Risk prediction scores for mortality, cerebrovascular, and heart disease among Chinese people with type 2 diabetes

Authors and affiliations

Jianchao Quan[1], Deanette Pang[2], Tom K. Li[1], Cheung Hei Choi[3], Shing Chung Siu[4], Simon Y. Tang[5], Nelson M. Wat[6], Jean Woo[7], Zheng Yi Lau[2], Kelvin B. Tan[2†], Gabriel M. Leung[1†]
[†] Joint senior authors

Author affiliations

[1] Division of Health Economics, Policy and Management, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China

[2] Ministry of Health, Singapore

[3] Queen Elizabeth Hospital, Hong Kong, China

[4] Department of Medicine & Rehabilitation, Tung Wah Eastern Hospital, Hong Kong, China

[5] Tuen Mun Hospital, Hong Kong SAR, China

[6] Caritas Medical Centre, Hong Kong SAR, China

[7] Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

Corresponding author

Dr. Jianchao Quan   Email: jquan@hku.hk

DISCLOSURE SUMMARY

The authors have nothing to disclose.

Risk prediction scores for mortality, cerebrovascular, and heart disease among Chinese people with type 2 diabetes

1 # Abstract

2 **CONTEXT**

3 Risk scores for cardiovascular and mortality outcomes have not been commonly applied in Chinese

4 populations.

5 **OBJECTIVE**

6 To develop and externally validate a set of parsimonious risk scores (HKU-SG) to predict the risk of

7 mortality, cerebrovascular disease and ischemic heart disease among Chinese people with type 2

8 diabetes; and compare HKU-SG risk scores to other existing ones.

9 **DESIGN**

10 Retrospective population-based cohorts drawn from Hong Kong Hospital Authority health records

11 from 2006 to 2014 for development and Singapore Ministry of Health records from 2008 to 2016 for

12 validation. Separate 5-year risk scores were derived using Cox proportional hazards models for each

13 outcome.

14 **SETTING**

15 Study participants were adults with type 2 diabetes aged 20 years or over, consisting of 678,750

16 participants from Hong Kong and 386,425 participants from Singapore.

17 **MAIN OUTCOME MEASURES**

18 Performance was evaluated by discrimination (Harrell's c-index), and calibration plots comparing

19 predicted against observed risks.

20 **RESULTS**

21 All models had fair external discrimination. Among the risk scores for the diabetes population, ethnic-

22 specific risk scores (HKU-SG and JADE) performed better than UKPDS and RECODe models.

23 External validation of the HKU-SG risk scores for mortality, cerebrovascular disease and ischemic

24 heart disease had corresponding c-indices of 0.778, 0.695, and 0.644. The HKU-SG models appeared

25 well calibrated on visual plots, with predicted risks closely matching observed risks.

26 **CONCLUSIONS**

27 The HKU-SG risk scores were developed and externally validated in two large Chinese population-

28 based cohorts. The parsimonious use of clinical predictors compared to previous risk scores could

29 allow wider implementation of risk estimation in diverse Chinese settings.

30

31 **Keywords:** type 2 diabetes mellitus, risk score, mortality, cardiovascular disease

Abbreviations

32	GFR	Glomerular filtration rate
33	HKU	University of Hong Kong
34	HKU-SG	University of Hong Kong-Singapore
35	ICD-9	International Classification of Disease, 9th Edition
36	ICPC	International Classification of Primary Care
37	JADE	Joint Asia Diabetes Evaluation
38	RECODe	Risk Equations for Complications Of type 2 Diabetes
39	UKPDS	UK Prospective Diabetes Study
40	Urine ACR	Urine albumin:creatinine ratio

Introduction

Diabetes increases the risk of many diseases including cardiovascular disease, renal disease, cancer, and ultimately leads to premature mortality. Life expectancy of people with diabetes in China is estimated to be 8 to 10 years shorter than that of people without diabetes.[1] Predicting the absolute risk of adverse outcomes can guide decision-making for individuals and populations. Risk prediction models can also be used to assess the cost-effectiveness of different disease management options and the performance of treatment pathways, particularly when there is incomplete data on outcomes, or when the outcomes need to be extrapolated into the future.[2]

