

ORIGINAL INVESTIGATION

Open Access

Prediction model for high glycated hemoglobin concentration among ethnic Chinese in Taiwan

Kuo-Liong Chien^{1,2}, Hung-Ju Lin², Bai-Chin Lee², Hsiu-Ching Hsu², Ming-Fong Chen^{2*}

Abstract

Background: This study aimed to construct a prediction model to identify subjects with high glycated hemoglobin (HbA1c) levels by incorporating anthropometric, lifestyle, clinical, and biochemical information in a large cross-sectional ethnic Chinese population in Taiwan from a health checkup center.

Methods: The prediction model was derived from multivariate logistic regression, and we evaluated the performance of the model in identifying the cases with high HbA1c levels ($\geq 7.0\%$). In total 17,773 participants (age ≥ 30 years) were recruited and 323 participants (1.8%) had high HbA1c levels. The study population was divided randomly into two parts, with 80% as the derivation data and 20% as the validation data.

Results: The point-based clinical model, including age (maximal 8 points), sex (1 point), family history (3 points), body mass index (2 points), waist circumference (4 points), and systolic blood pressure (3 points) reached an area under the receiver operating characteristic curve (AUC) of 0.723 (95% confidence interval, 0.677- 0.769) in the validation data. Adding biochemical measures such as triglycerides and HDL cholesterol improved the prediction power (AUC, 0.770 [0.723 - 0.817], $P = < 0.001$ compared with the clinical model). A cutoff point of 7 had a sensitivity of 0.76 to 0.96 and a specificity of 0.39 to 0.63 for the prediction model.

Conclusions: A prediction model was constructed for the prevalent risk of high HbA1c, which could be useful in identifying high risk subjects for diabetes among ethnic Chinese in Taiwan.

Background

Poor control of type 2 diabetes, presenting as an elevated glycated hemoglobin (HbA1c) level, is associated with macro- and micro-vascular complications among patients with diabetes [1-3]. HbA1c, similar to fasting and post-challenge glucose levels, is a marker for monitoring glucose levels to prevent diabetic complications, such as retinopathy [4]. Furthermore, a high HbA1c level in the general population predicts a further risk of coronary heart disease [5]. Therefore, it is mandatory to construct a prediction model to identify individuals with a high HbA1c level in the general population, despite the low prevalence (1.3%) [6].

A prediction model using anthropometric, lifestyle, clinical and biochemical measures from routine examinations has been developed to identify high-risk individuals for diabetes in cross-sectional [7-10] and prospective

cohort studies [11-18]. These models appear to be effective in identifying people with a high risk of diabetes. However, the prediction model for a high HbA1c level is limited [6,19] and there is currently no data available on ethnic Chinese. In this study, a prediction model was constructed and its performance tested in detecting prevalent but unknown levels of high HbA1c in a large, cross-sectional ethnic Chinese population who were recruited from a health checkup program in Taiwan.

Methods

Subjects

This cross-sectional study involved 25,452 adult subjects who participated in the health checkup program at the Health Management Center of one tertiary hospital, from January 2003 to December 2006. The sampling strategy for the study population, including inclusion and exclusion criteria, is shown in Additional file 1, Figure S1. After excluding subjects with a history of diabetes with medication, cardiovascular disease, cancer, missing or duplicated data, and age less than 30 years, a total of

* Correspondence: mfchen@ntu.edu.tw

²Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan

Full list of author information is available at the end of the article

17,773 participants were recruited into the study. The study protocol has been described previously [20,21]. Briefly, details of socio-economic status, along with medical and medication histories were collected by questionnaires, and standardized clinical measure procedures were undertaken. The protocol was approved by the hospital's Institutional Research Board. Standardized physical examination procedures, such as anthropometric measures and blood pressure, were also performed [22,23]. Blood pressure was measured in a resting position by trained medical assistants, while body mass index (BMI) was calculated as weight (in kilograms)/square of height (in meters), and waist circumference was measured midline between the low costal margin and superior posterior iliac crest.

Blood Sampling and Analytic Methods

The procedures for blood sampling and analytic methods have been described in previous studies [22,24]. Briefly, blood samples were collected from each participant after fasting for at least 12 hours. Serum total cholesterol levels were measured using the CHOD-PAP method (Boehringer Mannheim, Germany). HDL cholesterol was measured following precipitation of apolipoprotein B-containing lipoproteins with phosphotungstic acid and magnesium ions (Boehringer Mannheim, Germany). Triglyceride concentrations were measured by the GPO-DAOS method (Wako Co., Japan). The aforementioned lipids were measured using a Hitachi 7450 automated analyzer (Hitachi, Japan). LDL-C concentrations were calculated using the Friedewald formula. CRP was measured by automated nephelometric immunoassay using a Beckman Array instrument (Beckman Array 360 system, Canada). All of the measurements were carried out in a single hospital with a coefficient of variation of 5%. HbA1c levels were measured by automatic high-performance liquid chromatography using a Bio-Rad HbA1c kit (Bio-Rad Diagnostic Group, Hercules, CA, USA) in the central laboratory of the hospital. Standardization using mass spectroscopy and capillary electrophoresis was used, and prepared mixtures of purified HbA1c and HbA0 were used as calibrators [25]. With regards to the cutoff level of HbA1c, abnormally high HbA1c levels were defined as 6.7% according to the sensitivity and specificity of diabetes diagnosis and diabetic retinopathy in one cross-sectional study [26]. In addition, a study by the UK Prospective Diabetes Study Group demonstrated that over 10 years, the mean HbA1c level in their intensive treatment group was 7.0% [1]. Moreover, an HbA1 level of 7.3% or greater is considered the cutoff value for screening diabetes in Pima Indians [27]. Therefore, we set the threshold for abnormally high HbA1c at 7%.

