A Retrospective Population Study to Develop a Predictive Model of Prediabetes and Incident Type 2 Diabetes Mellitus from a Hospital Database in Japan Between 2004 and 2015

Hai Wang
A Xin Zheng
C Zheng-Hai Bai
C Jun-Hua Lv
C Jiang-Li Sun
E Yu Shi
A Hong-Hong Pei

Background: Type 2 diabetes mellitus is a global public health problem. Prediabetes may be reversed by weight loss, diet, and lifestyle changes. However, without intervention, between 30–50% of individuals with prediabetes develop type 2 diabetes. This retrospective population study was conducted to develop a predictive model of prediabetes and incident type 2 diabetes mellitus using data from 2004 to 2015 from the DRYAD Japanese hospital database.

Material/Methods: A retrospective longitudinal population study was conducted using the DRYAD database from Murakami Memorial Hospital, Gifu, Japan, to construct a predictive model for prediabetes and incident type 2 diabetes mellitus in the population. Univariate analysis and multivariate analysis were performed to identify the variables that were associated with prediabetes. These variables were used to construct (75% samples) and verify (25% samples) the predictive model.

Results: From 2004 to 2015, a total of 11,113 cases were identified. Multivariate logistic regression analysis included the six variables of age, waist circumference, smoking history, the presence of fatty liver, fasting blood glucose (FBG), and glycated hemoglobin (HbA1c) level. Data were used to construct (75% samples) and verify (25% samples) the predictive model. The area under the receiver operating characteristic (ROC) curve (AUC) of the predictive model was 0.87 (0.85–0.89) in the training cohort and 0.87 (0.86–0.90) in the validation cohort.

Conclusions: A prognostic model based on six variables was predictive for incident type 2 diabetes mellitus and prediabetes in a healthy population in Japan.

MeSH Keywords: Decision Support Techniques • Diabetes Mellitus, Type 2 • Healthy People Programs

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/920880
Background

Type 2 diabetes mellitus is a global public health problem [1]. Prediabetes may be diagnosed on routine blood testing and by an increased fasting blood glucose (FBG) level [1,2]. Incident diabetes is diagnosed from the finding of impaired glucose tolerance (IGT) or impaired fasting glucose (IFG) results [1,2]. The incidence of Type 2 diabetes mellitus increases annually in both developed and developing countries. Worldwide, in 2008, the number of adult patients with diabetes was approximately 108 million, which rose to approximately 422 million in 2014 [2]. Diabetes caused 1.5 million deaths in 2012, and the effects of hyperglycemia resulted in an additional 2.2 million deaths by increasing the risk of cardiovascular disease and other diseases [3].

Prediabetes is associated with abnormal glucose homeostasis and hyperglycemia and is a risk factor for the development of type 2 diabetes mellitus [4]. The progression from prediabetes to type 2 diabetes mellitus can take months or years. However, it is important to recognize that prediabetes is reversible when diet and lifestyle changes are made that include weight loss and exercise [5]. Early intervention in the prediabetic population has been estimated to reduce the risk of type 2 diabetes mellitus by 40–70% [6,7]. However, because prediabetes may be undetected, approximately 30–50% of patients eventually develop type 2 diabetes mellitus [7,8].

If effective interventions could be implemented for high-risk patients before the development of prediabetes, these may further reduce the incidence of type 2 diabetes mellitus. However, there is a lack of effective methods to identify individuals who are likely to develop prediabetes. There is a need to develop simple predictive models to identify patients with prediabetes who are at increased risk of developing type 2 diabetes mellitus to implement preventive measures even earlier and reduce the prevalence of type 2 diabetes.

Routine health checks can include demographic data, measurement of serum lipids, body mass index (BMI), waist circumference, fasting blood glucose, and glycated hemoglobin. Population studies that use data from healthy individuals can identify early risk factors for disease. The DRYAD database is a public data repository, which contains clinical data from healthy individuals attending a routine medical examination (human dock) [8,9]. The clinical program and database were established from 1994 by Dr. Masahide Hamaguchi at Murakami Memorial Hospital, Gifu, Japan, to screen for chronic diseases and their risk factors and to promote public health [9]. Since 1994, the DRYAD program has evaluated more than 8,000 medical examinations annually, and 60% of participants underwent one to two medical examinations per year [9].