The risk of mortality and cardiovascular outcomes of people with type 2 diabetes has been assessed by the UK Prospective Diabetes Study (UKPDS), the Joint Asia Diabetes Evaluation (JADE), and the Risk Equations for Complications Of type 2 Diabetes (RECODe) risk scores based on cohorts from 1977-1991, 1995-2005 and 2001-2009 respectively.[3–8] Newer versions of the UKPDS models using data from the original cohorts have also been published (UKPDS 82: Outcomes Model 2).[9] However, improvements in medical care, lifestyles and socioeconomic development in the intervening years may well have influenced patient outcomes. The UKPDS and RECODe derivation data did not include Chinese participants. The older UKPDS scores (UKPDS 56 & 60) are known to inflate the observed risk among Chinese people with diabetes, while the newer version (UKPDS 82) and the RECODe scores have not yet been validated in Chinese populations.[6,7] The JADE score for Hong Kong Chinese was based on only 7,000 participants with a low number of events. Both the JADE and RECODe scores require more sophisticated laboratory data such as the urinary albumin-creatinine ratio which can limit general deployment across different settings, especially in resource-poor conditions.

To address this gap for Chinese patients with diabetes, we sought to develop and internally validate a new model (HKU risk score) on a very large population-based cohort of Chinese patients with diabetes in Hong Kong, and to externally validate a more parsimonious model (HKU-SG risk score)

66   in a large population-based Singapore cohort. We compared the performance of our newly developed

67   algorithms with the existing UKPDS, RECODe and JADE risk scores.

Methods

Setting

69   We used the Hong Kong Hospital Authority's electronic health records (Clinical Management

70   System) to identify all people with diabetes in the public health care system from 1 January 2006 to

71   31 December 2014. Data for Singapore were obtained from Ministry of Health administrative datasets

72   covering 1 January 2008 to 31 December 2016. We identified 678,750 participants in the Hong Kong

73   derivation cohort and 386,425 participants in the Singaporean validation cohort as aged 20 years or

74   over with a confirmed diagnosis of type 2 diabetes. Detailed description of the data sources and case

75   identification of type 2 diabetes can be found in Supplemental Appendix 1.[10]

Outcomes

76   Deaths were recorded by the relevant government immigration department cross-referenced by Hong

77   Kong or Singapore identity card numbers. Participants were censored on the cohort end date of 31

78   December 2014 for Hong Kong, and 31 December 2016 for Singapore. We used the following

79   diagnosis codes in admission and attendance records to ascertain ischemic heart disease (ICD-9 codes:

80   410-414 and ICPC codes: K74-K76),[3,7] and cerebrovascular disease (ICD-9 codes: 430-434, 436, and

81   ICPC: K90).[4,6] Thus, ischemic heart disease includes myocardial infarction, angina pectoris, coronary

82   atherosclerosis and aneurysms; while cerebrovascular disease includes intracranial hemorrhages (e.g.

83   subarachnoid, intracerebral), and occlusion of cerebral arteries. For the derivation of ischemic heart

84   disease and cerebrovascular disease models, we excluded participants with a previous episode of

85   ischemic heart disease or cerebrovascular disease respectively.

Predictors

86   For each participant, we determined the cohort entry date as the later of: date of diabetes diagnosis or

87   study start date (1 January 2006 for Hong Kong participants and 1 January 2008 for the Singapore

88   participants respectively). Candidate predictors were selected based on data availability, clinical

89   relevance, and variables included in other risk models such as the UKPDS, JADE and RECODe

90   models (Supplemental Table 1).[10] Clinical values were defined as the closest recorded value within 2

91   years of cohort entry date and are listed in the Supplemental Appendix 1.[10] All other predictor

92   variables were based on the latest recorded information at baseline. The duration of diabetes was

93   defined as the time from diagnosis until cohort entry, so an individual has a duration of 6 years if

94   diagnosed in the year 2000 and a duration of zero years if diagnosed in 2006 or later. Chronic kidney

95   disease was defined by ICD-9 codes 585-587, including chronic kidney disease (585), renal failure

96   (586) and renal sclerosis (587). We did not include time-varying covariates as the risk models are

97   intended to be used for initial assessment at baseline. The distribution of missing values is detailed in

98   Supplemental Table 2 and 3.[10] We applied multiple imputation to handle missing values in both the

99   derivation and validation cohorts by imputing five datasets using the R rms package. The regression-

100   based imputation model included the candidate predictors, an interaction term between age and

101   duration of diabetes, the Nelson-Aalen estimator and the event indicator.