Statistical analysis

The basic demographic, anthropometric measurements, lifestyle factors, and biochemical measures were described according to a high HbA1c concentration, defined as HbA1c \geq 7%. The constructed model for HbA1c \geq 6.5% was similar so that we reported the findings about HbA1c \geq 7%. Missing waist circumference data in the first year (2003) were imputed with the mean values of waist circumference due to the specific HbA1c status to improve the power of the prediction model. The study population was divided randomly into two parts, with 80% as the derivation data and 20% as the validation data.

Multivariate logistic models were used to predict the risk of a high HbA1c level in the derivation data. First, the Cambridge model [28], including variables of age, BMI, anti-hypertensive medication, family history, and smoking status was used to construct the model [6,19]. Second, an additional anthropometric measurement (waist circumference) and systolic blood pressure were incorporated into the model. History of hypertension medication was excluded due to non-significance in the model. This second prediction model was called the clinical model. Third, important biochemical indicators, including C-reactive protein (CRP), HDL cholesterol, and triglyceride concentrations were added into the model to construct the full biochemical model [29]. Fasting glucose was not included in the model to prevent over-correction by glucose concentration.

Based on results of the multivariate logistic models from the derivation data, two strategies for constructing the prediction model were applied. First, the coefficients for the prediction model from the derivation data were used directly, which is a common strategy in the literature [18,30-32]. By directly calculating the coefficients and individual variables, the individual risk was derived in the validation data. We provided the nomogram using Harrell's method [33]. Second, a point-based chart was constructed from the derivation data according to the strategy suggested by Sullivan and colleagues [34]. This strategy was as follows: continuous variables were organized into meaningful categories and the reference values for each variable were determined. We assigned a 5-year increase in age as the referent risk, and points associated with each of the categories of the risk factors were calculated by comparing with the referent risk. Therefore, an individual's risk was constructed from the validation data by the following formula: $Risk = 1/[1 + \exp(-\beta X)]$ where βX is the sum of the reference risk and the product of the 5-year risk constant and the individual points [34].

Performance of the proposed coefficient-based and point-based prediction models were compared with the Cambridge model [6,19,28]. The area under the receiver

operating characteristic curve (AUC) was used to compare the discriminatory capability among the models. A receiver operating characteristic curve is a graph of sensitivity versus 1-specificity (or false-positive rate) for various cut-off definitions of a positive diagnostic test result [35]. Statistical differences in the AUCs were compared using the method of DeLong et al [36].

Furthermore, the goodness-of-fit for all models was assessed based on the Hosmer-Lemeshow test [37]. The global summary statistics included the Brier score [38], twice the forecast-outcome-covariance (a measure of how accurately the forecast corresponds to the outcome, similar to R^2 in linear regression) [39], and discrimination (c statistic), which is the same as the AUC [40].

The simple points model was compared with other models using net reclassification improvement (NRI) and integrated discrimination improvement (IDI) statistics [41]. NRI was based on the reclassification tables and was calculated from a sum of differences between the "upward" movement in categories for event subjects and the "downward" movement of non-event subjects [41]. The NRI was presented according to the presumed risk categories of high HbA1c according to quartiles (0.6%, 1.2%, and 2.6%). The IDI was viewed as the difference between improvement in average sensitivity and any potential increase in average "one minus specificity". The statistic was a difference in Yates discrimination slopes between the new and old models [38,42].

All of the 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), Stata version 9.1 (Stata Corporation, College Station, Texas) and R <http://www.R-project.org>.

Results

Basic characteristics

Among the study participants, 323 cases (1.8%) had an HbA1c level $\geq 7\%$. Table 1 shows the basic demographic, clinical, lifestyle, socio-economic status and biochemical measures of the study participants. Participants with higher HbA1c levels were likely to be older, male, have a higher body mass index (BMI), waist circumference, blood pressure, cholesterol, triglycerides, CRP, and white blood cell count, and lower HDL cholesterol level. In addition, participants with higher HbA1c values were likely to take anti-hypertensive medication, have a higher rate of a positive family history of diabetes and current smoking status. The distribution of socio-economic information, such as marital status and job, was similar between participants with and without abnormal HbA1c levels. The distributions of most continuous and categorical variables were consistent in each gender, and there was no differential effect.

Constructing the prediction models

The results of the multivariate logistic regression models are listed in Table 2. Hypertension medication was not statistically significant and was excluded in further analyses. Waist circumference and BMI were both associated with abnormal HbA1c levels in the clinical model. Age, family history of diabetes, waist circumference, systolic blood pressure, and biochemical measures, including CRP and triglycerides, were significantly associated with higher HbA1c levels. HDL cholesterol was borderline inversely associated with higher HbA1c.

Regression coefficient-based and point-based prediction models based on the clinical and biochemical models were developed. A nomogram based on the clinical and biochemical models was constructed (Figure 1). Table 3 shows the point-based clinical model to estimate high HbA1c risk using the points system, derived from the coefficients of the clinical model: age (8 points), sex (1 point), family history (3 points), BMI (2 points), waist circumference (4 points), and systolic blood pressure (3 points). This approach allowed for the manual estimation of the risk of developing a higher HbA1c level for each individual (Table 3). The waist circumference-related point-based biochemical model, additionally including HDL cholesterol (3 points) and triglycerides (2 points), is shown in Additional file 1, Table S1. A cutoff point of 7 had a sensitivity of 0.76 to 0.96 and a specificity of 0.39 to 0.63 for both clinical and biochemical prediction models.

