External validation of the KORA S4/F4 prediction models for the risk of developing type 2 diabetes in older adults

Abbasi, Ali; Corpeleijn, Eva; Peelen, Linda M.; Gansevoort, Ron T.; de Jong, Paul E.; Gans, Rijk O.B. ; Rathmann, Wolfgang; Kowall, Bernd; Meisinger, Christine; Hillege, Johann

Published in:
European Journal of Epidemiology

DOI:
10.1007/s10654-011-9648-4

Publication date:
2012

Eur J Epidemiol (2012) 27:47–52
DOI 10.1007/s10654-011-9648-4

DIABETES MELLITUS

External validation of the KORA S4/F4 prediction models for the risk of developing type 2 diabetes in older adults: the PREVEND study

Ali Abbasi · Eva Corpeleijn · Linda M. Peelen · Ron T. Gansevoort · Paul E. de Jong · Rijk O. B. Gans · Wolfgang Rathmann · Bernd Kowall · Christine Meisinger · Hans L. Hillege · Ronald P. Stolk · Gerjan Navis · Joline W. J. Beulens · Stephan J. L. Bakker

Received: 6 October 2011 / Accepted: 23 December 2011 / Published online: 4 January 2012

Abstract Recently, prediction models for type 2 diabetes mellitus (T2DM) in older adults (aged ≥55 year) were developed in the KORA S4/F4 study, Augsburg, Germany. We aimed to externally validate the KORA models in a Dutch population. We used data on both older adults (n = 2,050; aged ≥55 year) and total non-diabetic population (n = 6,317; aged 28–75 year) for this validation. We assessed performance of base model (model 1: age, sex, BMI, smoking, parental diabetes and hypertension) and two clinical models: model 1 plus fasting glucose (model 2); and model 2 plus uric acid (model 3). For 7-year risk of T2DM, we calculated C-statistic, Hosmer–Lemeshow χ^2-statistic, and integrated discrimination improvement (IDI) as measures of discrimination, calibration and reclassification, respectively. After a median follow-up of 7.7 years, 199 (9.7%) and 374 (5.9%) incident cases of T2DM were ascertained in the older and total population, respectively. In the older adults, C-statistic was 0.66 for model 1. This was improved for model 2 and model 3 (C-statistic = 0.81) with significant IDI. In the total population, these respective C-statistics were 0.77, 0.85 and 0.85. All models showed poor calibration ($P < 0.001$). After adjustment for the intercept and slope of each model, we observed good calibration for most models in both older and total populations. We validated the KORA clinical models for prediction of T2DM in an older Dutch population, with discrimination similar to the development cohort. However, the models need to be corrected for intercept and slope to acquire good calibration for application in a different setting.

Keywords Type 2 diabetes · Prediction model · External validation · Update · Older adults

Abbreviations

ARICA	Atherosclerosis risk in communities
BMI	Body mass index
DESIR	Data from the epidemiological study on the insulin resistance syndrome
FINDRISC	Finnish diabetes risk score
HbA1c	Glycosylated hemoglobin
IDI	Integrated discrimination improvement
KORA	Cooperative health research in the region of Augsburg
PREVEND	Prevention of renal and vascular end stage disease

Electronic supplementary material The online version of this article (doi:10.1007/s10654-011-9648-4) contains supplementary material, which is available to authorized users.

A. Abbasi (✉) · E. Corpeleijn · H. L. Hillege · R. P. Stolk
Department of Epidemiology, University of Groningen, University Medical Center Groningen, Hanzeplein 1, P.O. Box 30.001, 9700 RB Groningen, The Netherlands
e-mail: a.abbasi@epi.umcg.nl

A. Abbasi · R. T. Gansevoort · P. E. de Jong · R. O. B. Gans · G. Navis · S. J. L. Bakker
Department of Internal Medicine, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

A. Abbasi · L. M. Peelen · J. W. J. Beulens
Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

W. Rathmann · B. Kowall
Institute of Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University, Düsseldorf, Germany

C. Meisinger
Helmholtz Zentrum München, German Research Center of Environmental Health, Institute of Epidemiology II, Neuherberg, Germany

Introduction

Type 2 diabetes is one of the major concerns in public health, becoming more prevalent worldwide in parallel with increasing rate of obesity and ageing [1]. There is evidence suggesting that diabetes can be prevented by diet and lifestyle modifications [2]. For this, individuals at risk of developing diabetes need to be accurately identified [3, 4]. In the rapidly growing group of older subjects, prediction and primary prevention of chronic complex diseases such as diabetes are important to aid in healthy aging [5].

