BMJ Open

Development of risk prediction equations for 5-year diabetes incidence using Japanese health check-up data: a retrospective cohort study

Shin Kawasoe ,[1] Takuro Kubozono,[1] Satoko Ojima,[1] Satoshi Yamaguchi,[1] Koji Higuchi,[1] Hironori Miyahara,[2] Koichi Tokushige,[2] Masaaki Miyata,[3] Mitsuru Ohishi[1]

To cite: Kawasoe S, Kubozono T, Ojima S, *et al*. Development of risk prediction equations for 5-year diabetes incidence using Japanese health check-up data: a retrospective cohort study. *BMJ Open* 2025;**15**:e097005. doi:10.1136/bmjopen-2024-097005

► Prepublication history for this paper is available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2024-097005).

Received 14 December 2024
Accepted 08 May 2025

[1]Department of Cardiovascular Medicine and Hypertension, Graduate School of Medical and Dental Sciences, Kagoshima University, Kagoshima, Japan
[2]Kagoshima Kouseiren Hospital, Kagoshima, Japan
[3]School of Health Sciences, Faculty of Medicine, Kagoshima University, Kagoshima, Japan

Correspondence to
Dr Takuro Kubozono;
kubozono@cepp.ne.jp

ABSTRACT

Objectives This study aimed to develop risk prediction equations for the 5-year incidence of diabetes among the Japanese population using health check-up data. We hypothesised that demographic and laboratory data from health check-ups could predict diabetes onset with high accuracy.

Design Retrospective cohort study.

Setting Data from a health examination in Japan between 2008 and 2016.

Participants Data were analysed from 31 084 participants aged 30–69 years. The presence of baseline diabetes and endocrine disease was included in the exclusion criteria, as were participants with missing data for the analysis. The study population was randomly divided into derivation and validation cohorts in a 1:1 ratio.

Primary outcome measures The primary outcome was the incidence of diabetes at the 5-year follow-up, defined as a fasting blood glucose level ≥126 mg/dL, glycosylated haemoglobin A1c (National Glycohemoglobin Standardization Program (NGSP)) ≥6.5%, or initiation of diabetes treatment. Predictor variables included age, sex, body mass index, blood pressure, underlying diseases, lifestyle factors and laboratory measurements. The primary measure was the area under the receiver operating characteristic curve (AUC) for the predictive equations.

Results In the derivation cohort, diabetes incidence was 5.0%. The prediction equation incorporating age, sex, body mass index, fasting blood glucose and glycosylated haemoglobin A1c showed good discriminatory ability with an AUC of 0.89, sensitivity of 0.81 and specificity of 0.81 in the validation cohort.

Conclusions The equation with laboratory measures effectively predicted the 5-year diabetes risk in the general Japanese population. It has potential clinical utility for identifying individuals at high risk of diabetes and guiding preventive interventions.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This study used a large sample size of over 30 000 participants, enhancing the statistical power and reliability of the risk prediction model for diabetes incidence.
⇒ The use of routine health check-up data allows for easy integration into existing clinical workflows, potentially increasing the applicability and practicality of the predictive model.
⇒ Retrospective cohort design limits control over data collection, potentially introducing unmeasured confounding factors.
⇒ The study was conducted within a single healthcare system in Japan, which may limit the generalisability of the findings to other populations or healthcare settings.

INTRODUCTION

Diabetes is a major health problem worldwide. According to the latest epidemiological data, the number of adults with diabetes globally reached approximately 537 million in 2021.[1] In Japan, a 2023 survey reported that the total number of patients currently receiving treatment for diabetes was 5.52 million.[2] These figures highlight the urgent need for effective prevention and management strategies to address the growing burden of diabetes both globally and in Japan. Complications of diabetes include cardiovascular disease, kidney disease, blindness, neuropathy and lower limb amputation, all of which significantly reduce patients' quality of life and increase the risk of death.[3 4] These complications not only affect individual health but also contribute to a substantial economic burden due to increased healthcare costs and loss of productivity.[5] Identifying individuals at high risk of developing diabetes and providing preventive interventions are important not only for maintaining individual health but also from a healthcare economic perspective.[6 7]

In some cases of diabetes, autoimmune mechanisms and genetic factors contribute significantly to its pathogenesis.[8 9] However, in many cases, lifestyle has a significant

impact on its development.[10] Obesity, physical inactivity, unhealthy diet and smoking are all well-documented risk factors for the development of type 2 diabetes.[11] To assess the risk of developing diabetes, we need to take a comprehensive approach that focuses on multiple risk factors, not just on a single risk factor.[12] In Japan, annual health check-ups to prevent lifestyle-related diseases have been recommended for people over 40 years of age since 2008, during which a wide range of anthropometric indices, lifestyle-related questionnaires and basic blood data are collected. These check-ups provide a unique opportunity for healthcare providers to identify individuals at high risk of developing diseases.

