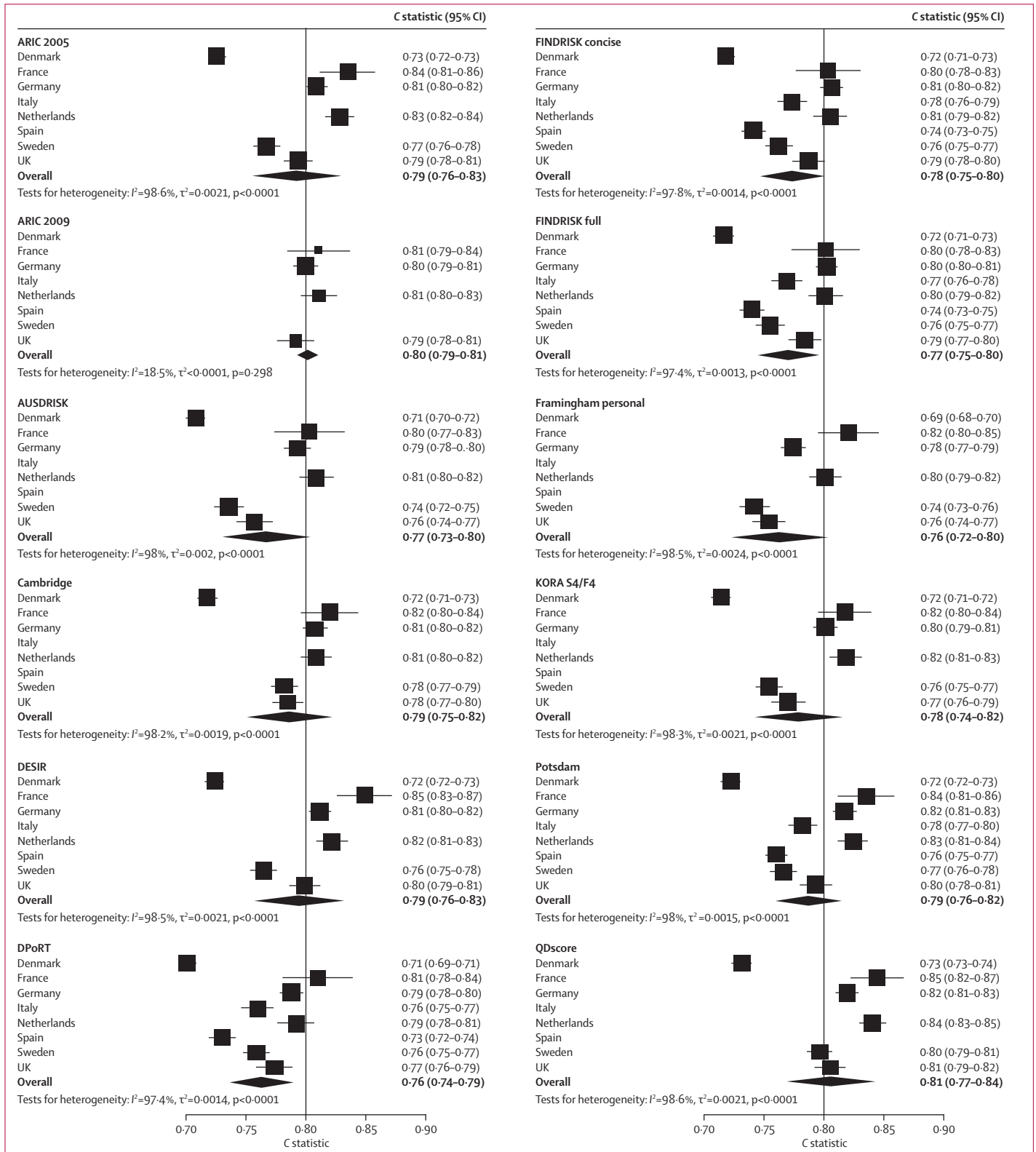
Figure 1: Discrimination of the different models for the prediction of incident type 2 diabetes at 10 years of follow-up overall and by country
Discrimination refers to the ability of the model to distinguish between participants who did and did not develop diabetes during 10 years of follow-up. ARIC=Atherosclerosis Risk in Communities. AUSDRISK=Australian Type 2 Diabetes Risk Assessment Tool. DESIR=Epidemiological Study on the Insulin Resistance Syndrome. DPoRT=Diabetes Population Risk Tool. FINDRISK=Finnish Diabetes Risk Score. KORA S4/F4=Cooperative Health Research in the Region of Augsburg (KORA), Survey 4.