Statistical Models

103   We developed the HKU and HKU-SG risk scores using the full data of Hong Kong cohort. To

104   account for non-linearity, we transformed the continuous variables using restricted cubic splines with

105   four knots placed at the respective 5th, 35th, 65th and 95th sample percentiles.[11] We derived separate

106   Cox proportional hazard models for predicting all-cause mortality, ischemic heart disease, and

107   cerebrovascular disease at five years. For each outcome, we fitted a Cox proportional hazard model

108   including all candidate predictors (full model), which was then approximated, blinded for the

109   outcome, into a more parsimonious model of the most predictive variables (HKU risk score). We

110   assigned the full model a R^2 value of 1.0 by design and then excluded candidate variables using a

111   backward elimination algorithm (ordinary least squares model in which the estimated linear predictor

112   from the full model was the outcome and all candidate variables were entered in exactly the same

113   manner as in the full model) to achieve an arbitrary level of $R^2>0.95$ for the HKU risk score.[11] This

114   method has been outlined by Harrell, and applied in the development of other risk scores.[12] We tested

115   for interaction terms between age and duration of diabetes, age and gender, age and smoking, gender

116 and smoking, which did not improve model performance, and thus were dropped from the final

117 models. The final predictor variables included in the HKU risk scores are age, duration of diabetes,

118 gender, smoking status, HbA1c, total and HDL-cholesterol, blood pressure, creatinine, hemoglobin,

119 white cell count, resting heart rate, pre-existing conditions (cancer, congestive heart failure, diabetes

120 complications), and medication for diabetes, hypertension and statins (Supplemental Table 1).[10] To

121 allow for wider applicability of the risk score to different clinical settings with more limited data

122 availability, we then developed the HKU-SG risk scores including all predictors common to both the

123 Hong Kong and Singapore datasets: age, duration of diabetes, gender, smoking status, body mass

124 index, systolic and diastolic blood pressure, HbA1c, LDL-cholesterol, pre-existing conditions (atrial

125 fibrillation and chronic kidney disease).

126 **Internal and external validation**

127 Discrimination was assessed by Harrell's c-index (an extension of the receiver operating characteristic

128 statistic to survival data ranging from 0 to 1, where higher values indicate better performance).[11] Each

129 risk model was internally validated using 100 bootstrap samples.[13,14] We used the Singapore dataset to

130 externally validate the HKU-SG models. We applied the UKPDS, RECODe and JADE risk scores on

131 the Hong Kong dataset to assess validity but were unable to validate on the Singaporean dataset due to

132 lack of certain data variables. We assessed calibration by comparing the observed 5-year event rates

133 with the risk score predictions, and plotted the mean predicted 5-year probability of events by

134 percentile against the observed event rate. To test sensitivity to missing data, we repeated the

135 validation using complete case analysis. We also analyzed a subset of Chinese ethnicity within the

136 Singapore validation cohort.

137 All analyses were carried out using R version 3.4 (R Foundation for Statistical Computing) with the

138 rms and Hmisc packages.[11] The analyses followed the framework for derivation and validation of

139 prediction models proposed by Harrell[11] and Royston and Altman[15] and reported in accordance with

140 the TRIPOD statement.[16] The study was approved by the Institutional Review Boards of all seven

141 Hospital Authority clusters in Hong Kong (HKWC, HKEC, KC/KEC, KWC, NTWC, NTEC).

Results

Baseline characteristics of the Hong Kong and Singapore study samples are shown in Table 1. The HKU and HKU-SG risk scores were developed from the Hong Kong cohort of 4,294,018 person-years for mortality, 3,868,798 person-years for cerebrovascular disease, and 3,741,576 for ischemic heart disease. During a mean follow-up of 6.3 years, 143,047/678,750 (21.1%) participants died. After excluding participants with previous history of cerebrovascular disease and ischemic heart disease respectively, 43,215/623,294 (6.9%) participants developed cerebrovascular and 54,365/610,647 (8.9%) participants developed ischemic heart disease during follow-up. Mean follow-up time was similar for the Hong Kong derivation and Singapore validation cohort. A flow diagram outlining the cohort selection is shown in Supplemental Figure 1.[10] Predictors included in the final HKU and HKU-SG 5-year risk scores are listed in Supplemental Table 1, with the hazard ratio of each variable shown in Supplemental Figure 2.[10] To predict individual risk scores, the final models are presented as an interactive web tool for ease of use: https://jquan.shinyapps.io/riskmodel. The full equations for the risk models are specified in Supplemental Appendix 2.[10]

All risk scores (new and existing) performed best in discrimination (c-index) with mortality outcomes followed by cerebrovascular disease and lowest with ischemic heart disease (Table 2 and Table 3). In internal validation of 100 bootstrap samples, the c-index of the HKU models ranged from 0.845 for mortality, 0.727 for cerebrovascular disease, to 0.717 for ischemic heart disease. The corresponding HKU-SG models with fewer predictors had lower discriminatory performance with c-index scores of 0.805 for mortality, 0.722 for cerebrovascular disease and 0.700 for ischemic heart disease.

