Int J Public Health (2010) 55:315–323
DOI 10.1007/s00038-010-0130-y

ORIGINAL ARTICLE

San Antonio heart study diabetes prediction model applicable to a Middle Eastern population? Tehran glucose and lipid study

Mohammadreza Bozorgmanesh · Farzad Hadaegh · Azadeh Zabetian · Fereidoun Azizi

Received: 12 November 2009 / Revised: 9 February 2010 / Accepted: 10 February 2010 / Published online: 9 March 2010

Abstract
Objectives To assess the validity of the San Antonio heart study (SAHS) diabetes prediction model in a large representative Iranian population.
Methods A risk function derived from data in the SAHS to predict the 7.5-year risk of diabetes, was tested for its ability to predict incident diabetes in 3,242 individuals aged ≥20 years. The performance or ability to accurately predict diabetes risk, of the SAHS function compared with the performance of risk functions developed specifically from the Tehran lipid and glucose study. Comparisons included goodness of fit, discrimination, and calibration.
Results The participants were followed for 6.3 years. The area under the receiver operating characteristic curve (AROC) for diabetes of SAHS model was 0.83 (95% CI 0.80–0.86). The model overestimated the risk of diabetes in TLGS population with the overall bias of 111%. After the recalibration, the model-predicted probability agreed well with the actual observed 6-year risk of diabetes.
Discussion and conclusion The American SAHS was a prediction model for diabetes with good discrimination in an Iranian target population after calibration.

M. Bozorgmanesh · F. Hadaegh (✉) · A. Zabetian · F. Azizi
Prevention of Metabolic Disorders Research Center,
Research Institute for Endocrine Sciences,
Shahid Beheshti University of Medical Sciences,
P.O. Box 19395-4763, Tehran, Islamic Republic of Iran
e-mail: fzhadaegh@endocrine.ac.ir

M. Bozorgmanesh
e-mail: mr_bozorgmanesh@endocrine.ac.ir

A. Zabetian
e-mail: zabetian@endocrine.ac.ir

F. Azizi
e-mail: azizi@endocrine.ac.ir

Keywords Prediction · Model · Diabetes · Validation · General practice · Screening

Introduction

The Middle East is expected to bear one of the world's greatest increases in the absolute burden of diabetes (hereafter diabetes) in the near future (Wild et al. 2004; Hadaegh et al. 2008). Some investigators have used cross-sectional data to develop models to detect prevalent diabetes (Colagiuri et al. 2002). Definitive evidence of benefit, however, is lacking on the role of early detection of undiagnosed diabetes in asymptomatic individuals as a strategy to reduce the personal, public, and economic cost of diabetes (Colagiuri et al. 2004). On the other hand, several recent clinical trials have shown that diabetes in high-risk individuals can be prevented or the onset delayed by lifestyle modifications or pharmaceutical intervention (Pan et al. 1997; Knowler et al. 2002). The evidence for diabetes prevention is based on identifying individuals with impaired glucose tolerance (IGT) using 2-h post-challenge plasma glucose (2h-PCPG) levels test, which is time consuming, costly, and inconvenient (McNeely et al. 2003). These findings have become the basis on which the investigations to use longitudinal data, from cohorts free of diabetes at baseline, for predicting the development of diabetes in 8–10 years. Models have included readily available demographic, anthropometric, and clinical risk factors and diagnostic tests (Stern et al. 2002; Lindstrom and Tuomilehto 2003). Individualized primary prevention in high-risk individuals to prevent diabetes is a feasible and attractive way to reduce diabetes-related morbidity and mortality. Therefore, accurately predicting diabetes may help targeted interventions to prevent diabetes in every day clinical practice, and facilitate the efficient allocation of resources in public health policy; the major challenge in this case would be to identify those with high risk of developing diabetes in the future (Stern et al. 2002; Lindstrom and Tuomilehto 2003; McNeely et al. 2003). Previous models have been shown to have the same performances, with a sensitivity of 70–75% and specificity of 55–70% (Lindstrom and Tuomilehto 2003; Glumer et al. 2004). The majority of models have been developed and validated in Caucasian population and it has been shown that they could not be necessarily applicable to other populations of different ethnic origin (Glumer et al. 2006). Currently proposed models yielded low validity when applied to a new population, most likely due to differences in population characteristics. Models to predict future risk for diabetes should be simple, parsimonious, and accurate. The choice of predictors depends in part on how the model will be implemented. To choose a model, the availability of risk factor data in the clinical setting, the optimal cut-point to define high-risk individuals, and the simplicity of the model should be considered (Herman 2009). Performance of diabetes risk models should be assessed in the target population where they will be applied.

