New risk-scoring system including non-alcoholic fatty liver disease for predicting incident type 2 diabetes in East China: Shanghai Baosteel Cohort

Guang-Yu Chen1†, Hai-Xia Cao1†, Feng Li2, Xiao-Bo Cai3, Qing-Hong Ao4, Yan Gao4, Jian-Gao Fan1*

1Center for Fatty Liver Disease, Department of Gastroenterology, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, 2Department of Gastroenterology, Zhongshan Hospital, Fudan University, 3Department of Gastroenterology, Shanghai First People’s Hospital, and 4Center for Health Care, Shanghai Baoshan Iron & Steel Co., Shanghai, China

Keywords
Diabetes, Non-alcoholic fatty liver disease, Proportional hazard models

*Correspondence
Jian-Gao Fan
Tel.: +86-21-25077340
Fax: +86-21-25076160
E-mail address: fattyliver2004@126.com

J Diabetes Investig 2016; 7: 206–211
doi: 10.1111/jdi.12395

ABSTRACT

Aims/Introduction: The present study aimed to explore the incidence of type 2 diabetes, and to develop a risk-scoring model for predicting diabetes among the adult health check-up population in East China.

Materials and Methods: Participants from the Shanghai Baosteel Cohort (age \( \geq 20 \) years) without diabetes at baseline were recruited in a 6-year follow-up study. In order to explore risk factors for diabetes, this cohort was categorized into two groups: new diabetes and no diabetes. Three models were developed by Cox regression analysis. The model accuracy was assessed using the area under the receiver operating characteristic curve.

Results: A total of 6,542 individuals were included in the Shanghai Baosteel Cohort Study. Of them, 368 (5.6%) developed type 2 diabetes at the end of the follow-up period. Cox regression analysis found a close association between incident type 2 diabetes and several risk factors including non-alcoholic fatty liver diseases at baseline. The Shanghai Baosteel Score including advanced age (2 points), hypertriglyceridemia (2 points), obesity (2 points), non-alcoholic fatty liver diseases (2 points) and impaired fasting glucose (3 points) had a good diagnostic performance with estimated area under the receiver operating characteristic curve (0.724), sensitivity (57.9%) and specificity (72.2%) at a cut-off point of >3.

Conclusions: A risk-scoring system including non-alcoholic fatty liver diseases can help identify individuals at a high risk of diabetes in the East Chinese population.

INTRODUCTION

Type 2 diabetes has become a serious worldwide public health problem over the past decades, and has increased the burden of patients, their families and the healthcare system. The global prevalence of type 2 diabetes in 2013 was 8.3% in adults, with current projections estimating 592 million diabetic people by 2035\(^1\). The prevalence of type 2 diabetes is much higher in low- and middle-income countries than developed countries. According to the Diabetes Atlas, China had 98.4 million diabetes patients in 2013, the largest diabetic population in the world\(^1\). The direct medical cost of type 2 diabetes in 2008 reached $9.1 billion, the increased economic burden is thus a significant challenge for China\(^2\). There has been great interest in predicting incident type 2 diabetes in the general population in order to prevent the epidemic of this disease\(^3\).

According to the American Diabetes Association criteria, glycated hemoglobin (HbA1c) is an important diagnostic tool for diabetes. However, it is not routinely used in health check-up centers or community hospitals. Therefore, it is pivotal to know whether the incidence of diabetes can be predicted with risk factors including impaired fasting glucose (IFG). Based on...
epidemiological research, the risk factors associated with the onset of type 2 diabetes are not clear. Most of the existing diabetes risk prediction models are based on a cross-sectional study evaluating subjects with the associated risk factors of diabetes. Although there are established models based on cohort studies, most are studies of Western populations, with just a few in Asian populations. Recently, some predicting models were set up in Chinese adults, but they are all from North China.

Epidemiological studies have shown that non-alcoholic fatty liver disease (NAFLD)-related biomarkers (such as serum alanine aminotransferase and gamma-glutamyltransferase) and ultrasound-based NAFLD can be used to predict type 2 diabetes. However, to date, few studies have focused on the association between dyslipidemia, NAFLD and type 2 diabetes. In the present large-scale cohort study, we established a new simple scoring system to explore whether hyperlipidemia and NAFLD can predict the onset of type 2 diabetes in Shanghai, China.