The HKU-SG risk scores were externally validated using the Singapore study sample comprising 2,363,508 person-years of follow-up for mortality, 2,055,127 person-years for cerebrovascular disease and 1,747,518 person-years for ischemic heart disease. In external validation (Table 2), the discriminative performance for mortality (c-index 0.772) was higher than that of cerebrovascular disease (c-index 0.695) and ischemic heart disease (c-index 0.644). There was little difference in discriminative performance results from sensitivity analyses using complete case analysis of the

168 Singapore validation sample, or by restricting the Singapore validation sample to the Chinese

169 ethnicity subgroup (Supplemental Table 4).[10]

170 As the Singapore data sample lacked the requisite variables, the Hong Kong derivation sample was

171 used for external validation of the JADE, RECODe and UKPDS risk scores. In contrast to the HKU-

172 SG results, there were differences in discriminative performance of the JADE, RECODe and (to a

173 lesser extent) UKPDS risk scores between external validation of the imputed data samples compared

174 to complete case analysis (Table 3). The JADE, RECODe and UKPDS risk scores all performed

175 worse in complete case analysis. For example, the c-index of JADE risk scores dropped from 0.841 to

176 0.793 for mortality, 0.730 to 0.716 for cerebrovascular disease, and 0.685 to 0.655 for ischemic heart

177 disease. The JADE risk score derived in, thus calibrated to, Hong Kong Chinese consistently

178 displayed better discriminative performance than the RECODe and UKPDS risk scores.

179 For calibration performance, the HKU-SG models showed close correspondence between predicted

180 and observed risk at 5 years with the Singapore validation cohort across virtually the entire

181 distribution for mortality and cerebrovascular disease (Figure 1). There was some underestimation of

182 the observed risk of ischemic heart disease by the HKU-SG risk score. The JADE risk score for

183 mortality showed close correspondence with the observed risk of mortality among the Hong Kong

184 cohort, but overestimates the risk of cerebrovascular disease (at higher risk percentiles) and ischemic

185 heart disease (Figure 2). In comparison, the calibration performance of the RECODe and UKPDS risk

186 scores showed substantial overestimation of the observed risks of mortality and cerebrovascular

187 disease (Figure 2). The newer UKPDS 82 risk score underestimated the risk of ischemic heart disease

188 while the older UKPDS 56 risk scores substantially overestimates the observed risk. The newer

189 UKPDS risk scores had better calibration than the older UKPDS risk scores. Poor correspondence

190 between the predicted and observed risks for cerebrovascular disease and ischemic heart disease is

191 particularly apparent in the higher risk percentiles.

192

Discussion

We developed and validated (internally and externally) new risk scores for mortality, cerebrovascular disease, and ischemic heart disease in Chinese patients with type 2 diabetes in Hong Kong and Singapore. These contemporaneous HKU and HKU-SG risk scores update previous risk scores developed from a 1995-2005 Chinese cohort; since then improvements in medical care, lifestyles and socioeconomic development in the intervening years may well have influenced disease outcome trajectories. The UKPDS and RECODe risk scores were developed in Western settings with few Asian participants and previously had not been validated in large Chinese populations.

The new HKU and HKU-SG risk scores use simple clinical variables that are readily available, avoiding more sophisticated laboratory investigations such as the urinary albumin:creatinine ratio and microalbuminuria, that have limited the common usage of previous JADE, RECODe and UKPDS risk scores in Chinese and resource-poor populations. We were unable to externally validate the JADE, RECODe and UKPDS risk scores on the Singapore validation sample due to lack of data on the required predictors. Generally, for all the models, discrimination and calibration performance was highest for mortality and lowest for ischemic heart disease. In spite of using a simpler set of predictors, the discriminative performances of the HKU-SG risk scores on external validation were comparable to JADE and RECODe, and superior to UKPDS. External validation of the HKU-SG model yielded robust results with virtually identical discriminative performance (c-indices) using complete case analysis or Chinese-only subgroup. In contrast, complete cases analyses of the JADE, RECODe and UKPDS risk scores resulted in substantially lower discriminative performance (c-indices), exhibiting much greater sensitivity to missing data.

