A Quick Self-assessment Tool to Identify Individuals at High Risk for Type 2 Diabetes in the Chinese General Population

Jing Xie¹, Dongsheng Hu², Dahai Yu³, Chung-Shiuan Chen⁴, Jiang He⁴, Dongfeng Gu³

¹ Centre for Eye Research Australia, Department of Ophthalmology, University of Melbourne, Australia
² Department of Epidemiology, Shenzhen University School of Medicine, Shenzhen, Guangdong, PR China
³ Department of Evidence-Based Medicine, Cardiovascular Institute and Fu Wai Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, PR China
⁴ Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA

Correspondence to:
Dongsheng Hu, MD, PhD
Department of Epidemiology
Shenzhen University School of Medicine
3688 Nanhai Avenue
Shenzhen, Guangdong 518060
PR China
Email: dongsheng_hu@hotmail.com

Word count:
Abstract: 250
Main text: 2335
Licence Statement

"The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in JECH editions and any other BMJPGL products to exploit all subsidiary rights, as set out in our licence (http://jech.bmj.com/ifora/licence.pdf)."
Abstract

**Background:** Current available tools for identifying individuals at high risk for type 2 diabetes can be invasive, costly and time-consuming. This study aims to develop and validate a self-assessment tool for identifying individuals at high risk for type 2 diabetes in the Chinese general population.

**Methods:** A cross-sectional survey was conducted from 2000 to 2001 in a nationally representative sample of 15,540 Chinese adults aged 35 to 74 years. The diabetes risk level (DRL) was assessed by classification and regression tree (CART) analysis using four predictors: age, body mass index, waist-hip ratio (WHR) and waist circumference (WC).

**Results:** The significant predictors for type 2 diabetes were WHR and age for women, and WC and age for men. The categories generated by CART analysis stratified women into 8 DRLs and men into 5 DRLs. The prevalence of type 2 diabetes increased with the increasing of DRLs in both women and men. A DRL ≥ 6 predicted type 2 diabetes status with a sensitivity of 0.61 (95% confidence interval [CI]: 0.55, 0.67), specificity of 0.71 (95% CI: 0.70, 0.73) in women, and a DRL ≥ 3 predicted type 2 diabetes status with a sensitivity of 0.59 (95% CI: 0.52, 0.65), specificity of 0.63 (95% CI: 0.62, 0.65) in men.

**Conclusions:** This study demonstrates that application of the DRLs has identified a substantial proportion of individuals with type 2 diabetes in the Chinese general population. It suggests that there is a great potential of applying the self-assessment tool in health care limited settings.

**Key words:** Type 2 diabetes; diabete risk level; predictor

**Abbreviations**

AUC    Area Under the receiver operating characteristics Curve  
CI     Confidence Interval  
BMI    Body Mass Index  
CART   Classification and Regression Tree  
DRL    Diabetes Risk Level
NPV  Negative Predictive Value  
PPV  Positive Predictive Value  
ROC  Receiver Operating Characteristics  
WHR  Waist-Hip Ratio  
WC  Waist Circumference
Introduction

The rising prevalence of type 2 diabetes has become a major public health problem worldwide and it has been predicted to affect 366 million people by the year 2030, of which at least 40 million people would be from China alone. Approximately a 1% increase in the prevalence of type 2 diabetes will result in 1 million more patients in China, which impose a massive social-economic burden in the community.

Studies have reported that type 2 diabetes is associated with a number of genetic, environmental and lifestyle factors. Recent intervention studies have also clearly shown that type 2 diabetes could be prevented or delayed through simple lifestyle modification among high-risk individuals. Therefore, earlier detection of individuals at high risk for type 2 diabetes is a high priority for primary prevention. However, current available screening tools for identifying individuals at high risk for type 2 diabetes can be invasive, costly and time-consuming. Furthermore, most diabetes risk scores have been developed in western populations, south Asian population and Arabs with application in clinical practice/primary care services. Such predicting tools are lacking in the Chinese population. We also note that different risk scores are not extrapolated across populations. In addition, some of these risk scores used laboratory tests and required history of medication, which limited their use in China due to limited health care resources, especially in private-financing health care services. Easy self-assessment anthropometric measures have been suggested to be good indicators for type 2 diabetes in different populations, and might be used as alternatives for developing screening tools.

