Development and Validation of an All-Cause Mortality Risk Score in Type 2 Diabetes

The Hong Kong Diabetes Registry

Xilin Yang, PhD; Wing Yee So, MBChB, FRCP; Peter C. Y. Tong, MBBS, PhD, FRCP; Ronald C. W. Ma, MBChB, MRCP; Alice P. S. Kong, MBChB, FRCP; Christopher W. K. Lam, PhD, HonFRCPath; Chung Shun Ho, PhD; Clive S. Cockram, MD, FRCP; Gary T. C. Ko, MD, FRCP; Chun-Chung Chow, MBBS, FRCP; Vivian C. W. Wong, MBBS, FRCOG; Juliana C. N. Chan, MD, FRCP

Background: Diabetes reduces life expectancy by 10 to 12 years, but whether death can be predicted in type 2 diabetes mellitus remains uncertain.

Methods: A prospective cohort of 7583 type 2 diabetic patients enrolled since 1995 were censored on July 30, 2005, or after 6 years of follow-up, whichever came first. A restricted cubic spline model was used to check data linearity and to develop linear-transforming formulas. Data were randomly assigned to a training data set and to a test data set. A Cox model was used to develop risk scores in the test data set. Calibration and discrimination were assessed in the test data set.

Results: A total of 619 patients died during a median follow-up period of 5.51 years, resulting in a mortality rate of 18.69 per 1000 person-years. Age, sex, peripheral arterial disease, cancer history, insulin use, blood hemoglobin levels, linear-transformed body mass index, random spot urinary albumin-creatinine ratio, and estimated glomerular filtration rate at enrollment were predictors of all-cause death. A risk score for all-cause mortality was developed using these predictors. The predicted and observed death rates in the test data set were similar ($P > .70$). The area under the receiver operating characteristic curve was 0.85 for 5 years of follow-up. Using the risk score in ranking cause-specific deaths, the area under the receiver operating characteristic curve was 0.95 for genitourinary death, 0.85 for circulatory death, 0.85 for respiratory death, and 0.71 for neoplasm death.

Conclusions: Death in type 2 diabetes mellitus can be predicted using a risk score consisting of commonly measured clinical and biochemical variables. Further validation is needed before clinical use.

Arch Intern Med. 2008;168(5):451-457
from other hospitals within the region are assessed, at the earliest, 4 to 6 weeks after discharge and account for fewer than 10% of all referrals. Once a diabetic subject is enrolled, he or she will be observed until death. Approval was obtained from the Chinese University of Hong Kong Clinical Research Ethics Committee. Written informed consent was obtained from all patients for data analysis and research purposes.

All-cause death on or before July 30, 2005, was recorded or otherwise censored on July 30, 2005. Mortality data from the Hong Kong Death Registry were retrieved, and causes of death were cross-checked with hospital admissions recorded in the Hong Kong Hospital Authority computer system. These databases were matched by a unique identification number, the Hong Kong identity card number. The latter is compulsory for all residents of Hong Kong and is used by all government departments and many other organizations for personal files. Death is further classified into cause-specific deaths according to the International Classification of Diseases, Ninth Revision.

From 1993 to 2005, a total of 7920 patients with diabetes were enrolled. Of them, 332 with type 1 diabetes (defined as acute presentation with diabetic ketoacidosis, heavy ketonuria [>3+], or continuous requirement of insulin within 1 year of diagnosis) and 5 with uncertain type 1 diabetes status were excluded from the analysis. Data from the remaining 7538 patients with type 2 diabetes, who were predominantly of Chinese ethnicity, were analyzed.