Calibration performance of the HKU-SG risk scores in the external Singapore validation sample were superior to the calibration of JADE, RECODe and UKPDS risk scores in the Hong Kong sample. Close correspondence was particularly observed for the HKU-SG mortality and cerebrovascular disease risk scores. The JADE model for mortality also showed close calibration with the observed risks in the Hong Kong cohort sample. Consistent with other studies, the UKPDS and RECODe risk

219   scores tended to overestimate (the former more so than the latter) the risk of cerebrovascular disease

220   and ischemic heart disease compared to the observed risk resulting in suboptimal calibration.[6,7]

221   Our study had a number of limitations. Both the development and validation samples were drawn

222   from population-based health records, which increases generalizability, but necessarily have less

223   complete ascertainment of clinical data compared to the idealized settings of clinical trials or

224   dedicated disease registers used by the previous risk models.

225   Comparisons of discrimination and calibration performance by the new HKU-SG risk scores against

226   JADE, RECODe and UKPDS should be interpreted with caution as we were unable to perform

227   external validation of all models on the same Singapore validation sample. The JADE, RECODe and

228   UKPDS risk scores were externally validated on the Hong Kong sample. While mortality and

229   cerebrovascular disease risks were similar in the Hong Kong and Singapore samples, rates of ischemic

230   heart disease were higher in Singapore relative to Hong Kong. This is consistent with previous

231   observations of a greater burden of coronary heart disease and related deaths in Singapore relative to

232   Hong Kong.[17–19] This likely led to the HKU-SG risk score underestimating the observed risk of

233   ischemic heart disease in Singapore as shown in the calibration plots.

234   China has the highest absolute disease burden of diabetes in the world accounting for more than one

235   in four diabetes cases globally.[20] In China's progress towards its recently promulgated *Healthy China*

236   *2030* strategic plan,[21] there is a case for diabetes services to be strengthened nationally. By calculating

237   individual-level risk using routinely recorded data, these risk prediction scores can be used to evaluate

238   population health status. These new risk scores can be immediately implemented in the clinical setting

239   for Chinese patients with diabetes using the online risk stratification tool. We took advantage of two

240   large population-based datasets from Hong Kong and Singapore to develop and validate more

241   parsimonious risk scores for the Chinese population, with the aim of providing wider implementation

242   among populations in less developed settings or where clinical data are less readily available. We

243   envisage the HKU-SG risk scores will be useful as a decision support for Chinese patients with

244   diabetes to assess their disease risk.

Acknowledgements

References

1. Bragg F, Holmes MV, Iona A, et al. Association Between Diabetes and Cause-Specific Mortality in Rural and Urban Areas of China. *JAMA*. 2017;317(3):280. doi:10.1001/jama.2016.19720

2. Halpern MT, Luce BR, Brown RE, Geneste B. Health and Economic Outcomes Modeling Practices: A Suggested Framework. *Value Health*. 1998;1(2):131-147. doi:10.1046/j.1524-4733.1998.120131.x

3. Stevens RJ, Kothari V, Adler AI, et al. The UKPDS risk engine: a model for the risk of coronary heart disease in Type II diabetes (UKPDS 56). *Clin Sci*. 2001;101(6):671–679.

4. Kothari V, Stevens RJ, Adler AI, et al. UKPDS 60: Risk of Stroke in Type 2 Diabetes Estimated by the UK Prospective Diabetes Study Risk Engine. *Stroke*. 2002;33(7):1776-1781. doi:10.1161/01.STR.0000020091.07144.C7

5. Yang X, So WY, Tong PC, et al. Development and validation of an all-cause mortality risk score in type 2 diabetes: The Hong Kong Diabetes Registry. *Arch Intern Med*. 2008;168(5):451–457.

