

# Development and Validation of an Equation to Predict the Incidence of Coronary Heart Disease in Patients with Type 2 Diabetes in Japan

Yasunari Yamashita (✉ [dp18205@st.kitasato-u.ac.jp](mailto:dp18205@st.kitasato-u.ac.jp))

Kitasato University School of Pharmacy: Kitasato Daigaku Yakugakubu Daigakuin Yakugaku Kenkyuka  
<https://orcid.org/0000-0002-4006-2770>

**Gaku Inoue**

Kitasato Institute Hospital: Kitasato Kenkyujo Byoin

**Yoichi Nozaki**

Kitasato University: Kitasato Daigaku

**Rina Kitajima**

Kitasato University: Kitasato Daigaku

**Kiyoshi Matsubara**

Kitasato University School of Pharmacy: Kitasato Daigaku Yakugakubu Daigakuin Yakugaku Kenkyuka

**Takeshi Horii**

Kitasato University Hospital: Kitasato Daigaku Byoin

**Junichi Mohri**

Kitasato University Hospital: Kitasato Daigaku Byoin

**Koichiro Atsuda**

Kitasato University Hospital: Kitasato Daigaku Byoin

**Hajime Matsubara**

Kitasato Institute Hospital: Kitasato Kenkyujo Byoin

---

## Research note

**Keywords:** JJ risk engine, type 2 diabetes, coronary heart disease, discrimination, calibration, hypoglycemia, C-statistic, box plot, outlier

**Posted Date:** July 1st, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-669020/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Research Notes on November 25th, 2021. See the published version at <https://doi.org/10.1186/s13104-021-05844-w>.

# Abstract

**Objective:** In the diabetes treatment policy after the Kumamoto Declaration 2013, it is difficult to accurately predict the incidence of complications in patients using the JJ risk engine. This study was conducted to develop a prediction equation suitable for the current diabetes treatment policy using patient data from Kitasato Institute Hospital (Hospital A) and to externally validate the developed equation using patient data from Kitasato University Hospital (Hospital B). Outlier tests were performed on the patient data from Hospital A to exclude the outliers. Prediction equation was developed using the patient data excluding the outliers and was subjected to external validation.

**Results:** By excluding outlier data, we were able to develop a new prediction equation for the incidence of coronary heart disease (CHD) as a complication of type 2 diabetes, incorporating the use of antidiabetic drugs with a high risk of hypoglycemia. This is the first prediction equation in Japan that incorporates the use of antidiabetic drugs. We believe that it will be useful in preventive medicine for treatment for people at high risk of CHD as a complication of diabetes or other diseases. In the future, we would like to confirm the accuracy of this equation at other facilities.

## Introduction

It is important to prevent the development of diabetic complications during diabetes treatment [1]. One of the methods to predict the incidence of diabetic complications is risk engine, which is used to personalize medicine for patients. Currently, several risk engines have been developed to predict the incidence of diabetic complications [1–5]. In 2012, the Japan Diabetes Complications Study (JDCS)/The Japanese Elderly Diabetes Intervention Trial (J-EDIT) risk engine (JJ risk engine) was developed to accurately predict macro- and microvascular complications in Japanese patients with type 2 diabetes [1]. However, after the Kumamoto Declaration 2013, this risk engine was evaluated only through internal validation, without considering hypoglycemia prevention as the priority, and the implementation of external validation has been a challenge [1, 6]. In our previous study, we externally validated the prediction accuracy of the JJ risk engine using data from patients with type 2 diabetes at Hospital A. The results showed that the prediction of the JJ risk engine and the actual frequency of diabetic complications in Hospital A diverged [7]. Although the cause of this discrepancy is unknown, one reason may be the change in diabetes treatment to one that emphasizes hypoglycemia prevention [6, 8–10]. Therefore, we conclude that it is difficult to accurately predict the complication rates in patients using the JJ risk engine based on the diabetes treatment policies after the Kumamoto Declaration of 2013 [7].

Besides hypoglycemia, other risk factors for coronary heart disease (CHD) include aging, hypertension, hyperlipidemia, obesity, and chronic kidney disease (CKD) [11, 12]. Outlier tests were conducted for each risk factor, because prediction without outliers is more accurate than prediction with outliers [13, 14]. Therefore, in this study, we developed a new prediction equation that is more accurate and suitable for the patient population of Hospital A and externally validated the new prediction equation using patient data from Hospital B.

# Methods

## Patients:

Selection criteria:

The subjects were patients with type 2 diabetes who visited Hospital A or Hospital B from January 2013 to December 2013 and continued treatment for the following 5 years until 2018.

