

A prediction model for type 2 diabetes risk among Chinese people

K. Chien · T. Cai · H. Hsu · T. Su · W. Chang ·
M. Chen · Y. Lee · F. B. Hu

Received: 13 March 2008 / Accepted: 4 November 2008 / Published online: 5 December 2008
© Springer-Verlag 2008

Abstract

Aims/hypothesis A range of prediction rules for the onset of type 2 diabetes have been proposed. However, most studies have been conducted in white groups and it is not clear whether these models apply to Asian populations. The purpose of this study was to construct a simple points model for predicting incident diabetes among Chinese people.

Electronic supplementary material The online version of this article (doi:10.1007/s00125-008-1232-4) contains supplementary material, which is available to authorised users.

K. Chien · F. B. Hu (✉)
Department of Nutrition,
Harvard School of Public Health,
677 Huntington Avenue,
Boston, MA 02115, USA
e-mail: frank.hu@channing.harvard.edu

K. Chien
Institute of Preventive Medicine, College of Public Health,
National Taiwan University,
Taipei, Taiwan

K. Chien · H. Hsu · T. Su · M. Chen · Y. Lee (✉)
Department of Internal Medicine,
National Taiwan University,
No. 7, Chung-San South,
Taipei 100, Taiwan
e-mail: ytleee@ntu.edu.tw

T. Cai
Department of Biostatistics,
Harvard School of Public Health,
Boston, MA, USA

W. Chang
Department of Emergency Medicine,
National Taiwan University Hospital,
Taipei, Taiwan

Methods We estimated the 10 year risk of diabetes in a cohort study of middle-aged and elderly participants who were free from diabetes at baseline. Cox regression coefficients were used to construct the simple points model and the discriminatory ability of the resulting prediction rule was determined using AUC and net reclassification improvement and integrated discrimination improvement statistics. Fivefold random splitting was used to test the internal validity and obtain bootstrap estimates of the AUC. **Results** Of the 2,960 participants without diabetes at the baseline examination, 548 developed type 2 diabetes during a median 10 year follow-up period. Age (four points), elevated fasting glucose (11 points), body mass index (eight points), triacylglycerol (five points), white blood cell count (four points) and a higher HDL-cholesterol (negative four points) were found to strongly predict diabetes incidence in a multivariate model. The estimated AUC for the model was 0.702 (95% CI 0.676–0.727). This model performed better than existing prediction models developed in other populations, including the Prospective Cardiovascular Münster, Cambridge, San Antonia and Framingham models for diabetes risk.

Conclusions/interpretation We have constructed a model for predicting the 10 year incidence of diabetes in Chinese people that could be useful for identifying individuals at high risk of diabetes in the Chinese population.

Keywords Prediction model · Type 2 diabetes

Abbreviations

IDI integrated discrimination improvement
NRI net reclassification improvement
PROCAM Prospective Cardiovascular Münster

Introduction

Prevention of diabetes and its associated complications has become a major public health priority worldwide. Recent clinical trials demonstrated that lifestyle interventions in individuals with impaired glucose tolerance can substantially delay the development of diabetes [1, 2], providing a rationale for the identification of high-risk individuals so as to implement early lifestyle intervention strategies to prevent diabetes. The prediction models for the risk of diabetes can help to guide screening and interventions and to predict diabetes occurrence [3–5]. Routinely available and easily collected clinical and lifestyle-related information has been found to be effective for identifying diabetes cases [6–12]. In addition, a prediction model has been developed in Mexican Americans [13] and further tested in Japanese Americans [14]. This San Antonia model included a variety of diabetes risk factors, including age, sex, ethnicity, fasting glucose, systolic blood pressure, HDL-cholesterol, body mass index and family history, to generate a prediction points system and appeared to have a predictive power similar to that of the diagnostic criteria for metabolic syndrome [13, 15]. Several other prediction models have also been developed, primarily in white populations [3, 5, 16, 17]. However, risk scores developed in white people may not apply to other ethnic groups [18]. Therefore, we specifically developed a prediction model for type 2 diabetes using a community-based cohort of middle-aged and elderly Chinese in Taiwan.