6. Yang X, So W-Y, Kong APS, et al. Development and Validation of Stroke Risk Equation for Hong Kong Chinese Patients With Type 2 Diabetes: The Hong Kong Diabetes Registry. *Diabetes Care*. 2007;30(1):65-70. doi:10.2337/dc06-1273

7. Yang X, So W-Y, Kong APS, et al. Development and Validation of a Total Coronary Heart Disease Risk Score in Type 2 Diabetes Mellitus. *Am J Cardiol*. 2008;101(5):596-601. doi:10.1016/j.amjcard.2007.10.019

8. Basu S, Sussman JB, Berkowitz SA, Hayward RA, Yudkin JS. Development and validation of Risk Equations for Complications Of type 2 Diabetes (RECODe) using individual participant data from randomised trials. *Lancet Diabetes Endocrinol*. 2017;5(10):788-798. doi:10.1016/S2213-8587(17)30221-8

9. Hayes AJ, Leal J, Gray AM, Holman RR, Clarke PM. UKPDS Outcomes Model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82. *Diabetologia*. 2013;56(9):1925-1933. doi:10.1007/s00125-013-2940-y

10. Quan, Jianchao; Pang, Deanette; Li, Tom K.; Choi, Cheung Hei; Siu, Shing Chung; Tang, Simon Y.; et al. (2019): Risk prediction scores for mortality, cerebrovascular, and heart disease among Chinese people with type 2 diabetes - Supplementary Information. figshare. Journal contribution. https://doi.org/10.6084/m9.figshare.9888923.v2

11. Harrell FE. *Regression Modeling Strategies*. 2nd ed. Cham: Springer International Publishing; 2015.

12. Hijazi Z, Oldgren J, Lindbäck J, et al. The novel biomarker-based ABC (age, biomarkers, clinical history)-bleeding risk score for patients with atrial fibrillation: a derivation and validation study. *The Lancet*. 2016;387(10035):2302-2311. doi:10.1016/S0140-6736(16)00741-8

283   13.   Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for
284   development and an ABCD for validation. *Eur Heart J*. June 2014:ehu207.
285   doi:10.1093/eurheartj/ehu207

286   14.   Moons KGM, Kengne AP, Woodward M, et al. Risk prediction models: I. Development, internal
287   validation, and assessing the incremental value of a new (bio)marker. *Heart*. 2012;98(9):683-
288   690. doi:10.1136/heartjnl-2011-301246

289   15.   Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods.
290   *BMC Med Res Methodol*. 2013;13:33. doi:10.1186/1471-2288-13-33

291   16.   Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable
292   prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*.
293   2015;350:g7594. doi:10.1136/bmj.g7594

294   17.   Dwyer T, Emmanuel SC, Janus ED, Wu Z, Hynes KL, Zhang C. The emergence of coronary
295   heart disease in populations of Chinese descent. *Atherosclerosis*. 2003;167(2):303-310.
296   doi:10.1016/S0021-9150(03)00008-X

297   18.   Singapore Heart Foundation. Singapore - Statistics - About The Heart & Heart Disease.
298   http://www.myheart.org.sg/article/about-the-heart-and-heart-disease/statistics/singapore/75.
299   Published 2017. Accessed December 21, 2017.

300   19.   Department of Health. HealthyHK - Public Health Information and Statistics of Hong Kong -
301   Coronary Heart Diseases.
302   http://www.healthyhk.gov.hk/phisweb/en/healthy_facts/disease_burden/major_causes_death/cor
303   onary_heart_disease/. Published 2017. Accessed December 21, 2017.

304   20.   International Diabetes Federation. *IDF Diabetes Atlas, 8th Edn*. Brussels, Belgium: International
305   Diabetes Federation; 2017. http://www.diabetesatlas.org/. Accessed May 18, 2016.

306   21.   Tan X, Liu X, Shao H. Healthy China 2030: A Vision for Health Care. *Value Health Reg Issues*.
307   2017;12:112-114. doi:10.1016/j.vhri.2017.04.001