Risk for development of diabetes is appreciably higher in older subjects than in younger subjects. Several prediction models, including the Finnish (FINDRISC), the Atherosclerosis Risk in Communities (ARIC), the Framingham, Data from the Epidemiological Study on the Insulin Resistance Syndrome (DESIR) and Cambridge diabetes risk scores, have been developed in middle-aged populations and validated in other populations [4, 6–11]. As there are indications of 'reverse epidemiology' in older populations [12, 13], it is questionable whether risk scores developed in middle-aged populations can be extrapolated to older subjects. Moreover, it has been shown that predictive value of common risk factors declines with ageing [14]. Recently, the Cooperative Health Research in the Region of Augsburg (KORA) S4/F4 models have been developed to specifically predict the risk of type 2 diabetes in older subjects [15]. Because some risk scores that showed less performance when external validation was attempted, it is important that risk scores are validated in an independent population before they are brought into clinical practice [15, 16].

Therefore, we aimed to investigate the performance of the KORA models to predict incident type 2 diabetes in a large sample of non-diabetic Dutch adults, in particular older adults. We assessed performance of the model in terms of discrimination, calibration, recalibration and reclassification.

Methods

Population and design of the derivation study

The KORA S4/F4 study is a community-based cohort of 2,656 individuals (aged 55–74 years) living in the area of Augsburg, Germany in 1999. Details of the study design, recruitment, and procedures have been published elsewhere [17]. Among 887 individuals who participated for a median follow-up of 7-years, 91 (10.5%) incident cases of type 2 diabetes were observed in the KORA cohort [15, 17]. The KORA data set was used to compare baseline characteristics with those of the validation cohort.

Population and design of the validation study

We used data from the Prevention of Renal and Vascular Endstage Disease (PREVEND) study. The PREVEND study is a community-based prospective cohort of 8,592 inhabitants (aged 28–75 years) of the city of Groningen, The Netherlands who were screened for baseline measurements between 1997 and 1998. Details of the study design, recruitment, and measurements have been published elsewhere [18]. From the baseline cohort, we excluded 295 participants who had diabetes and 1,980 with missing data on clinical characteristics or data on follow-up, leaving 6,317 non-diabetic total population and a sample of 2,050 older adults (aged ≥55 years old) for this prospective validation analysis. The latter was used for primary validation while the former was used for secondary validation in a population with a much larger age range. All participants gave written informed consent prior to study inclusion. The PREVEND cohort complied with the Declaration of Helsinki and was approved by the medical ethics committee in The Netherlands.

Outcome, predictors and measurements

The main outcome was incidence of type 2 diabetes which was classified if one or more of the following criteria were met: fasting plasma glucose ≥7.0 mmol/l (126 mg/dl); non-fasting sample plasma glucose ≥11.1 mmol/l (200 mg/dl); self-report of a physician diagnosis of type 2 diabetes; pharmacy-registered use of glucose-lowering agents [19]. To estimate the predicted 7-year risk for type 2 diabetes in our cohort, we calculated the linear predictors of the KORA prediction models [15]. The base model (model 1) included data on age, sex, parental diabetes, body mass index (BMI), smoking status and hypertension. The clinical KORA models included additional data on fasting glucose, serum uric acid and HbA1c [15]. As data on HbA1c was unavailable, we validated a reported clinical model with data on fasting glucose (model 2) [15]. Moreover, the authors were asked to provide a clinical model with data on fasting glucose and uric acid (model 3), presented in Supplementary Table 1.

