Development and validation of prediction models for the risks of diabetes-related hospitalization and in-hospital mortality in patients with type 2 diabetes

Tsai-Chung Li, Chia-Ing Li, Chiu-Shong Liu, Wen-Yuan Lin, Chih-Hsueh Lin, Sing-Yu Yang, Jen-Huai Chiang, Cheng-Chieh Lin

PII: S0026-0495(18)30042-8
DOI: doi:10.1016/j.metabol.2018.02.003
Reference: YMETA 53739

To appear in:

Received date: 6 October 2017
Accepted date: 9 February 2018

Please cite this article as: Tsai-Chung Li, Chia-Ing Li, Chiu-Shong Liu, Wen-Yuan Lin, Chih-Hsueh Lin, Sing-Yu Yang, Jen-Huai Chiang, Cheng-Chieh Lin , Development and validation of prediction models for the risks of diabetes-related hospitalization and in-hospital mortality in patients with type 2 diabetes. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Ymeta(2018), doi:10.1016/j.metabol.2018.02.003

Development and validation of prediction models for the risks of diabetes-related hospitalization and in-hospital mortality in patients with type 2 diabetes

Tsai-Chung Li, PhD,[1,2]†, Chia-Ing Li, PhD,[1,2]†, Chiu-Shong Liu, MD,[3,4,5], Wen-Yuan Lin, MD, PhD,[3,5], Chih-Hsueh Lin, MD,[3,5], Sing-Yu Yang, BS,[1], Jen-Huai Chiang, MD,[6], Cheng-Chieh Lin, MD, PhD,[3,4,5*]

1. Department of Public Health, College of Public Health, China Medical University, Taichung, Taiwan
2. Department of Healthcare Administration, College of Medical and Health Science, Asia University, Taichung, Taiwan
3. School of Medicine, College of Medicine, China Medical University, Taichung, Taiwan
4. Department of Medical Research, China Medical University Hospital, Taichung, Taiwan
5. Department of Family Medicine, China Medical University Hospital, Taichung, Taiwan
6. Management Office for Health Data, China Medical University Hospital, Taichung, Taiwan

* Correspondence to: Cheng-Chieh Lin, China Medical University, 91 Hsueh-Shih Road, Taichung, 40421, Taiwan, Tel: 886-4-22052121ext. 1013, Fax: 886-4-22060248, e-mail: cclin@mail.cmuh.org.tw

†: equally contributed as first authors.

Short title: Risk score prediction model for diabetes-related hospitalization and in-hospital mortality

Abstract

Objectives: Diabetes is a major cause of hospitalization and in-hospital mortality. However, a scoring system that can be used to identify diabetic patients at risk of diabetes-related hospitalization and in-hospital mortality is lacking.

Methods: We included 32,653 patients in this retrospective cohort study. All recruited patients had type 2 diabetes, were 30–84 years of age, and were enrolled in the National Diabetes Care Management Program over the period of 2001–2003. We used the Cox proportional hazard regression model to derive risk scores. The predictive accuracy of the models were evaluated using receiver operating characteristic curves. We conducted the Hosmer–Lemeshow test to assess the agreement between predicted and observed risks.

Results: Over a follow-up period of eight years, 6,243 patients were hospitalized for diabetes-related events, and 2,048 deaths were registered in hospital records. For the one-, three-, five-, and eight-year periods, the areas under the curve (AUC) for diabetes-related hospitalization in the validation set were 0.80, 077, 0.76, and 0.74, respectively. The corresponding values for in-hospital mortality in the validation set were 0.87, 080, 0.77, and 0.76. The goodness-of-fit test showed that the predicted and observed probabilities in the one-, three-, five-, and eight-year periods were similar for diabetes-related hospitalization and in-hospital mortality in the validation set (all p values > 0.05).

Conclusion: We developed models for the estimation of the risks of diabetes-related hospitalization and in-hospital mortality in patients with type 2 diabetes. The models may be used to identify diabetic patients who are at high risk for hospital admission and in-hospital mortality.

Key words: diabetes-related hospitalization; in-hospital mortality; type 2 diabetes; prediction model

Background

In recent decades, prevalence of diabetes has dramatically increased in developed and developing countries. Diabetes is an ambulatory care-sensitive condition (ACSC). Timely, effective, and efficient outpatient care can decrease hospitalization risks for patients with ACSCs [1, 2]. Epidemiological evidence suggests hyperglycemia determined by glycated hemoglobin A1C (HbA1c) and glucose variability are associated with ACSC hospitalizations [3, 4] In addition to ACSC hospitalizations, in-hospital mortality is an important performance index for quality improvement.

The development of a prediction model for hospitalization or in-hospital mortality in patients with type 2 diabetes will facilitate the prevention of hospitalizations or in-hospital mortality that may consequently reduce health care spending and improve the quality of care as well as quality of life. In addition, it can facilitate the screening of high-risk patients, help guide preventative interventions, and plan interventions and future health care needs. Only one study has been conducted on patients with type 2 diabetes with a focus on heart failure hospitalization [5] and one in-hospital mortality [6]. The latter study is limited by its small sample size and failure to consider several important diabetes-related variables, such as HbA1c and glycemic variation. No prior study has developed a prediction model for diabetes-related hospitalization, including hypoglycemia in addition to diabetes. The development of prediction models for diabetes-related hospitalization and in-hospital mortality in Chinese patients with type 2 diabetes is urgently needed to address the aforementioned research gap. Thus, we aimed to develop prediction models for acute diabetes-related hospitalization and in-hospital mortality using a nationwide cohort in Taiwan.

Methods

Data sources

The National Diabetes Care Management Program (NDCMP), which was established by the National Health Insurance (NHI) Program in 2001, enrolls patients with type 1 and type 2 diabetes. The NDCMP is a nurse case management program. A nationwide retrospective cohort study was conducted among patients with type 2 diabetes who were enrolled in the NDCMP over the period of 2001–2004. The date of entry was defined as the index date. The subjects were followed up until 31 December 2011 and were monitored for withdrawal from the NHI program, death, or development of events (diabetes-related hospitalization or in-hospital mortality). Patients with less than one year of follow-up were excluded. We used the NDCMP and National Health Insurance Research Database (NHIRD) database to construct a cohort of patients with type 2 diabetes. Each individual in Taiwan has a unique personal identification number (PIN). For security and privacy purposes, the identity data of patients who are enrolled in the NHIRD and NDCMP are scrambled cryptographically through the same approach. All NHI datasets can be interlinked through the scrambled PIN of each patient. We combined the datasets of NDCMP and NHIRD, including the details of ambulatory care orders of NHIRD from 2001–2004, to acquire information on the baseline characteristics of the patients. Baseline characteristics included socio-demographic factors; duration of type 2 diabetes; age of onset, diabetes-related factors and biomarkers; comorbidity; and types of anti-diabetes, hypertension, cardiovascular, and hyperlipidemia medications. We then used the 2001–2011 inpatient and outpatient databases of NHIRD to obtain data on subsequent events starting beginning from one year after the index date to 2011. Events one year after the index date were counted to rule

out the possibility of reverse causality. The proportion of enrollees withdrawing from the NHI program is low given the comprehensive coverage of the NHI program. Therefore, bias arising from loss to follow-up is negligible.

Study subjects

Initially, 63,084 enrolled patients with diabetes were identified from the NDCMP program, named as Taiwan Diabetes Study, on the basis of the criteria of the American Diabetes Association (International Classification Disease, Ninth Revision, Clinical Modification, abbreviated as ICD-9CM; Code of 250). This NDCMP program aims to increase the quality of diabetes care by increasing monitoring frequency, providing continuity of care, and decreasing diabetes-related complications. Patients who were included in the study had to have at least one year of follow-up to enable the calculation of visit-to-visit variations in HbA1c. In addition, patients must not have missing information regarding baseline characteristics, comorbidities, and laboratory blood test results. We excluded patients who had type 1 diabetes (ICD-9-CM; code 250.x1/x3) and gestational diabetes (n = 2,108) and those aged under 30 years or above 85 years (n = 1,025). We also excluded participants with missing data on socio-demographic factors, lifestyle behaviors, and blood biochemical indices, as well as those with less than one year of follow-up (n = 27,298) to rule out the possibility of reverse causality from the analysis. A total of 32,653 participants were retained for the analysis (15,213 men and 17,440 women) (Figure 1). These 32,653 participants were randomly assigned to a derivation set (n = 21,769) or a validation set (n = 10,884) at a 2:1 ratio. Ethics approval was obtained from the Ethical Review Board of the China Medical University Hospital.