The 7.5-year diabetes risk prediction equation with simple clinical variables developed using data from the San Antonio Heart Study (SAHS), and has been reported to have a good discriminatory capacity. The utility of the San Antonio diabetes prediction model (SADPM) has been tested in some populations such as, the German cross-sectional cohort, the Japanese American prospective cohort, and most recently in a Chinese population, with conflicting results on external validity in populations of different ethnic backgrounds (McNeely et al. 2003; Rathmann et al. 2005; Chien et al. 2009). All variables used in this equation [age, sex, ethnicity, systolic blood pressure, high-density lipoprotein cholesterol (HDL-C), body mass index (BMI), and family history of diabetes and fasting plasma glucose (FPG)] were available in the Tehran lipid and glucose study (TLGS) and the age of the participants of the SAHS (mean 42.6–44.8 years) was close to those of TLGS (41.6 years); we, therefore, assessed the applicability of SAHS diabetes prediction model (SADPM) to an Iranian adult population.

Methods

Study population

In 1999–2001 the TLGS recruited a population-based cohort of more than 15,000 residents of district 13 of Tehran aged over 3 years, as a representative of Tehran population to participate in a baseline study. The rationale and design of the study has been described elsewhere (Azizi et al. 2009). Among the subjects $\geq$20 years ($n = 10{,}368$) we excluded individuals assigned to intervention study ($n = 3{,}931$), those with prevalent diabetes mellitus (using oral hypoglycemic agents or insulin, baseline FPG $\geq$ 7.0 mmol/l or 2h-PCPG $\geq$11.1 mmol/l, $n = 698$), and those with incomplete data on diabetes status ($n = 623$) or baseline clinical measurements ($n = 98$). Remaining 5,018 individuals were followed up to the second (2002–2005) and third (2005–2008) TLGS examinations, $\sim$3 years apart, for an average period of 6.3 years (Fig. 1). Participants who did not attend follow-up study ($n = 1{,}297$) and those who left the study before the end of the follow-up period without first developing diabetes ($n = 479$) were excluded from the final analysis leaving 3,242 participants (participation rate = 65%).The screening method in the baseline as well as the follow-up visits was the same and included measurement of both FPG and 2h-PCPG. The main reasons for lack of attendance at follow-up examinations, despite repeated calls, were either migration or other personal reasons. Informed written consent was obtained from all subjects and the Ethical Committee of Research Institute for Endocrine Sciences approved this study.

Clinical, anthropometric, and laboratory measurements

The baseline examination included information on age, sex, family history of diabetes and medication use. Height, weight, waist circumference (WC), blood pressure, plasma glucose level, total and HDL-C, and serum triglycerides (TGs) levels were measured by using previously reported methods (Azizi et al. 2009); BMI calculated as weight (kg) divided by height (m) squared. The standard 2h-PCPG was performed for all individuals $\geq$20 years, not on anti-diabetic drugs. Serum low-density lipoprotein cholesterol (LDL-C) levels were estimated by the Friedwald formula for those with serum TG <400 mg/dl (Friedwald et al. 1972).

Definition of variables and outcomes

Participants were classified as developing new diabetes during follow-up if they had FPG $\geq$7 mmol/l, or 2h-PCPG $\geq$11.1 mmol/l or taking anti-diabetic medication (Genuth et al. 2003).

Statistical analysis

Mean and standard deviation were used to describe continuous variables, while frequency distributions were obtained for categorical variables. The Student's t test was used to assess the significance of associations for continuous data, χ^2 for categorical, and Mann–Whitney U statistic for ordinal variables.

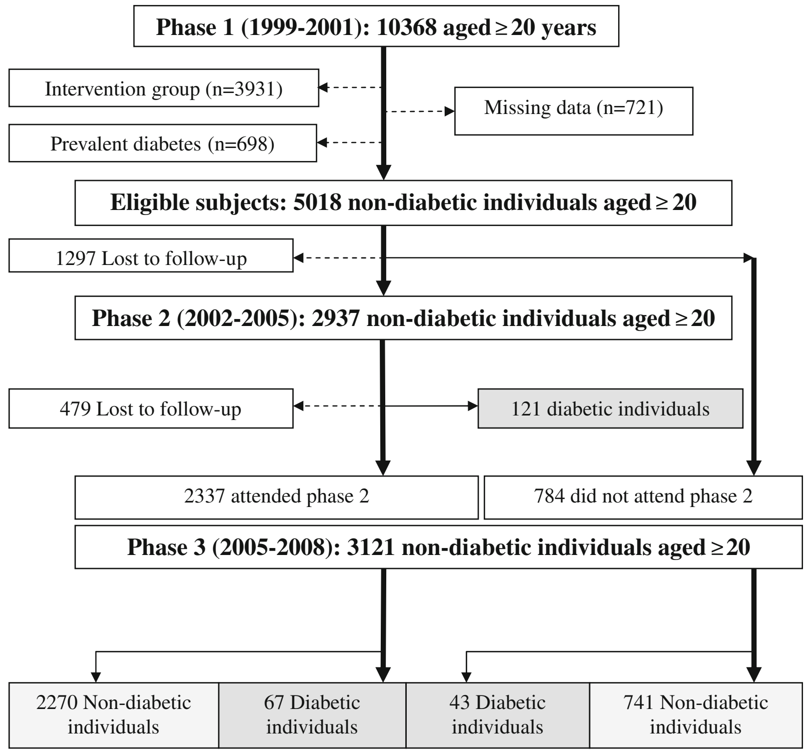

Fig. 1 Study population: included, excluded, lost to the follow-up and events, and censored individuals

Our primary focus was to examine the transferability of the SADPM to a representative population of Iran. The performance of the SADPM among the TLGS cohort was assessed according to three evaluations: re-estimation, discrimination, and calibration.