Data analysis

To externally validate these models, we assessed the discrimination and calibration performances in our cohort [20]. The discrimination performance denotes to what extent the model distinguishes between individuals with and without the outcome. Discrimination was expressed as the C-statistic with 95% confidence interval, where a value of 1 implies a perfect discrimination and a value of 0.5 implies performance no better than chance. We compared

the C-statistics of the clinical models to that of the base model as reference. The calibration compares predicted risks with observed risks. We applied the Hosmer–Lemeshow χ^2 test to evaluate the calibration performance. A lower χ^2 value with a non-significant P value represents good calibration. Also, calibration was visually checked by comparing the predicted probabilities versus observed incident cases of diabetes in each decile of predictions [20]. To recalibrate the prediction models, we used the original KORA models and applied logistic regression to derive the intercept and the calibration slope of each model in the total and older populations, and separately for women and men (Supplementary method part 1) [21]. We used these intercepts and slopes by fitting a model with the original linear predictor as the only covariate in the PREVEND data set. We multiplied each linear predictor by the calibration slope and added the calibration intercept to each original model [21]. Thereafter, we added data on waist circumference, a non-invasive risk factor for diabetes [11], to the re-calibrated models and assessed if this could improve predictive ability. We examined improvement of diabetes prediction in terms of discrimination, calibration and integrated discrimination improvement (IDI), a measure of reclassification (Supplementary method part 2) [20]. The analyses were performed separately in the older adults and in total PREVEND population. All the statistical analyses were carried out using Statistical Package for Social Sciences version 18 (SPSS Inc, Chicago, Illinois, USA), Stata software version 10.0 (Stata-Corp LP, College Station, TX, USA) and R-2.11.0 for Windows (http://cran.r-project.org/).

Results

In the older adults, we observed 199 (9.7%) incident cases of type 2 diabetes during follow-up for a median of 7.7 years. In the total population, we observed 374 (5.9%) cases during this follow-up. Baseline participants' characteristics of the KORA and PREVEND cohorts (aged $\geq$55 years) are shown in Supplementary Table 2. Participants of PREVEND were more likely to be male, older, more likely to be smoker and to have hypertension, but had lower BMI, lower parental history of diabetes and had lower fasting glucose and serum uric acid than participants of KORA.

Table 1 depicts the performance of the KORA models in terms of discrimination and calibration. In the older adults, a relatively low discriminative ability was observed for the base model (C-statistic = 0.66), being lower than the original C-statistic of 0.76 [15]. This was significantly improved for both models 2 and 3 (C-statistic = 0.81), being comparable with the original C-statistic of 0.81 [15]. The discriminative ability was not significantly different between these clinical models ($P = 0.78$). The base and both clinical models did not show good calibration ($P < 0.001$ for 7-year risk). When we tested the performance of each model in the total population, we observed a better discriminative ability for the base (C-statistic of 0.77; $P < 0.001$) and both clinical models (both C-statistics = 0.85; $P < 0.001$). Similarly, the base and both clinical models did not show good calibration ($P < 0.001$ for 7-year risk) in the total population. After adjustment for the calibration intercept and the calibration slope of each model, good calibration was observed for the clinical models ($P > 0.05$ for 7-year risk, Table 1), but not for the base model. Figure 1 (A, B) depicts the agreement between the predicted 7-year risk and observed risk of type 2 diabetes in each decile of predictions before and after recalibration. Of note, the predictive probability of model 3 was deviated from the ideal line in the older adults above 15% risk (Figure 1 A); indeed, the 7-year risk was underestimated for this risk category. The IDIs were significant when we compared the prediction performance of the clinical models to that of the base model ($P < 0.001$). In a subsequent analysis, we stratified total population by gender. We observed that all KORA models showed better discrimination performance in women than in men (Table 1). Addition of waist circumference improved predictive ability of the base model in the total population (C-statistic = 0.79; $P < 0.001$), but not in the older population separately (C-statistic = 0.67; $P = 0.30$). Addition of waist circumference did not improve predictive ability of model 3, neither in the total population (C-statistic = 0.85; $P = 0.11$) nor in the older population separately (C-statistic = 0.81; $P = 0.41$).

Discussion

In this external validation study, we prospectively assessed performance of the KORA models to predict the risk of developing type 2 diabetes in an independent Dutch population. We found that the prediction models with clinical data on glucose with or without uric acid performed well and this was much better than the base model in terms of discrimination and reclassification in the older adults. Moreover, we observed a similar pattern but with higher discriminative abilities in the total population. All models showed poor calibration performance. The calibration was good after adjustment for the intercept and the slope of clinical models, but not for the base model.