Therefore, this retrospective population study was conducted to develop a predictive model of prediabetes and incident type 2 diabetes mellitus using data from 2004 to 2015 from the DRYAD Japanese hospital database.

Material and Methods

Study design and data source

This retrospective population study was conducted using data from the DRYAD database (https://datadryad.org), obtained from Dr. Murakami, Memorial Hospital, Gifu, Japan [9]. The study included the analysis of anonymized data and did not require patient consent. The aim of the study was to construct a predictive model to identify patients at high risk of type 2 diabetes mellitus from a healthy patient population.

Inclusion and exclusion criteria

From 2004 to 2015, there were 20,944 medical examination cases, of which 16,283 met the inclusion criteria, 5,170 cases were excluded, and 11,113 cases were included in the study. The patients who participated in the medical examination program at Murakami Memorial Hospital from 2004 to 2015 were included who had repeat diagnostic tests and examinations for diabetes mellitus with a follow-up of ≥5 years. The exclusion criteria included a diagnosis of diabetes at the baseline examination (n=323), missing data (n=863), known liver disease (n=416), alcohol consumption >60 gm/day for men and >40 gm/day for women (n=739), medication use (n=2,321), and a fasting blood glucose (FBG) ≥6.1 mmol/L (n=808).

Clinical and biochemical data

Type 2 diabetes mellitus was diagnosed as a glycated hemoglobin level (HbA1c) ≥6.5% or fasting blood glucose (FBG) ≥7 mmol/L, or as self-reported diabetes [10]. The indicators collected during the initial medical examination included age, gender, body mass index (BMI), smoking history, degree of exercise, waist circumference, fatty liver, alcohol consumption, alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma glutamyl transpeptidase (GGT), total cholesterol, high-density lipoprotein (HDL), triglyceride, glycated hemoglobin (HbA1c), fasting blood glucose (FBG), systolic blood pressure, and diastolic blood pressure. Fatty liver was diagnosed by the findings on abdominal ultrasonography performed by a trained technician. In the follow-up period, fasting blood glucose or HbA1c was used to diagnose diabetes.

Statistical analysis

Study data in the two groups were presented as the mean±standard deviation (SD), and data counts were presented...
as numerical values and percentages. Univariate analysis and multivariate analysis were performed to identify the variables that were associated with incident type 2 diabetes mellitus. The variables with a P-value <0.05 in multivariate analysis were used to construct the predictive model using 75% of the total sample size. The remaining 25% was used to verify the model. Statistical analysis was performed using EmpowerStats version 2019 epidemiology software (www.empowerstats.com) and R software. P<0.05 was considered to represent statistical significance.

Results

Clinical and demographic data

From 2004 to 2015, a total of 11,113 patients who met the study inclusion and exclusion criteria were included in the population study. There were 8,296 cases in the training cohort and 2,817 cases in the validation cohort. The mean age of the training group and the validation group were 43.52±8.50 years and 43.40±8.45 years, respectively. The female to male ratio between the two groups was 3762/4534 and 1229/1588, respectively. The mean BMI of the training group and the validation group were 22.13±3.11 kg/m² and 22.18±3.08 kg/m², respectively (Table 1).

Univariate analysis and multivariate Cox regression analysis

Univariate analysis showed that the variables of age, gender, body mass index (BMI), waist circumference, smoking, fatty liver, alcohol consumption, fasting blood glucose (FBG), systolic blood pressure, diastolic blood pressure, ALT, triglyceride, HbA1c, AST, exercise, GGT and HDL were associated with incident diabetes. Multivariate Cox regression analysis showed
that six variables of age, gender, waist circumference, smoking, fatty liver, fasting blood glucose, and HbA1c were associated with incident diabetes. The hazard ratio (HR) values of multivariate Cox regression for age, gender, waist circumference, smoking, fatty liver, fasting blood glucose and HbA1c were 1.04 (1.02–1.05), 0.62 (0.45–0.86), 1.03 (1.00–1.05), 1.76 (1.40–2.21), 2.21 (1.71–2.87), 3.64 (2.88–4.59), and 4.62 (3.66–5.84), respectively (Table 2).