Table 1. Baseline characteristics of study participants

Characteristics	Mortality				Cerebrovascular disease				Ischemic heart disease			
	Hong Kong		Singapore		Hong Kong		Singapore		Hong Kong		Singapore	
n	678,750		386,425		623,294		345,303		610,647		304,362	
Number of outcome events (n, *%*)	143,047	(21.1)	58,906	(15.2)	43,215	(6.9)	24,177	(7.0)	54,365	(8.9)	35,391	(11.6)
Follow-up time, years (mean, *SD*)	6.3	(3.3)	6.1	(2.9)	6.2	(3.3)	6.0	(2.9)	6.1	(3.3)	5.7	(3.0)
Total follow-up time, person-years	4,294,018	-	2,363,508	-	3,868,798	-	2,055,127	-	3,741,576	-	1,747,518	-
Age, years (mean, *SD*)	63.6	(13.6)	60.8	(13.1)	62.8	(13.5)	59.9	(12.9)	62.8	(13.5)	59.2	(12.9)
20-49	104,502	(15.4)	73,333	(19.0)	102,485	(16.4)	70,868	(20.5)	101,713	(16.7)	67,248	(22.1)
50-59	163,753	(24.1)	106,255	(27.5)	157,705	(25.3)	98,941	(28.7)	154,714	(25.3)	88,942	(29.2)
60-69	166,114	(24.5)	106,658	(27.6)	154,747	(24.8)	94,938	(27.5)	150,638	(24.7)	82,626	(27.1)
70-79	157,229	(23.2)	69,032	(17.9)	138,008	(22.1)	57,415	(16.6)	134,069	(22.0)	46,951	(15.4)
80+	87,152	(12.8)	31,147	(8.1)	70,349	(11.3)	23,141	(6.7)	69,513	(11.4)	18,595	(6.1)
Female (n, *%*)	337,523	(49.7)	187,921	(48.6)	311,285	(49.9)	168,826	(48.9)	309,588	(50.7)	155,991	(51.3)
Chinese ethnicity (n, *%*)	-[a]	-[a]	245,791	(63.6)	-[a]	-[a]	218,175	(63.2)	-[a]	-[a]	198,123	(65.1)
Duration of DM, years (mean, *SD*)	3.8	(6.3)	1.4	(2.3)	3.6	(6.2)	1.3	(2.2)	3.6	(6.1)	1.1	(2.0)
Current smoker (n, %)	76,638	(11.0)	24,892	(6.0)	71,122	(11.0)	22,749	(7.0)	68,699	(11.0)	20,190	(7.0)
Body mass index, kg/m^2 (mean, *SD*)	25.8	(4.2)	26.8	(5.0)	25.8	(4.3)	26.9	(5.1)	25.8	(4.3)	26.8	(5.1)
Hb$_{A1c}$, % (mean, *SD*)	7.6	(1.9)	7.6	(1.8)	7.6	(1.9)	7.7	(1.9)	7.6	(1.9)	7.7	(1.9)
Systolic blood pressure, mmHg (mean, *SD*)	138.5	(19.8)	132.6	(16.8)	138.4	(19.6)	132.3	(16.7)	138.6	(19.6)	132.5	(16.6)
Diastolic blood pressure, mmHg (mean, *SD*)	76.3	(11.5)	75	(12.0)	76.5	(11.5)	75.1	(11.9)	76.7	(11.5)	75.4	(11.7)
LDL-cholesterol, mmol/L (mean, *SD*)	3.1	(1.0)	3.0	(0.9)	3.1	(1.0)	3.0	(0.9)	3.1	(1.0)	3.0	(0.9)
Presence of pre-existing medical conditions (n, *%*)												
Atrial fibrillation	26,572	(3.9)	15,395	(4.0)	19,133	(3.1)	10,537	(3.1)	18,016	(3.0)	5,918	(1.9)
Ischemic heart disease	68,103	(10.0)	82,063	(21.2)	58,866	(9.4)	66,370	(19.2)	-	-	-	-
Chronic kidney disease	15,072	(2.2)	42,505	(11.0)	12,562	(2.0)	33,979	(9.8)	10,652	(1.7)	24,667	(8.1)
Cerebrovascular accidents	55,456	(8.2)	41,122	(10.6)	-	-	-	-	46,219	(7.6)	25,429	(8.4)

[a] We did not have recorded information on ethnicity from the Hong Kong Hospital Authority. The Hong Kong resident population was estimated to be 93.6% of Chinese ethnicity in 2011 by Census and Statistics Department.[22]

Table 2. Internal and external validation of HKU and HKU-SG model: model performance measured by Harrell's c-index

	Internal validation using Hong Kong cohort[a]	External validation using Singapore cohort	
		Imputation	Complete cases
Mortality			
HKU	0.845 (0.845-0.846)	-	-
HKU-SG	0.805 (0.804-0.806)	0.772 (0.770-0.774)	0.778 (0773-0.785)
Cerebrovascular disease			
HKU	0.727 (0.725-0.729)	-	-
HKU-SG	0.722 (0.720-0.725)	0.695 (0.690-0.699)	0.701 (0.692-0.710)
Ischemic heart disease			
HKU	0.717 (0.715-0.719)	-	-
HKU-SG	0.700 (0.698-0.702)	0.644 (0.640-0.647)	0.635 (0.627-0.643)

[a] Optimism corrected C-index

(95% confidence interval)