Outcome ascertainment

In-hospital mortality was set as the primary outcome measure and was determined through record linkage with the inpatient care dataset and was confirmed by enrollment dataset in the NHIRD. The second outcome measure was diabetes-related hospitalization, which was defined as acute admission to an acute hospital, excluding elective admissions or planned rehabilitation based on major discharge diagnoses code from ICD-9: diabetes (ICD-9-CM codes 250) [7]. Hospitalization was determined through record linkage with the inpatient care dataset. The time of follow-up began with recruitment (index date) and ended with death during the hospitalization period or a new hospitalization event, withdrawal from the NHI program, or end of follow-up on December 31, 2011.

Covariates

Data on other baseline chronic medical conditions were retrieved from outpatient and inpatient claim data at the 24-month period preceding the cohort entry. Instead of a 12-month period, a 24-month period was used because some chronic medical conditions are not common and a long period was required so as not to miss any diagnosis. The histories of acute hypertension (ICD-9-CM codes 404-405), stroke (ICD-9-CM codes 431-438), cardiovascular disease (ICD-9 code 410 to 413, 414.01 to 414.05, 414.8 and 414.9), peripheral arterial disease (ICD-9-CM codes 443.9), peripheral neuropathy (ICD-9-CM codes 356), diabetes retinopathy (ICD-9-CM codes 362.0), peripheral circulatory disturbance (ICD-9-CM codes 250.7), chronic kidney disease (ICD-9 codes 585), traumatic amputation (ICD-9-CM codes 895.x-897.x), ketoacidosis (ICD-9-CM codes 250.1), postural hypotension (ICD-9-CM codes 458), arterial embolism and thrombosis (ICD-9-CM codes 444), and hyperlipidemia (ICD-9-CM codes 272) before the index date were identified as comorbidities.

Socioeconomic factors included age and gender. Age was treated as a continuous variable, which is the default in the algorithm used in the Framingham Heart study. Gender was categorized as male and female. Lifestyle behaviors included smoking (yes, no), alcohol consumption (yes, no), duration of diabetes (continuous), body mass index (BMI; body weight divided by height^2, kg/m^2), and early onset of diabetes (yes if age of onset > 45 years, no). Blood biochemical indexes consisted of HbA1c, fasting plasma glucose (FPG), triglyceride (TG), high-density lipoprotein (HDL), low-density lipoprotein (LDL), serum glutamic-pyruvic transaminase (SGPT), and chronic kidney disease (CKD) (yes if the estimated glomerular filtration rate < 60 [mL/min/1.73 m^2], no). The coefficient of variation (CV) for HbA1c or FPG measurements from outpatient visits within the first year of the index date for each patient was calculated for those with more than two HbA1c or FPG measurements performed in the first year. The CV of HbA1c or FPG was divided by the square root of the ratio of total visits divided by total visits minus 1 [8] to account for the likelihood that the number of visits might affect glucose variation.

Data on the use of medications prescribed for disease treatment were derived for the 12-month period preceding the cohort entry. We identified the subjects' outpatient prescriptions within one year of their enrollment to define their use of anti-diabetes, anti-hypertension, anti-hyperlipidemia, and cardiovascular medications. Anti-diabetes medications were classified into no medication, sulfonylureas (SUs), including SU monotherapy and SUs plus other oral agents, non-SU oral agents, insulin, insulin plus SUs, and insulin plus non-SU oral agents. A patient is defined as a user of a specific medication if his/her number of prescription days for this specific medication is greater than 90 days.

Statistical analysis

Categorical variables were presented as proportions, and continuous variables were presented as means and standard deviations (SDs). We calculated the standardized effect sizes to measure the differences in the baseline characteristics of the derivation and validation sets. Cox's proportional hazard models were adopted to estimate the hazard ratios of the predictive variables to develop prediction models of diabetes-related hospitalization and in-hospital mortality in the derivation set and to assess the models' predictive accuracy in the validation set. We selected independent variables that resulted in the "best" model through the following steps [9]: First, we performed a careful univariable analysis of each variable. Second, we selected variables with univariable tests of p-value<0.25 [10, 11] as candidates for our multivariable model. Third, candidate variables were entered simultaneously into the multivariable model. To assess whether some co-morbidity variables were highly collinear, we estimated their regression coefficients and compared their significance. Only one such highly correlated variable remained in the multivariate Cox model. Finally, after refining the main effect model, we checked the assumption of Cox's proportional hazard model for all variables in our multivariate model. Then, we further examined the interactions between independent variables.

We followed the steps proposed in the Framingham Heart study [12] to develop the predictive model for determining the risk scores for diabetes-related hospitalization and in-hospital mortality. The seven steps were as follows: (1) estimation of the parameters of the multivariable Cox's proportional hazard model; (2) categorization of the risk factors and determination of their reference values W_i; (3) assignment of a score for each category to determine the referent risk factor

profile with a base category of 0 for each risk factor; (4) determination of the distance from the base category to each category in regression units; (5) setting the constant B, the number of regression units that reflect 1 point in the final point system; (6) calculation of the number of points for each category of each risk factor, where $\text{Point}_{ij} = \beta_i (W_{ij} - W_{iREF})/B$; and (7) determination of the prediction risks for all possible total scores. The risks of diabetes-related hospitalization and in-hospital mortality were calculated with the following equation: $1 - P_0^{\exp(\sum \beta_i \times X_i - \sum \beta_i \times \overline{M}_i)}$, where P_0 is the baseline diabetes-related hospitalization or in-hospital death-free probability, β_i is the regression coefficient for X_i, and $\overline{M_i}$ is the mean level of X_i. Predictive accuracy was assessed through a receiver operating characteristic (ROC) curve analysis. The area under curve (AUC) was used to determine the discriminatory ability of the predictive model. Goodness-of-fit tests were performed by comparing the observed and predicted events of diabetes-related hospitalization or in-hospital mortality using the Hosmer–Lemeshow x^2 test. We used the multiple imputation method to impute missing data as sensitivity analysis to assess whether our results were sensitive to missing data. For internal validation, we corrected the potential for over-fitting or "optimism" by using a 1000-time bootstrap resampling approach. The agreement between the model-predicted probabilities and the observed probabilities was determined to assess model calibration. The intercept was calculated through the calibration-in-large method to assess whether the predictions were systematically too low or too high. An intercept value of 0 indicated the absence of systematic deviation in the estimated predicted probabilities. Furthermore, the estimated calibration slope was used to assess the extremeness of the predicted probabilities. A slope value approaching 1 indicated no overfitting in the model. We performed statistical analysis using SAS version

9.4 (SAS Institute Inc., Cary, NC). The level of significance was set at two-tailed p-value<0.05.

Results

After eight years of follow-up, 6,243 diabetes-related hospitalization cases and 2,048 in-hospital deaths occurred. The baseline characteristics of the cases were summarized in accordance with the derivation and validation sets (Table 1). The 21,769 patients in the derivation set had a mean age of 61.1 years (SD of 10.8 years), 46.5% of these patients were male, and 1378 patients died during follow-up period. Cancer was the leading cause of death (n=301 [21.84%]), followed by CVD (n=220 [15.97%]), pneumonia (140 [10.16%]), diabetes (30 [2.18%]), and CKD (14 [1.02%]). The 10,884 patients in the validation set had a mean age of 61.1 years (SD of 10.8 years), 46.7% of these patients were male and 670 patients died during follow-up period. Cancer was again the leading cause of death (n=133 [19.85%]), followed by CVD (n=135 [20.15%]), pneumonia (78 [11.64%]), diabetes (15 [2.24%]), and CKD (7 [1.04%]). All standardized differences were less than 0.1, indicating a negligible difference in the mean or the prevalence of baseline characteristics between the derivation and validation sets.