METHODS

Data Selection

The Shanghai Baosteel Study included a cohort consisting of apparently healthy employees who underwent health check-ups every 2 years. A total of 12,640 individuals underwent a baseline health examination during the period from 1995 to 1996, and 7,147 individuals were enrolled in this study who underwent a reexamination 6 years after the initial examination in 2001–2002.

They are followed up every 2 years. Over 5% of 7,147 individuals who had diabetes at the baseline examination (n = 334) or with missing data on baseline characteristics (n = 116) were excluded. In addition, a total of 155 participants with high alcohol consumption (n = 58), the ethanol intake per week was more than 140 g in men, 70 g in women, in the past 12 months), cancer (n = 51), hepatitis C virus infection (n = 12), autoimmune liver disease (n = 2) and missing data on the following follow-up health check-ups (n = 49) were excluded. Consequently, a total of 6,542 participants were recruited into the study for the risk of diabetes.

A questionnaire was administered including age, sex, alcohol consumption, personal medical history and so on. Anthropometric data were collected including height, weight, blood pressure and so on. Body mass index (BMI) was then calculated by dividing weight in kilograms (kg) by height in meters squared (m²). Obesity is defined as BMI ≥25 kg/m². Blood samples were obtained for biochemical tests including fasting plasma glucose (FPG), total cholesterol (TC) and triglycerides (TG) after an overnight fast of 12 h. Hypertriglyceridemia was defined as serum TG ≥1.7 mmol/L and hypercholesteremia as serum TC ≥2.6 mmol/L. The diagnosis of NAFLD was decided by type B ultrasound criteria, including hepatorenal echo contrast, liver brightness, deep attenuation and vascular blurring, using a 3.5-MHz probe. The definition of type 2 diabetes incidence was based on a blood test at follow up, according to the American Diabetes Association criteria, as well as diagnosis and/or receipt of diabetes medication during follow up.

The investigations were carried out by trained nurses from the Center for Health Care of Shanghai Baosteel. The study protocol followed the Ethical Guidelines for Clinical/Epidemiological Studies of the National Health and Family Planning Commission of the People’s Republic of China in accordance with the Declaration of Helsinki, and received ethical approval from the institutional review boards of all participating institutions. Informed consent was obtained from all participants.

Statistical Analysis

Baseline characteristics were summarized separately in new diabetes and no diabetes at follow up, and compared using unpaired t-tests for continuous variables and \( \chi^2 \)-tests for categorical variables. The relative risks (RR) and 95% confidence intervals (95% CI) were calculated using proportional hazards models adjusted for age and sex.

To develop the risk scores for predicting the 6-year incidence of diabetes, we estimated point scores from the \( \beta \) coefficients of three multivariate Cox proportional hazards models. The first model, including the continuous variable, was based on age, TG, BMI and FPG. The second model, including the binary variable, was based on age group (age <55 years or ≥55 years), hypertriglyceridemia, obesity, NAFLD and IFG. Then the third model, a scoring system, was developed for each significant variable in the second model, and a score was assigned to each variable based on the regression coefficients by 10 and rounding to the nearest integer. The Shanghai Baosteel Score (SBS) was calculated as the sum of points for each variable in the third model. A Cox hazard regression model with a significant incidence of diabetes-related variables was set up by the following formula: risk score = \( X_1 \times \beta_1 + X_2 \times \beta_2 + \ldots + X_p \times \beta_p \).

For each model, the predictive performance of the risk score was evaluated with respect to the area under the curve (AUC) in a receiver operating characteristic curve (ROC), sensitivity, specificity, positive likelihood ratio (sensitivity / [1-specificity]) and negative likelihood ratio ([1-sensitivity] / specificity) were calculated. The cut-off score that gave the maximum sum of sensitivity and specificity was taken as the optimum.

All statistical tests were two-sided with a type I error of 0.05, and P-values <0.05 were considered statistically significant. Statistical analysis was carried out using SPSS 13.0 software (SPSS Inc, Chicago, IL, USA).