Methods

Study design and participants The National Taiwan University Hospital Committee Review Board approved the study protocol, details of which have been published previously [19, 20]. Briefly, the Chin-Shan Community Cardiovascular Cohort Study began in 1990 by recruiting 1,703 men and 1,899 women of Chinese ethnicity aged 35 years old or older from the Chin-Shan township, 30 km north of metropolitan Taipei, Taiwan. Information about anthropometry, lifestyle and medical conditions was assessed by interview questionnaires in 2 year cycles for the initial 6 years; the validity and reproducibility of the collected data and measurements have been reported in detail elsewhere [21]. Individuals with baseline incomplete blood data ($n=41$), diagnosis of diabetes (fasting glucose ≥ 7.0 mmol/l or with a history of hypoglycaemic medication, $n=473$), or history of cardiovascular disease or cancer ($n=170$) were excluded from this investigation. After these exclusions, we included 2,960 individuals in this study. During the follow-up from 1990 to 2000, 548 individuals developed diabetes defined by fasting glucose levels ≥ 7.0 mmol/l or by the use of oral hypoglycaemic or insulin medication. The response rate of the cohort par-

ticipants was 85.7% at the end of the study. All participants give informed consent.

Body mass index was calculated from weight (kg)/height² (m²). Blood pressure was measured twice in the right arm using a mercury sphygmomanometer with the individual seated comfortably, arms supported and positioned at the level of the heart after resting for 10 min. Hypertension was defined according to the criteria established by the Seventh Joint National Committee [22]: systolic blood pressure ≥ 140 mmHg; diastolic blood pressure ≥ 90 mmHg; or history of hypertension medication. Family history was defined by any parental or first-degree sibling history of diabetes.

Measurement of biochemical markers The procedure for blood collection has been reported elsewhere [23, 24]. Briefly, all venous blood samples drawn after a 12 h overnight fast were immediately refrigerated and transported within 6 h to the National Taiwan University Hospital. Serum samples were then stored at -70°C before batch assay for levels of total cholesterol, triacylglycerol, and HDL-cholesterol. Standard enzymatic tests for serum cholesterol and triacylglycerol were used (Merck 14354 and 14366, Merck, Darmstadt, Germany). HDL-cholesterol levels were measured in supernatant fractions after the precipitation of specimens with magnesium chloride phosphotungstate reagents (Merck 14993). LDL-cholesterol concentrations were calculated as total cholesterol minus cholesterol in the supernatant fraction by the precipitation method (Merck 14992) [25]. Blood samples for glucose analysis were drawn into glass test tubes each containing 80 mol/l fluoride/oxalate reagent. After centrifugation at $1,500\times g$ for 10 min, the glucose level in the supernatant fraction was measured by enzymatic assay (Merck 3389) using an Eppendorf 5060 autoanalyzer. The peripheral-blood-cell analysis was carried out using the blood-cell counter (Sysmex Cell Counter NE-8000, TOA Medical Electronics, Kobe, Japan).

Statistical analysis We used a multivariate Cox proportional hazards model to establish a parsimonious model for predicting risk of diabetes. This model included six significant predictors: age, body mass index, white blood cell count, triacylglycerol, HDL-cholesterol and fasting glucose. Because sex, family history of diabetes, hypertension and systolic blood pressure are biologically important predictors of the risk of diabetes [26], we examined the incremental predictive value of adding these variables to the above model. In addition, lifestyle factors such as physical activity, smoking and drinking alcohol were also tested to derive the parsimonious model. However, the likelihood ratio test suggested that adding these variables into the model did not improve prediction beyond the concise

Table 1 Diabetes risk factors at baseline according to disease status in the study participants and the estimated coefficients and relative risk (95% CI) from the multivariate Cox model

Variable	Total (N=2,960)		No diabetes (n=2,412)		New diabetes (n=548)		p value ^a	Multivariate Cox model			p value ^b	
	Mean	SD	Mean	SD	Mean	SD		β	RR	95% CI		
Age (years)	54	12.3	54	12.4	54	11.7	0.93	0.008	1.01	1.00	1.02	0.043
Body mass index (kg/m ²)	23.3	3.3	23	3.2	24.7	3.6	<0.0001	0.082	1.09	1.06	1.11	<0.0001
White blood cell count (10 ⁹ /l)	6.2	1.6	6.2	1.6	6.5	1.6	<0.0001	0.086	1.09	1.03	1.15	0.002
Triacylglycerol (mmol/l)	1.31	0.88	1.23	0.79	1.64	1.17	<0.0001	0.153	1.17	1.07	1.27	0.0003
HDL-cholesterol (mmol/l)	1.25	0.32	1.27	0.32	1.16	0.32	<0.0001	-0.367	0.69	0.50	0.95	0.024
Fasting glucose (mmol/l)	5.64	0.55	5.58	0.53	5.88	0.55	<0.0001	0.695	2.01	1.71	2.35	<0.0001

^a Significance level for the difference between new diabetes and no diabetes

^b Multivariate adjusted significance level
β, estimated regression coefficient

model. Thus, our final model did not include sex, hypertension or lifestyle factors.