To our best of knowledge, there are few studies that derive and validate prediction models for of the risk of developing type 2 diabetes in the older adults. The main strengths of our study were including a large population-based cohort, available data on blood sampling in each screening visit and pharmacy registry, and applying the latest standards of prediction research. Some limitations

Table 1 Predictive performance of the KORA models in the PREVEND study

Prediction model (cases/total)	No. of predictors	C-statistic (95% CI)	IDI (*P* value) (not recalibrated model)[2]	$\chi 2$ statistic (*P* value) (not recalibrated model)	Calibration intercept	Calibration slope	$\chi 2$ statistic (*P* value) (recalibrated model)	IDI (*P* value) (recalibrated model)[a]
Age $\geq$ *55*: 199/2,050								
Base model	6	0.66 (0.63−0.70)	Ref.	45.35 (<0.001)	−0.953	0.590	19.20 (0.02)	Ref.
Model 2	7	0.81 (0.78−0.84)	0.084 (<0.001)	43.17 (<0.001)	0.096	0.858	10.02 (0.35)	0.120 (<0.001)
Model 3	8	0.81 (0.78−0.84)	0.093 (<0.001)	39.20 (<0.001)	−0.167	0.747	3.03 (0.96)	0.108 (<0.001)
Whole population: 374/6,317								
Base Model	6	0.77 (0.75−0.79)	Ref.	137.87 (<0.001)	−0.645	0.636	23.74 (0.004)	Ref.
Model 2	7	0.85 (0.83−0.87)	0.043 (<0.001)	492.14 (<0.001)	0.134	0.702	4.58 (0.87)	0.091 (<0.001)
Model 3	8	0.85 (0.83−0.87)	0.045 (<0.001)	674.70 (<0.001)	−0.043	0.633	12.00 (0.21)	0.084 (<0.001)
Women: 155/3,282								
Base model	6	0.83 (0.80−0.85)	Ref.	64.89 (<0.001)	−0.082	0.808	64.89 (<0.001)	Ref.
Model 2	7	0.88 (0.85−0.91)	0.061 (<0.001)	236.57 (<0.001)	0.452	0.790	5.35 (<0.80)	0.114 (<0.001)
Model 3	8	0.88 (0.86−0.91)	0.072 (<0.001)	21.49 (0.01)	0.166	0.709	8.46 (0.49)	0.109 (<0.001)
Men: 219/3,035								
Base model	6	0.71 (0.68−0.75)	Ref.	134.22 (<0.001)	−0.976	0.516	19.69 (0.02)	Ref.
Model 2	7	0.81 (0.79−0.84)	0.033 (<0.001)	321.84 (<0.001)	−0.095	0.633	5.51 (0.79)	0.074 (<0.001)
Model 3	8	0.81 (0.78−0.84)	0.064 (<0.001)	473.80 (<0.001)	−0.228	0.566	5.98 (0.74)	0.100 (<0.001

Base model (model 1) included data on age, sex, parental diabetes, BMI, smoking status and hypertension. Model 2 additionally included fasting glucose, and model 3 additionally included fasting glucose and uric acid. A higher discrimination C-statistic, non-significant *P* value of the calibration χ^2 statistic and significant *P* value of IDI represent better performance for each model

[a] IDI denotes integrated discrimination improvement, and is calculated as difference of the mean predicted risk between two models for those who developed outcome and those who did not develop outcome. The base model was considered as the reference