So far, no screening tool for type 2 diabetes has been reported in the Chinese population, which results in a large proportion of undiagnosed type 2 diabetes in China. In the current study, we aim to develop and validate a self-assessment tool for identifying individuals at high risk for type 2 diabetes in the Chinese general population.

Research Design and Methods

Study Population
Our analyses are based on data collected from the International Collaborative Study of Cardiovascular Disease in Asia (The InterASIA) conducted in China. The method of this study has been described in detail elsewhere. In brief, a four-stage stratified sampling method was used to select a nationally representative sample of the general population in China from 2000 to 2001. A total of 19,012 persons were randomly selected from 20 primary sampling units (street districts in urban areas or townships in rural areas). Of the 19,012 persons, 15,838 participants (15,540 participants aged 35 to 74 years) completed the survey and examination (response rate: 83%). The current study included 994 persons with type 2 diabetes and 13,129 persons with normal fasting glucose aged 35 to 74 years. Participants with impaired fasting glucose (n=1,121) and those with missing fasting glucose measurements (n=296) were excluded from this analysis.

The InterASIA study was approved by the Institutional Review Board at the Tulane University Health Sciences Center and the ethics committee and other relevant regulatory bodies in China. Informed consent was obtained from each participant.

Data Collection

Data collection was conducted in examination centers at local health stations or community clinics by trained research staff using a standard questionnaire. Information included age, gender, self-reported history of diabetes, and the previous diagnosis and treatment of diabetes.

Anthropometric Measurements

Body weight, height, waist circumference (WC), and hip circumference were measured by trained observers according to a standard protocol in the InterASIA. Body mass index (BMI) was calculated as weight (in kilograms) divided by height (in meters) squared. Waist-hip ratio (WHR) was calculated as WC divided by hip circumference.

Laboratory Measurements

Blood specimen samples were obtained to determine plasma glucose after at
least 12 hours of fasting and collected by use of vacuum tubes containing sodium fluoride. The serum glucose was measured by using the modified hexokinase enzymatic method (HITACHI automatic clinical analyzer, Model 7060, Japan).

Participants without a prior diagnosis of diabetes were categorized according to the American Diabetes Association diagnostic criteria as follows: undiagnosed diabetes (fasting glucose $\geq 7.0$ mmol/l) and impaired fasting glucose (IFG: 6.1-6.9 mmol/l) \textsuperscript{29}. Diabetes was defined as self-reported history of diabetes plus undiagnosed diabetes. Obesity was defined as BMI of 28 kg/m$^2$ or higher, overweight was defined as BMI of 24 or higher but less than 28 \textsuperscript{30}.

**Data Analyses**

Analysis was conducted in women and men separately. Before building models, the data set was randomly split into two subset samples, 50.2\% (n=7,083) of the data for the training sample and the remaining 49.8\% (n=7,040) for the testing sample. Continuous variables were expressed as median (25th percentile, 75th percentile) and categorical variables as n (%). Chi-square test was used to analyze categorical data. Differences in continuous variables were evaluated by Mann Whitney Wilcox on test. Classification and regression tree (CART) analysis was used to develop a simple self-assessment tool for identifying individuals at high risk for type 2 diabetes. This procedure repeatedly partitioned the data to create subgroups with highly homogenous outcomes \textsuperscript{31}. CART constructs a tree that will separate the data in the "best" way by finding binary splits on variables; finds the best splitting variable and the best splitting point at each stage. The decision rule was determined by the Gini criterion, a measure of variability within the new subgroups \textsuperscript{32}. The target variable has a value of 0 or 1 depending on diabetes status (0 if absent, or 1 if present). The explanatory variables selected to predict the risks of type 2 diabetes consist of four easily attainable self-assessment continuous variables (age, BMI, WC and WHR). Nodes in the CART analysis were constrained to have a minimum size of 400 records in parent nodes and 200 records in final child nodes. The proportion of subjects having type 2 diabetes was reported in each node of the tree. A risk stratification system (diabetes risk level,
A multivariable logistic regression model including the same predictors was also developed. The predictive performance of CART analysis was evaluated by using the area under the receiver operating characteristics (ROC) curve (AUC), sensitivity, specificity, the positive predictive value (PPV), and the negative predictive value (NPV).

All analyses were performed using STATA version 9.2 (Stata Corporation, College Station, TX, USA) except the CART analysis was conducted using SPSS 16.0 statistical software package (SPSS, Chicago, IL, USA). A p-value <0.05 was considered statistically significant.