We constructed the categorisation point system according to the concise model using the methods suggested by Sullivan and colleagues [27]. First, we organised the continuous variables into meaningful categories and determined reference values for each variable. Second, we determined the referent risk-factor profile by assigning the median value in each category and estimated how far each category was from the referent in regression units. Then we set a constant to reflect the increase in risk associated with a 10 year increase in age and decided points associated with each of the risk-factor categories. The point totals ranged from -15 to 32. Finally, we constructed the individual's risk from the formula:

$$\text{Risk} = 1 - S_0(t)^{\exp(\sum \beta X - \sum \beta \bar{X})}$$

where $S_0(t)$ was the average survival at time t (e.g. $t=10$ years) or the survival rate at the mean values of the risk factor, βX values were approximated from the sum of baseline risk and product of point totals and the constant. $\beta \bar{X}$ values were the sum of the products of the regression coefficients and means or the proportions of the variables.

We also constructed the regression coefficient-based model by assigning β values as estimated regression coefficients.

We conducted the internal validation of the simple points model and obtained a bias-corrected estimate of the AUC using a fivefold cross-validation procedure [28]. We randomly split the data into five equal parts. For $k=1, \dots, 5$, we used the k th part as the validation dataset and the remaining four parts as the training dataset. For each partition of the validation and training sets, we obtained the coefficients from the training set and assigned the points from the coefficients in the validation set to evaluate performance by estimating the AUC for the corresponding simple points model. Then we evaluated the overall performance of the

points model by averaging the AUC estimates obtained from the five different partitions. The ranges of AUC across the five partitions indicated the stability of the prediction model [29]. To account for the variability in estimating the model

Table 2 Simple points system according to the concise model

Risk factor	Category	Points
Age (years)	35–44	-1
	45–54	0
	55–64	1
	65–74	2
	≥75	3
Body mass index (kg/m ²)	<21	-3
	21–22	0
	22–24	2
	24–26	4
	≥26	8
White blood cell count (×10 ⁹ /l)	<4.9	-1
	4.9–5.7	0
	5.7–6.4	1
	6.4–7.5	2
	≥7.5	4
Triacylglycerol (mmol/l)	<0.70	-1
	0.70–0.92	0
	0.93–1.19	0
	1.19–1.71	1
	≥1.72	5
HDL-cholesterol (mmol/l)	<0.98	1
	0.98–1.13	0
	1.14–1.29	-1
	1.30–1.49	-2
	≥1.50	-4
Fasting glucose (mmol/l)	<5.16	-5
	5.16–5.43	0
	5.44–5.71	3
	5.72–6.10	6
	≥6.11	11

variables and the AUC, we used the bootstrap method to construct 95% confidence intervals for the AUC. The standard-error estimates and the confidence intervals were obtained based on 1,000 bootstrap samples.

Finally, we compared the performance of the proposed prediction model with that of various prediction models derived from other populations, including Cambridge [8–10], Prospective Cardiovascular Münster (PROCAM) [30], San Antonio [13, 14] and Framingham [5]. AUC was used to compare the discriminatory capabilities of these models and our simple points model. An AUC curve is a graph of sensitivity vs 1–specificity (or false-positive rate) for various cut-off definitions of a positive diagnostic test result [31]. We listed the sensitivity and specificity for the best cut-off values from various models. Statistical differences in the AUCs were compared using the method of DeLong et al. [32]. In addition, we assessed the goodness of fit for all models based on the Hosmer–Lemeshow test [33]. The global summary statistics, including Yates slope [34, 35], Brier score [34] and discrimination *C* statistics [36], were calculated in various models. Moreover, we compared the simple points model with other models by using the net reclassification improvement (NRI) and integrated discrimination improvement (IDI) statistics [37]. The NRI statistic was based on the reclassification tables and was calculated from a sum of differences between the ‘upward’ movement in categories for event participants and the ‘downward’ movement in those for non-event participants [37]. We presented the NRI according to the a priori risk categories of diabetes (0–15%, 15–20%, 20–25%, and $\geq 25\%$). The IDI can be interpreted as a difference between improvement in average sensitivity and any potential increase in average ‘1–specificity’, and the statistic was a difference in Yates discrimination slopes between the new and old models [34, 35].