Performance measures of the prediction model

The performance of the prediction models, including the Cambridge (Additional file 1, Table S2), coefficient-based and point-based clinical, and biochemical models were compared using different measures (Table 4). The clinical models had a fair discrimination ability with an AUC of 0.712 (95% confidence interval [CI] 0.664-0.760), and 0.723 (95% CI, 0.677-0.769) in the coefficient-based and point-based models, respectively. Adding biochemical measures improved the prediction (AUC of 0.773 [95% CI, 0.726-0.821] and 0.770 [95% CI, 0.723-0.817] for the coefficient-based and point-based models, respectively). Moreover, in the biochemical models, the AUCs were the highest and the Brier scores were the lowest, and they were likely to have a smaller Hosmer-Lemeshow chi-square and high P values, indicating a good calibration ability for a high HbA1c level. Figure 2 shows the AUCs of the various prediction models in the validation data. The Hosmer-Lemeshow chi-square values indicated a goodness-of-fit for these prediction models.

The performance ability between the various prediction models was tested using NRI and IDI statistics (Table 5). The clinical coefficient-based model was

Table 1 Basic demographic, clinical, lifestyle, and biochemical characteristics of the study population, specified by HbA1c concentration

| | HbA1c < 7% n = 17450 | | HbA1c ≥ 7% n = 323 | | P value |
|-----------------------------|-------------------------|------|-----------------------|-------|---------|
| | Mean | SD | Mean | SD | |
| Age, year | 51.0 | 10.9 | 56.6 | 10.2 | < .0001 |
| BMI, kg/m ² | 23.8 | 3.2 | 25.8 | 3.7 | < .0001 |
| Waist, cm | 83.5 | 9.1 | 89.6 | 9.9 | < .0001 |
| Systolic BP, mmHg | 122.6 | 16.0 | 131.6 | 16.1 | < .0001 |
| Diastolic BP, mmHg | 73.1 | 10.5 | 77.5 | 10.2 | < .0001 |
| Fasting glucose, mg/dL | 90.6 | 10.2 | 172.0 | 54.7 | < .0001 |
| Postprandial glucose, mg/dL | 117.8 | 48.1 | 167.7 | 99.1 | < .0001 |
| Total cholesterol, mg/dL | 203.7 | 36.8 | 220.7 | 47.3 | < .0001 |
| Triglyceride, mg/dL | 119.1 | 75.1 | 193.0 | 151.9 | < .0001 |
| HDL-cholesterol, mg/dL | 44.7 | 44.1 | 40.3 | 8.9 | < .0001 |
| LDL-cholesterol, mg/dL | 119.6 | 32.5 | 134.8 | 43.5 | < .0001 |
| CRP, mg/dL | 0.16 | 0.40 | 0.31 | 0.66 | < .0001 |
| Uric acid, mg/dL | 6.05 | 1.53 | 6.03 | 1.55 | 0.77 |
| White blood cells | 5.43 | 1.49 | 6.30 | 1.76 | < .0001 |
| HbA1c, % | 5.42 | 0.37 | 8.86 | 1.98 | < .0001 |
| | | % | % | | |
| Gender | women | 44.9 | 31.6 | | < .0001 |
| | men | 55.2 | 68.4 | | |
| BMI group | | | | | < .0001 |
| BMI < 25 | | 67.4 | 43.0 | | |
| 25~30 | | 29.1 | 45.2 | | |
| BMI ≥ 30 | | 3.5 | 11.8 | | |
| Medication history | | | | | |
| Hypertension | | 12.7 | 19.2 | | 0.001 |
| Diuretics usage | | 1.9 | 4.3 | | 0.001 |
| Lipid lowering | | 3.2 | 3.7 | | 0.59 |
| Family history of diabetes | | | | | < .0001 |
| | None | 71.3 | 61.3 | | |
| | Second relatives | 5.4 | 4.0 | | |
| | First relatives | 23.3 | 34.7 | | |
| Current smoking | Yes | 13.9 | 20.7 | | 0.0004 |
| Alcohol drinking | Yes | 55.6 | 50.2 | | 0.05 |
| Marital status | Unmarried | 11.5 | 9.5 | | 0.55 |
| | Married | 87.7 | 89.6 | | |
| | Separate | 0.3 | 0.6 | | |
| | Unknown | 0.5 | 0.3 | | |
| Job | Manual work | 5.4 | 5.3 | | 0.23 |
| | Business | 24.0 | 21.7 | | |
| | Government, Teacher | 21.3 | 17.3 | | |
| | Housework | 10.1 | 12.4 | | |
| | No job | 4.6 | 6.5 | | |
| | Service | 5.8 | 5.6 | | |
| | Student | 0.4 | 0.0 | | |
| | Other job | 28.6 | 31.3 | | |