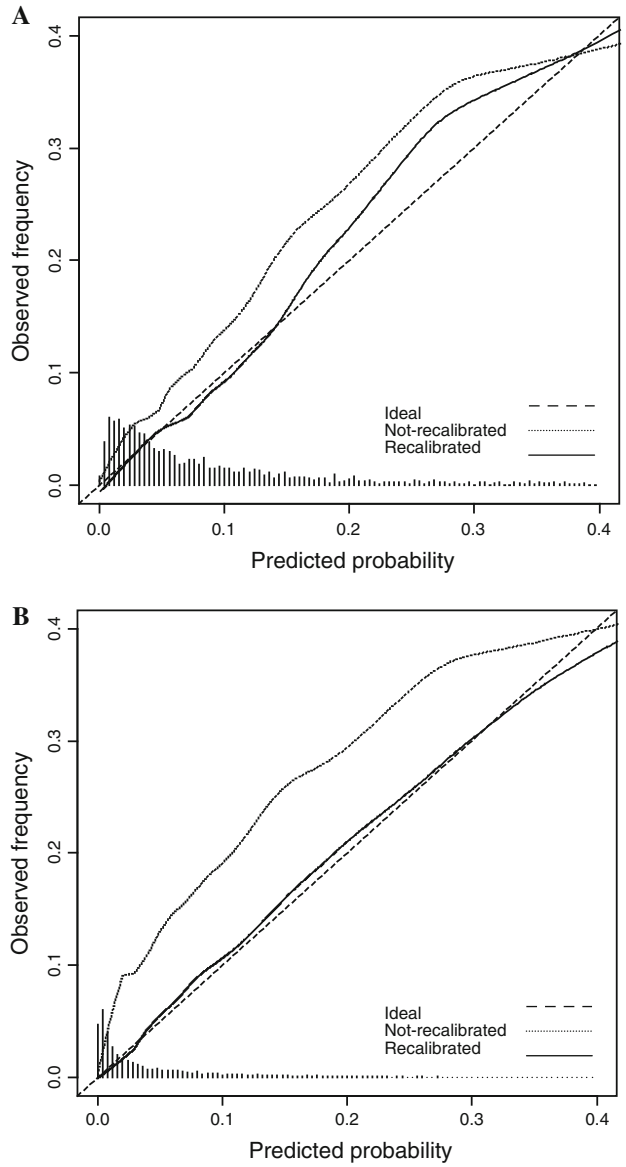

Fig. 1 Comparison between predicted risk versus observed diabetes frequency in the PREVEND cohort according to the KORA model 3, Model 3, included data on the base KORA model plus glucose and uric acid for the risk prediction of diabetes in the PREVEND cohort. **A** depicts calibration plots in the older adults, aged ≥55 years (n = 2,050). **B** depicts calibration plots in the total population (n = 6,317). The dashed line represents an ideal calibration (with intercept 0 and slope 1); the dotted line is for the not-recalibrated model and the solid line is after adjustment for the intercept and slope

should be addressed. First, we excluded the individuals with missing data at baseline or during follow-up. However, the baseline characteristics of excluded participants were similar to those who were included in our analysis. Therefore, it is less likely that this might have led to selection bias. Moreover, both derivation and validation data sets were gathered among Whites and our findings need to be further evaluated in other populations.

Our findings were consistent with previous validation studies in which performance of other prediction models for type 2 diabetes were tested in independent populations [22, 23]. Of note, predictive performance of models is often decreased in the validation sample. Several differences between derivation and validation samples might explain this change of performance. These include differences in healthcare systems, methods of measurement, and patients' characteristics [22]. Our protocol to measure the predictors and incident cases of type 2 diabetes was very comparable with the KORA study during similar follow-up time and period. Regardless possible differences in healthcare systems, characteristics of the participants of PREVEND were remarkably different from the participants of KORA.

When we calculated the C-statistic as a discrimination measure, we observed good ability (C-statistic ≥ 0.81) of clinical KORA models to distinguish between incident cases of type 2 diabetes and those who remained free of diabetes in both older and total populations. This was comparable for both clinical models and much higher than the base model. This cannot be explained by differences in the incidence of type 2 diabetes, as the C-statistic is hardly affected by different incidences of the outcome. A better discriminative ability of KORA models in the total population than in the older adults might be explained by a difference in case mix between KORA and PREVEND cohorts, less heterogeneity among the older adults of PREVEND, effect of predictors and difference in regression coefficients of KORA models [23]. For example, variables like age, BMI, or hypertension discriminate better between cases and non-cases in the total population [14] because low BMI or hypertension is more frequent in the younger adults who develop diabetes less often.

Both base and clinical KORA models showed poor calibration both in the older and total populations. In other words, the mean predicted risk by the KORA models were significantly different from the observed risk of type 2 diabetes in the PREVEND cohort. One explanation for this is different incidence of type 2 diabetes between KORA and PREVEND cohorts. To further assess the calibration performance of each models, we used logistic recalibration of the original prediction models [21]. After this adjustment, both KORA clinical models showed good calibration.

In conclusion, addition of fasting glucose improved performance of KORA base model in both older and total populations in terms of discrimination and reclassification. Further addition of uric acid did not matter much. Both base and clinical models showed poor calibration. After correction for the intercept and the slope, most KORA models showed good calibration, indicating that there is often a need to adapt prediction models before application in a different setting.