RESULTS

Participants’ Characteristics

Baseline characteristics of the study population are presented in Table 1. Of 6,542 individuals, 5,617 were men (85.9%), with an average age of 35.3 ± 10.0 years. The average BMI was 22.6 kg/m². At baseline, the prevalence of obesity, hypertriglyceridemia, hypercholesteremia, and fatty liver were 20.7%,
Table 1 | Baseline characteristics by incident diabetes status and univariate analyses of the relative risk for diabetes

| Characteristic | Total (n = 6,542) | New diabetes (n = 368) | No diabetes (n = 6,174) | RR (95% CI) | P-value |
|---------------|------------------|------------------------|-------------------------|-------------|---------|
| Sex, % (men)  | 85.9             | 90.2                   | 85.6                    | 1.91 (1.34–2.70) | 0.014   |
| Age (years)   | 35.3 ± 10.0      | 39.4 ± 11.2            | 35.1 ± 9.8              | 1.04 (1.03–1.05) | <0.001  |
| Height (cm)   | 170.2 ± 66       | 169.8 ± 64             | 170.2 ± 66              | 0.98 (0.96–1.00) | 0.224   |
| Weight (cm)   | 65.6 ± 96        | 70.6 ± 97              | 65.3 ± 95               | 1.05 (1.04–1.06) | <0.001  |
| BMI (kg/m²)   | 22.6 ± 2.9       | 24.5 ± 3.2             | 22.5 ± 2.8              | 1.20 (1.16–1.24) | <0.001  |
| FPG (mmol/L)  | 5.48 ± 0.56      | 5.85 ± 0.62            | 5.46 ± 0.55             | 3.13 (2.60–3.77) | <0.001  |
| TC (mmol/L)   | 4.41 ± 0.90      | 4.74 ± 1.02            | 4.39 ± 0.89             | 1.27 (1.13–1.43) | <0.001  |
| TG (mmol/L)   | 1.43 ± 1.08      | 2.03 ± 1.47            | 1.39 ± 1.05             | 1.33 (1.25–1.41) | <0.001  |
| Obesity (%)   | 20.7             | 42.1                   | 194                     | 2.38 (1.93–2.95) | <0.001  |
| Hypercholesteremia (%) | 8.3 | 16.0 | 7.8 | 1.49 (1.09–2.04) | <0.001 |
| Hypertriglyceridemia (%) | 26.1 | 46.7 | 24.9 | 2.16 (1.76–2.67) | <0.001 |
| IFG (%)       | 14.4             | 39.9                   | 12.8                    | 3.65 (2.96–4.51) | <0.001  |
| NAFLD (%)     | 3.2              | 13.0                   | 2.6                     | 4.01 (2.96–5.44) | <0.001  |

Data are % or mean ± standard deviation. P-values for continuous outcomes were based on a t-test and χ²-test for categorical variables. Relative risk (RR) and 95% confidence interval (CI) were adjusted for age and sex by proportional hazards models. BMI, body mass index; FPG, fasting plasma glucose; HR, hazard ratio; IFG, impaired fasting glucose; NAFLD, non-alcoholic fatty liver disease TC, total cholesterol; TG, triglycerides.

26.1%, 8.3% and 3.2%, respectively. During the 6-year follow up, 368 participants (5.6%) developed type 2 diabetes, and the 100-person year incidence of diabetes was 0.93. All participants were categorized into two groups, the baseline characteristics of the two groups are shown in Table 1. The prevalence of NAFLD was higher in the new diabetes group (13.0%) compared with the no diabetes group (2.6%).

Risk Factors for Diabetes

Table 1 shows the relative risks associated with the onset of diabetes, adjusted for age and sex by proportional hazards models. Baseline BMI, IFG, TG, TC and the presence of NAFLD were associated with the incident of diabetes after adjusted for age and sex. Of them, IFG and NAFLD at baseline were strong risk factors for type 2 diabetes at the end of the 6-year follow-up period. The risk of diabetes was fourfold higher in the NAFLD group than in the participants without NAFLD. However, IFG was considered as the strongest risk factor after adjusting confounding factors (Table 2). Using Cox multiple regression analysis, age was also the factor strongly related to the incidence of diabetes, while serum TC level was ruled out (Table 2).