It would be clinically useful to create risk models for predicting diabetes using variables measured during these health check-ups. Various risk models have been developed to predict the onset of diabetes, incorporating factors such as age, body mass index (BMI), family history, blood pressure (BP) and lipid profile.[13 14] However, the applicability of these models to the Japanese population remains uncertain owing to differences in genetic, environmental and lifestyle factors. Several studies have proposed diabetes prediction models for the Japanese population, demonstrating reasonable predictive accuracy.[13 15–19] However, some models lack rigorous validation, raising concerns about their generalisability.

The purpose of this study is to develop predictive equations specific to the Japanese population that calculate the risk of developing diabetes 5 years later based on individual health examination information. Additionally, this study emphasises the importance of model calibration to enhance reliability and clinical utility. These equations may be useful for lifestyle interventions and careful monitoring at the individual level and for the development of preventive methods such as medical resource allocation at the population level.

METHODS

Study population

We used data from individuals aged 30–69 years who had annual health check-ups at Kagoshima Kouseiren Hospital between April 2008 and March 2016. The majority of participants had check-ups annually, as recommended by Japanese health welfare policy. First, the first visit during the period was defined as baseline data, and second, data from 5 years after the baseline were picked up to obtain outcomes. To develop models for predicting future diabetes incidence, we excluded individuals who already had diabetes at the baseline time point. Diabetes at baseline was defined based solely on clinical and laboratory data collected at the initial health check-up, not based on time-series information or medical history prior to the baseline year. Specifically, individuals were excluded if they met any of the following criteria at baseline: (1) use of antidiabetic medications (oral antihyperglycaemic drugs, insulin or GLP-1 receptor agonists), (2) fasting blood glucose (FBG) ≥126 mg/dL or (3) HbA1c (National Glycohemoglobin Standardization Program (NGSP)) ≥6.5%. Only participants who did not meet any of these criteria at their first health check-up were included in the analysis cohort. We also excluded individuals who were being treated for or had a history of endocrine disorders, which could have a direct impact on blood glucose levels. We also excluded those with missing data used in the analysis. Participants were randomly divided equally into two groups, one as a derivation cohort to develop equations to predict diabetes 5 years later and the other as a validation cohort to assess the validity of the equation obtained from the derivation cohort.

The data were anonymised, and all participants were given the option to opt out of the study. This study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Ethics Committees of the Graduate School of Medical and Dental Sciences, Kagoshima University.

Patient and public involvement

Patients and the public were not directly involved in the design, conduct or analysis of this study. However, the study was designed to address outcomes that are relevant to patients and the findings will be disseminated to participants via newsletters and website.

Risk factors and definition of outcome

Height and weight were measured using reliable measuring devices, and BMI was calculated as weight (kg) divided by height squared (m^2). BP was measured after participants had been sitting quietly for 5 min. Self-administered questionnaires were used to collect information on smoking, alcohol consumption, habitual exercise and family diabetes history, which was then classified as follows: smokers (currently smoking) or non-smokers (never smoked or smoked in the past), chance drinkers (≤10 days per month) or usual drinkers (>10 days per month), habitual exercisers (>30 min per day) and those with a family history of diabetes. Information on the presence of hypertension, dyslipidaemia and hyperuricaemia, as well as medications for these conditions, was also collected with a self-administered questionnaire. Blood samples were taken after an overnight fast. Serum triglyceride, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, blood glucose, uric acid and creatinine levels were measured using standard laboratory methods. The diseases were defined as follows: hypertension, use of antihypertensive agents or systolic blood pressure ≥140 mm Hg, diastolic blood pressure ≥90 mm Hg according to the 2019 Japanese Society of Hypertension guidelines; dyslipidaemia, use of lipid-lowering agents, serum triglycerides ≥150 mg/dL, low-density lipoprotein cholesterol ≥140 mg/dL or high-density lipoprotein cholesterol <40 mg/dL; and hyperuricaemia, use of uric acid lowering agents, serum uric acid >7.0 mg/dL.

The outcome was defined as diabetes after 5 years. In some cases, HbA1c levels were measured using the

Japanese Diabetes Society (JDS) guidelines. In research and clinical contexts, conversion of HbA1c values from the JDS reference to the NGSP reference is often necessary to standardise measurements and compare data internationally. The commonly used conversion formula is as follows[20]:

$$\text{HbA1c (NGSP)} = 1.02\text{HbA1c (JDS)} + 0.25$$

This conversion equation was established based on comparative studies and calibration methodologies to align the JDS measurements with the NGSP reference, ensuring consistency and accuracy in HbA1c reporting. Diabetes was defined as any of the following: (1) use of antidiabetic agents (oral antihyperglycaemic drugs, insulin and GLP-1 receptor agonists) and (2) HbA1c (NGSP) >6.5%.