We also ran a clinical model that included age, sex, BMI, family history and antihypertensive medication, but not requiring blood test results, and used the AUC to compare the predictive ability of this model with the model that included laboratory-based measures.

All statistical tests were two-sided with a type I error of 0.05, and *p* values < 0.05 were considered statistically significant. Analyses were performed with SAS version 9.1 (SAS Institute, Cary, NC, USA) and Stata version 9.1 (Stata Corporation, College Station, TX, USA).

Results

Of the 2,960 participants without diabetes at the baseline examination, 548 developed type 2 diabetes during a 10 year follow-up period. Among the 548 participants with incident diabetes, 396 were not receiving pharmacological treatment for diabetes and were given a diagnosis exclusively on the

basis of plasma glucose levels that met the American Diabetes Association criteria (≥ 7.0 mmol/l). Of the 136 participants with confirmed pharmacological treatment for diabetes, 78 also met plasma glucose criteria for diabetes. The baseline characteristics of study participants and the results of a Cox multivariate analysis that included age, body mass index, white cell count, triacylglycerol, HDL-cholesterol and fasting glucose level are shown in Table 1. Participants who developed incident diabetes also tended to have increased systolic blood pressure and a higher prevalence of family history of diabetes, but these two variables were not significantly associated with risk of diabetes after adjustment for other covariates. Therefore, they were not included in the final model.

We developed a simple points system to estimate the diabetes risk using the baseline survival function at 10 years and the coefficients of the concise model (Table 2): age (four points), elevated fasting glucose (11 points), body mass index (eight points), triacylglycerol (five points), white blood cell count (four points) and a higher HDL-cholesterol (negative four points). This approach allowed manual estimation of 10 year risk of developing diabetes for each individual, as shown in Table 3.

By using the simple points system to estimate risk of incident diabetes during a 10 year follow-up interval, we determined that 42% of the sample had a risk below 20%, 28% had risk a 20–30% risk, and 30% had a risk of 30% or higher. This simple points model has a good discriminatory ability with an AUC of 0.702 (95% CI 0.676–0.727). Adding sex, hypertension and family history of diabetes did not improve the predictive power (AUC 0.700, 95% CI 0.675–

Table 3 The total points and absolute risk function for the categorical model in the simple points model

Points total	Risk	Points total	Risk	Points total	Risk
-15	0.05	1	0.16	17	0.46
-14	0.05	2	0.18	18	0.48
-13	0.06	3	0.19	19	0.51
-12	0.06	4	0.20	20	0.54
-11	0.07	5	0.22	21	0.57
-10	0.07	6	0.23	22	0.59
-9	0.08	7	0.25	23	0.62
-8	0.09	8	0.26	24	0.65
-7	0.09	9	0.28	25	0.68
-6	0.10	10	0.30	26	0.71
-5	0.11	11	0.32	27	0.73
-4	0.11	12	0.34	28	0.76
-3	0.12	13	0.36	29	0.79
-2	0.13	14	0.39	30	0.81
-1	0.14	15	0.41	31	0.84
0	0.15	16	0.43	32	0.86

Table 4 Sensitivity, specificity, best Youden index and likelihood ratios of the optimal cut-off value for the risk of diabetes in each model

Model	Optimal cut-off value	Sensitivity	Specificity	Youden index	Correctly classified	LR+	LR–
PROCAM	(≥ -14.4)	0.66	0.56	1.21	0.58	1.48	0.62
Cambridge	(≥ -2.1)	0.72	0.40	1.12	0.46	1.20	0.70
Framingham	(≥ -2.9)	0.55	0.72	1.27	0.69	1.96	0.62
San Antonia	(≥ -0.2)	0.48	0.78	1.26	0.72	2.16	0.67
Logistic, coefficient-based	(≥ -1.5)	0.69	0.62	1.31	0.63	1.81	0.50
Simple points	(≥ 13)	0.52	0.78	1.30	0.73	2.36	0.62
Cox, coefficient-based	(≥ 6.7)	0.62	0.70	1.31	0.68	2.02	0.55