The significant baseline predictors of univariate and multivariate Cox's proportional hazard models for diabetes-related hospitalization and in-hospital mortality are shown in Table 2. The following significant predictors of diabetes-related hospitalization and in-hospital mortality were the same but had different risk estimates: age, BMI, hospitalization status one year prior at baseline, FPG-CV, HbA1c-CV, creatinine, total-cholesterol-to-HDL ratio, stroke, diabetes retinopathy, and anti-diabetes medications. Risk scores were assigned to each of the final predictors (Table 3) in accordance with the results of Cox's proportional hazard

models. The significant predictors of diabetes-related hospitalization included the following: age (−2–8 points), duration of type 2 diabetes (0–6 points), BMI ≥ 22.5 kg/m^2 (−2 point), hospitalization status one year prior at baseline (5 points), HbA1c ≥ 7% (3 points), FPG-CV (0–3 points), HbA1c-CV (0–2 points), creatinine (0–5 points), total-cholesterol-to-HDL ratio (0–2 points), stroke (2 points), diabetes retinopathy (4 points), anti-diabetes medications (0–11 points), and cardiovascular medications (1 point). The significant predictors of in-hospital mortality included age (−2–8 points), male (2 points), BMI ≥ 22.5 kg/m^2 (−1 point), hospitalization status one year prior at baseline (1 point), HbA1c-CV>17.5 % (2 points), creatinine (0–2 points), total cholesterol to HDL ratio (0–2 points), stroke (1 point), diabetes retinopathy (1 point), hypoglycemia (1 point), and anti-diabetes medications (0-2 point).

Diabetes-related hospitalization risk scores ranged from −4 to 52, and in-hospital mortality risk scores ranged from −3 to 21. The risks of diabetes-related hospitalization ranged from 0.29% to 96.92% within a one-year period, from 0.76% to 99.99% within a three-year period, from 1.12% to 100% within a five-year period, and from 1.65% to 100% within an eight-year period. In-hospital mortality ranged from 0.01% to 53.80% within a one-year period, from 0.05% to 99.66% within a three-year period, from 0.12% to 100% within a five-year period, and from 0.25% to 100% within an eight-year period (Supplemental table 1).

As shown in Figure 2, the AUCs of diabetes-related hospitalization for one-year, three-year, five-year, and eight-year periods were 0.80, 077, 0.76 and 0.74, respectively, in the validation set. These values indicated that the diabetes-related hospitalization predictive model has good predictive ability. Figure 3 shows that the AUCs of in-hospital mortality for one-year, three-year, five-year, and eight-year periods were 0.87, 0.80, 0.77, and 0.76, respectively, in the validation set. These

values demonstrated that the predictive model of in-hospital mortality showed good discriminatory ability. The predicted number of diabetes-related hospitalization cases or in-hospital deaths according to deciles were similar to the observed events in the one-, three-, five-, and eight-year risk predictions for the validation set (Supplemental Figures 1–2). The goodness-of-fit test showed good calibration for both predictive models (all p-values>0.05).

Discussion

To the best of our knowledge, this study is the first to establish prediction models for the risks of diabetes-related hospitalization and in-hospital mortality in Taiwanese patients with type 2 diabetes. We combined traditional risk factors and diabetes-related biomarkers to develop valid prediction models for ACSC hospitalizations among patients with type 2 diabetes. The two models showed acceptable and good discrimination and calibration in the validation sets. Thus, these two tools may be applied to decrease the costs of hospital admission by enabling timely intervention and appropriate care management.

Although the impact of risk factors for hospital admission is complex, some factors have clearly defined effects. Therefore, prediction models for the risk of hospital admission have been developed for various settings and populations. The AUC values of the most current prediction models in the literature for the risk of hospital admission are between 0.61–0.71, indicating that these models have poor or average performances; these models are designed for either older or elderly general populations [13-15] or patients with obstructive lung disease [16]. Moreover, these models suggest the need for detecting important predictors that have a major role in hospitalization risk. The four existing models reported in the literature do not include blood-based biomarkers and drug-related information. By contrast, our proposed

models focus on the prediction of diabetes-related hospitalization and in-hospital mortality in patients with type 2 diabetes by including blood-based biomarkers and drug-related information. The importance of HbA1c as a predictor for in-hospital mortality had been demonstrated in patients with ischemic stroke [17]. Our study further demonstrate variation in FPG and HbA1c are important predictors for hospitalization and in-hospital mortality. Our prediction models for short-term diabetes-related hospitalization and in-hospital mortality showed AUC values greater than 0.80, which indicated they had good discrimination ability to identify diabetic patients at high risk for hospital admission and interventions. This favorable feature of the models may decrease avoidable hospital admissions and enable the provision of case management in the most appropriate settings.

Before we developed prediction models, we have examined the effects of all variables carefully in the multivariate model for their plausibility by comparing their effects with those in literature for supporting evidence in light of consistency. We found the effects of all variables are consistent with those reported in literature except for BMI. We found the effects of BMI had conflict findings across different ethnic groups. A study exploring the relationship between BMI and all-cause mortality in patients with type 2 diabetes from the Nurses' Health Study and Health Professionals Follow-up Study, they reported a U-shape relationship across BMI categories [18]. It indicates that both extreme BMI values (i.e., BMI: 18.5-22.4, 30.0-34.9, and >=35 kg/m^2) were associated with increased risks of mortality. On the contrary, A prior study found a protective effect for type 2 diabetes patients with greater values of BMI [2], which is consistent with our findings. In this study analyzing data from National Health Interview Survey in Taiwan, they reported BMI≥24 kg/m^2 was associated with decreased risk of all-cause mortality in patients with diabetes compared with those

with BMI <24 kg/m^2. The possible explanation for the protective effect of BMI ≥25 kg/m^2 in our study is that the majority of patients in BMI ≥25 kg/m^2 were between 25-30 kg/m^2 (about 80%), and its effect was similar to those with normal weight, which was consistent with those observed in Tobias et al. The effect of BMI ≥25 kg/m^2 became protective when comparing with those with underweight who had the highest risks. We found there were some common and specific factors for hospitalization and in-hospital mortality. The factors in these two prediction models can provide individual estimates of risks and serve as guidance for the clinical management of high risk patients. The use of the risk scores for risk stratification can be helpful to clinicians or health policy providers to provide interventions such as diets and medication to reduce hospitalization and in-hospital death risks and health costs.

A previous study has shown that patients with diabetes are more likely to face frequent and longer hospitalization than those without diabetes. Approximately 25% of patients with diabetes have been hospitalized at least once [19]. Given the specific clinical features of diabetes, research on prediction models for hospitalization risk should focus on diabetic populations. A study from Hong Kong developed a risk-scoring system for predicting hospitalization for heart failure in patients with type 2 diabetes [5]. A study from Spain study developed tools for the prediction of in-hospital mortality in patients with type 2 diabetes [6]. However, the former study is limited given that it only included hospitalization events as a result of heart failure. Furthermore, the latter study included a small sample size and ignored several important diabetes-related variables, such as duration of diabetes, age of onset, and biomarkers of HbA1c and glycemic variation. Our study included FPG-CV and HbA1c-CV to improve the predictive ability of the models. In our study, most of the

included predictors, such as age, BMI, HbA1c, stroke, hospitalization during the previous 12 months, and insulin use, are supported by previous predictive models for hospitalization [5, 6, 13, 15, 20-22]. In addition, other diabetes-related indicators [7, 23-26], such as FPG-CV [23], HbA1c-CV [26], creatinine [25], total cholesterol to HDL ratio [25], and hypoglycemia [24], have been reported as significant predictors of hospitalization. Thus, the outcomes of hospitalizations in patients with type 2 diabetes may be avoided through the regulation of blood sugar, lipid profile, and serum creatinine. Regular blood test monitoring among patients with type 2 diabetes is also essential to decrease the likelihood of avoidable hospitalization.

We conducted a sensitivity analysis with a multiple imputation method to impute missing data. A total of 52,623 patients with type 2 diabetes were included in the sensitivity analysis. In the sensitivity analysis, the AUCs of the one-year, three-year, five-year, and eight-year risks of diabetes-related hospitalization were 0.79, 0.76, 0.75, and 0.73, respectively, whereas those of in-hospital mortality were 0.84, 0.78, 0.77, and 0.75, respectively. The results of the ROC curves obtained through the sensitivity analysis were similar to those in the original analysis, thereby demonstrating the robustness of our results. The internal validation of the performance of the present model was assessed on the basis of 1000 samples from a bootstrap resampling approach. The optimism-corrected calibration intercepts were 0.03 and 0.02, and the corresponding slopes were 0.92 and 0.91 for diabetes-related hospitalization risk and in-hospital mortality, respectively. These statistics indicated the good calibration of the present model. Moreover, the shrinkage of regression coefficients is no longer necessary in the prediction model.