Acknowledgments This work was supported by the Netherlands Heart Foundation, Dutch Diabetes Research Foundation and Dutch Kidney Foundation. This research was performed within the framework of CTMM, the Center for Translational Molecular Medicine (www.ctmm.nl); project PREDICCt (grant 01C-104-07). We thank Prof. Dr. L.T.W. de Jong-van den Berg and Dr. S.T. Visser from the Department of Social Pharmacy, Pharmacoepidemiology and Pharmacotherapy, Groningen University Institute for Drug Exploration, University Medical Center Groningen and University of Groningen, The Netherlands for providing the data on use of glucose-lowering agents according to central pharmacy registration. The Diabetes Cohort Study was funded by a German Research Foundation project grant to W. Rathmann (DFG;RA 459/2–1). The KORA research platform and the KORA Augsburg studies are financed by the Helmholtz Zentrum München, German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education, Science, Research and Technology and by the State of Bavaria. We thank the field staff in Augsburg who were involved in the conduct of the studies. The German Diabetes Center is funded by the German Federal Ministry of Health and the Ministry of Innovation, Science, Research and Technology of the State of North Rhine Westphalia. None of the study sponsors had a role in study design; in data collection, analysis, or interpretation; in writing the report; or in the decision to submit for publication.

Conflict of interest No duality of interest relevant to this manuscript is declared.

References

1. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care. 2004;27:1047–53.
2. Diabetes Prevention Program Research Group. 10-year follow-up of diabetes incidence and weight loss in the diabetes prevention program outcomes study. Lancet. 2009;374:1677–86.
3. Schmidt MI, Duncan BB, Bang H, et al. Identifying individuals at high risk for diabetes: the atherosclerosis risk in communities study. Diabetes Care. 2005;28:2013–8.
4. 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.
5. Stolk RP, Hutter I, Wittek RP. Population ageing research: a family of disciplines. Eur J Epidemiol. 2009;24:715–8.
6. Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. Diabetes Care. 2003;26:725–31.
7. Kahn HS, Cheng YJ, Thompson TJ, Imperatore G, Gregg EW. Two risk-scoring systems for predicting incident diabetes mellitus in U.S. adults age 45–64 years. Ann Intern Med. 2009;150:741–51.
8. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB Sr. Prediction of incident diabetes mellitus in middle-aged adults: the framingham offspring study. Arch Intern Med. 2007;167:1068–74.
9. 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.
10. 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.
11. Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. BMJ. 2011;343:d7163.
12. Vischer UM, Safar ME, Safar H, et al. Cardiometabolic determinants of mortality in a geriatric population: is there a 'reverse metabolic syndrome'? Diabetes Metab. 2009;35:108–14.
13. Weiss A, Boaz M, Beloosesky Y, Kornowski R, Grossman E. Body mass index and risk of all-cause and cardiovascular mortality in hospitalized elderly patients with diabetes mellitus. Diabet Med. 2009;26:253–9.
14. de Ruijter W, Westendorp RG, Assendelft WJ, et al. Use of framingham risk score and new biomarkers to predict cardiovascular mortality in older people: population based observational cohort study. BMJ. 2010;338:a3083.
15. 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;27:1116–1123.
16. Rathmann W, Martin S, Haastert B, et al. KORA study group. performance of screening questionnaires and risk scores for undiagnosed diabetes: the KORA survey 2000. Arch Intern Med. 2005;165:436–41.
17. Rathmann W, Haastert B, Icks A, et al. High prevalence of undiagnosed diabetes mellitus in Southern Germany: target populations for efficient screening. The KORA survey 2000. Diabetologia. 2003;46:182–9.
18. Lambers Heerspink HJ, Brantsma AH, de Zeeuw D, et al. Albuminuria assessed from first-morning-void urine samples versus 24-h urine collections as a predictor of cardiovascular morbidity and mortality. Am J Epidemiol. 2008;168:897–905.
19. Abbasi A, Corpeleijn E, Postmus D, et al. Plasma procalcitonin and risk of type 2 diabetes in the general population. Diabetologia. 2011;54:2463–5.
20. 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.
21. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. J Clin Epidemiol. 2008;61:76–86.
22. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009;338:b605.
23. Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. Am J Epidemiol. 2010;172:971–80.