LR+, likelihood ratio for a positive test result; LR–, the likelihood ratio for a negative test result

0.725). In addition, the regression-coefficient-based model has a similar AUC value (0.701, 95% CI 0.675–0.726). The optimal cut-off value for the simple points model was set as 13, with a sensitivity of 0.52 and a specificity of 0.78 (Table 4). We found that the simple points model had the highest proportions of correctly classified and best likelihood ratio values among all models, and had a Youden index value similar to those of the coefficient-based models. The clinical model based on anthropometric measures and medication had a lower predictive ability (AUC 0.646, 95% CI 0.621–0.672) than the simple points model that included blood test results ($p < 0.001$).

The within-study model validation was assessed by the aforementioned fivefold cross-validation procedure. The AUCs for the five partitions ranged from 0.664 to 0.711, indicating a moderately high reliability of discrimination for the model in repeated random-sample subsets. The AUCs from the PROCAM, Cambridge, San Antonia and Framingham models were significantly lower than that from our simple points model (Fig. 1).

Table 5 presents summary statistics for the performance of models in terms of predicting diabetes risk in the cohort dataset. The simple points model had the highest Yates' slope and C statistics, indicating good discriminatory ability. When comparing the predicted and observed risks, we observed a non-significant p value for the Hosmer–Lemeshow statistics for the simple points model, indicating good calibration. The San Antonia model was most similar to the simple points model, with the smallest NRI and IDI values and the highest correlation coefficient. In addition, we calculated the NRI according to different risk categories and the results were similar. We found the NRI values were similar in the quartiles and the 15–20–25% risk categories.

Discussion

Statement of principal findings Using a community-based cohort study, we developed a simple points model to predict 10 year risk of type 2 diabetes in a Chinese population based

on six variables: age, body mass index, white blood cell count, triacylglycerol, HDL-cholesterol and fasting glucose levels. These values could be relatively easily obtained in clinical practice and the points system we developed is simple to use. The availability of the simple clinical tool to predict future risk of disease, as has been the case for prediction of coronary heart disease, should improve the prediction of diabetes risk, identify high-risk populations and enhance preventive strategies.

Existing diabetes risk functions Several diabetes-prediction models have previously been developed in various populations. In cross-sectional studies conducted in the USA and Europe, prediction models based on clinical information and lifestyle-related factors have appeared to be useful for identifying undiagnosed diabetes cases and high HbA_{1c} levels in screening populations [7, 8, 38, 39]. For example, the Cambridge model has been applied successfully to identify individuals with high HbA_{1c} levels [9, 10]. In addition, a recent cohort study showed that the Cambridge model was useful for identifying individuals with a higher

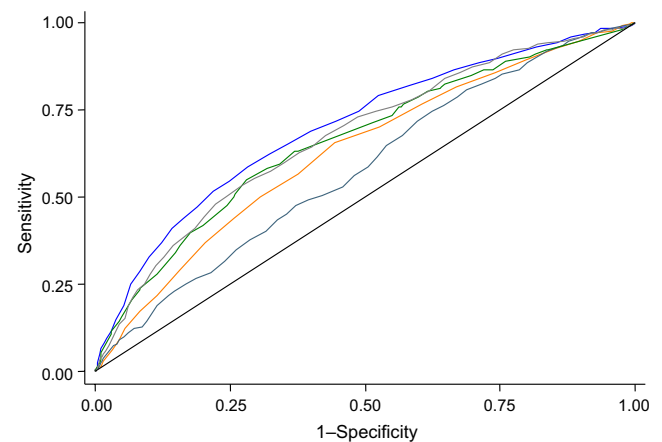


Fig. 1 Receiver-operating characteristic curves for various models applied to the study population. Blue, simple (AUC 0.701); grey, San Antonia (AUC 0.675); green, Framingham (AUC 0.662); orange, PROCAM (AUC 0.631); dark blue, Cambridge (AUC 0.581); black, reference (AUC 0.500)

Table 5 Summary statistics comparing risk prediction algorithms with prediction based on covariates in the PROCAM, Cambridge, Framingham, San Antonio and the simple points algorithms for the study cohort ($N=2960$)