CLINICAL MEASUREMENTS

Details of assessment methods, definitions, and laboratory assays have been described elsewhere. Clinical examination and laboratory investigations were performed after at least 8 hours of fasting. We used the abbreviated Modification of Diet in Renal Disease (MDRD) formula recalibrated for Chinese patients to estimate glomerular filtration rate (eGFR) expressed in milliliters per minute per 1.73 meters squared:

$$
egGR = 186 \times \left( \frac{SCR \times 0.011}{1.154} \right)^{1-0.203} \times \left( \frac{Age \times 0.742}{Female \times 1.233} \right)$$

where SCR is serum creatinine expressed as micromoles per liter (originally expressed as milligrams per liter) and 1.233 is the adjusting coefficient for Chinese. Peripheral arterial disease (PAD) was defined by lower limb amputation, revascularization for PAD, or absence of foot pulses as confirmed by an ankle-brachial ratio of less than or equal to 0.90 measured by Doppler ultrasound examination.

STATISTICAL ANALYSES

A commercially available statistical software system (SAS release 9.10; SAS Institute Inc, Cary, North Carolina) was used to perform the statistical analysis. A restricted cubic spline with 4 knots was used in univariate and multivariate Cox proportional regression analyses to check the linearity of predictors at enrollment. For variables that significantly violated linearity assumption (as indicated by the restricted cubic spline-curves), simple algebra formulas were used to improve linearity, whenever possible, to derive a risk score.

Split-half validation was used to develop the risk score. A random number between 0 and 1 was computer generated, and the data set was randomly divided into 2 data sets using a cutoff point of 0.5. The training data set was used to develop the model, and the test data set was used to validate the developed all-cause mortality risk score. In the training data set, after the linearity of predictors at enrollment were examined using univariate and multivariate analysis, Cox proportional regression analysis with a backward algorithm ($P < .30$ for stay) was used to remove predictors with a $P$ value greater than or equal to .30. The Akaike Information Criterion (AIC) has been shown to be asymptotically equivalent to the cross-validation criterion and the bootstrap. Because an automatic AIC algorithm in SAS PROC PHREG was not available, an alternative method, the Stepwise AIC Subset Blanket method, was used to select a group of variables for the predictive model. Let $LR_p$ denote the likelihood ratio $\chi^2$ and $p$ denote the number of the predictors in the final model; the estimated shrinkage is $(LR_p)/LR$, and a shrinkage below 0.85 raises concern of overfitting.

The candidate baseline variables used for model selection included clinical complications of sensory neuropathy, retinopathy, PAD, coronary heart disease (CHD), stroke, and cancer as well as conventional risk factors of age, sex, smoking status, duration of diabetes, systolic blood pressure, glycated hemoglobin (HbA1c), body mass index (BMI [calculated as weight in kilograms divided by height in meters squared]), high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, triglycerides, and spot urinary albumin-creatinine ratio (ACR). Other parameters included $eGFR$, blood hemoglobin levels (expressed in milligrams per deciliter), and white blood cell count, which predicted cardiovascular complications in our previous analyses. Treatment variables included lipid-lowering drugs, angiotensin-converting enzyme inhibitors, angiotensin II receptor blockers, other antihypertensive drugs, oral antidiabetic drugs, diet only, and insulin use at enrollment.

Based on the risk curve analysis, BMI, high-density lipoprotein cholesterol levels, white blood cell count, and ACR exhibited “regular” nonlinearity (eg, being symmetrical or having a threshold effect); therefore, linear transformations were performed for these variables before modeling. The proportional hazards assumption and functional form were checked using a supremum test, which was implemented using the ASSESS statement in the SAS procedure PROC PHREG. A $P$ value of less than .05 was considered to violate proportional hazards or to indicate that improvement in transformations remains possible. The formulas of risk scores and the 5-year probability of events from the Cox proportional hazard models have been previously described.

Validation of the risk equation was performed using the test data set. Calibration was checked using the Hosmer and Lemeshow test. The data were divided into deciles of the predicted absolute risk of all-cause death. The $\chi^2$ test (8 degrees of freedom) was constructed using the predicted and observed numbers of death stratified by deciles of predicted absolute risk over 5 years of follow-up. A $P$ value of less than .05 indicates a significant difference between the predicted and observed rates of all-cause death, suggesting poor calibration.