Exclusion criteria:

Patients who refused to participate in the study or had a history of any of the following diseases were excluded: angina, myocardial infarction, stroke, peripheral arterial disease, familial hypercholesterolemia, familial type III hyperlipidemia, nephrotic syndrome, renal diseases other than diabetic nephropathy, microhematuria, pre-proliferative and proliferative retinopathy, or major ocular diseases (e.g., glaucoma, dense cataract, or a history of cataract surgery).

This study was conducted in accordance with the Ethical Guidelines for Medical and Health Research Involving Human Subjects. The Kitasato University Kitasato Institute Hospital, Research Ethics Committee, approved the study (control number: 20051 and 20051-2) and provided permission to review patient records and use the corresponding data. The option to opt-out of the study was provided to the patients at the start of the study (2021).

## Statistical analysis

We developed a prediction equation based on the Cox proportional hazard model using patient data from Hospital A [15]. The backward stepwise method was used for the selection of variables [16].

Discrimination:

It is an index that evaluates how accurately the presence or absence of an event can be predicted by a prediction model. The C-statistic, which is calculated based on the receiver operating characteristic (ROC) curve, is used as a criterion for measuring the predictive accuracy [17, 18].

Calibration:

It is an index to measure the degree of agreement between the prediction by the model and the actual outcome. The significance probability calculated using the Hosmer–Lemeshow test is used as the criterion for predictability. The significance level was set at 0.05 ( $p < 0.05$ ) [18, 19].

External validation:

We developed prediction equations using Hospital A data and then performed external validation using Hospital B data.

Outlier testing using box plots:

Outlier tests with box plots were performed to reduce the impact of outliers of each risk factor on the prediction accuracy.

A box-and-whisker diagram of data for age, total cholesterol, HDL cholesterol, body mass index (BMI), urinary albumin, upper quartile (Q3), and lower quartile (Q1) was generated, and values that deviated from the range of  $Q1 - (Q3 - Q1)$  to  $Q3 + (Q3 - Q1)$  were considered outliers [20].

Although blood pressure was measured at the time of medical examination, it was excluded from the risk factors in this study because the time of measurement varied among subjects [21].

R version 2.5.1 (<http://www.r-project.org>, library Design, Hmisc, ROCR) was used to determine discrimination and calibration, whereas the ROC curve, Hosmer–Lemeshow test, and box plot were used for calculation [18–20].

## Results

There were 572 and 285 patients in Hospitals A and B, respectively. The baseline characteristics of the patients are presented in Table 1.

Table 1  
Baseline characteristics

Characteristics	Kitasato University Kitasato Institute Hospital			Kitasato University Hospital		
	Hospital A			Hospital B		
	All patients (n = 572)	Male (n = 384)	Female (n = 188)	All patients (n = 285)	Male (n = 169)	Female (n = 116)
Age (years)	64.5 ± 10.5	63.9 ± 10.5	65.6 ± 10.5	62.3 ± 12.4	61.0 ± 12.5	64.3 ± 12.0
Height (cm)	163.8 ± 8.8	168.0 ± 6.4	154.9 ± 5.9	162.0 ± 9.2	167.4 ± 6.8	154.0 ± 6.0
Weight (kg)	67.3 ± 14.0	71.4 ± 13.4	58.9 ± 10.9	69.1 ± 15.6	73.3 ± 14.8	62.8 ± 14.6
Simple retinopathy, n (confirmed/total)	46/572	35/385	11/187	81/285	41/169	40/116
Atrial fibrillation, n (confirmed/total)	10/572	8/385	2/187	9/285	4/169	5/116
Presence of exercise habits, n (confirmed/total)	91/572	63/385	28/187	17/285	14/169	3/116
Current smoker, n (confirmed/total)	84/572	66/385	18/187	41/285	31/169	10/116
Disease duration (years)	12.0 ± 8.3	12.7 ± 8.6	10.6 ± 7.6	13.1 ± 10.9	13.1 ± 10.8	13.0 ± 11.0
HbA1c (NGSP%)	7.3 ± 2.8	7.3 ± 3.3	7.1 ± 0.8	8.1 ± 1.8	8.0 ± 1.8	8.2 ± 1.8
Systolic blood pressure (mmHg)	126 ± 14	126 ± 14	126 ± 14	132 ± 15	130 ± 14	134 ± 15
Total cholesterol (mg/dL)	189 ± 31	184 ± 30	198 ± 30	198 ± 37	203 ± 39	192 ± 34
HDL cholesterol (mg/dL)	63 ± 17	60 ± 17	69 ± 17	58 ± 16	56 ± 17	60 ± 15
Urinary albumin (mg/gCr)	66 ± 245	65 ± 239	69 ± 257	186 ± 593	215 ± 684	143 ± 423
Medicine [sulfonylurea (SU) and/or insulin], n (confirmed/total)	338/572	239/385	99/185	195/285	115/169	80/116
SU, n (confirmed/total)	352/572	276/385	76/185	67/285	40/169	27/116