	Overall	Men	Women	Age <60 years	Age ≥60 years	BMI <25 kg/m²	BMI ≥25 kg/m²	Waist circumference ≥102 cm in men or ≥88 cm in women	Waist circumference <102 cm in men or <88 cm in women
C statistics									
ARIC 2005	0·79 (0·76–0·83)*	0·75 (0·73–0·78)*	0·81 (0·77–0·84)*	0·81 (0·77–0·85)*	0·73 (0·69–0·76)*	0·73 (0·66–0·80)*	0·71 (0·69–0·73)*	0·66 (0·65–0·67)†	0·75 (0·70–0·79)*
ARIC 2009	0·80 (0·79–0·81)	0·79 (0·74–0·83)*	0·81 (0·80–0·82)	0·82 (0·81–0·83)	0·74 (0·71–0·77)‡	0·71 (0·68–0·73)	0·73 (0·71–0·74)	0·66 (0·63–0·69)‡	0·75 (0·73–0·76)
AUSDRISK	0·77 (0·73–0·80)*	0·73 (0·70–0·76)*	0·78 (0·74–0·82)*	0·78 (0·74–0·82)*	0·70 (0·68–0·72)‡	0·69 (0·64–0·74)*	0·71 (0·68–0·73)*	0·64 (0·63–0·66)†	0·69 (0·65–0·73)*
Cambridge model	0·79 (0·75–0·82)*	0·76 (0·73–0·79)*	0·80 (0·76–0·83)*	0·80 (0·76–0·84)*	0·72 (0·70–0·74)*	0·70 (0·65–0·75)*	0·70 (0·68–0·72)*	0·66 (0·64–0·68)*	0·74 (0·69–0·79)*
DESIR model	0·79 (0·76–0·83)*	0·73 (0·71–0·75)*	0·81 (0·77–0·85)*	0·81 (0·77–0·85)*	0·74 (0·71–0·77)*	0·71 (0·65–0·76)*	0·72 (0·69–0·74)*	0·65 (0·63–0·67)*	0·71 (0·67–0·76)*
DPoRT model	0·76 (0·74–0·79)*	0·73 (0·71–0·75)*	0·78 (0·75–0·81)*	0·78 (0·75–0·81)*	0·70 (0·68–0·73)*	0·63 (0·59–0·67)*	0·68 (0·66–0·69)*	0·64 (0·63–0·65)‡	0·72 (0·68–0·75) *
FINDRISK concise	0·78 (0·75–0·80)*	0·75 (0·72–0·78)*	0·79 (0·76–0·81)*	0·79 (0·76–0·82) *	0·71 (0·69–0·73)*	0·68 (0·64–0·71)*	0·70 (0·69–0·72)*	0·63 (0·62–0·65)*	0·72 (0·68–0·75)*
FINDRISK full model	0·77 (0·75–0·80)*	0·74 (0·72–0·77)*	0·78 (0·76–0·81)*	0·78 (0·75–0·81)*	0·71 (0·68–0·73)*	0·68 (0·65–0·72)*	0·70 (0·69–0·72)*	0·63 (0·62–0·65)§	0·71 (0·67–0·74)*
Framingham personal	0·76 (0·72–0·80)*	0·74 (0·70–0·77)*	0·78 (0·74–0·82)*	0·78 (0·74–0·83)*	0·69 (0·67–0·73)*	0·64 (0·57–0·70)*	0·65 (0·63–0·60)*	0·61 (0·60–0·63)†	0·70 (0·65–0·75)*
KORA model	0·78 (0·74–0·82)*	0·75 (0·71–0·79)*	0·79 (0·75–0·82)*	0·80 (0·76–0·84)*	0·72 (0·69–0·74)*	0·73 (0·67–0·78)*	0·70 (0·68–0·73)*	0·66 (0·65–)·68)‡	0·75 (0·70–0·79)*
Potsdam model	0·79 (0·76–0·82)*	0·74 (0·72–0·77)*	0·81 (0·78–0·83)*	0·80 (0·77–0·83)*	0·72 (0·70–0·75)*	0·74 (0·68–0·79)*	0·72 (0·70–0·73)*	0·66 (0·65–0·68)§	0·75 (0·71–0·78)*