In survival analysis, the overall $C$ index can be regarded as a natural extension of the area under the receiver operating characteristic curve (aROC) and therefore as a measure of discrimination. However, it remains arbitrary in the selection of cutoff points using a trade-off between the sensitivity and specificity of risk scores derived from survival models. On the other hand, direct application of the aROC in survival models may be problematic because the aROC depends on follow-up time. More recently, several groups have developed algorithms to calculate time-specific sensitivity, specificity, and aROC. In this study, we used the Chambless method to calculate aROC, sensitivity, and specificity over 5 years of follow-up for checking of discrimination and selection of cutoff points. The discriminatory ability of a risk score of all-cause death is related to its ranking ability for cause-specific deaths. We therefore used the same method to estimate aROC, sensitivity, and specificity of the risk scores for cause-specific deaths over 5 years of follow-up.
RESULTS

CHARACTERISTICS OF STUDY POPULATION

At enrollment, the median age of the cohort was 57 years (interquartile range [IQR], 47-67 years), with a median duration of diabetes of 6 years (IQR, 1-11 years). During a median follow-up period of 5.51 years (IQR, 2.98-7.82 years), 10.13% (n=768) of the cohort died, resulting in a mortality rate of 18.68 (95% confidence interval, 17.37-19.99) per 1000 person-years.

As age and BMI were shown to violate the proportional hazards assumption after 6 years of follow-up, all patients followed up over 6 years were censored at 6 years, and the censored data were used to develop a 5-year all-cause mortality risk score. The training data set contained 3775 (mortality: 8.0%, or 303 deaths) and 3808 (mortality: 8.3%, or 316 deaths) patients, respectively, during the 6 years of follow-up (Table 1).

PREDICTING MODELS

Age, sex, PAD, history of cancer, insulin use, blood hemoglobin levels, and linear-transformed BMI, ACR, and eGFR were selected by the optimal AIC algorithm to be the predictors of all-cause mortality. Their hazard ratios and coefficients with 95% confidence intervals as well as assumption checking results are listed in Table 2. The linear transformations of these selected predictors in the fitted model were adequate, and the proportional hazards assumptions of the predictors were not violated. Based
on the estimated $\beta$ coefficients, the risk score and 5-year probability of all-cause mortality are constructed as follows:

\[
\text{Mortality Risk Score} = 0.0586 \times \text{Age} - 0.7049 \\
+ 0.1078 \times |\text{BMI} - 26.0| - 0.0165 \\
+ \log_{10}(1 + \text{ACR})^{1.1} \\
- 0.1913 \times \text{Hemoglobin} \\
+ 0.5698 \times \text{PAD} \\
+ 1.3384 \times \text{History of Cancer} \\
+ 0.7035 \times \text{Use of Insulin}
\]

where 0.9768 is the shrinkage of the predicting model obtained in the training data set.

**VALIDATION OF RISK SCORE**

In the test data set, the predicted and observed death rates over 5 years of follow-up were not different ($P > .70$) (Figure 1). The aROC was 0.845 for predicting all-cause death over 5 years of follow-up. Using the suggested cutoff point of 0.6873, the sensitivity was 75.4% and the specificity was 75.7%. Using the risk score (not the 5-year probability) in ranking cause-specific deaths, the aROC was 0.952 for genitourinary death, 0.854 for circulatory death, 0.849 for respiratory death, and 0.712 for neoplasm death (Figure 2). (The sensitivity and specificity of risk scores at different cutoff points for all-cause death and the 4 major cause-specific deaths, as well as a calculator of the risk score and the 5-year all-cause death probability, are available from the corresponding author.) The suggested cutoff point provides good discrimination between the high- and low-risk groups for death (Figure 3). If the PAD term was removed from the 5-year probability equation, the predicted and observed death rates remained similar ($\chi^2 = 4.6962; P > .70$) and the aROC for 5 years of follow-up was 0.845.