Constructing the Prediction Model

Based on Cox regression, we developed three different models to estimate the risk of diabetes (Table 2). Model 1 was based on age, BMI, TG and FPG. Model 2 was created by binary

Table 2 | Results of Cox regression analyses predicting newly detected diabetes: The content of three predictive models

| Variables | Model 1 | Model 2 | SBS |
|-----------|---------|---------|-----|
| Age (years) | 0.025   | 0.069 |    |
| Age range (0, <55 years; 1, ≥55 years) | – | – |    |
| BMI (kg/m²) | 0.132   | 0.568 |    |
| Obesity (0, BMI <25; 1, ≥25 kg/m²) | – | – |    |
| TG (mmol/L) | 0.154   | 0.568 |    |
| Hypertriglyceridemia (0, no; 1, yes) | – | – |    |
| FPG (mmol/L) | 0.974   | 0.442 |    |
| IFG (0, no; 1, yes) | – | – |    |
| NAFLD (0, no; 1, yes) | – | – |    |

Model 1 score (continuous variable) = 0.025 × age + 0.154 × TG + 0.132 × BMI + 0.974 × FPG. Model 2 score (binary variable) = 0.569 × age (0, age <55 years; 1, age ≥55 years) + 0.442 × TG (0, no; 1, yes) + 0.568 × obesity (0, BMI <25; 1, BMI ≥25 kg/m²) + 0.773 × non-alcoholic fatty liver disease (0, no; 1, yes) + 1.148 × IFG (0, no; 1, yes). Model 3 Score (SBS) = age(0, age <55 years; 1, age ≥55 years) × 2 + TG (0, no; 1, yes) × 2 + obesity (0, BMI <25; 1, BMI ≥25 kg/m²) × 2 + non-alcoholic fatty liver disease (0, no; 1, yes) × 2 + IFG (0, no; 1, yes) × 3. BMI, body mass index; FPG, fasting plasma glucose; HR, hazard ratio; IFG, impaired fasting glucose; NAFLD, non-alcoholic fatty liver disease; P, points scored; SBS, Shanghai Baosteel Score; TG, triglycerides.
variable (age range, obesity, hypertriglyceridemia, IFG and NAFLD). Model 3 was a risk score (SBS) based on model 2 including age range (2 points), hypertriglyceridemia (2 points), obesity (2 points), NAFLD (2 points) and IFG (3 points), a total of 11 points. The AUC for models 1, 2, and 3 were 0.727, 0.752 and 0.724, respectively (Figure 1). The optimal cut-off value of the SBS was over 3 points. Among all participants, 24.8% had a risk score more than SBS. The sensitivity and specificity for predicting incident diabetes was 57.9% and 77.2%, respectively (Table 3).

DISCUSSION
It is well known that the early detection and prevention of type 2 diabetes are very important to reduce the health burden worldwide, especially in developing countries and low-income countries18, 19. In the present study, we examined the risk factors for newly developing type 2 diabetes among individuals in the Shanghai Baosteel Cohort, and we developed a new risk-score system based on demographic and clinical data. To our knowledge, this is the first study to develop a scoring system to predict type 2 diabetes in East China. Participants enrolled in

---

**Table 3** | Screening performance of the developed diabetes risk scores for predicting future type 2 diabetes

| Score  | Area under the ROC curve (95% CI) | Optimal cut-off value | Sensitivity (%) | Specificity (%) | LR+  | LR−  |
|--------|---------------------------------|-----------------------|-----------------|-----------------|------|------|
| Model 1 | 0.727 (0.698–0.755)             | ≥0.671                | 58.42           | 77.1            | 2.55 | 0.54 |
| Model 2 | 0.752 (0.725–0.779)             | ≥0.864                | 65.76           | 72.71           | 2.41 | 0.47 |
| SBS    | 0.724 (0.695–0.752)             | >3                    | 57.88           | 77.23           | 2.54 | 0.55 |

CI, confidence interval; LR+, likelihood ratio for a positive test result; LR−, the likelihood ratio for a negative test result; ROC, receiver operating characteristics curve; SBS, Shanghai Baosteel Score.
the present study were mostly healthy workers from the Shanghai Baosteel Limited company.