QDScore	0·81 (0·77–0·84)*	0·78 (0·74–0·82)*	0·81 (0·77–0·85)*	0·82 (0·78–0·86)*	0·75 (0·72–0·77)*	0·75 (0·69–0·80)*	0·72 (0·70–0·75)*	0·68 (0·66–0·70)*	0·77 (0·72–0·82)*
Ratios of expected to observed rates									
ARIC 2005	1·06 (1·04–1·08)	1·08 (1·02–1·14)‡	1·06 (0·97–1·15)*	1·02 (0·98–1·06)†	1·11 (1·03–1·19)§	2·22 (1·74–2·68)*	0·83 (0·74–0·93)*	0·74 (0·65–0·82)*	1·49 (1·34–1·64)*
ARIC 2009	1·04 (1·01–1·08)	1·01 (0·84–1·17)*	1·13 (0·96–1·30)*	1·20 (1·14–1·27)	0·79 (0·69–0·88)‡	2·90 (2·60–3·21)	0·69 (0·56–0·81)*	0·55 (0·56–0·65)*	1·72 (1·56–1·88)*
AUSDRISK	1·00 (0·98–1·02)	1·11 (1·02–1·18)§	0·93 (0·86–1·00)*	1·03 (0·99–1·08)	0·97 (0·87–1·02)§	2·29 (1·70–2·81)*	0·75 (0·67–0·83)*	0·77 (0·68–0·86)*	1·31 (1·17–1·45)*
Cambridge model	1·40 (1·28–1·52)*	1·76 (1·59–1·94)*	1·14 (0·98–1·30)*	1·29 (1·17–1·42)*	1·53 (1·30–1·77)*	1·27 (0·99–1·55)*	1·42 (1·30–1·53)*	1·30 (1·20–1·41)*	1·54 (1·34–1·73)*
DESIR model	1·06 (1·02–1·11)‡	1·06 (1·03–1·08)	1·06 (0·99–1·13)§	1·18 (1·12–1·24)‡	0·88 (0·83–0·92)	1·35 (1·07–1·64)*	1·02 (0·98–1·06)†	1·13 (1·05–1·22)*	0·99 (0·89–1·10)*
DPoRT model	1·34 (1·29–1·39)‡	1·58 (1·51–1·64)†	1·14 (1·11–1·17)	1·44 (1·39–1·49)†	1·07 (0·95–1·18)*	1·58 (1·33–1·83)*	1·30 (1·23–1·36)*	1·00 (0·92–1·07)*	1·80 (1·67–1·93)*
FINDRISK concise	0·98 (0·96–0·99)	0·85 (0·79–0·91)*	1·11 (1·02–1·19)*	1·00 (0·97–1·03)	0·94 (0·88–1·01)§	1·28 (1·04–1·51)*	0·93 (0·90–0·97)§	1·08 (1·03–1·13)*	0·84 (0·79–0·90)*
FINDRISK full model	0·98 (0·96–0·99)	0·89 (0·84–0·95)*	1·07 (1·00–1·15)*	1·00 (0·97–1·03)	0·94 (0·88–1·00)§	1·37 (1·12–1·62)*	0·92 (0·88–0·96)*	1·06 (1·01–1·11)*	0·88 (0·82–0·93)*
Framingham personal	0·99 (0·98–1·02)	0·78 (0·70–0·87)*	1·22 (1·09–1·35)*	1·12 (1·07–1·16)	0·81 (0·76–0·86)†	1·67 (1·30–2·05)*	0·88 (0·82–0·94)*	0·73 (0·66–0·79)*	1·36 (1·23–1·50)*
KORA model	1·00 (0·97–1·02)	1·09 (1·06–1·12)	0·91 (0·84–0·99)*	0·82 (0·74–0·91)*	1·26 (1·15–1·37)*	1·70 (1·37–2·02)*	0·87 (0·82–0·92)§	0·77 (0·73–0·80)†	1·31 (1·21–1·42)*
Potsdam model	0·98 (0·97–1·00)	1·06 (1·01–1·11)†	0·93 (0·85–1·02)*	0·90 (0·85–0·94)†	1·17 (1·08–1·25)*	1·38 (1·12–1·64)*	0·93 (0·88–0·97)*	0·96 (0·90–1·03)*	1·04 (0·94–1·14)*
QDScore	1·06 (1·04–1·08)	1·08 (1·05–1·11)*	1·04 (1·01–1·07)	1·04 (0·99–1·10)‡	1·22 (1·07–1·36)*	1·40 (1·15–1·65)*	0·99 (0·93–1·05)*	0·87 (0·80–0·94)*	1·29 (1·19–1·39)*