Table 2 Regression coefficients, standard errors and significant levels of various covariates in the two prediction models among the derivation data

| Variable | Clinical | | | Biochemical | | |
|----------------------------------|---------------------|-------|---------|---------------------|-------|---------|
| | Estimated parameter | SEM | P | Estimated parameter | SEM | P |
| Intercept | -12.906 | 0.775 | < .0001 | -11.668 | 0.922 | < .0001 |
| Sex, men vs. women | 0.351 | 0.152 | 0.021 | 0.174 | 0.160 | 0.28 |
| Age, +1 year | 0.042 | 0.006 | < .0001 | 0.043 | 0.006 | < .0001 |
| BMI, +1 kg/m ² | 0.076 | 0.031 | 0.014 | | | |
| Waist, +1 cm | 0.024 | 0.012 | 0.046 | 0.036 | 0.008 | < .0001 |
| Family history | 0.710 | 0.138 | < .0001 | 0.724 | 0.140 | < .0001 |
| Smoking history | 0.433 | 0.173 | 0.012 | 0.209 | 0.178 | 0.24 |
| Systolic blood pressure, +1 mmHg | 0.017 | 0.004 | < .0001 | 0.016 | 0.004 | 0.0003 |
| CRP, +1 mg/dL | | | | 0.229 | 0.079 | 0.004 |
| HDL, +1 mg/dL | | | | -0.018 | 0.008 | 0.029 |
| Triglyceride, +1 mg/dL | | | | 0.004 | 0.001 | < .0001 |

compatible with the Cambridge model (Set 1), and the biochemical models outperformed the Cambridge models due to a significant increase in IDI and NRI values (Set 2). The biochemical coefficient-based model had a better performance than the clinical coefficient-based model by NRI value (Set 3). In addition, the clinical point-based model and clinical coefficient-based model had similar performance measures (Set 4). Finally, the biochemical coefficient-based model outperformed the biochemical point-based model (Set 5), and the clinical and biochemical point-based models had a similar performance (Set 6). The sensitivity, specificity and cut-off values for the clinical and biochemical models were listed in Additional file 1, Table S3. Our findings support that the clinical and biochemical point-based models were excellent models for identifying individuals with a high HbA1c value.

Discussion

This study confirms and extends the results of a previous study [9] that showed significantly associated risk factors for a high HbA1c concentration in an adult population. Coefficient-based and point-based prediction models for clinical practice were constructed. The biochemical model incorporated several clinical and lifestyle risk factors, as well as biochemical measures, in order to provide a feasible and practical tool for detecting high HbA1c levels. The availability of simple clinical tools to predict the future risk of disease, such as those for predicting coronary heart disease [43-45], can improve predicting the risk of high HbA1c, identify high-risk populations, and enhance preventive strategies.

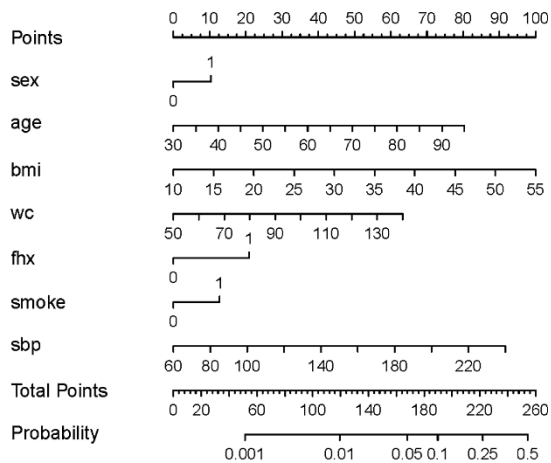
Thomas and colleagues investigated 7452 45-year-old British adults to compare the predictive power of the Cambridge risk score and body mass index for elevated HbA1c levels [19]. They found that the Cambridge

prediction model and body mass index had a similar identifying power for diabetes risk. Park *et al.* collected clinical data from 6567 adults in the European Prospective Investigation of Cancer-Norfolk cohort and showed that the Cambridge model performed well in predicting high HbA1c levels [Park, 2002 [6]]. In several other cross-sectional studies conducted in the US and Europe, prediction models based on clinical information and lifestyle-related factors have appeared to be useful in screening and identifying undiagnosed diabetes cases and high HbA1c levels among various populations [10,28,46,47].

The Cambridge risk model, including age, sex, steroid or antihypertensive medication, smoking, family history, and body mass index, has a specificity of 78% and a sensitivity of 51% to screen cases with HbA1c > = 7.0% [6]. In another screening project for a high HbA1c level defined as 7.0%, the Cambridge risk model was proven to be a good performance measure [19]. Although different criteria for defining high HbA1c levels for identifying undiagnosed diabetes and screening high risk cases, including 6.5% [9], 7.0% [6] and 7.5% [27], are available, we focused on a cutoff point of 7.0% in this study.

Based on these observations, this study provided a better prediction model than the Cambridge model for predicting a high HbA1c level among ethnic Chinese. We showed that some biochemical measures, especially components related to metabolic syndrome and inflammation such as triglycerides, low HDL cholesterol, and CRP, provide additional information for predicting a high HbA1c level. Aside from fasting glucose and MI, metabolic variables such as high triglycerides and low HDL were found to be strong predictors of type 2 diabetes. These variables also included risk functions developed in other populations [10,48] and were related to a high HbA1c level. Our study did not support a role for

Upper: Clinical model



Lower: Biochemical model

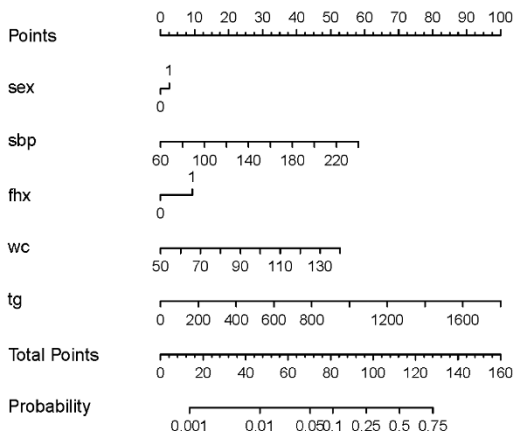


Figure 1 Nomogram to calculate the probability of high glycated hemoglobin (HbA1c > 7%) using the clinical (upper) and biochemical (lower) models. In the clinical models, sex (women as 0, men as 1), age, body mass index (BMI), waist circumference (WC), family history of diabetes (FHx), smoking, and systolic blood pressure (SBP) are calculated by reading from the point scale. In the biochemical model, only triglyceride (TG) is calculated. The total point score is then translated into probability of high HbA1c using the bottom scales, including total points and probability. For example, the probability of high HbA1c with a total point score of 170 is then 0.06, according to the two bottom lines. The participants can be classified according to the absolute probabilities accordingly.