Statistical analysis

With the exception of blood glucose and lipid levels, which were expressed as median (first quartile, third quartile), all continuous variables (age, BMI, BP and laboratory tests) were expressed as mean±SD. Proportions (percentages) were used to express categorical variables, such as illnesses and lifestyle factors. Student's unpaired t-test, Wilcoxon test and χ^2 test were used to examine differences between the derivation and validation cohorts for normally distributed continuous variables, skewed-distribution continuous variables and categorical data, respectively.

Univariate logistic regression analysis was performed in the derivation cohort to estimate the ORs and 95% CIs for the incidence of diabetes after 5 years for each variable. Two different models were considered in the multivariate analysis: one model included eight factors that did not require blood sampling (age, sex, BMI, current smoking, usual alcohol drinking, exercise habits, family history of diabetes and hypertension) as covariates, and the other model was composed of 12 factors that additionally included four factors that required blood sampling (dyslipidaemia, hyperuricaemia, FBG and HbA1c). Factors significantly associated with diabetes after 5 years were extracted, and two risk prediction equations were developed using the beta coefficient for each factor.

The two risk prediction equations were then applied to the validation cohort to calculate the area under the curve (AUC) using receiver operating characteristic (ROC) curve analysis to determine the cut-off values, sensitivity and specificity. Calibration plots were constructed to assess the agreement between the predicted and actual observed probabilities. All statistical analyses were conducted using the JMP Pro V. 16 (SAS Institute, Cary, NC, USA), and calibration was done with R V. 4.2.2 (The R Foundation for Statistical Computing, Vienna, Austria) and the RMS package. Statistical significance was set at $p<0.05$.

RESULTS

Baseline characteristics of the study population

Figure 1 shows a flowchart of the inclusion and exclusion criteria for this study. The study included 33 173 individuals aged 30–69 who had undergone a health check-up during the study period and for whom data were also available 5 years later. A total of 2089 participants were excluded (baseline diabetes, 1391; baseline endocrine disease, 176; and participants with missing variables, 552). Finally, we used the data of 31 084 participants (age, 54.9±10.0 years; male, 50.1%) for analysis.

Table 1 shows the baseline characteristics for the derivation and validation cohorts. The mean age of the derivation cohort was 54.9±10.0 years and 49.8% were men, while the mean age of the validation cohort was 54.8±10.1 years and 50.3% were men. There were no significant differences in baseline characteristics between the two cohorts. After 5 years (median, 5.0 years; first quartile, 4.7 years; third quartile, 5.6 years), we found 770 (5.0%) and 778 (5.0%) cases of diabetes in the derivation and validation cohorts, respectively.

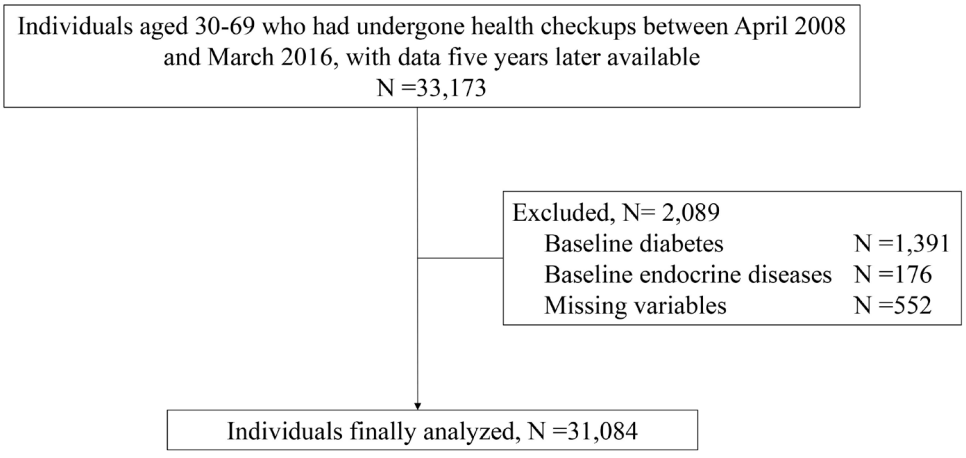

Figure 1 Flowchart of the study population. A total of 2089 individuals were excluded, and data from 31 084 individuals were used in the final analysis.