ARIC=Atherosclerosis Risk in Communities. AUSDRISK=Australian Type 2 Diabetes Risk Assessment Tool. DESIR=Epidemiological Study on the Insulin Resistance Syndrome. DPoRT=Diabetes Population Risk Tool. FINDRISK=Finnish Diabetes Risk Score. KORA S4/F4=Cooperative Health Research in the Region of Augsburg (KORA), Survey 4. *$p_{heterogeneity}$<0·0001. †$p_{heterogeneity}$<0·05. ‡$p_{heterogeneity}$<0·01. §$p_{heterogeneity}$<0·001.

Table 3: Discrimination and calibration of models for the prediction of incident type 2 diabetes at 10 years of follow-up overall and by subgroup

were measured (for discrimination) and in baseline diabetes risk (for calibration) across countries. The FINDRISK concise model[6] was previously tested in five cohorts from Europe, North America, and Australia.[35] Discrimination varied by country, with C statistics between 0·63 and 0·78,[35] which is consistent with our
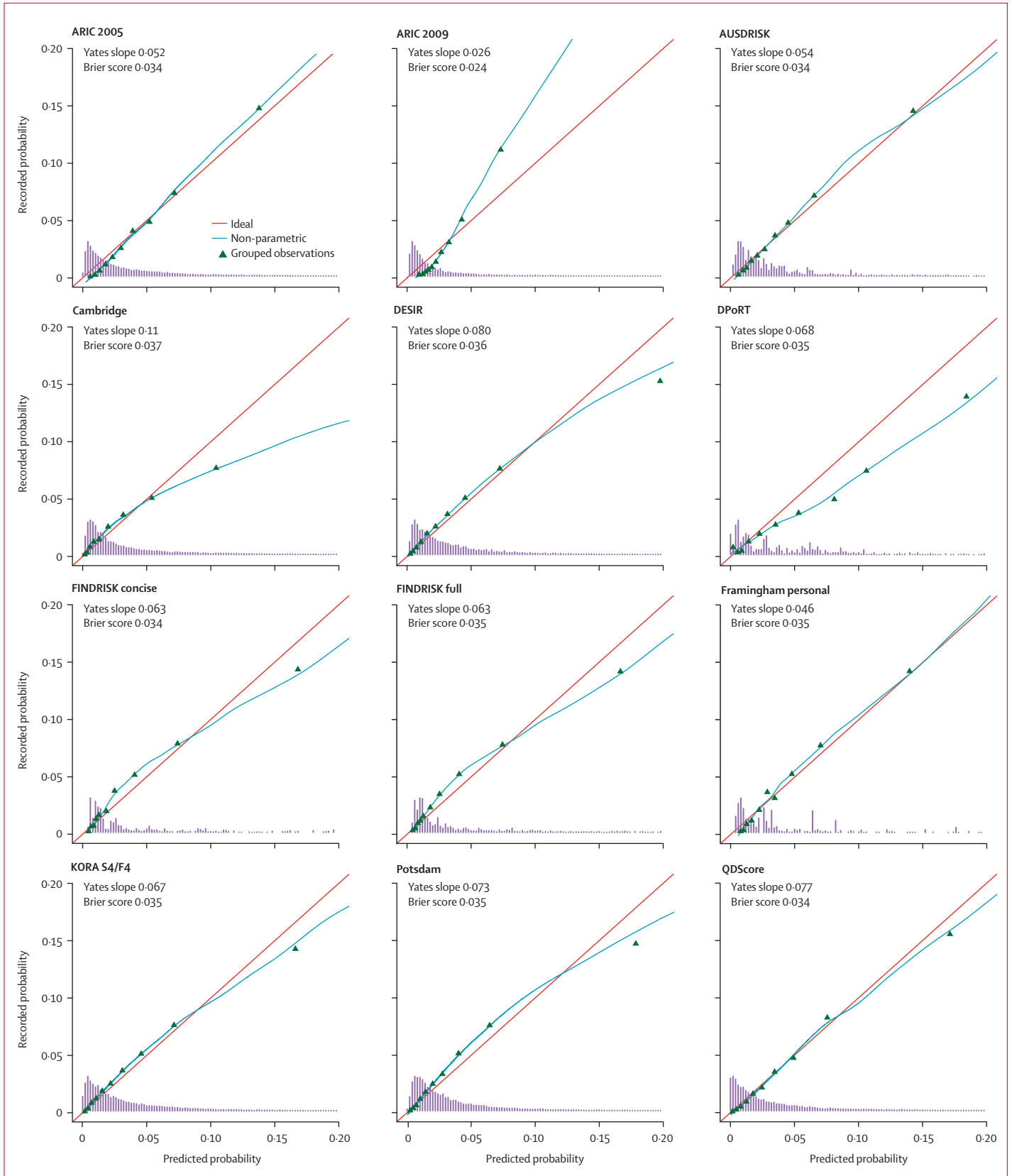
results. Therefore, risk scores might not always be generalisable to populations other than the development population.