The variables included in model 1 were age, BMI, serum TG and FPG. These variables are easily obtained in check-up centers and community hospitals. When adding NAFLD to model 1, the predictive ability increased from 0.727 to 0.752. The risk-score system based on model 2 had good discrimination ability (AUC 0.724, 95% CI 0.695–0.752). The scoring system produced by the present study includes simple clinical factors that could be assessed anywhere in a primary healthcare institution; that is, age, obesity, serum TG and IFG. By using this risk score, most clinicians or healthcare providers will be able to carry out screening in a large population. People at high risk are recommended to undergo further examinations, such as the oral glucose tolerance test and HbA1c test, and seek advice from a specialist for early diagnosis of diabetes. In our opinion, these individuals should be advised to develop healthy lifestyles, such as exercise and calorie restriction.

During the past years, many risk scores have been developed elsewhere. Most of them were developed in Caucasians, and contain variables that might not be readily available in other populations. von Eckardstein et al., based on 6.3-year follow up in a German population, established the Prospective Cardiovascular Munster model to predict the incidence of diabetes with an AUC of 0.79. In addition, the San Antonio model was established by Sterne to predict the risk of the onset of diabetes in Mexican Americans and non-Hispanic whites with a follow-up period of 7.5 years. That model was very complicated and included many factors, such as age, sex, obesity, family history of diabetes, FPG, blood pressure, high-density lipoprotein cholesterol and other biological markers, which made it difficult to apply. Recently, Wilson et al., based on the Framingham Offspring Study, constructed another predictive model of diabetes including FPG, BMI, high-density lipoprotein cholesterol, TG and family history of hypertension. The present study confirms and extends the results of these previous studies that FPG, BMI and TG are significantly associated with the incidence of diabetes in the adult population. The present study also verified that NAFLD was closely related to the prevalence and incidence of diabetes, which was consistent with the recent reports.

However, there were several limitations that should be considered based on the present study. First, we did not validate this risk score or compare this score with other models. Second, 2-h postload glucose test and HbA1c were not included in our study, although they are important tools for the diagnosis of diabetes. In clinical practice, it is inconvenient, unavailable and expensive to obtain these data when people undergo a health check-up. Finally, we used type B ultrasound to diagnose fatty liver. It is inaccurate, as it cannot identify mild hepatic steatosis, but it is much more feasible and convenient. Histological diagnosis of NAFLD by liver biopsy is the golden standard, while it seems to be invasive and impractical in a health check-up and follow up.

In conclusion, we found three models that are effective in estimating the risk of diabetes in the Shanghai Baosteel Cohort Study, and developed a new risk score system based on these models. This model promotes public awareness of controlling obesity, dyslipidemia, NAFLD and therefore adopting a healthy lifestyle to prevent diabetes.

ACKNOWLEDGMENTS
This project was funded by the National Basic Research Program of China (973 Program) (2012CB517501), National Natural Science Foundation of China (81070322, 81270491, 81470840), Shanghai Science and Technology Commission Foundation (09140903500, 10411956300), and the 100 Talents Program of the Shanghai Board of Health (XBR20111007).

DISCLOSURE
The authors declare no conflict of interest.