Table 1 Baseline characteristics of study population in the derivation and validation cohorts

	Derivation cohort	Validation cohort	P value
	n=15 542	*n*=15 542	
Age, years	54.9±10.0	54.8±10.1	0.230
Men, *n* (%)	7737 (49.8)	7817 (50.3)	0.364
BMI, kg/m^2	23.3±3.3	23.2±3.2	0.238
SBP, mm Hg	124.4±18.7	124.1±18.9	0.188
DBP, mm Hg	76.8±11.5	76.5±11.3	0.075
Hypertension, *n* (%)	4291 (27.6)	4252 (27.4)	0.620
Dyslipidaemia, *n* (%)	6894 (44.4)	6934 (44.7)	0.644
Hyperuricaemia, *n* (%)	2183 (14.1)	2103 (13.6)	0.190
Current smoking, *n* (%)	1536 (9.9)	1568 (10.1)	0.545
Habitual drinking, *n* (%)	3059 (19.7)	3019 (19.4)	0.567
Habitual exercise, *n* (%)	3229 (31.6)	3335 (32.0)	0.603
Family history, *n* (%)	1229 (7.9)	1150 (7.4)	0.092
Triglyceride, mg/dL	90(65, 130)	90(64, 132)	0.998
LDL-C, mg/dL	122.1±30.8	122.9±31.2	0.067
HDL-C, mg/dL	60.7±14.9	60.8±14.9	0.719
BG, mg/dL	96(90, 103)	95(90, 103)	0.214
HbA1c (NGSP), %	5.5±0.3	5.5±0.3	0.255
UA, mg/dL	5.3±1.5	5.3±1.5	0.323

Baseline characteristics of study population in the derivation and validation cohorts. Continuous variables are expressed as mean±SD, except for triglyceride and blood glucose levels, which are expressed as median (first quartile, third quartile). Categorical variables, including cardiovascular risk factors and lifestyle variables, are expressed as number of subjects and proportions (percentages). Differences between the derivation and validation cohorts for normally distributed continuous variables, skewed-distribution continuous variables, and categorical variables were analysed using the Student's unpaired t-test, Wilcoxon test and χ^2 test, respectively.
BG, blood pressure; BMI, body mass index; DBP, diastolic blood pressure; HbA1c, glycosylated haemoglobin; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; NGSP, National Glycohemoglobin Standardization Program; SBP, systolic blood pressure; UA, uric acid.

Association between individual risk factors and diabetes after 5 years

We performed a logistic regression analysis on the population in the derivation cohort. The associations between diabetes after 5 years and possible risk factors are shown in table 2. In the univariable model, older age, male sex, greater BMI, current smoking, habitual alcohol drinking, family history of diabetes, hypertension, dyslipidaemia, hyperuricaemia, higher FBG and higher HbA1c were positively associated with the incidence of diabetes.

Development of diabetes prediction equations

The results of multivariable logistic regression analysis using the 12 factors that required blood sampling are shown in table 2. Male sex (OR 1.33, p=0.028), greater BMI (OR 1.09, p<0.001), greater FBG (OR 1.12, p<0.001) and greater HbA1c levels (OR 12.77, p<0.001) were positively associated with the risk of diabetes, whereas older age (OR 0.98, p=0.015) was negatively associated with the risk of diabetes. We developed the following equations using five factors that were significantly associated with the outcome: age (continuous, years), sex (categorical: male=1; female=0), BMI (continuous: kg/m^2), FBG (continuous: mg/dL) and HbA1c (NGSP) (continuous: %).

$$p = 1/(1+\exp(-(-30.3726 - 0.0184 \times \text{age} + 0.2876 \times \text{sex} + 0.0905 \times \text{BMI} + 0.1101 \times \text{FBG} + 2.5471 \times \text{HbA1c})))$$

The median probability obtained by applying the equation to the derivation cohort was 0.011 (interquartile 0.004, 0.037), the AUC of the ROC curve for diabetes after 5 years was 0.90, the sensitivity was 0.84, and the specificity was 0.82 at a cut-off value of 0.048. When the risk equation was applied to the validation cohort, the median probability was 0.011 (interquartile 0.004, 0.036), the AUC of the ROC curve for diabetes after 5 years was 0.89, and the sensitivity was 0.81 and specificity was 0.81 at a cut-off value of 0.044 (figure 2A). Figure 3A shows the calibration plot obtained when the risk equation was applied to the validation cohort, which confirmed good calibration. There was a tendency to slightly underestimate the incidence of diabetes after 5 years for scores between 0.1 and 0.5. Conversely, there was an overestimation for scores