Table 3 Simple points system according to the clinical model and the simple points clinical model and absolute risk function for High HbA1c (> = 7%)

| Risk factor | Category | Point | Points total | Absolute Risk |
|-------------------------------|-------------------------|-------|--------------|---------------|
| Age, yr | 30-39 | 0 | 0 | 0.002 |
| | 40-49 | 2 | 1 | 0.002 |
| | 50-59 | 4 | 2 | 0.002 |
| | 60-69 | 6 | 3 | 0.003 |
| | > = 70 | 8 | 4 | 0.004 |
| Sex | Women | 0 | 5 | 0.004 |
| | Men | 1 | 6 | 0.005 |
| Family history | No | 0 | 7 | 0.007 |
| | Yes | 3 | 8 | 0.008 |
| Current smoker | No | 0 | 9 | 0.01 |
| | Yes | 2 | 10 | 0.013 |
| BMI, kg/m2 | < 21.1 | 0 | 11 | 0.015 |
| | 21.1-22.7 | 0 | 12 | 0.019 |
| | 22.8-24.3 | 1 | 13 | 0.023 |
| | 24.4-26.2 | 1 | 14 | 0.029 |
| | > = 26.3 | 2 | 15 | 0.035 |
| | Waist circumference, cm | <77 | 0 | 16 |
| 77-82.9 | | 0 | 17 | 0.053 |
| 83-83.9 | | 1 | 18 | 0.064 |
| 84-89.9 | | 1 | 19 | 0.078 |
| > = 90 | | 2 | 20 | 0.094 |
| systolic blood pressure, mmHg | <109 | 0 | 21 | 0.114 |
| | 109-117.9 | 0 | | |
| | 118-125.9 | 1 | | |
| | 126-134.9 | 2 | | |
| | > = 135 | 3 | | |

LDL cholesterol in identifying a high HbA1c level in the multivariate model.

In our clinical model, a family history of diabetes and systolic blood pressure had significant predictive power for the risk of high HbA1c, and the results imply that these two factors, aside from BMI and waist circumference, should be checked to screen for high HbA1c. Our point-based model clearly showed that family history, BMI, waist circumference, and systolic blood pressure synergistically added to the risk of high HbA1c.

The choice of optimal threshold for defining a high HbA1c level is still inconclusive [26,27,49]. Engelgau and colleagues collected information on diabetic retinopathy and provided the diagnostic threshold of HbA1c as 6.7%. However, Hanson *et al.* used a similar strategy but argued that the best cutoff value should be 7.8%. We used 7.0% as the threshold following previous prediction model results [6,19] and because micro-vascular complications increase appreciably when HbA1c > 7.0%.

Table 4 Summary statistics for different prediction models based on covariates in the Cambridge, clinical, and biochemical model algorithms on the validation data

| | Area under ROC curve | Brier score* | 2*Forecast Outcome Covariance | Hosmer Lemeshow chi-square** | Hosmer Lemeshow P value** |
|--------------------------------|----------------------|--------------|-------------------------------|------------------------------|---------------------------|
| Cambridge | 0.691 | 0.0219 | 0.0004 | 14.6 | 0.07 |
| Clinical, coefficient-based | 0.712 | 0.0217 | 0.0007 | 12.8 | 0.12 |
| Clinical, points-based | 0.723 | 0.0220 | 0.0003 | 18.7 | 0.03 |
| Biochemical, coefficient-based | 0.773 | 0.0213 | 0.0013 | 8.8 | 0.36 |
| Biochemical, points-based | 0.770 | 0.0219 | 0.0003 | 3.8 | 0.87 |

*A low Brier score indicated a goodness-of-fit. ** Low Hosmer-Lemeshow chi-square and high P values indicated a goodness-of-fit. A higher area under the ROC area as well as 2*Forecast-outcome-covariance represented better performance. A lower Hosmer-Lemeshow chi-square value represented a goodness-of-fit model.

The clinical and biochemical models can be implemented easily, and we provide the nomogram for the coefficient-based models. However, calculations are still necessary for the absolute risk probability. Therefore, the point-based models, although with slightly poorer performance than the coefficient-based models, are likely to be implemented. Manual calculation of an individual's risk by summing the points in the point-based models is feasible, such that health professionals can use it in clinical practice. The point-based model using clinical measures may have a useful role in stratifying a population so that those at the highest risk are offered further testing and intervention. A high-risk approach for primary prevention on the risk of diabetes is recommended [50] and the prediction model may be a feasible tool. In addition, multifactorial treatment on risk factors, including weight control, lipids, blood pressure, and glucose level lowering, in patients with type 2 diabetes is a difficult task [51], especially for patients with pre-existing cardiovascular

diseases [52]. A high HbA1c level in patients indicates poor glycemic control, so aggressive intervention is necessary for patients with high HbA1c.