Re-estimation shows the contribution to the incident diabetes of predictors used in the SADPM in the TLGS. We re-estimated the SADPM by developing multiple logistic regression models to predict incident diabetes, using the same variables as those incorporated in SADPM. Following Stern et al. (2002), we refer to this model as the full model, which included age, sex, FPG and 2h-PCPG, systolic and diastolic blood pressures, total, LDL and HDL cholesterol levels, TGs, BMI, and family history of diabetes. The statistical significance of each interaction term between sex and other factors was tested and likelihood ratio test was used to globally compare models with and without interactions. The importance of the 2h-PCPG levels for predicting diabetes was assessed by comparing the full model with a nested model that excluded the 2h-PCPG. We also examined the SADPM, a simplified model based on widely recognized diabetes risk factors. The variables used in the clinical model were age, sex, fasting glucose level, systolic blood pressure, HDL cholesterol level, BMI, and family history of diabetes. This model was also examined with (or without) selected interactions, and with (or without) 2h-PCPG.

Calibration measures how closely predicted risk is in concordance with observed actual outcomes (Hanley and McNeil 1982). We estimated the risk of incident diabetes for each participant by applying the SADPM function to the TLGS data. SADPM-predicted risks were then used to divide the observations into deciles of predicted risk. Predicted and actual numbers of events for each decile were then calculated and plots were constructed. The chi-square test proposed by Hosmer and Lemeshow (2000) was used to test the hypothesis that the observed and the expected number of events in each decile are equal.

Discrimination is the ability of a prediction model to separate those who experience an events from those who do not. We quantified this by calculating the area under a receiver operating characteristic (ROC) curve (AROC) for the multivariable models and 2h-PCPG. The AROCs were compared implementing the approach suggested by DeLong (DeLong et al. 1988). To obtain unbiased estimates of predictive accuracy SPSS 17.0 was used to implement an internal bootstrap on a total of 20,000 random samples of 3,242 subjects and the ability of the SADPM to discriminate participants with future diabetes was tested.

Nested models were compared by using likelihood ratio testing as global measure of goodness of fit.

Overall bias in predicted diabetes incidence was estimated as

Table 1 Baseline characteristics

Predictor	Men (n = 1,368)	Women (n = 1,874)	P value
Age (years)	42.8 (14.0)	40.7 (12.5)	<0.001
Family history of diabetes (%)	24.3	28.3	0.010
Systolic blood pressure (mmHg)	119.8 (17.2)	116.8 (17.5)	<0.001
Diastolic blood pressure (mmHg)	78.2 (10.5)	77.3 (10. 1)	0.019
Body mass index (kg/m^2)	25.7 (3.9)	27.6 (4.8)	<0.001
Waist circumference (cm)	88.1 (10.9)	87.2 (12.3)	0.020
Height (cm)	170.2 (6.6)	157.0 (5.9)	<0.001
Weight (kg)	74.4 (12.3)	68.0 (12.2)	<0.001
Waist-to-hip ratio (%)	91.4 (6.9)	83.3 (7.9)	<0.001
Waist-to-height ratio (%)	51.9 (6.6)	55.6 (8.3)	<0.001
Total cholesterol (mmol/l)	5.3 (1.1)	5.4 (1.2)	<0.001
HDL-C (mmol/l)	0.4 (0.1)	0.5 (0.1)	<0.001
LDL-C (mmol/l)	1.5 (0.4)	1.5 (0.4)	0.028
Triglycerides (mmol/l)	2.0 (1.3)	1.7 (1.1)	<0.001
FPG (mmol/l)	5.1 (0.5)	5.0 (0.5)	<0.001
2h-PCPG (mmol/l)	5.7 (1.7)	6.1 (1.6)	<0.001

Data are presented as mean (SD) and percent for continuous and categorical variables, respectively

FPG fasting plasma glucose, *HDL-C* high density lipoprotein cholesterol, *LDL-C* low density lipoprotein cholesterol, *2h-PCPG* 2-h post-challenge plasma glucose

$$\text{Over all bias} = [(\text{Predicted incidence} - \text{Observed incidence})/(\text{Observed incidence})] \times 100.$$

When there is a systematic overestimation or underestimation of risk, transferring a prediction function from one setting to another needs a process known as recalibration. Because the SADPM was derived from a data of longer follow-up duration than that of our study, and hence, with higher incident diabetes, we performed a simple recalibration by multiplying the risk of each individual by the ratio of the SAHS diabetes incidence rate into the TLGS diabetes incidence rate.