Table 2 The associations between diabetes after 5 years and risk factors

	Univariable		Multivariable					
			Without blood sampling			With blood sampling		
Risk factors	OR (95% CI)	P value	OR (95% CI)	P value	β coefficient	OR (95% CI)	P value	β coefficient
Age, years	1.02 (1.01–1.02)	<0.001	1.02 (1.01–1.03)	0.002	0.0200	0.98 (0.97–0.99)	0.015	−0.0184
Male sex	1.75 (1.51–2.04)	<0.001	1.66 (1.34–2.07)	<0.001	0.5096	1.33 (1.03–1.72)	0.028	0.2876
BMI, kg/m^2	1.16 (1.14–1.18)	<0.001	1.16 (1.13–1.20)	<0.001	0.1522	1.09 (1.06–1.13)	<0.001	0.0905
Current smoking	1.82 (1.49–2.21)	<0.001	1.26 (0.72–2.20)	0.421		0.90 (0.49–1.65)	0.736	
Habitual drinking	1.37 (1.16–1.62)	<0.001	0.88 (0.56–1.40)	0.595		0.61 (0.38–1.00)	0.051	
Habitual exercise	1.21 (0.98–1.50)	0.083	1.08 (0.87–1.35)	0.489		1.13 (0.88–1.43)	0.342	
Family history	1.71 (1.37–2.14)	<0.001	0.71 (0.33–1.55)	0.391		0.45 (0.20–1.02)	0.056	
Hypertension	2.17 (1.87–2.51)	<0.001	1.53 (1.23–1.90)	<0.001	0.4244	1.15 (0.90–1.47)	0.263	
Dyslipidaemia	1.88 (1.62–2.18)	<0.001	–	–	–	1.17 (0.93–1.48)	0.189	
Hyperuricaemia	1.68 (1.40–2.01)	<0.001	–	–	–	0.96 (0.70–1.31)	0.780	
FBS, mg/dL	1.15 (1.14–1.16)	<0.001	–	–	–	1.12 (1.10–1.13)	<0.001	0.1101
HbA1c (NGSP), %	56.90 (43.88–73.78)	<0.001	–	–	–	12.77 (8.58–19.01)	<0.001	2.5471

The ORs and 95% CIs for 5 year diabetes incidence after 5 years using univariable and multivariable logistic regression analysis were shown. In multivariable model, the ORs were adjusted for the following variables: age, sex, body mass index and blood pressure categories, current smoking, habitual alcohol drinking, habitual exercise, family history, diabetes, dyslipidaemia and hyperuricaemia.
BMI, body mass index; BP, blood pressure; FBS, fasting blood sugar; NGSP, National Glycohemoglobin Standardization Program.

above 0.5 (although this was less than 2% of the total population).

Results of multivariate logistic regression analysis using eight factors that did not require blood sampling are also presented in table 2. Older age, male, greater BMI and hypertension were positively associated with diabetes risk. The following equation was developed using these four factors: age (continuous, years), gender (category: male=1, female=0), BMI (continuous: kg/m^2) and hypertension (category: hypertension=1, no hypertension=0).

$$p = 1/(1 + \exp(-(8.4968 + 0.0200 \times \text{age} + 0.5096 \times \text{sex} + 0.1522 \times \text{BMI} + 0.4644 \times \text{hypertension})))$$

The ROC curve for diabetes at 5 years obtained by applying this equation to the derivation cohort had an AUC of 0.70, a sensitivity of 0.78, and a specificity of 0.51 in

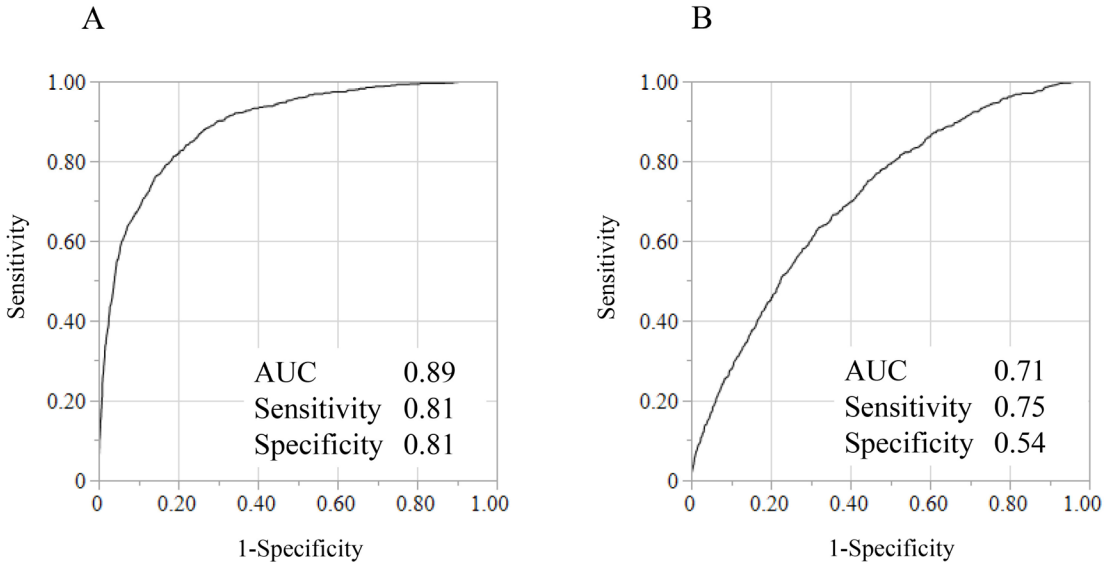

Figure 2 Receiver operating characteristics curves when the risk equations predicting 5 year diabetes incidence are applied to the validation cohort. Receiver operating characteristics curves when the risk equations predicting 5 year diabetes incidence are applied to the validation cohort are shown. (A) The equation that requires blood sampling and (B) The equation that does not require blood sampling. AUC, area under the curve.