Strengths and limitations

The advantages of this study include the nationwide population-based cohort

investigation and the inclusion of novel predictors of Hba1c-CV and FPG-CV. We used comprehensive health care claims data, which are reliable and practice-based information sources, to build the predictive models. In addition, we used a bootstrapping resampling approach to evaluate internal validation and assess the validity of the proposed risk prediction models.

This study has four limitations. First, predictors were measured at the baseline, and the possible time-varying effects of the predictors cannot be reported. Second, all patients with type 2 diabetes in this study were enrolled in a nationwide NDCMP in Taiwan. Thus the results of our study may not be generalized to other Asian populations because of different genetic background and health care systems. Genetic background plays an important role in incidence and complications of type 2 diabetes. In addition, healthcare systems may have different criteria/thresholds for hospitalization due to diabetes. These two factors may limit our external generalization. However, these two factors do not invalidate our results in predicting diabetes-related hospitalization and in-hospital mortality in Chinese persons with type 2 diabetes. For generalizing these two prediction models to other populations, external validation prior to their wide application should be evaluated. Third, information on hospitalization, comorbidity, and medication use was derived from claims databases. The direct input of such information would be required in future applications. In addition, the influence of different methodological approaches to data retrieval on the performance of the prognostic index remains unknown. Fourth, Glucagon-like peptide-1 (GLP-1) and Dipeptidyl peptidase-4 (DPP-4) inhibitors were not considered in the prediction model due to the time frame of the study.

Conclusion

Our study developed and validated two models that can effectively predict

diabetes-related hospitalizations and in-hospital mortality among patients with type 2 diabetes. These two validated predictive models are the first simple grading scales for diabetes-related hospitalizations and in-hospital mortality in Taiwan. These models may decrease the rates of avoidable hospitalizations or in-hospital mortality by helping primary care providers identify high-risk patients with diabetes who require interventions.

FUNDING INFO

This study was supported primarily by the Bureau of National Health Insurance (DOH94-NH-1007), the Ministry of Science and Technology of Taiwan (NSC 101-2314-B-039 -017-MY3 & NSC 102-2314-B-039-005-MY2 & MOST 104-2314-B-039-016, MOST 105-2314-B-039-021-MY3 & MOST 105-2314-B-039-025 -MY3) and Taiwan Ministry of Health and Welfare Clinical Trial and Research Center of Excellence (MOHW106-TDU-B-212-113004) and China Medical University (CMU105-S-22).

AUTHOR CONTRIBUTIONS

TCL: guarantor of integrity of the entire study, study design, manuscript editing and manuscript review. CIL: study concepts, data analysis, and manuscript editing. CSL: manuscript preparation, data collection and manuscript revision. WYL: manuscript preparation, data collection and manuscript revision. CHL: manuscript preparation, data collection and manuscript revision. SYY: data interpretation, data analysis and manuscript revision. JHC: data interpretation, data analysis and manuscript revision. CCL: study design, manuscript editing and manuscript review.

All authors made substantial contributions to the intellectual content of the manuscript and have approved the final submitted version.

ACKNOWLEDGEMENTS

None.

CONFLICT OF INTEREST

None.

References

[1] Billings J, Zeitel L, Lukomnik J, Carey TS, Blank AE, Newman L. Impact of socioeconomic status on hospital use in New York City. Health Aff (Millwood). 1993;12:162-73.

[2] Kornelius E, Huang CN, Yang YS, Lu YL, Peng CH, Chiou JY. Diabetes-related avoidable hospitalizations in Taiwan. Primary care diabetes. 2014;8:330-7.

[3] Rahbar S. The discovery of glycated hemoglobin: a major event in the study of nonenzymatic chemistry in biological systems. Ann N Y Acad Sci. 2005;1043:9-19.

[4] Warner EA, Ziboh AU. The effects of outpatient management on hospitalization for ambulatory care sensitive conditions associated with diabetes mellitus. South Med J. 2008;101:815-7.

[5] Yang X, Ma RC, So WY, Kong AP, Ko GT, Ho CS, et al. Development and validation of a risk score for hospitalization for heart failure in patients with Type 2 diabetes mellitus. Cardiovasc Diabetol. 2008;7:9.

[6] Ramirez-Prado D, Palazon-Bru A, Folgado-de-la Rosa DM, Carbonell-Torregrosa MA, Martinez-Diaz AM, Gil-Guillen VF. Predictive models for all-cause and cardiovascular mortality in type 2 diabetic inpatients. A cohort study. Int J Clin Pract. 2015;69:474-84.

[7] Li TC, Kardia SL, Li CI, Chen CC, Liu CS, Yang SY, et al. Glycemic control paradox: Poor glycemic control associated with higher one-year and eight-year risks of all-cause hospitalization but lower one-year risk of hypoglycemia in patients with type 2 diabetes. Metabolism: clinical and experimental. 2015;64:1013-21.

[8] Kilpatrick ES, Rigby AS, Atkin SL. A1C variability and the risk of microvascular complications in type 1 diabetes: data from the Diabetes Control and Complications Trial. Diabetes Care. 2008;31:2198-202.

[9] Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd ed: Wiley-Inter science Publication,; 2000.

[10] Mickey RM, Greenland S. The impact of confounder selection criteria on effect estimation. Am J Epidemiol. 1989;129:125-37.

[11] R.B. B, A.A. A. Comparison of Stopping Rules in Forward "Stepwise" Regression. Journal of the American Statistical Association. 1977;72:46-53.

[12] Sullivan LM, Massaro JM, D'Agostino RB, Sr. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. Stat Med. 2004;23:1631-60.

[13] Crane SJ, Tung EE, Hanson GJ, Cha S, Chaudhry R, Takahashi PY. Use of an electronic administrative database to identify older community dwelling adults at high-risk for hospitalization or emergency department visits: the elders risk assessment index. BMC Health Serv Res. 2010;10:338.

[14] O'Caoimh R, Gao Y, Svendrovski A, Healy E, O'Connell E, O'Keeffe G, et al. The Risk Instrument for Screening in the Community (RISC): a new instrument for predicting risk of adverse outcomes in community dwelling older adults. BMC Geriatr. 2015;15:92.

[15] Shelton P, Sager MA, Schraeder C. The community assessment risk screen (CARS): identifying elderly persons at risk for hospitalization or emergency department visit. Am J Manag Care. 2000;6:925-33.

[16] Beijers RJ, van den Borst B, Newman AB, Yende S, Kritchevsky SB, Cassano PA, et al. A Multidimensional Risk Score to Predict All-Cause Hospitalization in Community-Dwelling Older Individuals With Obstructive Lung Disease. J Am Med Dir Assoc. 2016;17:508-13.

[17] Tziomalos K, Dimitriou P, Bouziana SD, Spanou M, Kostaki S, Angelopoulou SM, et al. Stress hyperglycemia and acute ischemic stroke in-hospital outcome. Metabolism: clinical and experimental. 2017;67:99-105.

[18] Tobias DK, Pan A, Jackson CL, O'Reilly EJ, Ding EL, Willett WC, et al. Body-mass index and mortality among adults with incident type 2 diabetes. The New England journal of medicine. 2014;370:233-44.

[19] Donnan PT, Leese GP, Morris AD. Hospitalizations for people with type 1 and type 2 diabetes compared with the nondiabetic population of Tayside, Scotland: a retrospective cohort study of resource use. Diabetes Care. 2000;23:1774-9.

[20] McAna JF, Crawford AG, Novinger BW, Sidorov J, Din FM, Maio V, et al. A predictive model of hospitalization risk among disabled medicaid enrollees. Am J Manag Care. 2013;19:e166-74.

[21] Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. Am J Epidemiol. 2011;173:676-82.

[22] Schembri S, Anderson W, Morant S, Winter J, Thompson P, Pettitt D, et al. A predictive model of hospitalisation and death from chronic obstructive pulmonary disease. Respir Med. 2009;103:1461-7.

[23] Akirov A, Diker-Cohen T, Masri-Iraqi H, Shimon I. High Glucose Variability Increases Mortality Risk in Hospitalized Patients. J Clin Endocrinol Metab. 2017.