Data are expressed as means ± standard deviation

	Kitasato University Kitasato Institute Hospital			Kitasato University Hospital		
Insulin, n (confirmed/total)	104/572	79/385	25/185	131/285	76/169	55/116
Data are expressed as means ± standard deviation						

Patients who used either sulfonylurea (SU) drugs or insulin were considered medicine users. Among the variables, only medicine was found to have a value of  $p < 0.05$  ( $p = 0.03$ ). Therefore, only medicine was included as a variable in the prediction equation, and the prediction equation developed is as follows:

□K – medicine equation□

$$\lambda_t = \lambda_{0t} \times \exp \{ \beta \times \text{medicine} (0, 1) \}$$

$\lambda_t$ : Incidence rate by time  $t$ ;  $\lambda_{0t}$ : Baseline hazard for time  $t$ ;  $\beta$ : partial regression coefficient; medicine (0, 1): 1 for patients who used either SU or insulin, 0 for patients who used neither

The prediction equation using data from Hospital A resulted in a C-statistic of 0.734 and a calibration of  $p > 0.05$ , indicating no significant difference between the measured and predicted values. In contrast, external validation using data from Hospital B resulted in a C-statistic of 0.809 and a calibration of  $p < 0.05$ , indicating a significant difference between the measured and predicted values (Table 2).

Table 2  
Development and external validation of prediction equations

	n	$\beta$	HR	Discrimination; C-statistic (95% CI)	Calibration
Hospital A	572	1.39	4.0 (1.2–13.7)	0.734 (0.630–0.839)	0.292
Hospital B	285	-	-	0.809 (0.721–0.897)	0.006

Therefore, an outlier test using a box-and-whisker diagram was performed to improve the prediction accuracy. The outliers were age: 39 years or less; total cholesterol:  $\geq 266$  and  $\leq 96$ ; HDL cholesterol:  $\geq 105.5$ ; BMI:  $\geq 35$ ; and urine albumin:  $\geq 78.7$ . A total of 120 patients from Hospital A and 84 patients from Hospital B were excluded.

In the analysis of the variables after exclusion, the p-value for medicine was  $< 0.05$ , indicating a significant difference.

The developed prediction equation is as follows:

$$\lambda_{1825} = 0.01 \times \exp \{ 1.73 \times \text{medicine} (0, 1) \}$$

$\lambda_{1825}$ : Incidence rate within 5 years; medicine (0, 1): 1 for patients who used either SU or insulin, 0 for patients who used neither

The C statistic was 0.644 and the calibration was  $p > 0.05$ . There were no significant differences between the measured and predicted values. A total of 201 excluded patients from Hospital B were used for external validation; the C statistic was 0.750, and the calibration was  $p > 0.05$ . There were no significant differences between the measured and predicted values (Table 3).

Table 3  
Development and external validation of prediction equation after outlier testing

	n	$\beta$	HR	Discrimination; C-statistic (95% CI)	Calibration
Hospital A	452	1.73	5.6 (1.3–24.8)	0.644 (0.523–0.758)	0.974
Hospital B	201	-	-	0.750 (0.651–0.849)	0.322

## Discussion

After the outliers were excluded, a prediction equation (K-medicine equation) was developed for the incidence of CHD in patients with type 2 diabetes using patient data from Hospital A. Furthermore, after excluding outliers, the external validation using data from Hospital B showed that the C-statistic was moderate and the calibration was not significantly different, indicating a correct prediction.

Compared to the JJ risk engine, we incorporated the use of antidiabetic drugs into the prediction equation. In addition, while overseas risk engines have incorporated therapeutic drugs (dyslipidemia drugs) into the prediction equation [22], our study is the first to incorporate therapeutic drugs (diabetes drugs) into the prediction equation in Japan.

In contrast, validation using Hospital B patient data showed significant differences in the calibration, indicating an incorrect prediction. As per previous studies [11–14], the incorrect prediction might be owing to outliers of risk factors that influence the development of CHD, which affect the prediction accuracy.