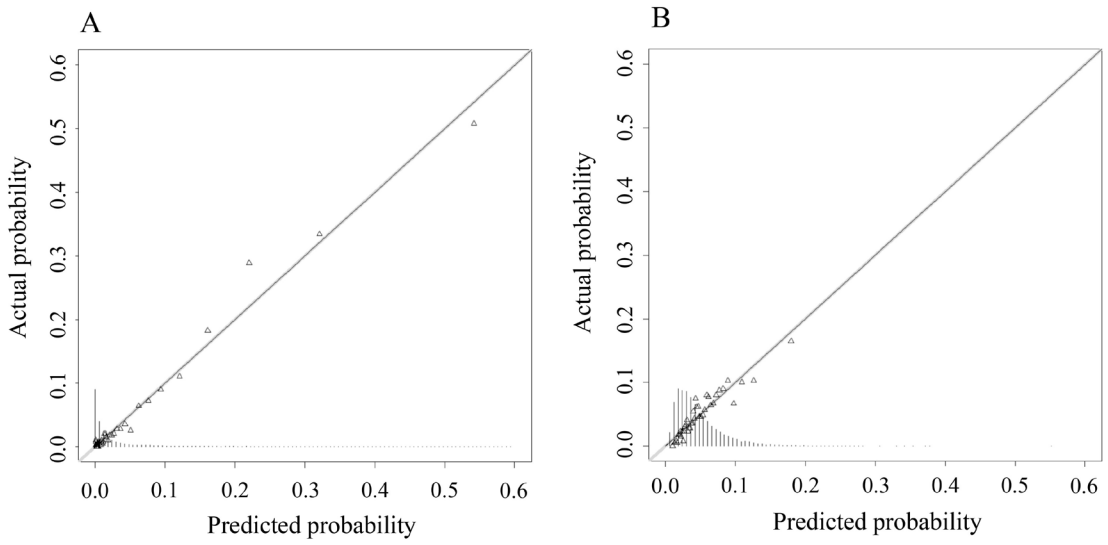

Figure 3 Calibration plots on predicting diabetes incidence after 5 years. Calibration when prediction equations were applied to the validation cohort. (A) The equation that requires blood sampling. (B) The equation that does not require blood sampling. The horizontal axis indicates the predicted probability and the vertical axis indicates the actual probability. The dashed line represents a perfect calibration.

the validation cohort, which had an AUC of 0.71, a sensitivity of 0.75 and specificity of 0.54 (figure 2B). Figure 3B shows the calibration plot, confirming good calibration.

DISCUSSION

We developed equations that predict the incidence of diabetes after 5 years using health check-up data. One equation consisted of five factors requiring blood sampling: age, gender, BMI, FBG and HbA1c, while the other one consisted of four factors not requiring blood sampling: age, gender, BMI and presence of hypertension. The former showed good discrimination performance with an AUC of 0.89 and good calibration, while the latter showed poor discrimination performance with an AUC of 0.71. We developed risk equations composed of simple indicators that are applicable to the Japanese population and are easy to use in clinical practice. The primary goal of diabetes prevention is to maintain the patient's quality of life and improve prognosis by preventing the development of macrovascular and microvascular complications. Since these complications develop over time after the onset of diabetes, it is reasonable and meaningful to predict risk and implement preventive interventions from young to middle age. For these reasons, this study aimed to develop a risk prediction model for a relatively young population aged 30–69 years.