**Table 2.** The results of univariate analysis and multivariate Cox regression analysis.

| Factor                  | Univariate analysis HR (95% CI), P-value | Multivariate analysis HR (95% CI), P-value |
|-------------------------|------------------------------------------|-------------------------------------------|
| Age                     | 1.06 (1.04–1.07), <0.001                 | 1.04 (1.02–1.05), <0.001                  |
| Gender                  |                                          |                                           |
| Female                  | Reference                                | Reference                                |
| Man                     | 2.51 (1.98–3.20), <0.001                 | 0.62 (0.45–0.86), <0.004                  |
| BMI                     | 1.24 (1.21–1.27), <0.001                 | 1.05 (0.98–1.12), 0.139                   |
| Waist circumference     | 1.09 (1.08–1.10), <0.001                 | 1.03 (1.00–1.05), 0.037                   |
| Smoking                 |                                          |                                           |
| No                      | Reference                                | Reference                                |
| Yes                     | 2.22 (1.80–2.73), <0.001                 | 1.76 (1.40–2.21), <0.001                  |
| Fatty liver             |                                          |                                           |
| No                      | Reference                                | Reference                                |
| Yes                     | 7.01 (5.70–8.62), <0.001                 | 2.21 (1.71–2.87), <0.001                  |
| Alcohol consumption     | 1.00 (1.00–1.00), 0.001                  | 1.00 (1.00–1.00), 0.992                  |
| Fasting blood glucose   |                                          |                                           |
| <5.6 mmol/L             | Reference                                | Reference                                |
| ≥5.6 mmol/L and <6.1 mmol/L | 8.98 (7.29–11.07), <0.001     | 3.64 (2.88–4.59), <0.001                  |
| Systolic blood pressure | 1.03 (1.03–1.04), <0.001                 | 1.00 (0.98–1.01), 0.630                   |
| Diastolic blood pressure| 1.05 (1.04–1.06), <0.001                 | 1.01 (0.98–1.03), 0.682                   |
| ALT                     | 1.01 (1.01–1.01), <0.001                 | 1.01 (1.00–1.02), 0.272                   |
| Triglyceride            | 1.79 (1.67–1.91), <0.001                 | 1.14 (1.00–1.30), 0.058                   |
| HbA1c                   |                                          |                                           |
| <5.7%                   | Reference                                | Reference                                |
| ≥5.7% and <6.5%         | 12.13 (9.80–15.03), <0.001               | 4.62 (3.66–5.84), <0.001                  |
| AST                     | 1.01 (1.01–1.01), <0.001                 | 1.00 (0.98–1.01), 0.814                   |
| Exercise                |                                          |                                           |
| No                      | Reference                                | Reference                                |
| Yes                     | 0.76 (0.57–1.02), 0.069                  | 0.99 (0.74–1.34), 0.963                   |
| GGT                     | 1.01 (1.01–1.01), <0.001                 | 1.00 (1.00–1.01), 0.167                   |
| HDL                     | 0.15 (0.11–0.21), <0.001                 | 0.68 (0.46–1.02), 0.060                   |

BMI – body mass index; FBG – fasting blood glucose; ALT – alanine aminotransferase; AST – aspartate aminotransferase; GGT – gamma glutamyl transpeptidase; HDL – high-density lipoprotein; HbA1c – glycated hemoglobin.

**Construction of the predictive model**

The six variables of age, gender, waist circumference, smoking, fatty liver, fasting blood glucose, and HbA1c were used to construct the predictive model and nomograms with 75% of the total sample. Finally, two predictive models were obtained. In model 1, gender was not associated with incident diabetes. In model 2, all of the variables of age, waist circumference, smoking, fatty liver, fasting blood glucose, and HbA1c were associated with incident diabetes. Therefore, model 2 was selected as the predictive model and was used to construct the
The hazard ratio (HR) of the area under the receiver operating characteristic (ROC) curve (AUC) of the predictive model in the training cohort was 0.87 (95% CI, 0.85–0.89), and the specificity and sensitivity of the predictive model were 0.79 and 0.82, respectively (Table 3, Figure 1).