In Hong Kong, the life expectancy at birth in 2005 was 78.8 years for men and 84.5 years for women. We further used the same set of predictors to fit 2 separate models to predict death before (premature death, $n=131$) and after the age of 70 years (late death, $n=172$). In the premature death model, age ($P = .16$), eGFR ($P = .85$), and PAD ($P = .30$) were no longer significant predictors, while in the latter model, blood hemoglobin levels ($P = .43$) and history of cancer ($P = .36$) became nonsignificant.

**COMMENT**

Diabetes causes an excess of death. In this analysis, the derived all-cause mortality risk score has good calibration and discriminatory ability, indicating that death can be predicted using commonly measured clinical and biochemical parameters. Using the same prospective co-

---

Table 2. Parameter Estimates of the All-Cause Mortality Risk Score for Chinese Patients with Type 2 Diabetes Using the Training Data Set

<table>
<thead>
<tr>
<th>Variables</th>
<th>Parameter Estimates</th>
<th>SEM</th>
<th>HR (95% CI)</th>
<th>P Value</th>
<th>P Value for Proportional Hazards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.0586</td>
<td>0.0063</td>
<td>1.06 (1.05-1.07)</td>
<td>&lt;.001</td>
<td>.18</td>
</tr>
<tr>
<td>Sex (1 if female, 0 if male)</td>
<td>-0.7049</td>
<td>0.1493</td>
<td>0.49 (0.37-0.66)</td>
<td>&lt;.001</td>
<td>.37</td>
</tr>
<tr>
<td>BMI−26.0</td>
<td>0.1078</td>
<td>0.0253</td>
<td>1.11 (1.06-1.17)</td>
<td>&lt;.001</td>
<td>.33</td>
</tr>
<tr>
<td>Log(1 + ACR)</td>
<td>0.1649</td>
<td>0.0310</td>
<td>1.18 (1.11-1.25)</td>
<td>&lt;.001</td>
<td>.26</td>
</tr>
<tr>
<td>eGFR (coded to 60 if eGFR &gt;60)</td>
<td>-0.0165</td>
<td>0.0068</td>
<td>0.98 (0.97-1.00)</td>
<td>.02</td>
<td>.52</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>-0.1913</td>
<td>0.0480</td>
<td>0.83 (0.75-0.91)</td>
<td>&lt;.001</td>
<td>.30</td>
</tr>
<tr>
<td>Peripheral arterial disease (1 if yes, 0 otherwise)</td>
<td>0.5698</td>
<td>0.1779</td>
<td>1.77 (1.25-2.51)</td>
<td>.001</td>
<td>.42</td>
</tr>
<tr>
<td>History of cancer (1 if yes, 0 otherwise)</td>
<td>1.3384</td>
<td>0.3900</td>
<td>3.81 (2.08-6.99)</td>
<td>&lt;.001</td>
<td>.36</td>
</tr>
<tr>
<td>Insulin use at enrollment (1 if yes, 0 otherwise)</td>
<td>0.7035</td>
<td>0.1394</td>
<td>2.02 (1.54-2.66)</td>
<td>&lt;.001</td>
<td>.55</td>
</tr>
</tbody>
</table>

Abbreviations: ACR, albumin-creatinine ratio; BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); CI, confidence interval; HR, hazard ratio; eGFR, estimated glomerular filtration rate (according to the Modification of Diet in Renal Disease formula recalibrated for Chinese patients); SEM, standard error of the mean.

*The $\chi^2$ value for likelihood ratio test, 387.50; degrees of freedom, 9; $P < .001$; and shrinkage of the model=\((387.50-9)/387.50=0.9768\).

bThe HR for $|\text{BMI}-26.0|$ indicates the relative risk per kilogram per square meter departure from 26.0 kg/m$^2$.

aThe HR for $|\text{BMI}−26.0|$ indicates the relative risk per kilogram per square meter departure from 26.0 kg/m$^2$. | 

©2008 American Medical Association. All rights reserved.