We noted some consistent trends across countries. Discrimination was generally low for the DPoRT model,[25] potentially because this model was developed to estimate diabetes risk in a population rather than for an individual, which could affect quality of the predictors and outcome data. The QDScore[29] was one of the best discriminatory models across countries. It is based on more predictors than other models are and so uses more information, and was developed from a large sample, providing precise estimates of the associations between predictors and outcomes. Calibration of the Cambridge model[23] was generally poor. Although validated for incident diabetes prediction,[36] this model was developed as a diagnostic model, which could have inflated estimates of the association between diabetes and predictors.

We recorded substantial variation in models' performance by major subgroups, with consistently better performance in women than in men. For the DESIR[24] and DPoRT[25] models and the QDScore,[29] this difference could be a result of the use of sex-specific coefficients that could efficiently capture the predictive information from risk factors. BMI and waist circumference are generally much stronger diabetes predictors in women than in men and could also explain these results.[37,38] In our study, the differences in discrimination linked to BMI were small. BMI and waist circumference might not be linearly related to diabetes risk.[37] By including linear or categorical terms of these predictors, most models might not completely capture the nature of the associations. Our finding that discrimination is mostly better in younger than in older participants is consistent with a previous validation of the KORA model.[39] However, stratification for a variable with discriminatory value by itself reduces discrimination.

The European guideline for the prevention of type 2 diabetes[10] and the International Diabetes Federation[9] recommend the use of reliable, straightforward, and practical prediction models to identify people at high risk of diabetes. Overall, our study shows that prediction models for type 2 diabetes are valid instruments for the identification of individuals at high risk. However, this finding cannot directly be translated into a recommendation to use a specific model, because the decision to adopt a prediction model in any specific setting is driven by many factors, of which performance is only one.

For individuals, highly accurate risk scores including biochemical testing are probably more relevant than non-invasive risk models are.[8] However, non-invasive risk scores could be used as part of the public health approach to diabetes prevention to identify individuals who should receive biochemical testing. Alternatively, a risk score could be used to identify a subgroup of the population for lifestyle intervention without additional testing. Mortality risk is increased in the large group of people who have positive risk scores, justifying direct action in this group.[40] Risk scores like those in the Cambridge model[23] and QDScore[29] in the UK, or the DPoRT model in Canada,[25] use data available from records in the countries where models were developed, and new data are not needed. Because such scores are likely to be applicable to anyone coming into contact with the health system, their performance could be better overall than that of another model with a slightly better discrimination. Nevertheless, a model for which a questionnaire is designed for self-screening at the population level (eg, the two FINDRISK models) might not be based on many responses, potentially overriding any performance advantage recorded in our analysis.[41]

Therefore, the decision to use a particular model could be country specific and depends on factors other than model performance, such as availability of measurements in the setting where the model is used. Additionally, prediction of diabetes is useful only when used to implement interventions to affect the outcomes of high-risk people.[4] Diabetes prevention trials have been mostly based on a high-risk status defined by blood tests. However, implementation studies are necessary to investigate effectiveness of such risk scores. A few implementation studies have used risk scores to select participants for interventions and have shown favourable effects on risk factor levels[42]—eg, a moderate weight loss effectively reduced risk of type 2 diabetes after 1 year of intervention in participants selected with the FINDRISK score.[43]