The predictive model was verified with the remaining 25% samples. The predictive model had high accuracy in the validation cohort. The HR of the AUC of the ROC curve of the predictive model in the validation cohort was 0.87 (95% CI, 0.86–0.90), and the specificity and sensitivity of the predictive model were 0.73 and 0.87, respectively (Table 3, Figure 2). A nomogram was developed based on the predictive model (Table 3, Figure 2).

The aim of this retrospective population study was to develop a predictive model of prediabetes and incident type 2 diabetes mellitus using data from 2004 to 2015 from the DRYAD Japanese hospital database [9]. The six variables identified included age, waist circumference, smoking history, the presence of fatty liver, fasting blood glucose (FBG), and the glycated hemoglobin (HbA1c) level. The predictive model developed had a high level of accuracy in the training model and the validation model, with the hazard ratio (HR) of the area under the receiver operating characteristic (ROC) curve (AUC) of 0.87 (0.85–0.89) and 0.87 (0.86–0.90), respectively.

In the present study, age was closely associated with the presence of incident diabetes, which is supported by previous studies. The findings from the 2017 Global Burden of Diseases study showed that the incidence of type 2 diabetes mellitus in China was closely related to age, with the age-standardized incidence rate increasing with age.
incidence rate rising by 0.92% (95% CI, 0.6–1.3%) for men and by 0.69% (95% CI, 0.3–1.0%) for women [3]. This finding may be explained by the function of the pancreatic β cells, which decrease with age, resulting in reduced insulin production [11]. Also, there has been increasing obesity in the elderly, especially abdominal obesity [12,13]. With age, there is a reduced number of insulin receptors on the fat cell membrane, and the binding affinity with insulin is reduced, resulting in reduced insulin sensitivity that leads to insulin resistance [14,15]. Physical activity in the elderly is also reduced, which is associated with impaired glucose tolerance with age [16].

Previous studies have shown that body mass index (BMI) and waist circumference are both correlated with incident type 2 diabetes mellitus [17]. However, when compared with BMI, waist circumference has a better predictive value for incident type 2 diabetes mellitus [18], which was also consistent with the findings from the present study. When waist circumference and BMI were both included in the multivariate regression analysis, this study found that waist circumference was associated with incident diabetes, but BMI was not. Therefore, waist circumference was used to construct the predictive model. Waist circumference is an indicator of visceral fat, while the BMI is an indicator of systemic fat. Visceral fat is a recognized risk factor for metabolic syndrome and is associated with the secretion of adipocytokines and inflammatory cytokines, which are associated with an increased incidence of type 2 diabetes mellitus [19,20].

In 2014, data from the InterAct Consortium showed that past smoking or current smoking were independent risk factors for diabetes mellitus [21]. The results showed that among men,
the risk of diabetes among former smokers and current smokers was 40% and 43% higher than that of non-smokers [21]. The increased risk in smokers was not affected by age, education level, region, exercise, and intake of alcohol, coffee, and meat [21]. The possible mechanism was that nicotine in tobacco resulted in vasoconstriction, stimulated adrenaline secretion, and lead to the fluctuation of blood pressure and blood glucose levels [22]. Also, smoking results in impaired insulin function and in insulin resistance [21,22].

Fatty liver, or hepatic steatosis, is a recognized risk factor for type 2 diabetes mellitus and occurs in about 30% of patients [23]. Patients with fatty liver can be asymptomatic, and physicians may be unaware of the association with diabetes mellitus, and patients are not routinely screened for fatty. In 2017, a study by Börjekström et al. in Sweden supported that the incidence of type 2 diabetes mellitus was significantly increased in patients with fatty liver [24]. Fatty liver is associated with impaired glycogen metabolism that results in hyperglycemia [25].

Fasting blood glucose (FBG) levels reflect basal insulin secretion levels and function [26]. Increased fasting blood glucose levels are associated with an increased risk of developing diabetes. Glycated hemoglobin (HbA1c) levels not only reflect blood sugar at a point in time but also reflect the average blood sugar from the past three months [27]. The higher the blood sugar, the higher the HbA1c. In 2018, Ruijgrok et al. reported the findings from a population study of 1,349 patients aged 50–75 years without diabetes at baseline in 1989, which showed that increased fasting blood glucose and HbA1c were closely associated with incident diabetes [28].