Several risk models have been reported to predict the development of diabetes in various countries and ethnic populations.[21 22] Since the prevalence of diabetes and the contribution of factors involved in its development are not the same for each racial group, the models need to be based on Japanese data to be applicable to the Japanese population. There have been few reports on diabetes prediction scores in Japanese patients. Sasai *et al* followed up with more than 53 000 people aged 40–69 years who underwent physical examinations and created a diabetes risk prediction sheet based on their results. This was the first report on diabetes prediction in Japan, but there has been no discussion on the discriminative power of the scores generated.[15] Heianza *et al* examined the discriminatory power of risk scores based on non-laboratory assessments alone and risk scores, including fasting glucose and HbA1c levels, based on data from 7654 non-diabetic patients aged 40–75 years. They reported that the risk score, including both fasting glucose and HbA1c, showed good discriminative power for 5 year diabetes onset (AUC 0.887, 95% CI 0.871 to 0.903).[16] Nanri *et al* created a score to predict the 3 year incidence of diabetes in more than 37 000 men and women aged 30 years or older and reported that the non-invasive model had an AUC of 0.734, and the invasive model had an AUC of 0.882.[17] In 2018, Hu *et al* reported a model predicting the 7 year risk of type 2 diabetes with AUC 0.73 for the non-invasive model and AUC 0.89 for the invasive model.[18] Regarding the prediction equation, there is only one report on the Japanese population by Xu *et al* in 2023.[19] They developed an equation to predict type 2 diabetes at 5 years in one cohort and applied it to another cohort, achieving an AUC of 0.643 in a non-invasive model and 0.845 in an invasive model including FBG and HbA1c levels. There was no significant advantage of the prediction equation created in this study that included items requiring blood sampling over the results of the external validation by Xu *et al*. However, our study advances the field by providing models with a focus on practical implementation and validated calibration. What is notable in common with these Japanese risk scores is that HbA1c and FBG levels have a very large impact on discriminability. For both risk scores, there were large AUC differences between the invasive and non-invasive models. While age, BMI

and family history are previously noted predictors of the risk of developing diabetes, FBG and HbA1c represent elevated blood glucose levels.[20] Both were considered to have a significant impact on predicting the development of diabetes.

Although the risk prediction equations developed in this study did not exceed the results of previous studies, they did reinforce the results of previous studies and highlighted an important aspect of diabetes prediction in the Japanese population.[23] When comparing the discriminative power of scores without blood sampling for diabetes incidence, the discriminative power was lower in the Japanese than in Westerners.[24] Asians are reported to have a lower insulin secretion capacity than Westerners and tend to develop diabetes from the mild obesity stage.[25 26] It may be more difficult in Asians than in Westerners to adequately predict the incidence of diabetes from anthropometric and non-invasive measurements.[27] Importantly, FBG and HbA1c levels are not known without blood tests, and the results of these studies underscore the importance of regular physical examinations in the Japanese population.[28] Elevated preprandial blood glucose and HbA1c levels indicate advanced glucose intolerance, but the equation might have been different if information on postprandial blood glucose, which indicates earlier glucose intolerance, had been added.[29]

There were some limitations to this study. First, as the data were not collected prospectively, future prospective observational studies should confirm the findings. Second, only individuals who had health check-ups at a single Japanese facility were eligible to participate. Furthermore, because the participants tended to be health-conscious, we were unable to completely eliminate the consequences of selection bias. Third, it was impossible to ascertain if adjusting for risk decreased the likelihood of acquiring diabetes. Fourth, the FBG and HbA1c levels were assessed just during the health check-ups; multiple measures were not taken. Fifth, there were no data on dietary-related substances associated with diabetogenesis, such as minerals and trace elements or detailed data on comorbidities. Lastly, we did not take into account markers of visceral fat such as fat mass or waist circumference.

In conclusion, we created simple and practical prediction equations based on age, sex, BMI, FBG and HbA1c values to predict the 5 year incidence of diabetes in a general Japanese population. Its prediction ability is reasonably good and reproducible. These prediction equations could be a simple and effective tool for identifying those at high risk of developing diabetes in the future and for providing education and counselling for prevention.

Acknowledgements We thank the medical staff at Kagoshima Kouseiren Hospital for their support with data collection.

Contributors TK is responsible for the overall content as guarantor. Research idea and study design: SK; data acquisition: HM and KT; data analysis/interpretation: SK, TK, HY, SM and KH; statistical analysis: SK, SO, SY and TK; supervision or mentorship: KT, MM and MO.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. The Kagoshima University Institutional Review Board and Kagoshima Kouseirin Hospital applies the restriction for public data sharing due to ethical and legal restrictions of the annual health check-up data containing sensitive information and that participant did not consent to public sharing. The deidentified data may be partly available upon ethical approval by request directed to Dr Takuro Kubozono (kubozono@m.kufm.kagoshima-u.ac.jp).