[24] Borzi V, Frasson S, Gussoni G, Di Lillo M, Gerloni R, Augello G, et al. Risk factors for hypoglycemia in patients with type 2 diabetes, hospitalized in internal

medicine wards: Findings from the FADOI-DIAMOND study. Diabetes research and clinical practice. 2016;115:24-30.

[25] Tomlin AM, Dovey SM, Tilyard MW. Risk factors for hospitalization due to diabetes complications. Diabetes research and clinical practice. 2008;80:244-52.

[26] Prentice JC, Pizer SD, Conlin PR. Identifying the independent effect of HbA1c variability on adverse health outcomes in patients with Type 2 diabetes. Diabet Med. 2016;33:1640-8.

Figure legends：

Figure 1: Flowchart for recruitment procedures of the predictive model for diabetes-related hospitalization

Figure 2. Receiver operating characteristic curve (ROC) for (A) 1-year (B) 3-year (C) 5-year (D) 8-year diabetes-related hospitalization risk in validation set.

Figure 3. Receiver operating characteristic curve (ROC) for (A) 1-year (B) 3-year (C) 5-year (D) 8-year in-hospital mortality risk in validation set.

Table 1. Baseline characteristic of the study population.

Variables	Derivation Set (n=21,769) MEAN±SD* or n (%)	Validation Set (n=10,884) MEAN±SD* or n (%)	Standardized effect size
Socio-demographic factors			
Age (years)	61.09±10.84	61.14±10.78	0.00
Gender			
Female	11640 (53.47)	5800 (53.29)	0.00
Male	10129 (46.53)	5084 (46.71)	-0.01
Smoking habit	3185 (14.63)	1641 (15.08)	-0.01
Alcohol drinking	1805 (8.29)	882 (8.10)	0.01
Age of diabetes onset (years)	54.56±10.94	54.57±10.92	0.00
Duration of type 2 diabetes (years)	6.56±6.36	6.61±6.38	-0.01
Body mass index (kg/m^2)	25.64±3.73	25.65±3.79	0.00
Obesity	7884 (36.22)	3944 (36.24)	0.00
Diabetes related factor and biomarker			
Systolic blood pressure (mm Hg)	134.6±17.55	134.85±17.71	-0.01
Diastolic blood pressure (mm Hg)	79.86±10.55	79.88±10.51	0.00
HbA1c level (%)	8.18±1.95	8.16±1.91	0.01
Fasting blood glucose (mm Hg)	171.65±65.22	171.67±63.54	0.00
Low-density lipoprotein (mg/dL)	117.47±31.19	117.29±31.2	0.01
High-density lipoprotein (mg/dL)	46.43±13.95	46.53±13.74	-0.01
Creatinine (mg/dL)	1.05±0.61	1.07±0.67	-0.03
SGPT (u/l)	31.98±31.54	32.35±36.23	-0.01
Total cholesterol (mg/dL)	195.76±41.70	195.44±41.94	0.01
Triglyceride (mg/dL)	171.5±129.70	169.95±129.51	0.01
eGFR (mg/dL)	73.96±22.06	73.40±22.25	0.03
Variation of fasting blood glucose (%)	32.3±26.05	32.51±26.54	-0.01
Variation of HBA1c (%)	16.86±15.27	16.79±14.98	0.00
Comorbidity			
Hypertension	9940 (45.66)	4995 (45.89)	0.00
Stroke	1099 (5.05)	561 (5.15)	0.00
Cardiovascular disease	1910 (8.77)	942 (8.65)	0.00
Peripheral arterial disease	188 (0.86)	80 (0.74)	0.01
Peripheral Neuropathy	176 (0.81)	93 (0.85)	0.00
Diabetes retinopathy	305 (1.40)	162 (1.49)	-0.01
Disease of peripheral circulatory disturbance	933 (4.29)	446 (4.10)	0.01
Hypoglycemia	896 (4.12)	437 (4.02)	0.01
Chronic kidney disease	229 (1.05)	108 (0.99)	0.01
Traumatic Amputation	4 (0.02)	0 (0.00)	-0.08
Ketoacidosis	160 (0.73)	99 (0.91)	-0.02
Postural hypotension	15 (0.07)	16 (0.15)	-0.03
Arterial embolism and thrombosis	34 (0.16)	17 (0.16)	0.00
Hyperlipidemia	5561 (25.55)	2762 (25.38)	0.00
Medication use			
Anti-diabetes medications			
No medication	1009 (4.64)	539 (4.95)	-0.01
SUs	16126 (74.08)	7999 (73.49)	0.01
Non-SUs	2531 (11.63)	1283 (11.79)	0.00
Insulin	759 (3.49)	384 (3.53)	0.00
Insulin+SUs	859 (3.95)	376 (3.45)	0.03
Insulin+non-SUs	485 (2.23)	303 (2.78)	-0.04
Hypertension medications	10917 (50.15)	5526 (50.77)	-0.01
Cardiovascular medications	7058 (32.42)	3426 (31.48)	0.02
Lipid medications	8316 (38.20)	4168 (38.29)	0.00

Outcome			
Diabetes-related hospitalization	4181 (19.21)	2062 (18.95)	0.01
In-hospital mortality	1378 (6.33)	670 (6.16)	0.00

*: SD = standard deviation; SUs: Sulfonylureas; non- SUs: nonsulfonylureas.

Table 2. Cox model measured hazard ratio and 95% confidence intervals of diabetes-related hospitalization and In-hospital mortality

	HR (95% CI)			
	Diabetes-related hospitalization		In-hospital mortality	
Variables	Crude	Adjusted	Crude	Adjusted
Socio-demographic factors				
Age (years)	1.04 (1.03, 1.04)***	1.03 (1.02, 1.03)***	1.09 (1.08, 1.10)***	1.08 (1.07, 1.09)***
Gender				
Female	1.00	1.00	1.00	1.00
Male	0.90 (0.85, 0.96)***	0.99 (0.93, 1.05)	1.68 (1.51, 1.87)***	1.82 (1.63, 2.03)***
Duration of type 2 diabetes				
0	1.00	1.00		
1-5	1.47 (1.29, 1.68)***	1.46 (1.27, 1.67)***		
6-10	2.26 (1.97, 2.60)***	1.75 (1.52, 2.02)***		
11-15	2.96 (2.57, 3.42)***	1.99 (1.71, 2.30)***		
16-20	3.76 (3.21, 4.40)***	1.95 (1.65, 2.30)***		
>20	4.48 (3.81, 5.26)***	2.07 (1.75, 2.46)***		
Body mass index (kg/m^2)				
<22.5	1.00	1.00	1.00	1.00
22.5-25	0.74 (0.67, 0.81)***	0.78 (0.71, 0.86)***	0.80 (0.69, 0.93)**	0.79 (0.68, 0.92)**
25-30	0.68 (0.63, 0.73)***	0.74 (0.68, 0.80)***	0.64 (0.56, 0.74)***	0.66 (0.57, 0.75)***
≥30	0.68 (0.61, 0.76)***	0.76 (0.68, 0.84)***	0.59 (0.49, 0.71)***	0.69 (0.57, 0.84)***
Hospitalization status one year prior to baseline				
No	1.00	1.00	1.00	1.00
Yes	2.48 (2.33, 2.64)***	1.89 (1.77, 2.02)***	2.02 (1.81, 2.27)***	1.39 (1.23, 1.58)***
Diabetes related factor and biomarker				
HbA1c (%)				
<7	1.00	1.00		
≥7	1.68 (1.56, 1.81)***	1.42 (1.31, 1.53)***		
Variation of fasting plasma glucose (%)				
<17.6	1.00	1.00	1.00	1.00
17.6-35.0	1.35 (1.25, 1.47)***	1.15 (1.06, 1.25)**	1.08 (0.94, 1.24)	1.04 (0.91, 1.19)
>35.0	2.03 (1.88, 2.19)***	1.44 (1.33, 1.56)***	1.38 (1.21, 1.57)***	1.19 (1.04, 1.37)*
Variation of HbA1c (%)				
<8.5	1.00	1.00	1.00	1.00
8.5-17.5	1.27 (1.17, 1.37)***	1.09 (1.01, 1.18)*	1.10 (0.96, 1.26)	1.06 (0.92, 1.21)
>17.5	1.46 (1.36, 1.58)***	1.29 (1.19, 1.39)***	1.27 (1.11, 1.44)***	1.23 (1.07, 1.41)**
Creatinine (mg/dL)				
Male:0.5-1.2; female: 0.7-1.5	1.00	1.00	1.00	1.00
Abnormal	2.84 (2.64, 3.06)***	1.95 (1.81, 2.11)***	2.98 (2.63, 3.38)***	1.89 (1.66, 2.16)***
Total cholesterol to HDL ratio				
Male<5; female<4.5	1.00	1.00	1.00	1.00
Male:5-9.5; female:4.5-7.0	1.18 (1.11, 1.26)***	1.12 (1.05, 1.20)***	1.06 (0.95, 1.18)	1.14 (1.02, 1.28)*
**Male>9.5; female>7.0	1.63 (1.38, 1.93)***	1.23 (1.03, 1.46)*	1.81 (1.38, 2.38)***	1.98 (1.50 2.61)***
Diabetes-related disorders				
Stroke				
No	1.00	1.00	1.00	1.00
Yes	2.18 (1.96, 2.43)***	1.24 (1.11, 1.39)***	2.45 (2.05, 2.93)***	1.37 (1.14, 1.65)***