To the best of our knowledge, this is the first study on a prediction model specifically developed for the risk of high HbA1c levels among ethnic Chinese. Because of the large sample size, estimates from our prediction models are stable as demonstrated by the internal validation study. Furthermore, the standardization and central laboratory mean the measurements are consistent throughout the study period. The homogeneous study participants provide a reliable estimate for the prediction model coefficients. We consider that these prediction models may be suitable for screening and identifying those at high risk of type 2 diabetes in the Asia-Pacific region.

This study had several limitations. First, the prevalent rate of high HbA1c levels was relatively low (1.8%) so that the predicted risk probability among the general population seems negligible. This low risk, however, may be underestimated. In fact, only 1.3% of participants had high HbA1c levels in the study conducted by Park *et al.* [6]. Therefore, identifying the high-risk population using the prediction model will be a useful tool for further prevention of diabetic complications. Second, the cross-sectional study design made causation difficult. Some anthropometric and lifestyle factors might have been influenced after the onset of diabetes. Our strategy to exclude existing diabetes cases was meant to reduce this reverse causation to as minimal as possible. Finally, we didn't include fasting plasma glucose in the biochemical model due to its high collinearity with HbA1c. In addition, the biochemical model was limited to the population who provided blood samples.

In conclusion, point-based prediction models were constructed to predict the prevalence of high HbA1c levels among ethnic Chinese. These simple clinical tools should help identify high-risk populations and improve prevention and treatment strategies for type 2 diabetes.

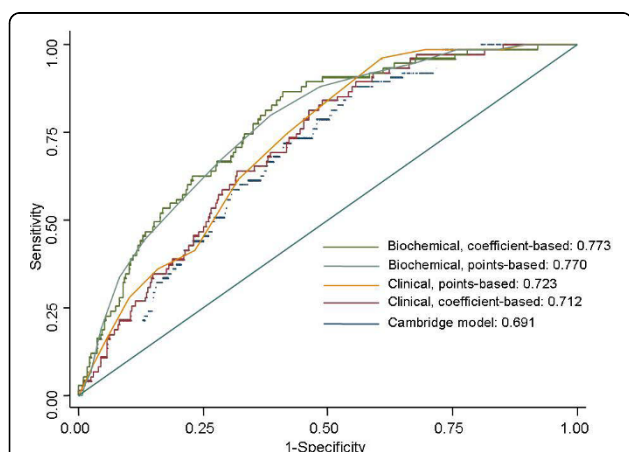


Figure 2 Areas under the ROC curves for the three prediction models in the validation data.

Table 5 Comparison of prediction performance by net reclassification improvement (NRI) and integrated difference improvement (IDI) with relative 95% confidence interval (CI) and significance levels

| Comparison set | Models | NRI* | 95% CI | P value | IDI | 95% CI | P value |
|----------------|--|--------|----------------|---------|---------|----------------|---------|
| 1 | Clinical coefficient-based vs. Cambridge | 0.066 | -0.054 - 0.186 | 0.28 | 0.007 | -0.001 - 0.014 | 0.051 |
| 2 | Biochemical coefficient-based vs. Cambridge | 0.294 | 0.141 - 0.447 | 0.0004 | 0.021 | 0.004 - 0.037 | 0.008 |
| 3 | Biochemical coefficient-based vs. clinical coefficient-based | 0.244 | 0.122 - 0.366 | 0.0002 | 0.014 | -0.002 - 0.030 | 0.053 |
| 4 | Clinical coefficient-based vs. clinical point-based | 0.090 | -0.022 - 0.203 | 0.23 | 0.009 | -0.001 - 0.019 | 0.057 |
| 5 | Biochemical coefficient-based vs. biochemical point-based | 0.362 | 0.262 - 0.462 | 0.0003 | 0.023 | 0.005 - 0.041 | 0.006 |
| 6 | Biochemical point-based vs. clinical point-based | -0.015 | -0.139 - 0.109 | 0.82 | 0.00003 | -0.001 - 0.001 | 0.49 |

*Presumed risk categories for NRI according to quartile values (0.6%, 1.2% and 2.6%)

Additional material

Additional file 1: Additional Tables & Figures.

Acknowledgements

The authors express their thanks to all of the participants in the study. The study was partly supported by the National Science Council, Taipei, Taiwan (grants NSC 96-2314-B-002-155 and NSC 95-2314-B-002 -125). The authors are also thankful to Dr. M. Pencina for kindly providing the SAS macro for computing NRI and IDI statistics.

Author details

¹Institute of Preventive Medicine, School of Preventive Medicine, National Taiwan University, Taipei, Taiwan. ²Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan.

Authors' contributions

KLC: collecting data, analyzing data, writing the draft, supervising the study; BCL: collecting data; HRL: collecting data; HCH: laboratory data measurements and quality control; MFC: obtaining funding, supervising the study. All authors have read and approved the manuscript.

Competing interests

This study has not been published or submitted elsewhere, and no ethical problems or conflicts of interest are declared.