A previous report showed that the risk of developing CHD was higher in patients who used SU drugs and insulin than in those who received dipeptidyl peptidase-4 inhibitors [23], indicating that SU drugs and insulin are risk factors for CHD.

In the selection of variables, there was a significant difference in the presence or absence of SU drugs or insulin use, and it was reasonable to include it in the variables of the prediction equation. The risk of CHD associated with the use of SU drugs and insulin is consistent with the results of previous reports [23]. This prediction equation indicates that the use of diabetes medications with a high risk of hypoglycemia influences the development of CHD as a complication of type 2 diabetes.

Because of outliers, about 21% and 30% of patients were excluded in Hospital A and Hospital B, respectively. As more than 70% of the patients in each institution remained after exclusion, we believe that our prediction formula is applicable to many patients with type 2 diabetes.

The unique feature of this study is that the prediction equation was developed using patient data from a medium-sized hospital in Japan, rather than large-scale clinical data. Furthermore, we were able to create a prediction equation and conduct external validation.

Exclusion of outlier data for age, total cholesterol, HDL cholesterol, BMI, and urinary albumin levels was a condition used in this prediction equation.

## Conclusion

We developed an equation to predict the incidence of CHD in patients with type 2 diabetes. Based on the prediction equation developed in this study, we believe that the use of diabetic drugs with a high risk of hypoglycemia influences the incidence of CHD as a complication of type 2 diabetes. Although this prediction equation is based on the patient population of Hospital A, we would like to confirm the accuracy of our prediction theory in other institutions in the future.

## Limitations

In this study, we did not examine the effects of LDL cholesterol and triglycerides because these two laboratory values were not provided for some patients.

Although blood pressure is a risk factor for CHD, it was not included in the outlier test in this study because hypertension was determined based on the results of several measurements [21]. Therefore, it was not possible to study the effects of blood pressure.

The relatively small number of people who developed CHD may have affected the reliability of the analysis [24].

## Abbreviations

CHD: coronary heart disease

ROC: receiver operating characteristic

HbA1c: glycated hemoglobin

NGSP: National Glycohemoglobin Standardization Program

HR: Hazard ratio

SU: sulfonylurea

LDL: low-density lipoprotein

## Declarations

### Ethics approval and consent to participate

This study was conducted in accordance with the Ethical Guidelines for Medical and Health Research Involving Human Subjects. The Kitasato University Kitasato Institute Hospital, Research Ethics Committee, approved the study (control number: 20051, 20051-2). For this retrospective study, we disclosed the purpose of the study to the participants and provided the option to opt-out before their participation. All participants provided written informed consent to participate in the study. In addition, the choice to opt-out provided participants an opportunity to refuse participation in the research.

### Consent for publication

Not applicable.

### Availability of data and material

The datasets generated and analyzed during the current study are not publicly available owing to the privacy of the research participants. Data are available from the corresponding author upon reasonable request; however, permission from the Kitasato University Kitasato Institute Hospital Research Ethics Committee is required.

### Competing interests

The authors declare that they have no competing interests.

### Funding

This research received no specific grant from any funding agency in public, commercial, or not-for-profit sectors.

### Authors' contributions

YY designed the study and wrote the initial draft of the manuscript. RK contributed to data collection and interpretation. KM, GI, and HM critically reviewed the manuscript. All authors read and approved the final manuscript.

### Acknowledgments

Not applicable.