Results

Comparison of the characteristics of the 7,083 participants used to develop the model (training sample) and the 7,040 participants used to test the model (testing sample) are shown in Table 1. The training sample and the testing sample were similar with respect to age, WHR, BMI, WC and the prevalence of type 2 diabetes. The overall prevalence of type 2 diabetes was 6.4% (95% confidence interval [CI]: 5.8%, 6.9%) and 6.7% (95% CI: 6.1%, 7.3%) for women and men, respectively.

Figure 1 and Figure 2 depict the classification tree along with the prevalence of type 2 diabetes in each child node. Of the four common easily known predictors (age, WHR, WC and BMI), WHR and age were selected by the CART analysis for women, whereas WC and age for men. WHR was proved to be the best predictor for women ($\chi^2 = 149.6, p < 0.001$) and WC for men ($\chi^2 = 64.01, p < 0.001$). Age was the second important predictor for type 2 diabetes in both women and men, respectively.

The final child nodes generated by CART analysis stratified women into 8 DRLs and men into 5 DRLs (Figure 1, 2). The prevalence of type 2 diabetes increased with the increasing of DRLs in both women and men as shown in Figure 3 (women: $\chi^2_{\text{trend}} = 2059.42, p<0.001$; men: $\chi^2_{\text{trend}} = 619.01, p<0.001$). Women with a DRL of 8 (WHR > 0.918) and men with a DRL of 5 (WC > 84 cm) are at the highest risk for type 2 diabetes.
Table 1 Socio-demographic and Clinical Characteristics of Study Participants in the Training Sample and the Testing Sample, by Gender

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Women</th>
<th>p</th>
<th>Men</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training sample</td>
<td>Testing sample</td>
<td></td>
<td>Training sample</td>
</tr>
<tr>
<td>Age (years, range)</td>
<td>48.2 (41.5, 57.4)</td>
<td>48.1 (41.0, 57.6)</td>
<td>0.59</td>
<td>48.8 (41.7, 58.7)</td>
</tr>
<tr>
<td>WC (cm)</td>
<td>77.0 (70.2, 84.2)</td>
<td>76.5 (70.0, 84.0)</td>
<td>0.37</td>
<td>80.8 (73.6, 88.0)</td>
</tr>
<tr>
<td>BMI (kg/m(^2))</td>
<td>23.4 (21.2, 26.0)</td>
<td>23.3 (21.2, 25.8)</td>
<td>0.35</td>
<td>23.4 (21.1, 25.7)</td>
</tr>
<tr>
<td>WHR (unit)</td>
<td>0.82 (0.75, 0.87)</td>
<td>0.82 (0.78, 0.87)</td>
<td>0.12</td>
<td>0.87 (0.82, 0.91)</td>
</tr>
<tr>
<td>Overweight (n (%))</td>
<td>1156 (31.5)</td>
<td>1154 (31.4)</td>
<td>0.95</td>
<td>1121 (32.8)</td>
</tr>
<tr>
<td>Overall obesity (n (%))</td>
<td>448 (12.2)</td>
<td>414 (11.3)</td>
<td>0.22</td>
<td>338 (9.9)</td>
</tr>
<tr>
<td>Central obesity (n (%))</td>
<td>1460 (39.8)</td>
<td>1419 (38.6)</td>
<td>0.32</td>
<td>1261 (37.0)</td>
</tr>
<tr>
<td>Type 2 Diabetes (n (%))</td>
<td>258 (7.3)</td>
<td>243 (6.6)</td>
<td>0.48</td>
<td>241 (7.1)</td>
</tr>
</tbody>
</table>

WC, waist circumference; BMI, Body mass index; WHR, waist to hip ratio.

Continuous variable presented as median (25th percentile, 75th percentile) and categorical variables as n (%).

* Diagnostic criteria for Chinese population: BMI ≥ 25kg/m\(^2\) and BMI < 28kg/m\(^2\) was defined as overweight.