Comparability between the participants and non-participants was supported by propensity score analysis. We developed a logistic regression models in which participation was predicted using baseline data on 5,018 participants regarding age, sex, systolic and diastolic blood pressure, TC, HDL-C, TG, BMI, FPG, 2h-PCPG, and family history of diabetes. We used the predicted values from these models (the propensity scores) in the main analysis models as one of the covariates and observed that propensity for participation was unrelated to diabetes development and that including participation propensity score covariate did not change models performances.

Results

Among 3,242 participants, who completed the follow-up (median 6.2 years), 231(7.1%) developed diabetes [Men 88 (6.4%), women 143 (7.6%)]. Baseline characteristics of the participants are represented in Table 1. Plausible interactions were tested between sex and anthropometric measures as well as systolic and diastolic blood pressure, TG, HDL-C, FPG, and 2h-PCPG, but none were significant. We observed that only effect of age was modified by sex and all other interaction terms were not statistically significant. The global likelihood ratio test showed that all model performed better while including interaction terms (all Ps < 0.05).

Table 2 represents the performances of models in respect of discrimination capacity (AROCs and their 95% CIs), calibration (Hosmer–Lemeshow goodness-of-fit statistics), and global goodness of fit (likelihood ratio test). Hosmer–Lemeshow test rejected the goodness of fit for the models with 2h-PCPG (both full and simple clinical). All multivariate models discriminated diabetic from nondiabetic participants better than 2h-PCPG as is evident in higher AROCs of the multivariate models. Among multivariate models, the lowest AROC belonged to the simple clinical model without 2h-PCPG that was significantly higher than that of 2h-PCPG. Full models with or without 2h-PCPG had lower -2log likelihood than simple clinical models (all Ps < 0.0001). Models including 2h-PCPG had lower -2log likelihood than those without 2h-PCPG.

Figure 2 shows that at the same false-positive rate, simple clinical model had slightly better discrimination than 2h-PCPG.

Applying SADPM directly to the TLGS population produced a high Hosmer–Lemeshow χ^2 ($\chi^2 = 202.4$, $P < 0.0001$), with overall bias (systematic overestimation) being 111% that indicates poor calibration. After recalibrating the SADPM by taking into account the differences

Table 2 Performance of predicting models for type 2 diabetes mellitus

	AROC (95% CI)	HL-χ^2 (P value)	P values for model comparison	
			Discrimination	Calibration
Models				
Full model (+2h-PCPG)	0.86 (0.83–0.89)	18.8 (0.016)		
Full model	0.84 (0.81–0.87)	2.6 (0.975)		
Simple clinical model (+2h-PCPG)	0.86 (0.83–0.89)	22.9 (0.003)		
Simple clinical model	0.84 (0.81–0.87)	7.9 (0.448)		
2h-PCPG	0.78 (0.75–0.82)	–		
Original SADPM	0.83 (0.80–0.86)	202.4 (<0.0001)		
Recalibrated SADPM[a]	0.83 (0.80–0.86)	12.5 (0.131)		
Comparison				
Full model (+2h-PCPG) versus full model			0.017	<0.0001
Simple clinical model (+2h-PCPG) versus simple clinical model			0.007	<0.0001
Full model versus simple clinical model (+2h-PCPG)			0.270	<0.0001
Full model versus simple clinical model			0.111	<0.0001
Simple clinical model versus 2h-PCPG			<0.001	–

AROC area under the receiver operating characteristic curve, *HL-χ^2* Hosmer–Lemeshow chi-square, *SAHS* San Antonio heart study diabetes prediction model, *2h-PCPG* 2-h post-challenge plasma glucose

[a] The risk for each individual was multiplied by the ratio of the SAHS diabetes incidence rate to the TLGS diabetes incidence rate

in underlying rates of developing diabetes, the function worked well in this population. Hosmer–Lemeshow χ^2 significantly decreased (12.5, $P = 0.13$) down to an acceptable range, and the fit of the SADPM improved (Figs. 3 and 4) but not more so than re-estimating risk in TLGS using identical variables (Table 2).

Since the duration of follow-up in SAHS and TLGS were not the same, we assessed the goodness of fit of the SAHS diabetes prediction model in a subsample of participants ($n = 558$) who have been followed for at least 7.5 years and observed that the calibration improved ($\chi^2 = 25.9$, $P = 0.0005$), but was not still acceptable.

The top quintiles of recalibrated model-predicted risk ($\geq$18%) identified 68% of participants who developed diabetes on follow-up (sensitivity). Proportion of participants without incident diabetes who were not in the top quintile of predicted risk was 84% (specificity). Among persons in top quintile of risk, 24% developed diabetes (positive predictive value) and among those not in top quintile 97% did not develop diabetes (negative predictive value).

The CIs of AROC for the SAHS model (0.83; 0.80–0.86) was validated by bootstrap procedure. In 95% of replications the AROCs were between 0.81 and 0.86.