Diabetes retinopathy				
No	1.00	1.00	1.00	1.00
Yes	3.40 (2.88, 4.02)***	1.61 (1.35, 1.91)***	2.21 (1.61, 3.04)***	1.71 (1.23, 2.37)**
Hypertension				
No			1.00	1.00
Yes			1.75 (1.57, 1.95)***	1.21 (1.08, 1.35)**
Hypoglycemia				
No			1.00	1.00
Yes			2.01 (1.63, 2.46)***	1.37 (1.11, 1.7)**
Medication use				
Anti-diabetes medications				
No medication	1.45 (1.17, 1.79)***	1.48 (1.20, 1.84)***	1.6 (1.22, 2.11)***	1.56 (1.18, 2.05)**
SUs	1.92 (1.69, 2.19)***	1.63 (1.43, 1.86)***	1.11 (0.92, 1.32)	1.08 (0.90, 1.30)
Non-SUs	1.00	1.00	1.00	1.00
Insulin	7.47 (6.35, 8.79)***	3.76 (3.18, 4.45)***	2.60 (1.99, 3.39)***	1.65 (1.26, 2.17)***
Insulin+SUs	6.36 (5.42, 7.47)***	3.97 (3.36, 4.68)***	1.37 (1.01, 1.86)*	1.23 (0.90, 1.68)
Insulin+non-SUs	5.09 (4.21, 6.16)***	2.68 (2.20, 3.27)***	2.23 (1.62, 3.07)***	1.88 (1.37, 2.60)***
Cardiovascular medications				
No	1.00	1.00		
Yes	1.60 (1.50, 1.70)***	1.18 (1.11, 1.26)***		

*:p<0.05; **:p<0.01; ***:p<0.001

HR: hazard ratio; CI: confidence intervals; SUs: Sulfonylureas; non- SUs: nonsulfonylureas.

Table 3. Parameter estimates of regression coefficient and risk socre of predictors for diabetes-related hospitalization from the final multivariate Cox's proportional hazards model.

Risk factor	Diabetes-related hospitalization			In-hospital mortality		
	$\hat{\beta}(\widehat{SE})$	P-value	Risk score	$\hat{\beta}(\widehat{SE})$	P-value	Risk score
Socio-demographic factors						
Age (years)	0.03 (0.002)	<0.001	-2 to 8	0.08 (0.003)	<0.001	-2 to 8
Gender						
Female	ref	ref	0	ref	ref	0
Male	-0.01 (0.03)	0.71	0	0.60 (0.06)	<0.001	2
Duration of type 2 diabetes						
0	ref	ref	0			
1-5	0.38 (0.07)	<0.001	3			
6-10	0.56 (0.07)	<0.001	4			
11-15	0.69 (0.08)	<0.001	5			
16-20	0.67 (0.08)	<0.001	5			
>20	0.73 (0.09)	<0.001	6			
Body mass index (kg/m^2)						
<22.5	ref	ref	0	ref	ref	0
22.5-25	-0.25 (0.05)	<0.001	-2	-0.24 (0.08)	0.002	-1
25-30	-0.31 (0.04)	<0.001	-2	-0.42 (0.07)	<0.001	-1
≥30	-0.28 (0.06)	<0.001	-2	-0.37 (0.10)	<0.001	-1
Hospitalization status one year prior to baseline						
No	ref	ref	0	ref	ref	0
Yes	0.64 (0.03)	<0.001	5	0.33 (0.06)	<.0001	1
Diabetes related factor and biomarker						
HbA1c (%)						
<7	ref	ref	0			
≥7	0.35 (0.04)	<0.001	3			
Variation of fasting plasma glucose (%)						
<17.6	ref	ref	0	ref	ref	0
17.6-35.0	0.14 (0.04)	0.001	1	0.04 (0.07)	0.58	0
>35.0	0.36 (0.04)	<0.001	3	0.18 (0.07)	0.01	0
Variation of HbA1c (%)						
<8.5	ref	ref	0	ref	ref	0
8.5-17.5	0.09 (0.04)	0.03	1	0.06 (0.07)	0.42	0
>17.5	0.25 (0.04)	<0.001	2	0.21 (0.07)	0.003	1
Creatinine (mg/dL)						
Male:0.5-1.2; female: 0.7-1.5	ref	ref	0	ref	ref	0
Abnormal	0.67 (0.04)	<0.001	5	0.64 (0.07)	<0.001	2
Total cholesterol to HDL ratio						
Male<5; female<4.5	ref	ref	0	ref	ref	0
Male:5-9.5; female:4.5-7.0	0.11 (0.03)	<0.001	1	0.13 (0.06)	0.02	0

Male>9.5; female>7.0	0.20 (0.09)	0.02	2	0.68 (0.14)	<0.001	2
Diabetes-related disorders						
Stroke	0.22 (0.06)	<0.001	2	0.32 (0.09)	<0.001	1
Diabetes retinopathy	0.47 (0.09)	<0.001	4	0.53 (0.17)	0.001	1
Hypertension				0.19 (0.06)	0.001	0
Hypoglycemia				0.32 (0.11)	0.003	1
Medication use						
Anti-diabetes medications						
No medication	0.39 (0.11)	<0.001	3	0.44 (0.14)	0.002	1
SUs	0.49 (0.07)	<0.001	4	0.08 (0.09)	0.39	0
Non-SUs	ref	ref	0	ref	ref	0
Insulin	1.32 (0.09)	<0.001	10	0.50 (0.14)	<0.001	1
Insulin+SUs	1.38 (0.08)	<0.001	11	0.21 (0.16)	0.19	1
Insulin+non-SUs	0.99 (0.10)	<0.001	8	0.63 (0.16)	<0.001	2
Cardiovascular medications						
No	ref	ref	0			
Yes	0.17 (0.03)	<0.001	1			

SUs: Sulfonylureas; non- SUs: nonsulfonylureas.