In the present study, six variables were identified and included in the predictive model of type 2 diabetes mellitus, which included age, smoking history, fatty liver, waist circumference, fasting blood glucose, and HbA1c. The model was stable with high predictive value, which was supported by the AUC of the training cohort and the validation cohort, which were 0.87 (95% CI, 0.85–0.89) and 0.87 (95% CI, 0.86–0.90), respectively. This model may be of clinical value and was supported by a nomogram. The model and the six-value nomogram may be used to evaluate the risk of incident type 2 diabetes mellitus for individuals. The nomogram includes a range of values and outcomes with points representing scores on the vertical line or linear predictor. This model may be used to screen patients with a high risk of type 2 diabetes mellitus at an early stage to result in lifestyle interventions, such as dietary changes, weight loss, and exercise to delay incident type 2 diabetes mellitus in the otherwise healthy population.

This study had several limitations. Retrospective data were used to construct the model and to validate the model, which may have resulted in bias. The model requires testing in future prospective studies to verify the predictive value.

Conclusions

This retrospective population study was conducted to develop a predictive model of prediabetes and incident type 2 diabetes mellitus using data from 2004 to 2015 from the DRYAD Japanese hospital database. A prognostic model based on six variables was predictive for incident type 2 diabetes mellitus and prediabetes in a healthy population in Japan. The variables included age, waist circumference, smoking history, the presence of fatty liver, fasting blood glucose (FBG), and the glycated hemoglobin (HbA1c) level.

Acknowledgments

The authors thank Dr. Masahide Hamaguchi for providing the data from the DRYAD database.

Conflict of interest

None.

References:

1. Hu S, Han M, Rezaei A et al: L-arginine modulates glucose and lipid metabolism in obesity and diabetes. Curr Protein Pept Sci, 2017; 18(6): 599–608
2. Nazir MA, Al Ghamdi L, Al Kadi M et al: The burden of diabetes, its oral complications and their prevention and management. Open Access Maced J Med Sci, 2018; 6(8): 1545–53
3. GBD 2017 DALYs and HALE Collaborators: Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. Lancet, 2018; 392(10159): 1859–922
4. Mohamed SF, Mwangi M, Mutua MK et al: Prevalence and factors associated with pre-diabetes and diabetes mellitus in Kenya: Results from a national survey. BMC Public Health, 2018; 18: 1215
5. Perreault L, Farch K: Approaching prediabetes. J Diabetes Complications, 2014; 28(2): 226–33
6. Goldberg RB, Mather K: Targeting the consequences of the metabolic syndrome in the Diabetes Prevention Program. Arterioscler Thromb Vasc Biol, 2012; 32(9): 2077–90
7. Tabak AG, Herder C, Rathmann W et al: Prediabetes: A high-risk state for diabetes development. Lancet, 2012; 379(9833): 2279–90
8. Ram J, Selvam S, Snehahita C et al: Improvement in diet habits, independent of physical activity helps to reduce incident diabetes among prediabetic Asian Indian men. Diabetes Res Clin Practice, 2014; 106(3): 491–95
9. Okamura T, Hashimoto Y, Hamaguchi M et al: Ectopic fat obesity presents the greatest risk for incident type 2 diabetes: A population-based longitudinal study. Int J Obes (Lond), 2019; 43(1): 139–48
10. Petersmann A, Nauck M, Müller-Wieland D et al: Definition, classification and diagnosis of diabetes mellitus. Exp Clin Endocrinol Diabetes, 2018; 126(7): 406–10
11. Tamura Y, Izumiyama-Shimomura N, Kimbara Y et al: Telomere attrition in obesity and diabetes. Curr Protein Pept Sci, 2017; 18(6): 599–608
12. Petersmann A, Nauck M, Trautwein C et al: Definition, classification, and diagnosis of diabetes mellitus and prediabetes. Exp Clin Endocrinol Diabetes, 2018; 126(7): 406–10
13. Tamura Y, Izumiyama-Shimomura N, Kimbara Y et al: Telomere attrition in obesity and diabetes. Curr Protein Pept Sci, 2017; 18(6): 599–608