REFERENCES
1. Guariguata L, Whiting DR, Hambleton J, et al. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res Clin Pract 2014; 103: 137–149.
2. Hu H, Sawhney M, Shi L, et al. A Systematic Review of the Direct Economic Burden of Type 2 Diabetes in China. Diabetes Ther 2015; 6: 7–16.
3. Noble D, Mathur R, Dent T, et al. Risk models and scores for type 2 diabetes: systematic review. BMJ 2011; 343: d7163.
4. Glumer C, Carstensen B, Sandbaek A, et al. A Danish diabetes risk score for targeted screening: the Inter99 study. Diabetes Care 2004; 27: 727–733.
5. Mohan V, Deepa R, Deepa M, et al. A simplified Indian Diabetes Risk Score for screening for undiagnosed diabetic subjects. J Assoc Physicians India 2005; 53: 759–763.
6. 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: 438–444.
7. Gao WG, Dong YH, Pang ZC, et al. A simple Chinese risk score for undiagnosed diabetes. Diabet Med 2010; 27: 274–281.
8. Xin Z, Yuan J, Hua L, et al. A simple tool detected diabetes and prediabetes in rural Chinese. J Clin Epidemiol 2010; 63: 1030–1035.
9. Aekplakorn W, Bunnag P, Woodward M, et al. A risk score for predicting incident diabetes in the Thai population. Diabetes Care 2006; 29: 1872–1877.
10. Sun F, Tao Q, Zhan S. An accurate risk score for estimation 5-year risk of type 2 diabetes based on a health screening population in Taiwan. Diabetes Res Clin Pract 2009; 85: 228–234.
11. Chien K, Cai T, Hsu H, et al. A prediction model for type 2 diabetes risk among Chinese people. Diabetologia 2009; 52: 443–450.
12. Rahman M, Simmons RK, Harding AH, et al. A simple risk score identifies individuals at high risk of developing Type 2
13. Schulze MB, Hoffmann K, Boeing H, 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: 510–515.

14. Bae JC, Rhee EJ, Lee WY, et al. Combined effect of nonalcoholic fatty liver disease and impaired fasting glucose on the development of type 2 diabetes: a 4-year retrospective longitudinal study. *Diabetes Care* 2011; 34: 727–729.

15. Fraser A, Thinggaard M, Christensen K, et al. Alanine aminotransferase, gamma-glutamyltransferase (GGT) and all-cause mortality; results from a population-based Danish twins study alanine aminotransferase, GGT and mortality in elderly twins. *Liver Int* 2009; 29: 1494–1499.

16. Park SK, Seo MH, Shin HC, et al. Clinical availability of nonalcoholic fatty liver disease as an early predictor of type 2 diabetes mellitus in Korean men: 5-year prospective cohort study. *Hepatology* 2013; 57: 1378–1383.

17. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2004; 27(Suppl 1): S5–S10.

18. Beran D. The impact of health systems on diabetes care in low and lower middle income countries. *Curr Diab Rep* 2015; 15: 591.

19. Mendis S, Chestnov O. Costs, benefits, and effectiveness of interventions for the prevention, treatment, and control of cardiovascular diseases and diabetes in Africa. *Prog Cardiovasc Dis* 2013; 56: 314–321.

20. von Eckardstein A, Schulte H, Assmann G. Risk for diabetes mellitus in middle-aged Caucasian male participants of the Prospective Cardiovascular Munster study: implications for the definition of impaired fasting glucose by the American Diabetes Association. Prospective Cardiovascular Munster. *J Clin Endocrinol Metab* 2000; 85: 3101–3108.

21. Stern MP, Fatehi P, Williams K, et al. Predicting future cardiovascular disease: do we need the oral glucose tolerance test? *Diabetes Care* 2002; 25: 1851–1856.

22. Wilson PW, Meigs JB, Sullivan L, et al. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern Med* 2007; 167: 1068–1074.

23. Bedogni G, Miglioli L, Masutti F, et al. Prevalence of and risk factors for nonalcoholic fatty liver disease: the Dionysos nutrition and liver study. *Hepatology* 2005; 42: 44–52.

24. Fan JG, Li F, Cai XB, et al. Effects of nonalcoholic fatty liver disease on the development of metabolic disorders. *J Gastroenterol Hepatol* 2007; 22: 1086–1091.

25. Fan JG, Farrell GC. Epidemiology of non-alcoholic fatty liver disease in China. *J Hepatol* 2009; 50(1): 204–210.

26. Gao X, Fan JG; Study Group of Liver and Metabolism, Chinese Society of Endocrinology. Diagnosis and management of non-alcoholic fatty liver disease and related metabolic disorders: consensus statement from the Study Group of Liver and Metabolism, Chinese Society of Endocrinology. *J Diabetes* 2013; 5: 406–415.