** Diagnostic criteria for Chinese population: BMI ≥ 28 kg/m\(^2\) was defined as obesity.

*** Diagnostic criteria for Chinese population: WC ≥ 85 cm for men or WC ≥ 80 cm for women was defined as obesity.

◊: Percentage of diabetes in training/testing sample as number of diabetes/total number in training or testing sample.
Sensitivity, specificity, PPV, NPV and likelihood ratios which predict the presence of type 2 diabetes for a threshold are shown in Table 2. A threshold of DRL $\geq 6$ in women identified 62% individuals with type 2 diabetes with a specificity of 72%, and a threshold of DRL $\geq 3$ in men identified 60% individuals with type 2 diabetes with a specificity of 63%.

The CART analysis was able to stratify similar results in both training and testing samples (Figure 1, 2, 4, 5). The AUCs in the training sample were not significantly different from those in the testing sample for both women and men (women: 0.71 vs. 0.73; men: 0.65 vs. 0.69), indicating a good prediction above chance. Other diagnostic statistics were also similar in the two samples (Table 2).

Multivariable (age, WHR, WC and BMI) logistic regression analyses show that age, WC and WHR were associated with type 2 diabetes in women

\[
p = \left(1 + e^{-(0.048 \times \text{Age} + 0.020 \times \text{WC} + 0.01 \times \text{WHR})}\right)^{-1},
\]

and age, WC, WHR, and BMI in men

\[
p = \left(1 + e^{-(0.033 \times \text{Age} + 0.019 \times \text{WC} + 3.691 \times \text{WHR} + 0.043 \times \text{BMI})}\right)^{-1}.
\]

There was no difference of AUCs between CART analysis and logistic regression model (data not shown).
<table>
<thead>
<tr>
<th>Diagnostic statistics (95% CI)</th>
<th>Women (Diabetic Risk levels ≥ 6)</th>
<th>Men (Diabetic Risk levels ≥ 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training sample</td>
<td>Testing sample</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>----------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.61 (0.55, 0.67)</td>
<td>0.62 (0.55, 0.68)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.71 (0.70, 0.73)</td>
<td>0.73 (0.71, 0.74)</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>0.14 (0.12, 0.16)</td>
<td>0.14 (0.12, 0.16)</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>0.96 (0.95, 0.97)</td>
<td>0.96 (0.96, 0.97)</td>
</tr>
<tr>
<td>Positive likelihood ratio</td>
<td>2.16 (1.94, 2.42)</td>
<td>2.27 (2.02, 2.54)</td>
</tr>
<tr>
<td>Negative likelihood ratio</td>
<td>0.54 (0.47, 0.64)</td>
<td>0.53 (0.45, 0.62)</td>
</tr>
</tbody>
</table>

CI, confidence interval
Discussion

In this study, a self-assessment tool with two simple predictors to identify individuals at high risk for type 2 diabetes was developed and successfully validated by CART analysis in a Chinese general population. The final child nodes generated by CART analysis stratified women into 8 DRLs and men into 5 DRLs. Nearly two-thirds of individuals with type 2 diabetes have been identified with a reasonable specificity.

As a developing country, the cost on primary care has already been a colossal burden in China. Moreover, awareness, treatment and control of diabetes are relatively low among Chinese adults. Therefore, it will be of great value to develop a simple tool for identifying individuals at high risk for type 2 diabetes. To the best of our knowledge, no screening tool for diabetes has been developed previously in the Chinese population. As the DRLs generated by CART analysis in this study can be easily applied in general population and identify individuals with diabetes, there is a great potential of its application in primary care settings in resource limited countries.

The predictive performance and discriminative ability of this simple tool is highly comparable to those diabetes risk scores, which composed of different set of clinical variables. The AUC of DRLs generated by CART analysis in our study is similar to those in Indian (AUC=0.73), Thai (AUC=0.74) and Danish (AUC=0.76). However, those diabetes risk scores included at least four variables, and also required medical information. Even higher AUCs (German: 0.84; US: 0.85; UK: 0.80; US: 0.82, Spain: 0.83 for men, 0.87 for women; and middle east: 0.83) have been reported if more risk factors included, such as smoking, alcohol consumption, and biochemical tests. It is well known that the more variables included in the prediction tool, the less applicable in the general population. Given that diabetes could be prevented or delayed by changing life style and physical activity patterns, it seems more sensible to target individuals at high risk by simple programme, which could be run easily in the general public.