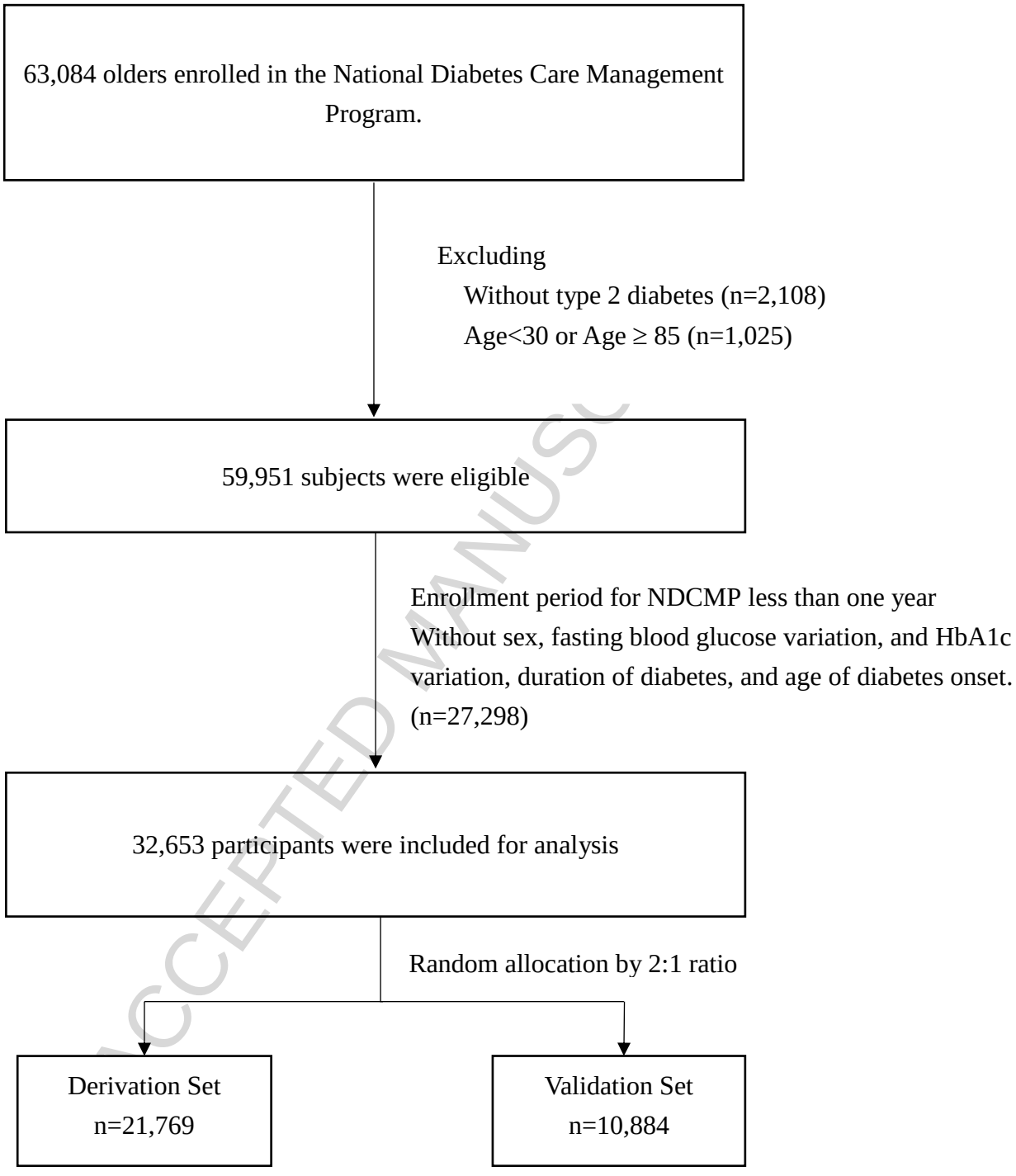
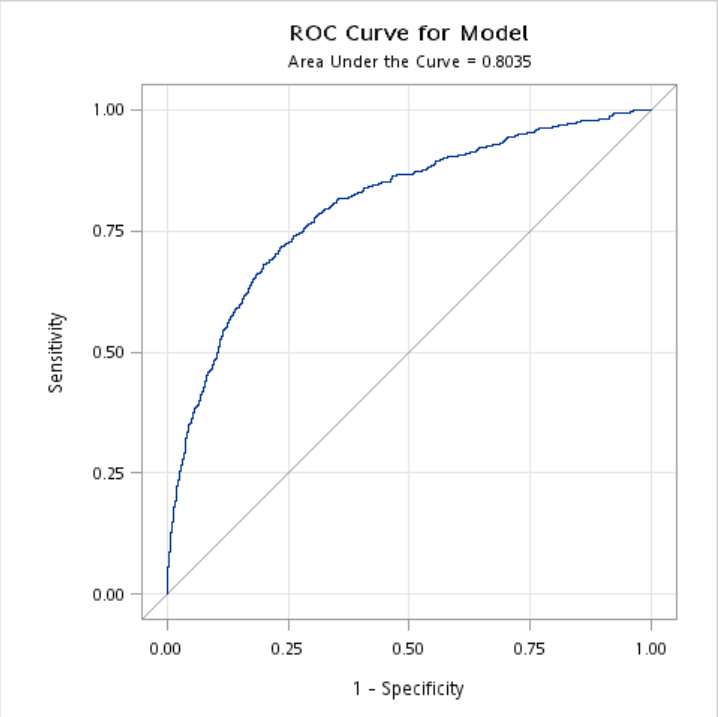
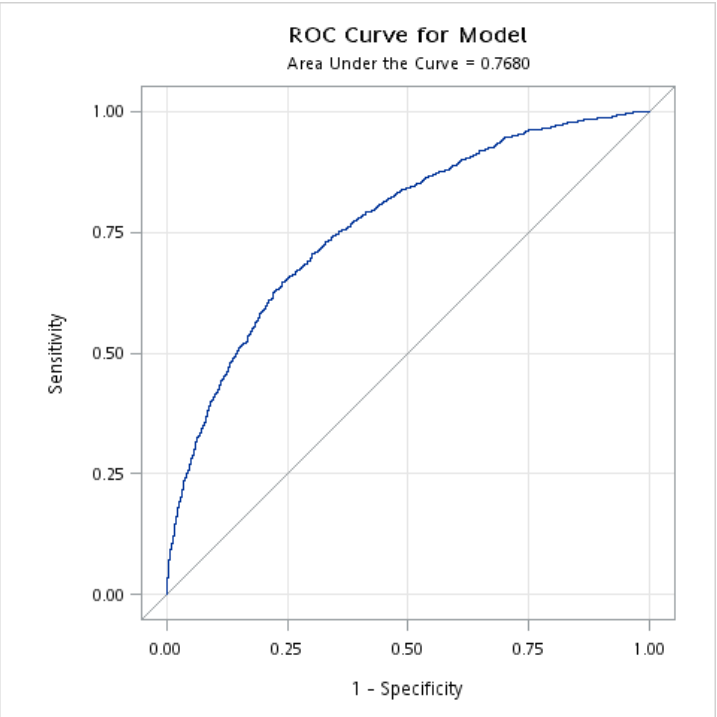


Figure 1: Flowchart for recruitment procedures of the predictive model for diabetes-related hospitalization

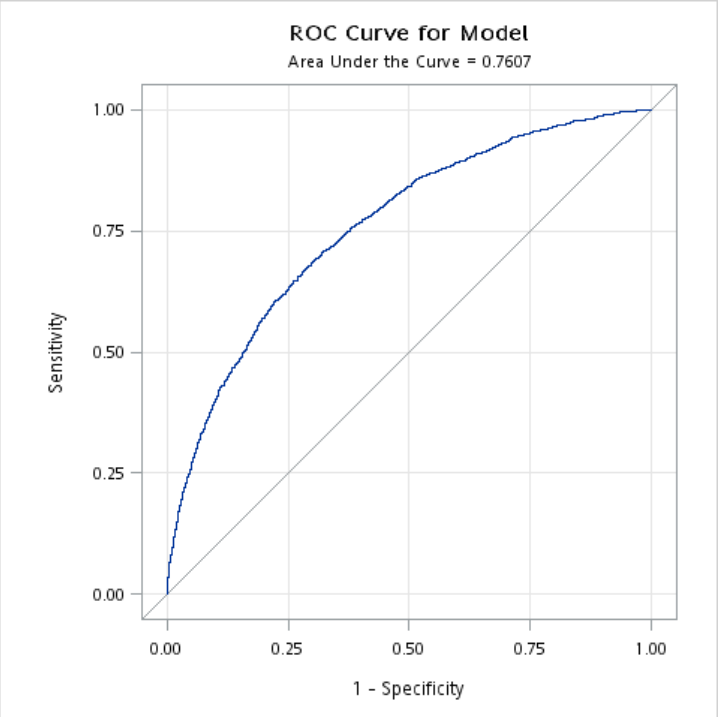
(A)



(B)



(C)



(D)