12. Boateng GO, Adams EA, Odei Boateng M et al: Obesity and the burden of health risks among the elderly in Ghana: A population study. PLoS One, 2017; 12(11): e0186947

13. Jésus P, Guerchet M, Pilleron S et al: Undernutrition and obesity among elderly people living in two cities of developing countries: Prevalence and associated factors in the EDAC study. Clin Nutr ESPEN, 2017; 21: 40–50

14. de Mutsert R, Gast K, Widyra R et al: Associations of abdominal subcutaneous and visceral fat with insulin resistance and secretion differ between men and women: The Netherlands Epidemiology of Obesity Study. Metab Syndr Relat Disord, 2018; 16(1): 54–63

15. Fang P, Yu M, Zhang L et al: Baicalin against obesity and insulin resistance through activation of AKT/AS160/GLUT4 pathway. Mol Cell Endocrinol, 2018; 448: 77–86

16. Naufahu J, Elliott B, Markiv A et al: High-intensity exercise decreases IP6K1 muscle content and improves insulin sensitivity (S2) in glucose-intolerant individuals. J Clin Endocrinol Metab, 2018; 103(4): 1479–90

17. Foghsgaard S, Andreasen C, Vedtofte L et al: Nonalcoholic fatty liver disease is prevalent in women with prior gestational diabetes mellitus and independently associated with insulin resistance and waist circumference. Diabetes Care, 2017; 40(1): 109–16

18. Hou X, Chen S, Hu G et al: Stronger associations of waist circumference and waist-to-height ratio with diabetes than BMI in Chinese adults. Diabetes Res Clin Pract, 2019; 147: 9–18

19. Karatzi K, Moschonis G, Polychronopoulou MC et al: Cutoff points of waist circumference and trunk and visceral fat for identifying children with elevated inflammation markers and adipokines: The Healthy Growth Study. Nutrition, 2016; 32(10): 1063–67

20. Saito T, Murata M, Otani T et al: Association of subcutaneous and visceral fat mass with serum concentrations of adipokines in subjects with type 2 diabetes mellitus. Endocr J, 2012; 59(1): 39–45

21. InterAct Consortium, Spijkerman AM, van der A DL, Nilsson PM et al: Smoking and long-term risk of type 2 diabetes: The EPIC-InterAct study in European populations. Diabetes Care, 2014; 37(12): 3164–71

22. Scott AL, Pranckevicius NA, Nurse CA, Scott GR: Regulation of catecholamine release from the adrenal medulla is altered in deer mice (Peromyscus maniculatus) native to high altitudes. Am J Physiol Regul Integr Comp Physiol, 2019; 317(3): R407–17

23. Bølling S: The epidemiology of non-alcoholic fatty liver disease. Liver Int, 2017; 37(Suppl. 1): 81–84

24. Björkström K, Stål P, Hultcrantz R, Hagström H: Histologic scores for fat and fibrosis associate with development of type 2 diabetes in patients with nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol, 2017; 15(9): 1461–68

25. Irimia IM, Meyers CM, Segvich DM et al: Lack of liver glycogen causes hepatic insulin resistance and steatosis in mice. J Biol Chem, 2017; 292(25): 10455–64

26. Tafayli H, Lee S, Arslanian S: Declining beta-cell function relative to insulin sensitivity with increasing fasting glucose levels in the nondiabetic range in children. Diabetes Care, 2010; 33(9): 2024–30

27. Wilding J, Godec T, Khunti K et al: Changes in HbA1c and weight, and treatment persistence, over the 18 months following initiation of second-line therapy in patients with type 2 diabetes: Results from the United Kingdom Clinical Practice Research Datalink. BMC Med, 2018; 16(1): 116

28. Ruijgrok C, Dekker IM, Beulens JW et al: Size and shape of the associations of glucose, HbA1c, insulin and HOMA-IR with incident type 2 diabetes. The Hoorn Study. Diabetologia, 2018; 61(1): 93–100