Downloaded From: https://jamanetwork.com/ by a Non-Human Traffic (NHT) User on 02/02/2020
hhort, we have developed risk scores to predict stroke (aROC, 0.77) and end-stage renal disease (aROC, >0.94). In this analysis, the all-cause death risk score achieved a discrimination of up to 0.85, a level near excellence. In a Chinese community-based cohort (n = 30 121), the risk score for hard CHD events had only moderate discrimination (C statistic, 0.71 for men and 0.74 for women). In another cohort of Chinese male steel workers (n=4000), the aROCs were 0.76, 0.78, and 0.82 for CHD, ischemic, and hemorrhage stroke, respectively. Consistent with our previous findings, the good discrimination of the risk score was largely attributable to death related to genitourinary system diseases (mainly end-stage renal disease) and circulatory and respiratory diseases. Of note, while we observed different predictors for death before and after the age of 70 years, albuminuria was a common risk factor in both groups, supporting this factor for intervention in all age groups. Calibration is another important indicator of performance of risk scores. Our risk score achieved excellent calibration, possibly because of the use of spline-guided linear transformation and optimal AIC model selection approach, which resulted in little shrinkage of the fitted model.

Lee et al30 developed a prognostic index in community-dwelling older US adults (≥50 years) based on functional measures such as bathing, walking, managing money, and pushing large objects. In this analysis, we used clinical and biochemical measurements that are recommended for periodic assessments in persons with diabetes24 and identified age, male sex, BMI, ACR, eGFR, PAD, blood hemoglobin levels, use of insulin, and history of cancer as independent predictors for death. The validity of the risk score remained high even after the exclusion of PAD, which might not be routinely assessed in busy clinics. In support of our findings, both low and high BMI and microalbuminuria are strong predictors for all-cause mortality in both diabetic and nondiabetic subjects. Fried et al25 verified that kidney function predicted all-cause mortality, while others have reported the association between PAD and CHD-related death. Anemia is now confirmed to be a multiplier for death and multiple comorbidities in diabetic and nondiabetic subjects.28

In the United Kingdom Prospective Diabetes Study, glycemic control was the most powerful predictor for progression in albuminuria, which in turn predicted all-cause mortality.25 In this study, although HbA1c was significant in univariate analysis, the significance was heavily confounded by adjustment for either ACR or insulin use. As continuous variables, low-density lipoprotein cholesterol (P = .54) and triglyceride (P = .76) levels were not significant in univariate analysis but were significant in both univariate and multivariate spline Cox models. The latter is a more robust method that can detect nonlinear associations. However, because we were not able to adequately linear transform these variables, the nontransformed variables were not selected in the final model in our computing of the risk score.

Although originally developed in a white middle-class population, the Framingham CHD prediction score30 has acceptable performance when applied to a general US population, including men and women as well as white and black individuals. Liu et al31 applied the Framingham risk score to Mainland Chinese subjects and found that the original Framingham functions systematically overestimated the absolute CHD risk in Chinese individuals, although the discrimination of the Framingham risk score for Chinese subjects was as good as a risk score derived from the Chinese study cohort. Furthermore, all factors in the equation have been shown to predict death and major events in different populations. Taken together, our risk score for death may also be useful in predicting death in other Chinese and possibly non-Chinese populations with type 2 diabetes, with or without recalibration.