Received: 7 July 2010 Accepted: 27 September 2010

Published: 27 September 2010

References

- Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). UK Prospective Diabetes Study (UKPDS) Group. *Lancet* 1998, **352**(9131):837-853.
- Stratton IM, Adler AI, Neil HA, Matthews DR, Manley SE, Cull CA, Hadden D, Turner RC, Holman RR: Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. *Bmj* 2000, **321**(7258):405-412.
- Khaw KT, Wareham N, Luben R, Bingham S, Oakes S, Welch A, Day N: Glycated haemoglobin, diabetes, and mortality in men in Norfolk cohort of European prospective investigation of cancer and nutrition (EPIC-Norfolk). *Bmj* 2001, **322**(7277):15-18.
- Liu QZ, Pettitt DJ, Hanson RL, Charles MA, Klein R, Bennett PH, Knowler WC: Glycated haemoglobin, plasma glucose and diabetic retinopathy: cross-sectional and prospective analyses. *Diabetologia* 1993, **36**(5):428-432.
- Schulze MB, Shai I, Manson JE, Li T, Rifai N, Jiang R, Hu FB: Joint role of non-HDL cholesterol and glycated haemoglobin in predicting future coronary heart disease events among women with type 2 diabetes. *Diabetologia* 2004, **47**(12):2129-2136.
- Park PJ, Griffin SJ, Sargeant L, Wareham NJ: The performance of a risk score in predicting undiagnosed hyperglycemia. *Diabetes Care* 2002, **25**(6):984-988.
- Al-Lawati JA, Tuomilehto J: Diabetes risk score in Oman: A tool to identify prevalent type 2 diabetes among Arabs of the Middle East. *Diabetes Res Clin Pract* 2007, **77**(3):438-444.
- Glumer C, Vistisen D, Borch-Johnsen K, Colagiuri S: Risk scores for type 2 diabetes can be applied in some populations but not all. *Diabetes Care* 2006, **29**(2):410-414.
- Ramachandran A, Snehalatha C, Vijay V, Wareham NJ, Colagiuri S: Derivation and validation of diabetes risk score for urban Asian Indians. *Diabetes Res Clin Pract* 2005, **70**(1):63-70.
- Kanaya AM, Wassel Fyr CL, de Rekeneire N, Shorr RI, Schwartz AV, Goodpaster BH, Newman AB, Harris T, Barrett-Connor E: Predicting the development of diabetes in older adults: the derivation and validation of a prediction rule. *Diabetes Care* 2005, **28**(2):404-408.
- 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**(10):1068-1074.
- Schulze MB, Hoffmann K, Boeing H, Linseisen J, Rohrmann S, Mohlig M, Pfeiffer AF, Spranger J, Thamer C, Haring HU, *et al*: An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. *Diabetes Care* 2007, **30**(3):510-515.
- Wannamethee SG, Shaper AG, Lennon L, Morris RW: Metabolic syndrome vs Framingham Risk Score for prediction of coronary heart disease, stroke, and type 2 diabetes mellitus. *Arch Intern Med* 2005, **165**(22):2644-2650.
- Silventoinen K, Pankow J, Lindstrom J, Jousilahti P, Hu G, Tuomilehto J: 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* 2005, **12**(5):451-458.
- McNeely MJ, Boyko EJ, Leonetti DL, Kahn SE, Fujimoto WY: 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* 2003, **26**(3):758-763.
- Lindstrom J, Tuomilehto J: The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 2003, **26**(3):725-731.
- Stern MP, Williams K, Haffner SM: Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test? *Ann Intern Med* 2002, **136**(8):575-581.
- von Eckardstein A, Schulte H, Assmann G: 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* 2000, **85**(9):3101-3108.
- Thomas C, Hypponen E, Power C: Type 2 diabetes mellitus in midlife estimated from the Cambridge Risk Score and body mass index. *Arch Intern Med* 2006, **166**(6):682-688.
- Chien KL, Lee BC, Hsu HC, Lin HJ, Chen MF, Lee YT: Prevalence, agreement and classification of various metabolic syndrome criteria among ethnic Chinese: A report on the hospital-based health diagnosis of the adult population. *Atherosclerosis* 2008, **196**(2):764-771.
- Chien KL, Hsu HC, Lee YT, Chen MF: Renal function and metabolic syndrome components on cardiovascular and all-cause mortality. *Atherosclerosis* 2008, **197**(2):860-867.
- Chien KL, Hsu HC, Chao CL, Lee BC, Chen MF, Lee YT: Interaction of obesity, metabolic syndrome and Framingham risk on steatohepatitis among healthy Taiwanese: population-based nested case-control study. *Cardiovasc Diabetol* 2006, **5**(1):12.