Model	Yates' slope (%) ^a	Brier score (%) ^a	Discrimination C statistic ^a	Hosmer–Lemeshow p value	NRI (%) ^b	IDI (%) ^b	Correlation coefficient ^b
PROCAM	2.3	14.8	0.630	0.008	36.3	6.58	0.563
Cambridge	1.1	14.9	0.580	0.001	45.3	7.82	0.340
Framingham	5.2	14.4	0.662	0.002	19.2	3.74	0.775
San Antonio	6.3	14.2	0.675	0.032	8.3	2.65	0.812
This study	8.9	13.8	0.701	0.874	–	–	–

^a Lower Brier scores and higher Yates' slope as well as higher discrimination C statistics represented better performance

^b Compared with this study model

risk of type 2 diabetes during follow-up [40]. Other cohort-based risk models, such as the FINDRISC (Finnish Diabetes Risk Score) [3], PROCAM [30], San Antonio [13] and Framingham models [5], were developed to predict incident diabetes in different populations. We did not compare our model with FINDRISC or the German risk score [17, 41] because dietary information, which is needed for risk scoring, was not collected from our cohort.

Using age, body mass index, fasting glucose, HDL-cholesterol, family history and hypertension status, von Eckardstein and colleagues developed a model (PROCAM) to predict the incidence of diabetes during 6.3 years of follow-up in a German population [30]. Stern and colleagues also constructed the San Antonio model to predict diabetes risk in Mexican Americans and non-Hispanic whites during a 7.5-year follow-up period [13] and this model was validated among Japanese Americans during a 10 year follow-up period [14]. The San Antonio model included biomarkers such as fasting glucose, blood pressure and HDL-cholesterol, in addition to age, sex, obesity and family history of diabetes. Recently, Wilson and colleagues developed a simple points diabetes-prediction model for the Framingham Offspring Study [5]. This model included fasting glucose, body mass index, HDL-cholesterol, family history of diabetes, triacylglycerol and hypertension. Among these models, only the San Antonio model resembled our simple points model in terms of predictive power (see Electronic supplementary material [ESM] Table 1).

Most of the variables included in our model were similar to those in previous risk functions. Moreover, white cell count, a marker of chronic inflammation, was incorporated into our prediction model. In addition to fasting glucose and BMI, metabolic variables such as high triacylglycerol and low HDL-cholesterol were found to be strong predictors of type 2 diabetes in our study. These variables in our model have also been included in prediction model developed for other populations [39, 42]. Moreover, we found that BMI had a discriminatory power similar to that of waist circumference in our model. Although family history of

diabetes and hypertension were important predictors of diabetes in other studies, adding these variables to our risk scores did not improve risk prediction. In addition, we did not include lifestyle factors such as physical activity and smoking because they were not significant predictors of diabetes after adjusting for other risk factors.

Strengths and weaknesses of the study To our knowledge, this is the first diabetes prediction model specifically developed for a Chinese population. Because of the large sample size, the estimates from our prediction models were found to be stable, as demonstrated by the internal validation study. Also, the use of a community-based population could reduce the possibility of selection bias. However, several potential limitations of this study should be mentioned.

First, the discriminatory capability of our simple points model was only moderately high (the AUCs were ~0.70), and somewhat lower than that of other prediction models from other populations.

Second, we did not include extensive biomarker data in the model. Adding other biomarkers such as insulin resistance may improve the discriminatory ability. However, these variables are more difficult to measure and interpret in clinical practice.

Third, our cohort has a higher diabetes incidence rate than Chinese women in Shanghai, China [43]. The high risk of diabetes among our study participants may be explained by older age, higher body mass index and a single measurement for case diagnosis [44].

Fourth, fasting plasma glucose but not post-challenge glucose was used to define the incidence of diabetes and to exclude individuals with undiagnosed diabetes at baseline. This is likely to have led to some error in estimation of the risk of diabetes and hence the performance of our model. Although fasting glucose is widely measured in routine health checkups in Taiwan [45], our model may be less practical in other populations in which blood test results for white cell count and fasting glucose, triacylglycerol and cholesterol are not routinely available or easily obtained.

In conclusion, we have constructed a simple points model for predicting the 10 year incidence of diabetes, and this model performed significantly better than other existing diabetes prediction models within the study population of ethnic Chinese people. This simple clinical tool should help identify high-risk populations and improve preventive and treatment strategies for the Chinese population.