Several authors have debated the limitations of risk factors (or weighted risk factors such as risk scores) as prognostic tools, especially at the individual level.32,33 In our attempt to use risk factors or risk scores to identify high-risk individuals for intervention, the cutoff point of individual risk factors or weighted risk factors such

Figure 2. The receiver operating characteristic curves (ROCs) of risk scores for all-cause death and cause-specific deaths in the test data set over 5 years of follow-up. The linked-dot curve indicates all-cause death (area under the ROC [aROC], 0.845); curve 1 (the highest curve), death due to diseases of the circulatory system (aROC, 0.852); curve 2, death due to diseases of the respiratory system (aROC, 0.854); curve 3, death due to diseases of the genitourinary system (aROC, 0.854); and curve 4 (the lowest curve), death due to neoplasm (aROC, 0.712).

Figure 3. Cumulative death probability in patients with an all-cause death risk score at or above the suggested cutoff point and below the cutoff point in the test data set over 6 years of follow-up (P < .001, log-rank test).
as risk scores should be able to separate the affected from the unaffected individuals. These risk scores, if internally and externally validated, are also useful in revealing the relative importance of individual risk factors and the complexity of their interactions in predicting outcomes. Furthermore, these risk scores may serve as an initial tool with which to identify high-risk individuals using readily available clinical and laboratory parameters for referral to specialized centers for more specific diagnostic testing and intensified management. In this regard, several cohort-based and randomized studies have shown the marked benefits of multidisciplinary disease management on mortality and morbidity rates. Therefore, in agreement with the authors' comments on the Framingham risk score, the limited predicting performance of a risk score at an individual level does not negate its values in advancing our knowledge of the disease, in generating a new hypothesis, and in calling for more research to improve its predictive values for clinical outcomes.

Our study has several limitations. First, apart from the variables in the equation, other factors such as low-density lipoprotein cholesterol and triglyceride levels might be important, although we were unable to devise adequate linear transformations for these variables in Cox models. The complicated relationships may also explain their noninclusion in the model in the first place. Second, we do not have migration data, although patients with chronic diseases such as diabetes often do not have medical insurance coverage and are less likely to emigrate to or be accepted by other countries as permanent residents. Therefore, major biases due to immigration are most unlikely. Third, despite the registry being set up in a hospital clinic, because of the lack of a comprehensive health insurance policy and integrated primary health care system in Hong Kong, the majority of patients, especially those with chronic illnesses, are treated in public hospitals, where care is heavily subsidized. In 2000, the Department of Health of Hong Kong conducted a survey and reported that more than 90% of patients diagnosed with diabetes were treated in the public health sector. The representativeness is further supported by an annualized rate of 16.43 per 1000 person-years for mortality and 14.08 per 1000 person-years for incident CHD. These figures are similar to those reported in several community-based databases.

In conclusion, we have developed an all-cause mortality risk score in type 2 diabetes with good accuracy in calibration and discrimination in the test data set. Given the highly prevalent and treatable nature of many of the risk predictors, our derived risk score may be clinically useful after it is adequately validated by other cohorts.

Accepted for Publication: October 1, 2007.

Correspondence: Juliana C. N. Chan, MD, FRCP, 9/F Department of Medicine and Therapeutics, Prince of Wales Hospital, Chinese University of Hong Kong, Shatin, Hong Kong SAR, China (jchan@cuhk.edu.hk).

Author Contributions: Drs Yang and So contributed equally to the manuscript. Study concept and design: Cockram, Ko, and Chan. Acquisition of data: So, Tong, Ma, Kong, Lam, Ho, Chow, Wong, and Chan. Analysis and interpretation of data: Yang, Tong, Cockram, and Chan.

Acknowledgments: This study was partially supported by an MSD University Grant and the Hong Kong Foundation for Research and Development in Diabetes, established under the auspices of the Chinese University of Hong Kong. Additional Contributions: L. Y. Tse, MBBS, Disease Surveillance and Health Promotion, Department of Health, Hong Kong Government, gave assistance and advice. The study patients were recruited and treated by the medical and nursing staff at the Prince of Wales Hospital Diabetes Center.

REFERENCES


13. So WY, Kong AP, Ma RC, et al. Glomerular filtration rate, cardiovascular endpoints,