Our study has some limitations. Some of the approaches used to account for predictors that were completely missing (ie, proxy variables, predictor omission, and country's exclusion) could marginally decrease discrimination of the models. It is now widely accepted that deletion of participants with missing values (who are frequent in large studies) yields biased results.[39] Therefore, we applied robust methods to deal with missing data.[44] The large sample and the multicountry nature of our study has allowed us to assess and compare 12 models across countries for the first time (panel). However, we did not include non-European participants. Therefore, our results might not be generalisable to non-European countries or other

Figure 2: **Calibration curves for each model for the prediction of incident type 2 diabetes at 10 years of follow-up for the total cohort**
Calibration of a model describes the extent to which the expected probability of diabetes matches the recorded probability of diabetes during follow-up. The ideal calibration (perfect agreement) is graphically represented by the dotted diagonal line at 45°. The vertical lines at the bottom of the graph depict the frequency distribution of the calibrated probabilities of diabetes. Grouped observations are for groups of participants across increasing deciles of predicted risk. ARIC=Atherosclerosis Risk in Communities. AUSDRISK=Australian Type 2 Diabetes Risk Assessment Tool. DESIR=Epidemiological Study on the Insulin Resistance Syndrome. DPoRT=Diabetes Population Risk Tool. FINDRISK=Finnish Diabetes Risk Score. KORA S4/F4=Cooperative Health Research in the Region of Augsburg (KORA), Survey 4.

ethnic groups. Finally, case definition was based on a clinical diagnosis of diabetes. Because diabetes can remain undetected, false-negative cases could have been present. Such false negatives could lead to an underestimation of the C statistic and affect the incidence of diabetes, affecting calibration of the risk scores. However, we adjusted for such effects of diabetes incidence by recalibrating the models.

References

1 International Diabetes Federation. IDF Diabetes Atlas, 5th edn. Brussels: International Diabetes Federation, 2011.

2 Li G, Zhang P, Wang J, et al. The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. *Lancet* 2008; **371:** 1783–89.

3 Gillies CL, Abrams KR, Lambert PC, et al. Pharmacological and lifestyle interventions to prevent or delay type 2 diabetes in people with impaired glucose tolerance: systematic review and meta-analysis. *BMJ* 2007; **334:** 299.

4 Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. *BMJ* 2011; **343:** d7163.

5 Buijsse B, Simmons RK, Griffin SJ, Schulze MB. Risk assessment tools for identifying individuals at risk of developing type 2 diabetes. *Epidemiol Rev* 2011; **33:** 46–62.

6 Lindstrom J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 2003; **26:** 725–31.

7 Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med* 2011; **9:** 103.

8 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.

9 Alberti KG, Zimmet P, Shaw J. International Diabetes Federation: a consensus on Type 2 diabetes prevention. *Diabet Med* 2007; **24:** 451–63.

10 Paulweber B, Valensi P, Lindstrom J, et al. A European evidence-based guideline for the prevention of type 2 diabetes. *Horm Metab Res* 2010; **42** (suppl 1)**:** S3–36.

11 Langenberg C, Sharp S, Forouhi NG, et al. Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. *Diabetologia* 2011; **54:** 2272–82.

12 Kaaks R, Slimani N, Riboli E. Pilot phase studies on the accuracy of dietary intake measurements in the EPIC project: overall evaluation of results. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* 1997; **26** (suppl 1)**:** S26–36.

13 Wareham NJ, Jakes RW, Rennie KL, et al. Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Public Health Nutr* 2003; **6:** 407–13.

14 Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Stat Med* 2006; **25:** 3474–86.

15 Hosmer DW, Lemeshow S. Applied logistic regression. New York: Wiley, 1989.

16 Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA. Validation of the Gail et al model of breast cancer risk prediction and implications for chemoprevention. *J Natl Cancer Inst* 2001; **93:** 358–66.

17 Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; **21:** 128–38.

18 Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. Rotterdam: Springer, 2009.

19 Janssen KJ, Vergouwe Y, Kalkman CJ, Grobbee DE, Moons KG. A simple method to adjust clinical prediction models to local circumstances. *Can J Anaesth* 2009; **56:** 194–201.

20 Schmidt MI, Duncan BB, Bang H, et al. Identifying individuals at high risk for diabetes: The Atherosclerosis Risk in Communities study. *Diabetes Care* 2005; **28:** 2013–18.