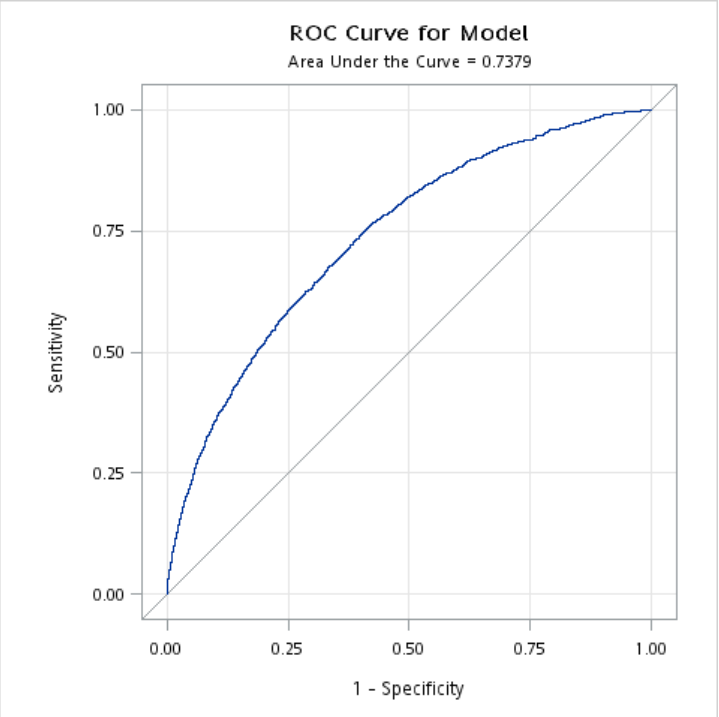
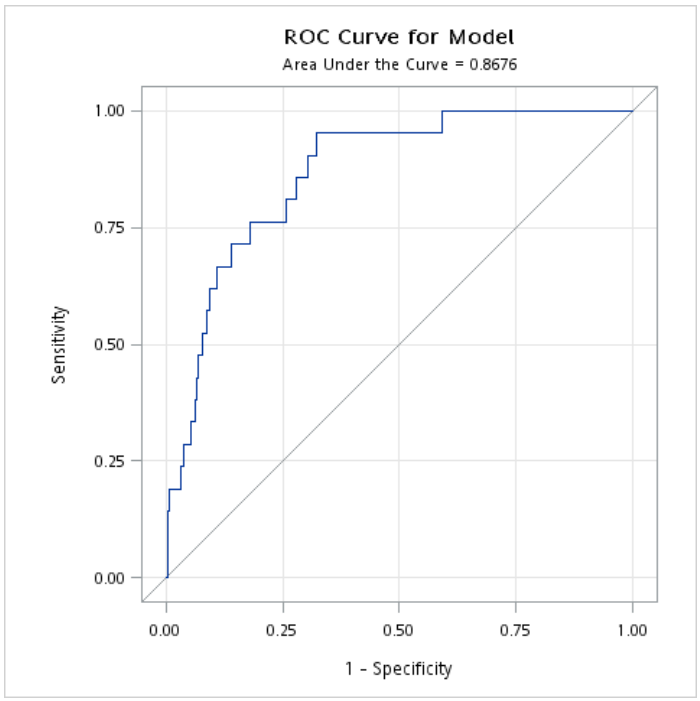
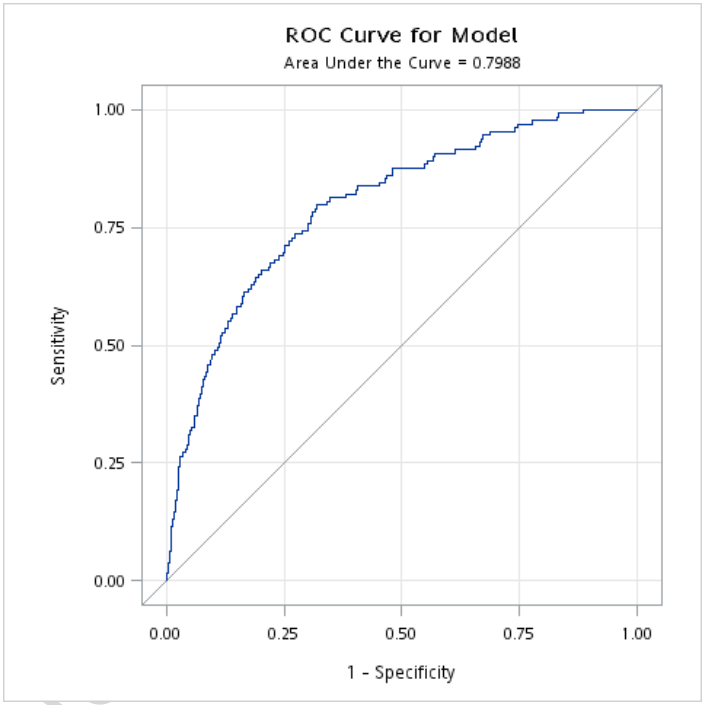


Figure 2. Receiver operating characteristic curve (ROC) for (A) 1-year (B) 3-year (C) 5-year (D) 8-year diabetes-related hospitalization risk in validation set

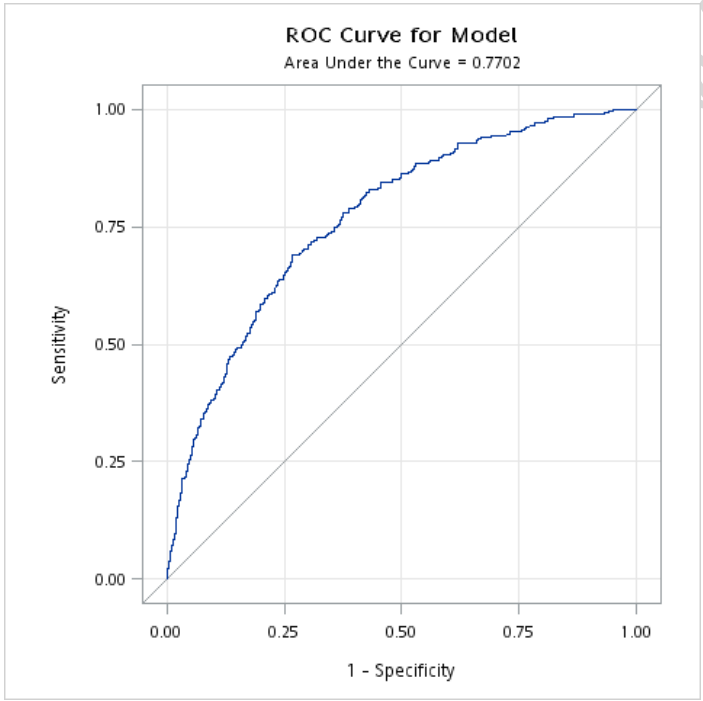
(A)



(B)



(C)



(D)

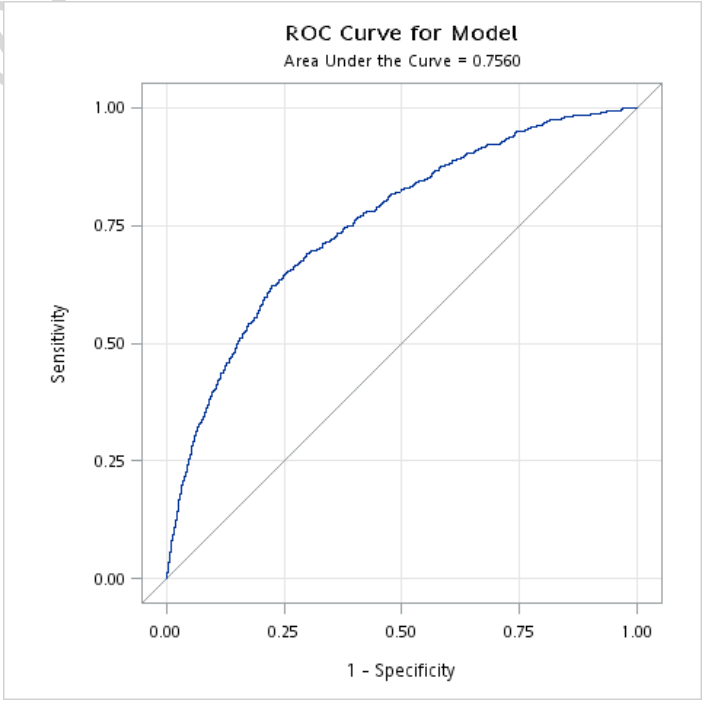


Figure 3. Receiver operating characteristic curve (ROC) for (A) 1-year (B) 3-year (C) 5-year (D) 8-year in-hospital mortality risk in validation set

Highlights

- Diabetes is a major cause of hospitalization and in-hospital mortality.
- A scoring system identifying diabetic patients at risk of such outcomes is lacking.
- Prediction models for such outcomes in patients with type 2 diabetes were proposed.
- Diabetes hospitalization prediction model's 1- and 3-year AUROC were 0.80 and 077.
- The corresponding values for in-hospital mortality were 0.87, and 080.