Discussion

Using data on patients free of diabetes at baseline, numerous prediction models have been developed which identified those at increased risk for incident diabetes in the future (Lindstrom and Tuomilehto 2003; Abdul-Ghani et al. 2007; Schulze et al. 2007, 2009; Wilson et al. 2007; Rahman et al. 2008; Chien et al. 2009; Kolberg et al. 2009). However, before widespread use of these instruments, there is a need to validate the models and risk scores in other populations, which has rarely been done.

To the best of our knowledge, this is the first study to examine the applicability of the SADPM in a Middle East population. The overall discrimination capacities of re-estimated and original simple clinical models were similar, though we do not have access to the original data and hence, the models could not be statistically compared. This observational study showed that the SADPM could not be directly applied to the TLGS population. After we recalibrated the model, however, the agreement between observed and predicted incident diabetes was good. Recalibration can correct overestimation or underestimation in ethnic groups with different underlying risk profile but does not produce a better model fit than re-estimating the equations with each site's data.

Ethnicity may explain some of the difference in the proportions of the SAHS and TLGS samples that developed diabetes. Parameters such as historical period, geographic location, methodological approach, disease spectrum, or follow-up interval, commonly affect prediction models and can have a role in limiting the generalizability of prediction systems (Justice et al. 1999).

Our results suggest that a simple clinical multivariate model, based on a panel of clinical characteristics deemed

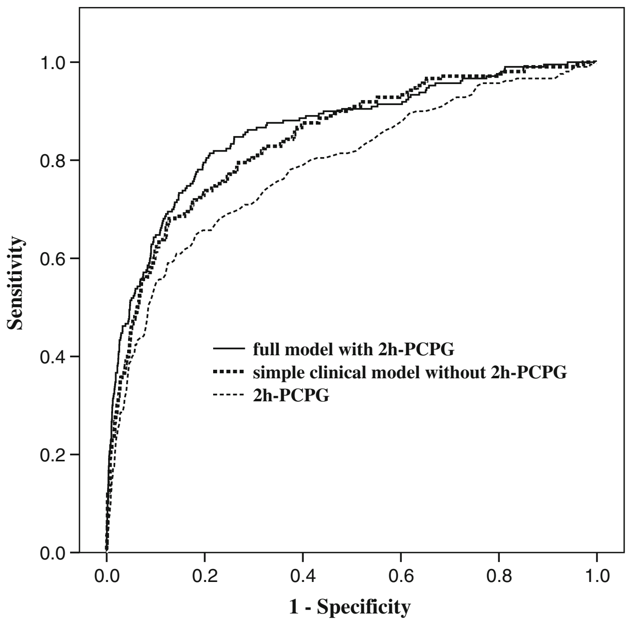

Fig. 2 AROCs for the full model with 2h-PCPG, simple clinical model without 2h-PCPG, and 2h-PCPG alone. *AROC* area under the receiver operating characteristic curve, *2h-PCPG* 2-h post-challenge plasma glucose

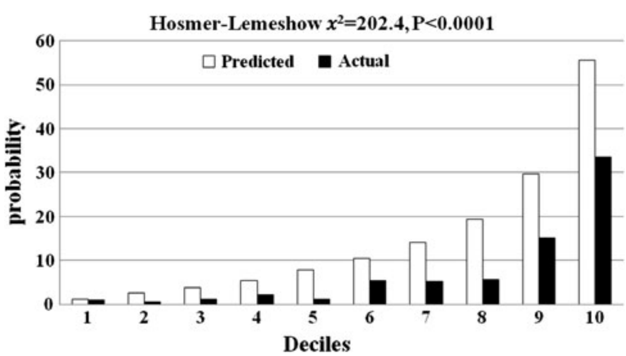

Fig. 3 Six-year predictions for type 2 diabetes: performance measures of original San Antonio Heart Study model for Tehran Lipid and Glucose Study. Calibration measures how closely predicted risk is in concordance with observed actual outcomes. The SADPM predicted risks were used to divide subjects into deciles of predicted risk for developing diabetes within 6 years. Plots were constructed showing predicted (*white*) and actual (*black*) event rates for each decile. A χ^2 statistic was calculated to compare the differences between predicted and actual event rates; small values indicated good calibration. Values exceeding 20 indicate significant lack of calibration ($P < 0.01$)

to be ordinarily available in a routine clinical setting, can be superior for identifying persons at high risk for future diabetes in comparison to the model relying exclusively on 2h-PCPG. Furthermore, addition of the 2h-PCPG value to the simple clinical model might challenge both the patients' and clinicians' acceptance of this model (Genuth et al. 2003). This finding was in line with other studies (Stern et al. 2002; Wilson et al. 2007). The 2h-PCPG test is relatively costly, inconvenient, labor-intensive, and time-