Acknowledgements We thank the participants in the Chin-Shan community and the cardiologists at the National Taiwan University Hospital for their assistance in this study. This study was supported partly by grants from the National Science Council in Taiwan. F. Hu's research is partly supported by the American Heart Association Established Investigator Award. We are thankful to M. Pencina for kindly providing the SAS macro for computing NRI and IDI statistics.

Duality of interest The authors declare that there is no duality of interest associated with this manuscript

References

- Tuomilehto J, Lindstrom J, Eriksson JG et al (2001) Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med* 344:1343–1350
- Knowler WC, Barrett-Connor E, Fowler SE et al (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 346:393–403
- Lindstrom J, Tuomilehto J (2003) The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 26:725–731
- Schmidt MI, Duncan BB, Bang H et al (2005) Identifying individuals at high risk for diabetes: The Atherosclerosis Risk in Communities study. *Diabetes Care* 28:2013–2018
- Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB Sr (2007) Prediction of Incident Diabetes Mellitus in Middle-aged Adults: The Framingham Offspring Study. *Arch Intern Med* 167:1068–1074
- Baan CA, Ruige JB, Stolk RP et al (1999) Performance of a predictive model to identify undiagnosed diabetes in a health care setting. *Diabetes Care* 22:213–219
- Glumer C, Carstensen B, Sandbaek A, Lauritzen T, Jorgensen T, Borch-Johnsen K (2004) A Danish diabetes risk score for targeted screening: the Inter99 study. *Diabetes Care* 27:727–733
- Griffin SJ, Little PS, Hales CN, Kinmonth AL, Wareham NJ (2000) Diabetes risk score: towards earlier detection of type 2 diabetes in general practice. *Diabetes Metab Res Rev* 16:164–171
- Park PJ, Griffin SJ, Sargeant L, Wareham NJ (2002) The performance of a risk score in predicting undiagnosed hyperglycemia. *Diabetes Care* 25:984–988
- Thomas C, Hyponen E, Power C (2006) Type 2 diabetes mellitus in midlife estimated from the Cambridge Risk Score and body mass index. *Arch Intern Med* 166:682–688
- Al-Lawati JA, Tuomilehto J (2007) Diabetes risk score in Oman: a tool to identify prevalent type 2 diabetes among Arabs of the Middle East. *Diabetes Res Clin Pract* 77:438–444
- Ramachandran A, Snehalatha C, Vijay V, Wareham NJ, Colagiuri S (2005) Derivation and validation of diabetes risk score for urban Asian Indians. *Diabetes Res Clin Pract* 70:63–70
- Stern MP, Williams K, Haffner SM (2002) Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test? *Ann Intern Med* 136:575–581
- McNeely MJ, Boyko EJ, Leonetti DL, Kahn SE, Fujimoto WY (2003) Comparison of a clinical model, the oral glucose tolerance test, and fasting glucose for prediction of type 2 diabetes risk in Japanese Americans. *Diabetes Care* 26:758–763
- Lorenzo C, Okoloise M, Williams K, Stern MP, Haffner SM (2003) The metabolic syndrome as predictor of type 2 diabetes: the San Antonio heart study. *Diabetes Care* 26:3153–3159
- Herman WH, Smith PJ, Thompson TJ, Engelgau MM, Aubert RE (1995) A new and simple questionnaire to identify people at increased risk for undiagnosed diabetes. *Diabetes Care* 18:382–387
- Schulze MB, Hoffmann K, Boeing H et al (2007) An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. *Diabetes Care* 30:510–515
- Glumer C, Vistisen D, Borch-Johnsen K, Colagiuri S (2006) Risk scores for type 2 diabetes can be applied in some populations but not all. *Diabetes Care* 29:410–414
- Chien KL, Hsu HC, Su TC, Chen MF, Lee YT, Hu FB (2007) Apolipoprotein B and non-high-density lipoprotein cholesterol and risk of coronary heart disease in Chinese. *J Lipid Res* 48:2499–2505
- Lee YT, Lin RS, Sung FC et al (2000) Chin-Shan Community Cardiovascular Cohort in Taiwan: baseline data and five-year follow-up morbidity and mortality. *Journal of Clinical Epidemiology* 53:836–846
- Lee YT, Sung FC, Lin RS et al (2001) Peripheral blood cells among community residents living near nuclear power plants. *Science of the Total Environment* 280:165–172
- Chobanian AV, Bakris GL, Black HR et al (2003) The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA* 289:2560–2572
- Chien KL, Lee YT, Sung FC, Hsu HC, Su TC, Lin RS (1999) Hyperinsulinemia and related atherosclerotic risk factors in the population at cardiovascular risk: a community-based study. *Clinical Chemistry* 45:838–846
- Chien KL, Sung FC, Hsu HC, Su TC, Chang WD, Lee YT (2005) Relative importance of atherosclerotic risk factors for coronary heart disease in Taiwan. *Eur J Cardiovasc Prev Rehabil* 12:95–101
- Wieland H, Seidel D (1983) A simple specific method for precipitation of low density lipoproteins. *Journal of Lipid Research* 24:904–909
- Schulze MB, Hu FB (2005) Primary prevention of diabetes: what can be done and how much can be prevented? *Annu Rev Public Health* 26:445–467
- Sullivan LM, Massaro JM, D'Agostino RB Sr (2004) Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med* 23:1631–1660
- Efron B, Tibshirani RJ (1994) An introduction to the bootstrap. Chapman & Hall/CRC, Boca Raton
- Cai T, Zheng Y (2007) Model checking for ROC regression analysis. *Biometrics* 63:152–163
- von Eckardstein A, Schulte H, Assmann G (2000) Risk for diabetes mellitus in middle-aged Caucasian male participants of the PROCAM study: implications for the definition of impaired fasting glucose by the American Diabetes Association. *Prospective Cardiovascular Munster. J Clin Endocrinol Metab* 85:3101–3108
- Hanley JA, McNeil BJ (1983) A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148:839–843
- DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44:837–845
- Hosmer DW Jr, Lemeshow S (1989) The multiple logistic regression model. In: *Applied Logistic Regression*. Wiley, New York, pp 25–37