The DRL in this study is very simple and also relatively easy to interpret. Compared to other diabetes risk scores, this DRL has a moderate sensitivity and a similar PPV (11-14%), but a high NPV (95-96%). For identifying individuals with diabetes, the DRL has a sensitivity of 61% with a specificity of 71% in women and a sensitivity of 60% with a specificity of 63% in men. According to NPV, individuals with lower DRL are unlikely to have diabetes (about 95% probability). This DRL emphasizes people with obesity, especially old people (women > 60 years.
and men > 50 years). Women with a DRL of 8 (WHR > 0.918) and men with a DRL of 5 (WC > 84 cm) will then need further blood tests and suggest to change to a healthier lifestyle for primary prevention. This information is very important for both the individuals and the health workers when making plans for action.

CART analysis has been proven to perform better than logistic regression model in developing diabetes risk scores. In the current study, CART analysis performed similar results as compared to the multiple logistic regression models in terms of sensitivities, specificities and AUCs. Results of CART analysis also confirmed that central adiposity (e.g., WHR and WC) are better predictors of type 2 diabetes than overall obesity (BMI), which is consistent with previous studies.

There are several advantages of the DRL. First, it is developed in a representative general Chinese population. Our study employed a multistage stratified random sampling method to select a representative national sample from the Chinese general population. Second, the development and validation of the model in 2 separate groups (randomized split sample) is an important strength, which lends support for robustness and generalizability of the results. Third, the final model is quite simple with only two items in a classification tree form. Compared to other diabetes risk scores derived from more variables and also included blood test for glucose, DRLs can be easily applied at home in the general population. Although adding laboratory results would be more predictive, they were not available in general public, and, thus, were not included in our analysis. Our goal in this study is to develop a parsimonious tool, and not primarily to maximize prediction. Potential limitations of the current analysis must be addressed. First, association between the independent variable and type 2 diabetes may not be causal given the cross-sectional nature of the current study. Second, our study did not include the younger age group.

**Conclusion**

The CART based analysis in the current study has created a simple robust tool with considerable discriminative ability and is easy to apply in the general public. We strongly believe that the public health implications of this decision tree are considerable. It is a cost-efficient and practical tool to identify individuals at high risk of type 2 diabetes in the general population and also offer a chance for early prevention measures.
What is already known on this subject?

- Type 2 diabetes is associated with a number of genetic, environmental and lifestyle factors and could be prevented or delayed through simple lifestyle modification among high risk individuals.
- Current available screening tools for identifying individuals at high risk for type 2 diabetes can be invasive, costly and time-consuming.
- Screening tool for type 2 diabetes is lacking in the Chinese population.

What this paper adds?

- The present study used two easy self-assessment anthropometric measures (waist-hip ratio and age for women, and waist circumference and age for men) to generate a new tool for identifying individuals at high risk for type 2 diabetes.
- This quick self-assessment tool has identified a substantial proportion of individuals with type 2 diabetes in the Chinese general population.

Funding:

The InterASIA study was funded by a contractual agreement between Tulane University, LA, USA, and Pfizer Inc., NY, USA. Several researchers employed by Pfizer Inc. were members of the Study Steering Committee that designed the study. However, the study was conducted, analyzed, and interpreted by the investigators independently of the sponsor.

Competing interests: None.

Ethics approval: The InterASIA study was approved by the Institutional Review Board at the Tulane University Health Sciences Center and the ethics committee and other relevant regulatory bodies in China.

Acknowledgements:

We would like to thank Dr Cong Sun from the Centre for Eye Research Australia, University of Melbourne, for reading this manuscript.

Contributors: All authors contributed to the paper. JX undertook the analysis and interpreted the results and draft the paper. D Hu, DHY, Dr. JH, Mrs CS C, and DF Gu critically revised the paper. All authors have seen and agreed the final draft. JX is guarantor for the paper.


(17) Cabrera de LA, Coello SD, Rodriguez Perez MC, Medina MB, Almeida GD, Diaz BB et al. A


(20) 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.


Figure 1 Predictors of Type 2 Diabetes and Risk Stratification by CART analysis for Women in the Training Sample

Figure 2 Predictors of Type 2 Diabetes and Risk Stratification by CART analysis for Men in the Training Sample

Figure 3 Type 2 Diabetes Prevalence Rates of Subgroups with Different Diabetic Risk Levels Developed by CART analysis in Women and Men

Figure 4 Predictors of Type 2 Diabetes and Risk Stratification by CART analysis for Women in the Testing Sample

Figure 5 Predictors of Type 2 Diabetes and Risk Stratification by CART analysis for Men in the Testing Sample