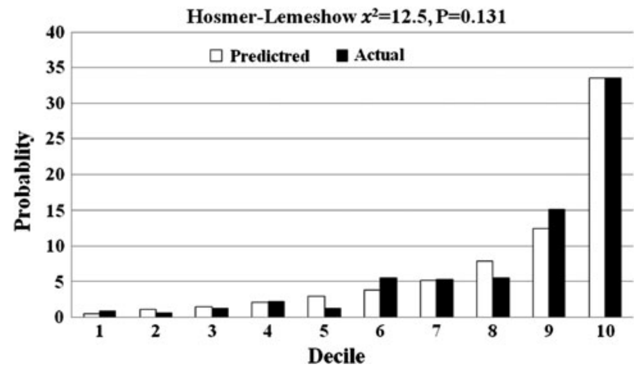

Fig. 4 Six-year predictions for type 2 diabetes: performance measures of recalibrated San Antonio Heart Study model for Tehran Lipid and Glucose Study. Calibration measures how closely predicted risk is in concordance with observed actual outcomes. The SADPM predicted risks were used to divide subjects into deciles of predicted risk for developing diabetes within 6 years. Plots were constructed showing predicted (*white*) and actual (*black*) event rates for each decile. A χ^2 statistic was calculated to compare the differences between predicted and actual event rates; small values indicated good calibration. Values exceeding 20 indicate significant lack of calibration ($P < 0.01$)

consuming; the reproducibility of the test has also been questioned (Genuth et al. 2003).

While developing a prediction model, one must accept a trade-off between sensitivity and specificity, which depends on the cut-point used to define a positive test result. We considered those in the top risk quintiles as having high risk of developing diabetes in the next 6 years (sensitivity 68% and specificity 84%). While predicting diabetes, tests with moderate sensitivity (about 60%) but high specificity (about 90%), repeated every 3 years, have been suggested to optimize the trade-off between disease detection and avoiding false-positive results (Johnson et al. 2005; Herman 2009).

Varieties of approaches have been undertaken in some studies to predict future risk of developing diabetes in various populations (Stern et al. 2002; Schulze et al. 2007). To choose a model the availability of risk factor data in the clinical setting, the optimal cut-point to define a positive test, and the simplicity of the model should be considered (Herman 2009).

The performance of SADPM in TLGS population was higher than Chinese population (AROC, 0.675) but lower than in Japanese population (AROC, 0.896). We have previously shown that waist-to-height ratio (WHtR) is superior to the other anthropometric measures of obesity in predicting diabetes (Hadaegh et al. 2009). WHtR, however, was not assessed in the SADPM, and hence, we were not able to examine if a model including WHtR from the SAHS better fitted data from the TLGS. Anthropometric measures of central adiposity and general obesity rank variably in predicting diabetes in different ethnic populations (Wang et al. 2005). Different levels of adjustment might contribute to different levels of predictability for diabetes of each anthropometric measure. Some studies recently compared different anthropometric measures in terms of their ability to predict diabetes and to determine whether predictive ability was modified by ethnicity. They observed that measures of central and overall adiposity predicted diabetes to a similar degree except for slight superiority for WHtR (Tulloch-Reid et al. 2003; MacKay et al. 2009; Nyamdorj et al. 2009). Wilson et al. (2007) demonstrated that there is virtually no difference in the model's predictive capability according to use of waist circumference or BMI. We re-estimated the SADPM, substituting the BMI with WHtR and observed no improvement in respect of predictive discrimination (AROC, 0.83) (data not shown).

The inclusion of blood pressure in diabetes prediction models is not new (Conen et al. 2007; Wilson et al. 2007). As part of the metabolic syndrome, diabetes and hypertension frequently occur together in an individual and are associated with obesity. Despite this close relationship between hypertension and diabetes, information on the relationship of blood pressure levels with the subsequent development of diabetes are scarce. It is unclear whether this hemodynamic characteristic has active impact on diabetes risk or is merely reflections of shared underlying physiologic factors (Kahn et al. 2009). Conen et al. in a population of initially healthy, middle-aged women, showed that blood pressure levels was strong predictor of incident diabetes. (Conen et al. 2007). Identifying individual with prevalent, already undiagnosed diabetes using risk factors such as hypertension could give rise to the reverse causality (chicken–egg), since hypertension and diabetes frequently coexist. This issue, however, would not be to the case in predictive models which are used to predict development of incident diabetes in the future; since examination of the risk factors precedes development of diabetes.

In cross-sectional studies where detecting prevalent (asymptomatic disease or undiagnosed) cases matter, the rationale for concentrating on parsimonious and simple models is clear (Herman 2009; Kahn et al. 2009). In prospective studies (event-free cohort), however, where prediction of incident disease is sought, simplicity is less important than accuracy (Herman 2009), particularly now that current technology has rendered computerized calculation or web-based calculation so prevalent that calculating even a more sophisticated risk score is possible (Schulze and Joost 2009). Nonetheless, simplicity with respect to measures used to develop a model can render the model more likely to be adopted in the clinical practice. The validation of the SADPM supported the conclusion that complex models are not needed to determine the risk of diabetes; the findings that could be more evident when viewed in the light of recent observations demonstrating that serum biomarkers and genetic risk factors were not able to substantially improve the predictability of the phenotype-based risk models (Talmud et al. 2010; Abdul-Ghani et al. 2007; Kahn et al. 2009; Schulze et al. 2009).