ORCID iD

Shin Kawasoe http://orcid.org/0009-0009-4277-4841

REFERENCES

1 International Diabetes Federation. IDF diabetes atlas, 2021. Available: https://diabetesatlas.org
2 Labour and Welfare, Japan. National Health and Nutrition Survey, 2023. Available: https://www.mhlw.go.jp/toukei/saikin/hw/kanja/23/index.html
3 Harding JL, Pavkov ME, Magliano DJ, *et al*. Global trends in diabetes complications: a review of current evidence. *Diabetologia* 2019;62:3–16.
4 ElSayed NA, Aleppo G, Aroda VR, *et al*. Prevention or delay of diabetes and associated comorbidities: standards of care in diabetes—2023. *Diabetes Care* 2023;46:S41–8.
5 Hsu WC, Araneta MRG, Kanaya AM, *et al*. BMI cut points to identify at-risk Asian Americans for type 2 diabetes screening. *Diabetes Care* 2015;38:150–8.
6 Bommer C, Sagalova V, Heesemann E, *et al*. Global economic burden of diabetes in adults: projections from 2015 to 2030. *Diabetes Care* 2018;41:963–70.
7 Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol* 2018;14:88–98.
8 Ilonen J, Lempainen J, Veijola R. The heterogeneous pathogenesis of type 1 diabetes mellitus. *Nat Rev Endocrinol* 2019;15:635–50.
9 Meigs JB. The genetic epidemiology of type 2 diabetes: opportunities for health translation. *Curr Diab Rep* 2019;19:62.
10 Kolb H, Martin S. Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes. *BMC Med* 2017;15:131.
11 Hu FB, Manson JE, Stampfer MJ, *et al*. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N Engl J Med* 2001;345:790–7.
12 Chatterjee S, Khunti K, Davies MJ. Type 2 diabetes. *Lancet* 2017;389:2239–51.
13 Kengne AP, Beulens JWJ, Peelen LM, *et al*. Non-invasive risk scores for prediction of type 2 diabetes (EPIC-InterAct): a validation of existing models. *Lancet Diabetes Endocrinol* 2014;2:19–29.
14 Tabák AG, Herder C, Rathmann W, *et al*. Prediabetes: a high-risk state for diabetes development. *Lancet* 2012;379:2279–90.
15 Sasai H, Sairenchi T, Irie F, *et al*. Development of a diabetes risk prediction sheet for specific health guidance. *Nihon Koshu Eisei Zasshi* 2008;55:287–94.
16 Heianza Y, Arase Y, Hsieh SD, *et al*. Development of a new scoring system for predicting the 5 year incidence of type 2 diabetes in Japan: the Toranomon Hospital Health Management Center Study 6 (TOPICS 6). *Diabetologia* 2012;55:3213–23.
17 Nanri A, Nakagawa T, Kuwahara K, *et al*. Development of risk score for predicting 3-year incidence of type 2 diabetes: Japan

epidemiology collaboration on occupational health study. *PLoS ONE* 2015;10:e0142779.
18 Hu H, Nakagawa T, Yamamoto S, *et al*. Development and validation of risk models to predict the 7-year risk of type 2 diabetes: The Japan epidemiology collaboration on occupational health study. *J Diabetes Investig* 2018;9:1052–9.
19 Xu Y. Development of a risk prediction equation for type 2 diabetes in Japanese population. *Diabetes Res Clin Pract* 2023;192:110002.
20 Ding L, Xu Y, Liu S, *et al*. Hemoglobin A1c and diagnosis of diabetes. *J Diabetes* 2018;10:365–72.
21 Buijsse B, Simmons RK, Griffin SJ, *et al*. Risk assessment tools for identifying individuals at risk of developing type 2 diabetes. *Epidemiol Rev* 2011;33:46–62.
22 Zhang M, Zhang H, Wang C, *et al*. Development and validation of a risk-score model for type 2 diabetes: a cohort study of a rural adult Chinese population. *PLoS ONE* 2016;11:e0152054.
23 Haneda M, Noda M, Origasa H, *et al*. Japanese clinical practice guideline for diabetes 2016. *Diabetol Int* 2018;9:1–45.
24 Onishi Y, Hayashi T, Sato KK, *et al*. Fasting and post-challenge glucose concentrations are associated with insulin secretion in Japanese subjects. *Diabet Med* 2015;32:66–73.
25 Wu JH, Kadowaki T. Dietary and lifestyle factors associated with risk of type 2 diabetes mellitus in Asian populations: A systematic review. *PLoS ONE* 2017;12.
26 Chan JC, Zhang Y, Ning G, *et al*. Diabetes in Asia and the Pacific: implications for the global epidemic. *Diabetes Care* 2018;41:1462–75.
27 Yoon KH, Lee JH. Epidemic obesity and type 2 diabetes in Asia. *Lancet* 2017;388:2477–86.
28 Kuwata H, Okamura T, Hayakawa T, *et al*. Risk of progression to diabetes among individuals with impaired fasting glucose and/or impaired glucose tolerance in a Japanese population. *J Diabetes Investig* 2015;6:456–63.
29 Kim MK, Han K, Koh ES, *et al*. The clinical utility of postprandial glucose measurements in predicting diabetes development: A 10-year community-based cohort study. *J Diabetes Investig* 2018;9:446–51.