The predictor variables in the HKU risk scores are age, duration of diabetes, gender, smoking status, HbA1c, total and HDL-cholesterol, blood pressure, creatinine, hemoglobin, white cell count, resting heart rate, pre-existing conditions (cancer, congestive heart failure, diabetes complications), and medication for diabetes, hypertension and statins. The HKU-SG risk scores include all predictors common to both the Hong Kong and Singapore datasets: age, duration of diabetes, gender, smoking status, body mass index, systolic and diastolic blood pressure, HbA1c, LDL-cholesterol, pre-existing conditions (atrial fibrillation and chronic kidney disease). For further details, see Supplemental Table 1.[10]

Table 3. External validation of JADE, RECODe, and UKPDS models: model performance measured by Harrell's c-index

	Hong Kong cohort	
Mortality	**With imputation**	**Complete cases**
JADE	0.841 (0.840-0.843)	0.793 (0.786-0.800)
RECODe	0.808 (0.807-0.810)	0.785 (0.778-0.793)
Cerebrovascular disease		
JADE	0.730 (0.727-0.733)	0.716 (0.707-0.724)
UKPDS 60	0.713 (0.710-0.715)	0.680 (0.673-0.685)
UKPDS 82	0.679 (0.676-0.681)	0.646 (0.640-0.650)
Ischemic heart disease		
JADE	0.685 (0.683-0.688)	0.655 (0.647-0.664)
UKPDS 56	0.676 (0.673-0.678)	0.655 (0.650-0.661)
UKPDS 82	0.654 (0.651-0.657)	0.648 (0.644-0.652)

(95% confidence interval)

Figure Legends

Figure 1. Mean predicted and observed risks of HKU-SG models at 5 years by percentile of predicted risk for the validation cohorts

HKU-SG models for the Singapore validation cohort

Figure 2. Mean predicted and observed risks of JADE, RECODe, and UKPDS models at 5 years by percentile of predicted risk for the validation cohorts

(a) Mortality outcomes for the Hong Kong cohort (JADE and RECODe models)

(b) Cerebrovascular disease outcomes for the Hong Kong cohort (JADE and UKPDS)

(c) Ischemic heart disease outcomes for the Hong Kong cohort (JADE and UKPDS)
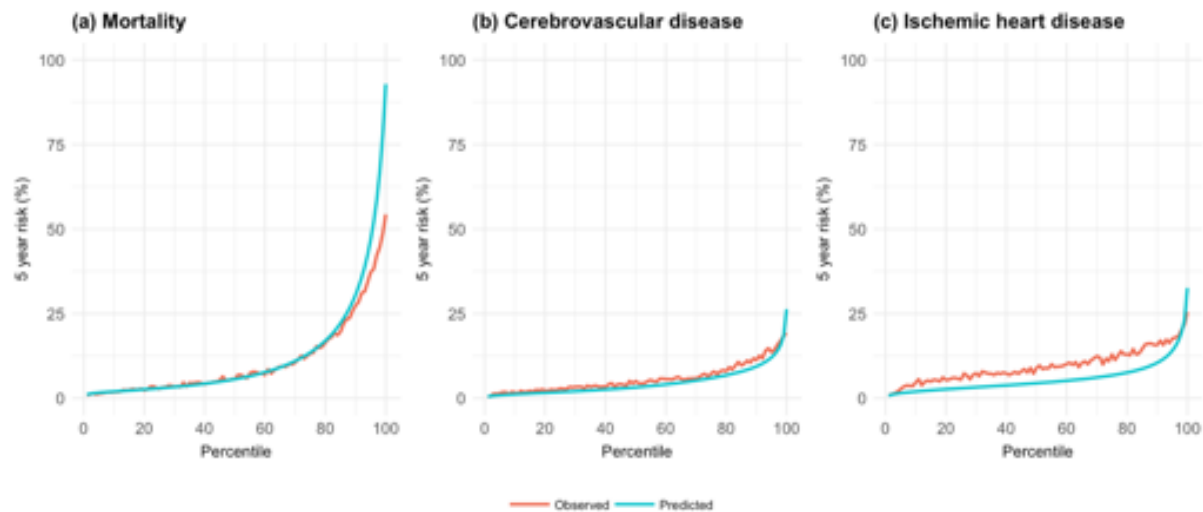
Figure 1.

Figure 2.

(a) Mortality
JADE   RECODe

B. Cerebrovascular disease
JADE   UKPDS 60   UKPDS 82

C. Ischemic heart disease
JADE   UKPDS 56   UKPDS 82

Observed   Predicted