Another diabetes risk score for predicting 5-year risk of diabetes was also developed in middle-aged Finnish population (Finnish diabetes risk score). They, however, did not use a formal 2h-PCPG at the beginning of their study; instead, the model identified cases by initiation of diabetes medications. Hence, it might not capture the full burden of diabetes including undiagnosed diabetes (Lindstrom and Tuomilehto 2003). Furthermore, the model has been recently tested in a German population and yielded a low validity for identifying undiagnosed diabetic cases (Rathmann et al. 2005).

The German diabetes risk score has been developed from European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study, which incorporated anthropometric, dietary, and life style risk factors including smoking and alcohol consumption. The model has been validated in three additional German populations and reported to be accurate for identifying high-risk or undiagnosed diabetic individuals (Schulze et al. 2007). Smoking and alcohol consumption are, however, value-laden and so prone to underreporting.

The estimates from our population models were found to be stable, as demonstrated by the internal validation study. Also, using a community-based population could reduce the possibility of selection bias.

There are several limitations which should be considered when examining the result of current study. First, the clinical model's predictive capability has been reported to attenuate as the follow-up duration increased to 10 years (McNeely et al. 2003). Pending further information from TLGS, applicability of model for long-term prediction of diabetes could not be guaranteed. Second, we were unable to validate other diabetes risk scores such as Finish diabetes risk score or German risk score due to lack of dietary information, which is required for risk scoring (Lindstrom and Tuomilehto 2003; Schulze et al. 2007). However, the SAHS model is superior over these models in that it predicts undiagnosed diabetes. Third, the participation rate in the current study was 64.6%. However, the selection bias is an unlikely explanation for these results since comparisons of participants and non-participants revealed modest difference, although statistical significance was often reached due to large numbers. Finally, the length of follow-up in this study was relatively short. Thus, while the risk score may be good at identifying those rapidly progressing to diabetes; participants with slower onset might be missed. Alternatively, longer follow-up might result in a stronger association between the risk and the incidence of diabetes if more of those with highest risk score keep onto develop diabetes.

Conclusion

These data suggest that performance of diabetes prediction models need to be assessed in the target population, where they will be ultimately be used. A recalibrated SAHS models, including a set of clinical variables, produced a relatively high prediction power of diabetes. Prediction models can be used in clinical setting, community setting, or self-use for pre-screening, screening or risk assessment/ prediction. The SADPM could be included as it works in clinical practice and public health practice with the aid of a hand-held, programmable calculator or a personal computer.

Acknowledgments This study was supported by grant No. 121 from the National Research Council of the Islamic Republic of Iran. We express our appreciation to the participants of district-13 of Tehran for their enthusiastic support in this study.

Conflict of interest statement None.

References

Abdul-Ghani MA, Williams K, DeFronzo RA, Stern M (2007) What is the best predictor of future type 2 diabetes? Diabetes Care 30:1544–1548. doi:10.2337/dc06-1331

Azizi F, Ghanbarian A, Momenan AA, Hadaegh F, Mirmiran P, Hedayati M et al (2009) Prevention of non-communicable disease in a population in nutrition transition: Tehran lipid and glucose study phase II. Trials 10:5. doi:10.1186/1745-6215-10-5

Chien K, Cai T, Hsu H, Su T, Chang W, Chen M et al (2009) A prediction model for type 2 diabetes risk among Chinese people. Diabetologia 52:443–450. doi:10.1007/s00125-008-1232-4

Colagiuri S, Cull CA, Holman RR (2002) Are lower fasting plasma glucose levels at diagnosis of type 2 diabetes associated with improved outcomes? Diabetes Care 25:1410–1417. doi:10.2337/diacare.25.8.1410

Colagiuri S, Hussain Z, Zimmet P, Cameron A, Shaw J (2004) Screening for type 2 diabetes and impaired glucose metabolism. Diabetes Care 27:367–371. doi:10.2337/diacare.27.2.367

Conen D, Ridker PM, Mora S, Buring JE, Glynn RJ (2007) Blood pressure and risk of developing type 2 diabetes mellitus: The Women's Health Study. Eur Heart J; 28:2937–2943. doi:10.1093/eurheartj/ehm400

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

Friedwald WT, Levy RI, Fredrickson DS (1972) Estimation of the concentration of low density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem 18:419–503

Genuth S, Alberti KG, Bennett P, Buse J, Defronzo R, Kahn R et al (2003) Follow-up report on the diagnosis of diabetes mellitus. Diabetes Care 26:3160–3167. doi:10.2337/diacare.26.11.3160

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. doi:10.2337/diacare.27.3.727