34. Yates J (1982) External correspondence: decomposition of the mean probability score. *Organizational Behavior and Human Performance* 30:132–156
35. Schmid C, Griffith J (1998) Multivariable classification rules: calibration and discrimination. In: Armitage P, Colton T (eds) *Encyclopedia of Biostatistics*. Wiley, Chichester
36. Cook NR (2007) Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 115:928–935
37. Pencina MJ, D'Agostino RBS, D'Agostino RBJ, Vasan RS (2008) Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Stat Med* 27:157–172
38. Diabetes Prevention Program Research Group (2005) Strategies to identify adults at high risk for type 2 diabetes: the Diabetes Prevention Program. *Diabetes Care* 28:138–144
39. Kanaya AM, Wassel Fyr CL, de Rekeneire N et al (2005) Predicting the development of diabetes in older adults: the derivation and validation of a prediction rule. *Diabetes Care* 28:404–408
40. Rahman M, Simmons RK, Harding AH, Wareham NJ, Griffin SJ (2008) A simple risk score identifies individuals at high risk of developing Type 2 diabetes: a prospective cohort study. *Fam Pract* 25:191–196
41. Silventoinen K, Pankow J, Lindstrom J, Jousilahti P, Hu G, Tuomilehto J (2005) The validity of the Finnish Diabetes Risk Score for the prediction of the incidence of coronary heart disease and stroke, and total mortality. *Eur J Cardiovasc Prev Rehabil* 12:451–458
42. D'Agostino RB Jr, Hamman RF, Karter AJ, Mykkanen L, Wagenknecht LE, Haffner SM (2004) Cardiovascular disease risk factors predict the development of type 2 diabetes: the insulin resistance atherosclerosis study. *Diabetes Care* 27:2234–2240
43. Villegas R, Liu S, Gao YT et al (2007) Prospective study of dietary carbohydrates, glycemic index, glycemic load, and incidence of type 2 diabetes mellitus in middle-aged Chinese women. *Arch Intern Med* 167:2310–2316
44. Chien KL, Chen MF, Hsu HC et al (2008) Plasma Uric Acid and the Risk of Type 2 Diabetes in a Chinese Community. *Clin Chem* 54:310–316
45. Chen TH, Chiu YH, Luh DL et al (2004) Community-based multiple screening model: design, implementation, and analysis of 42,387 participants. *Cancer* 100:1734–1743