## References

1. Tanaka S, Tanaka S, Imuro S, Yamashita H, Katayama S, Akanuma Y, et al. Predicting macro- and microvascular complications in type 2 diabetes: the Japan Diabetes Complications Study/the Japanese Elderly Diabetes Intervention Trial risk engine. *Diabetes Care*. 2013;36:1193-9.
2. Stevens RJ, Kothari V, Adler AI, Stratton IM; United Kingdom Prospective Diabetes Study (UKPDS) Group. The UKPDS risk engine: a model for the risk of coronary heart disease in Type II diabetes (UKPDS 56). *Clin Sci (Lond)*. 2001;101(6):671-9. Erratum in: *Clin Sci (Lond)*. 2002;102(6):679.
3. Shao H, Fonseca V, Stoecker C, Liu S, Shi L. Novel risk engine for diabetes progression and mortality in USA: building, relating, assessing, and validating outcomes (BRAVO). *Pharmacoeconomics*. 2018;36:1125-34.
4. Montagud-Marrahi E, Molina-Andújar A, Pané A, Ruiz S, Amor AJ, Esmatjes E, et al. Impact of simultaneous pancreas-kidney transplantation on cardiovascular risk in patients With diabetes. *Transplantation*. 2021. doi: [10.1097/TP.0000000000003710](https://doi.org/10.1097/TP.0000000000003710).
5. Postmus D, van Veldhuisen DJ, Jaarsma T, Luttik ML, Lassus J, Mebazaa A, et al. The COACH risk engine: a multistate model for predicting survival and hospitalization in patients with heart failure. *Eur J Heart Fail*. 2012;14:168-75. doi: [10.1093/eurjhf/hfr163](https://doi.org/10.1093/eurjhf/hfr163).
6. Nakanishi S, Hirukawa H, Shimoda M, Tatsumi F, Kohara K, Obata A, et al. Verification of Kumamoto Declaration 2013 and Glycemic Targets for Elderly Patients with Diabetes in Japan for prevention of diabetic complications: A retrospective longitudinal study using outpatient clinical data. *J Diabetes Investig*. 2019;10:290-301.
7. Yamashita Y, Kitajima R, Matsubara K, Inoue G, Matsubara H. A retrospective study on the usefulness of the JJ risk engine for predicting the incidence rate of coronary heart disease in type 2 diabetes patients. *BMC Res Notes*. 2021;14:92.
8. Goto A, Arah OA, Goto M, Terauchi Y, Noda M. Severe hypoglycaemia and cardiovascular disease: systematic review and meta-analysis with bias analysis. *BMJ*. 2013;347:f4533.
9. *Japan Diabetes Society*. Treatment Guide for Diabetes, BUNKODO; 2020-2021.
10. Araki E, Haneda M, Kasuga M, Nishikawa T, Kondo T, Ueki K, et al. New glycemic targets for patients with diabetes from the Japan Diabetes Society. *Diabetol Int*. 2016;7:327-30.
11. *The Japan circulation society*. Shimamoto, K., Arai, H., Iso, H., & Ouchi, Y. (2012). Guidelines for the primary prevention of ischemic heart disease revised version (JCS 2012).
12. Hubert HB, Feinleib M, McNamara PM, Castelli WP. Obesity as an independent risk factor for cardiovascular disease: a 26-year follow-up of participants in the Framingham Heart Study. *Circulation*. 1983;67:968-77.
13. Nnamoko N, Korkontzelos I. Efficient treatment of outliers and class imbalance for diabetes prediction. *Artif Intell Med*. 2020;104:101815. doi: [10.1016/j.artmed.2020.101815](https://doi.org/10.1016/j.artmed.2020.101815).
14. Ferdowsi H, Jagannathan S, Zawodniok M. An online outlier identification and removal scheme for improving fault detection performance. *IEEE Trans Neural Netw Learn Syst*. 2014;25:908-19. doi: [10.1109/TNNLS.2013.2283456](https://doi.org/10.1109/TNNLS.2013.2283456).

15. Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. *Annu Rev Public Health*. 1999;20:145-57. doi: [10.1146/annurev.publhealth.20.1.145](https://doi.org/10.1146/annurev.publhealth.20.1.145).
16. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic research: Principles and quantitative methods*. Belmont: Lifetime Learning Publications; 1982.
17. Shintani A. *Everyone's medical statistics Complete mastery of basic theory and EZR in 12 days*. Kodansha; 2016.
18. Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361-87.
19. Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd ed. A Wiley-Interscience Publication; John Wiley & Sons, Inc. 2000.
20. Walfish S. A review of statistical outlier methods. *Pharm Technol*. 2006;30:11:82.
21. Umemura S, Arima H, Arima S, Asayama K, Dohi Y, Hirooka Y, et al. The Japanese Society of Hypertension guidelines for the management of hypertension (JSH 2019). *Hypertens Res*. 2019;42:1235-481. doi: [10.1038/s41440-019-0284-9](https://doi.org/10.1038/s41440-019-0284-9).
22. Rawshani A, Rawshani A, Franzén S, Sattar N, Eliasson B, Svensson AM, et al. Risk factors, mortality, and cardiovascular outcomes in patients with type 2 diabetes. *N Engl J Med*. 2018;379:633-44. doi: [10.1056/NEJMoa1800256](https://doi.org/10.1056/NEJMoa1800256).
23. O'Brien MJ, Karam SL, Wallia A, Kang RH, Cooper AJ, Lancki N, et al. Association of second-line antidiabetic medications with cardiovascular events among insured adults with type 2 diabetes. *JAMA Netw Open*. 2018;1:e186125. doi: [10.1001/jamanetworkopen.2018.6125](https://doi.org/10.1001/jamanetworkopen.2018.6125).
24. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med*. 2016;35:214-26.