23. Chien KL, Yang CY, Lee YT: **Major Gene Effects in Systolic and Diastolic Blood Pressure in the Families Receiving Health Examination in Taiwan.** *Journal of Hypertension* 2003, **21**:1-7.
24. Chien KL, Hsu HC, Su TC, Lee YT: **Consistency in genetic inheritance mode and heritability patterns of triglyceride vs. high density lipoprotein cholesterol ratio in two Taiwanese family samples.** *BMC Journal Genetics* 2003, **4**:7-16.
25. Finke A, Kobold U, Hoelzel W, Weykamp C, Miedema K, Jeppsson JO: **Preparation of a candidate primary reference material for the international standardisation of HbA1c determinations.** *Clin Chem Lab Med* 1998, **36**(5):299-308.
26. Engelgau MM, Thompson TJ, Herman WH, Boyle JP, Aubert RE, Kenny SJ, Badran A, Sous ES, Ali MA: **Comparison of fasting and 2-hour glucose and HbA1c levels for diagnosing diabetes. Diagnostic criteria and performance revisited.** *Diabetes Care* 1997, **20**(5):785-791.
27. Hanson RL, Nelson RG, McCance DR, Bear JA, Charles MA, Pettitt DJ, Knowler WC: **Comparison of screening tests for non-insulin-dependent diabetes mellitus.** *Arch Intern Med* 1993, **153**(18):2133-2140.
28. 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**(3):164-171.
29. Gaziano TA, Young CR, Fitzmaurice G, Atwood S, Gaziano JM: **Laboratory-based versus non-laboratory-based method for assessment of cardiovascular disease risk: the NHANES I Follow-up Study cohort.** *Lancet* 2008, **371**(9616):923-931.
30. Aekplakorn W, Bunnag P, Woodward M, Sritara P, Cheepudomwit S, Yamwong S, Yipintsoi T, Rajatanavin R: **A risk score for predicting incident diabetes in the Thai population.** *Diabetes Care* 2006, **29**(8):1872-1877.
31. Heidemann C, Hoffmann K, Spranger J, Klipstein-Grobusch K, Mhlig M, Pfeiffer AF, Boeing H: **A dietary pattern protective against type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study cohort.** *Diabetologia* 2005, **48**(6):1126-1134.
32. Schmidt MI, Duncan BB, Bang H, Pankow JS, Ballantyne CM, Golden SH, Folsom AR, Chambless LE: **Identifying individuals at high risk for diabetes: The Atherosclerosis Risk in Communities study.** *Diabetes Care* 2005, **28**(8):2013-2018.
33. Harrell FE: **Regression modeling strategies with applications to linear models, logistic regression, and survival analysis.** New York: Springer 2001.
34. Sullivan LM, Massaro JM, D'Agostino RB Sr: **Presentation of multivariate data for clinical use: The Framingham Study risk score functions.** *Stat Med* 2004, **23**(10):1631-1660.
35. Hanley JA, McNeil BJ: **A method of comparing the areas under receive operating characteristic curves derived from the same cases.** *Radiology* 1983, **148**:839-843.
36. DeLong ER, DeLong DM, Clarke-Pearson DL: **Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach.** *Biometrics* 1988, **44**(3):837-845.
37. Hosmer DW Jr, Lemeshow S: **The multiple logistic regression model.** *Applied logistic regression* New York: John Wiley & Sons, 1 1989, 25-37.
38. Yates J: **External correspondence: decomposition of the mean probability score.** *Organizational Behavior and Human Performance* 1982, **30**:132-156.
39. StatCorp: **Stata Statistical Software: Release 10.** Collee Station, TX: StataCorp LP 2007.
40. Cook NR: **Use and misuse of the receiver operating characteristic curve in risk prediction.** *Circulation* 2007, **115**(7):928-935.
41. Pencina MJ, D' Agostino RBS, D' Agostino RBJ, Vasan RS: **Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond.** *Stat Med* 2008, **27**(2):157-172.
42. Schmid C, Griffith J: **Multivariable classification rules: calibration and discrimination.** In *Encyclopedia of Biostatistics*. Edited by: Armitage P, Colton T. Chichester, U.K.: Wiley; 1998.
43. Wolf PA, D'Agostino RB, Belanger AJ, Kannel WB: **Probability of stroke: a risk profile from the Framingham Study.** *Stroke* 1991, **22**(3):312-318.
44. Wilson PWF, Castelli WP, Kannel WB: **Coronary risk prediction in adults (the Framingham heart study).** *American Journal of Cardiology* 1987, **59**:91G-94G.
45. D'Agostino RB Sr, Grundy S, Sullivan LM, Wilson P: **Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation.** *JAMA* 2001, **286**(2):180-187.
46. Glumer C, Carstensen B, Sandbaek A, Lauritzen T, Jorgensen T, Borch-Johnsen K: **A Danish diabetes risk score for targeted screening: the Inter99 study.** *Diabetes Care* 2004, **27**(3):727-733.
47. Group DPPR: **Strategies to identify adults at high risk for type 2 diabetes: the Diabetes Prevention Program.** *Diabetes Care* 2005, **28**(1):138-144.
48. D'Agostino RB Jr, Hamman RF, Karter AJ, Mykkanen L, Wagenknecht LE, Haffner SM: **Cardiovascular disease risk factors predict the development of type 2 diabetes: the insulin resistance atherosclerosis study.** *Diabetes Care* 2004, **27**(9):2234-2240.
49. Davidson MB, Schriger DL, Peters AL, Lorber B: **Relationship between fasting plasma glucose and glycosylated hemoglobin: potential for false-positive diagnoses of type 2 diabetes using new diagnostic criteria.** *Jama* 1999, **281**(13):1203-1210.
50. Lauritzen T, Griffin S, Borch-Johnsen K, Wareham NJ, Wolfenbuttel BH, Rutten G: **The ADDITION study: proposed trial of the cost-effectiveness of an intensive multifactorial intervention on morbidity and mortality among people with Type 2 diabetes detected by screening.** *Int J Obes Relat Metab Disord* 2000, **24**(Suppl 3):S6-11.
51. Mengual L, Roura P, Serra M, Montasell M, Prieto G, Bonet S: **Multifactorial control and treatment intensity of type-2 diabetes in primary care settings in Catalonia.** *Cardiovasc Diabetol* 2010, **9**:14.
52. Fu AZ, Qiu Y, Radican L, Yin DD, Mavros P: **Pre-existing cardiovascular diseases and glycemic control in patients with type 2 diabetes mellitus in Europe: a matched cohort study.** *Cardiovasc Diabetol* 2010, **9**:15.

doi:10.1186/1475-2840-9-59

Cite this article as: Chien et al.: Prediction model for high glycated hemoglobin concentration among ethnic Chinese in Taiwan. *Cardiovascular Diabetology* 2010 **9**:59.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