21 Kahn HS, Cheng YJ, Thompson TJ, Imperatore G, Gregg EW. Two risk-scoring systems for predicting incident diabetes mellitus in US adults age 45 to 64 years. *Ann Intern Med* 2009; **150:** 741–51.

22 Chen L, Magliano DJ, Balkau B, et al. AUSDRISK: an Australian Type 2 Diabetes Risk Assessment Tool based on demographic, lifestyle and simple anthropometric measures. *Med J Aust* 2010; **192**: 197–202.

23 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**: 164–71.

24 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.

25 Rosella LC, Manuel DG, Burchill C, Stukel TA. A population-based risk algorithm for the development of diabetes: development and validation of the Diabetes Population Risk Tool (DPoRT). *J Epidemiol Community Health* 2011; **65**: 613–20.

26 Wilson PW, Meigs JB, Sullival 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**: 1068–74.

27 Rathmann W, Kowall B, Heier M, et al. Prediction models for incident type 2 diabetes mellitus in the older population: KORA S4/F4 cohort study. *Diabet Med* 2010; **27**: 1116–23.

28 Schulze MB, Hoffmann K, Boeing H, et al. An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. *Diabetes Care* 2007; **30**: 510–15.

29 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; **338**: b880.

30 Mann DM, Bertoni AG, Shimbo D, et al. Comparative validity of 3 diabetes mellitus risk prediction scoring models in a multiethnic US cohort: the Multi-Ethnic Study of Atherosclerosis. *Am J Epidemiol* 2010; **171**: 980–88.

31 Collins GS, Altman DG. External validation of QDSCORE(R) for predicting the 10-year risk of developing type 2 diabetes. *Diabet Med* 2011; **28**: 599–607.

32 Lin JW, Chang YC, Li HY, et al. Cross-sectional validation of diabetes risk scores for predicting diabetes, metabolic syndrome, and chronic kidney disease in Taiwanese. *Diabetes Care* 2009; **32**: 2294–96.

33 Schmid R, Vollenweider P, Waeber G, Marques-Vidal P. Estimating the risk of developing type 2 diabetes: a comparison of several risk scores: the Cohorte Lausannoise study. *Diabetes Care* 2011; **34**: 1863–68.

34 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**: 188–89.

35 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.

36 Rahman M, Simmons RK, Harding AH, Wareham NJ, Griffin SJ. A simple risk score identifies individuals at high risk of developing type 2 diabetes: a prospective cohort study. *Fam Pract* 2008; **25**: 191–96.

37 Vazquez G, Duval S, Jacobs DR Jr, Silventoinen K. Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis. *Epidemiol Rev* 2007; **29**: 115–28.

38 Abdullah A, Peeters A, de Courten M, Stoelwinder J. The magnitude of association between overweight and obesity and the risk of diabetes: a meta-analysis of prospective cohort studies. *Diabetes Res Clin Pract* 2010; **89**: 309–19.

39 Abbasi A, Corpeleijn E, Peelen LM, et al. External validation of the KORA S4/F4 prediction models for the risk of developing type 2 diabetes in older adults: the PREVEND study. *Eur J Epidemiol* 2012; **27**: 47–52.

40 Spijkerman A, Griffin S, Dekker J, Nijpels G, Wareham NJ. What is the risk of mortality for people who are screen positive in a diabetes screening programme but who do not have diabetes on biochemical testing? Diabetes screening programmes from a public health perspective. *J Med Screen* 2002; **9**: 187–90.

41 Wareham NJ, Griffin SJ. Risk scores for predicting type 2 diabetes: comparing axes and spades. *Diabetologia* 2011; **54**: 994–95.

42 Uusitupa M, Tuomilehto J, Puska P. Are we really active in the prevention of obesity and type 2 diabetes at the community level? *Nutr Metab Cardiovasc Dis* 2011; **21**: 380–89.

43 Saaristo T, Moilanen L, Korpi-Hyovalti E, et al. Lifestyle intervention for prevention of type 2 diabetes in primary health care: one-year follow-up of the Finnish National Diabetes Prevention Program (FIN-D2D). *Diabetes Care* 2010; **33**: 2146–51.

44 Janssen KJ, Donders AR, Harrell FE Jr, et al. Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol* 2010; **63**: 721–27.