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. doi:10.2337/diacare.29.02.06.dc05-0945

Hadaegh F, Bozorgmanesh MR, Ghasemi A, Harati H, Saadat N, Azizi F (2008) High prevalence of undiagnosed diabetes and abnormal glucose tolerance in the Iranian urban population: Tehran Lipid and Glucose Study. BMC Public Health 8:176. doi:10.1186/1471-2458-8-176

Hadaegh F, Shafiee G, Azizi F (2009) Anthropometric predictors of incident type 2 diabetes mellitus in Iranian women. Ann Saudi Med 29:194–200

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36

Herman WH (2009) Predicting risk for diabetes: choosing (or building) the right model. Ann Int Med 150:812–814. doi:10.1059/0003-4819-150-11-200906020-00010

Hosmer DW, Lemeshow S (2000). Applied logistic regression. Wiley-Interscience, New York

Johnson SL, Tabaei BP, Herman WH (2005) The efficacy and cost of alternative strategies for systematic screening for type 2 diabetes in the US population 45–74 years of age. Diabetes Care 28:307–311

Justice AC, Covinsky KE, Berlin JA (1999) Assessing the generalizability of prognostic information. Ann Intern Med 130:515–524

Kahn HS, Cheng YJ, Thompson TJ, Imperatore G, Gregg EW (2009) Two risk-scoring systems for predicting incident diabetes mellitus in US adults age 45 to 64 years. Ann Intern Med 150:741–751

Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA et al (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 346:393–403. doi:10.1056/NEJMoa012512

Kolberg JA, Jorgensen T, Gerwien RW, Hamren S, McKenna MP, Moler E et al (2009) Development of a type 2 diabetes risk model from a panel of serum biomarkers from the Inter99 cohort. Diabetes Care 32:1207–1212. doi:10.2337/dc08-1935

Lindstrom J, Tuomilehto J (2003) The diabetes risk score: a practical tool to predict type 2 diabetes risk. Diabetes Care 26:725–731. doi:10.2337/diacare.26.3.725

MacKay MF, Haffner SM, Wagenknecht LE, D'Agostino RB, Hanley AJG (2009) Prediction of type 2 diabetes using alternate anthropometric measures in a multi-ethnic cohort. Diabetes Care 32:956–958. doi:10.2337/dc08-1663

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

Nyamdorj R, Qiao Q, Soderberg S, Pitkaniemi JM, Zimmet PZ, Shaw JE et al (2009) BMI compared with central obesity indicators as a predictor of diabetes incidence in Mauritius. Obesity (Silver Spring) 17:342–348. doi:10.1038/oby.2008.503

Pan XR, Li GW, Hu YH, Wang JX, Yang WY, An ZX et al (1997) Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. Diabetes Care 20:537–544

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

Rathmann W, Martin S, Haastert B, Icks A, Holle R, Lowel H et al (2005) Performance of screening questionnaires and risk scores for undiagnosed diabetes: the KORA Survey 2000. Arch Intern Med 165:436–441

Schulze M, Joost H (2009) Rapid responses to: Henry S. Kahn, Yiling J. Cheng, Theodore J. Thompson, Giuseppina Imperatore, and Edward W. Gregg. Two risk-scoring systems for predicting incident diabetes mellitus in US adults age 45 to 64 years in Ann Intern Med 2009; 150:741–751: Integer points based scoring systems for diabetes prediction: The German Diabetes Risk Score (18 June). Retrieved 15 Jul 2009, from http://www.annals.org/cgi/eletters/150/11/741#114512

Schulze MB, Hoffmann K, Boeing H, Linseisen J, Rohrmann S, Mohlig M 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

Schulze MB, Weikert C, Pischon T, Bergmann MM, Al-Hasani H, Schleicher E et al (2009) Use of multiple metabolic and genetic markers to improve the prediction of type 2 diabetes: the EPIC-Potsdam Study. Diabetes Care 32:2116–2119. doi:10.2337/dc09-0197

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

Talmud PJ Hingorani AD Cooper JA, Marmot MG, Brunner EJ, Kumari M et al (2010) Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. BMJ 340:b4838. doi:10.1136/bmj.b4838

Tulloch-Reid MK, Williams DE, Looker HC, Hanson RL, Knowler WC (2003) Do measures of body fat distribution provide information on the risk of type 2 diabetes in addition to measures of general Obesity? Diabetes Care 26:2556–2561. doi:10.2337/diacare.26.9.2556

Wang Y, Rimm EB, Stampfer MJ, Willett WC, Hu FB (2005) Comparison of abdominal adiposity and overall obesity in predicting risk of type 2 diabetes among men. Am J Clin Nutr 81:555–563

Wild S, Roglic G, Green A, Sicree R, King H (2004) Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 27:1047–1053

Wilson PWF, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino R B Sr (2007) Prediction of incident diabetes mellitus in middle-aged adults: The Framingham offspring study. Arch Intern Med 167:1068–1074. doi:10.1001/archinte.167.10.1068