Development and validation of a predictive model for Chronic Kidney Disease progression in Type 2 Diabetes Mellitus based on a 13-year study in Singapore

Serena Low, Su Chi Lim, Xiao Zhang, Shiyi Zhou, Lee Ying Yeoh, Yan Lun Liu, Tavintharan Subramaniam, Chee Fang Sum

Aims: This study aims to develop and validate a predictive model for Chronic Kidney Disease (CKD) progression in Type 2 Diabetes Mellitus (T2DM).

Methods: We conducted a prospective study on 1582 patients with T2DM from a Diabetes Centre in regional hospital in 2002–2014. CKD progression was defined as deterioration across eGFR categories with ≥25% drop from baseline. The dataset was randomly split into development (70%) and validation (30%) datasets. Stepwise multivariable logistic regression was used to identify baseline predictors for model development. Model performance in the two datasets was assessed.

Results: During median follow-up of 5.5 years, 679 (42.9%) had CKD progression. Progression occurred in 467 (42.2%) and 212 patients (44.6%) in development and validation datasets respectively. Systolic blood pressure, HbA1c, estimated glomerular filtration rate and urinary albumin-to-creatinine ratio were associated with progression. Areas under receiving-operating-characteristics curve for the training and test datasets were 0.80 (95%CI, 0.77–0.83) and 0.83 (95%CI, 0.79–0.87). Observed and predicted probabilities by quintiles were not statistically different with Hosmer-Lemeshow $\chi^2$ 0.65 ($p = 0.986$) and 1.36 ($p = 0.928$) in the two datasets. Sensitivity and specificity were 71.4% and 72.2% in development dataset, and 75.6% and 72.3% in the validation dataset.

Conclusions: A model using routinely available clinical measurements can accurately predict CKD progression in T2DM.

© 2016 Elsevier Ireland Ltd. All rights reserved.

Keywords:
Diabetes mellitus
Chronic kidney disease
Progression

ARTICLE INFO

Article history:
Received 29 June 2016
Received in revised form
29 October 2016
Accepted 11 November 2016
Available online 22 November 2016

Keywords:
Diabetes mellitus
Chronic kidney disease
Progression

ABSTRACT

Aims: This study aims to develop and validate a predictive model for Chronic Kidney Disease (CKD) progression in Type 2 Diabetes Mellitus (T2DM).

Methods: We conducted a prospective study on 1582 patients with T2DM from a Diabetes Centre in regional hospital in 2002–2014. CKD progression was defined as deterioration across eGFR categories with ≥25% drop from baseline. The dataset was randomly split into development (70%) and validation (30%) datasets. Stepwise multivariable logistic regression was used to identify baseline predictors for model development. Model performance in the two datasets was assessed.

Results: During median follow-up of 5.5 years, 679 (42.9%) had CKD progression. Progression occurred in 467 (42.2%) and 212 patients (44.6%) in development and validation datasets respectively. Systolic blood pressure, HbA1c, estimated glomerular filtration rate and urinary albumin-to-creatinine ratio were associated with progression. Areas under receiving-operating-characteristics curve for the training and test datasets were 0.80 (95%CI, 0.77–0.83) and 0.83 (95%CI, 0.79–0.87). Observed and predicted probabilities by quintiles were not statistically different with Hosmer-Lemeshow $\chi^2$ 0.65 ($p = 0.986$) and 1.36 ($p = 0.928$) in the two datasets. Sensitivity and specificity were 71.4% and 72.2% in development dataset, and 75.6% and 72.3% in the validation dataset.

Conclusions: A model using routinely available clinical measurements can accurately predict CKD progression in T2DM.

© 2016 Elsevier Ireland Ltd. All rights reserved.

* Corresponding author. Fax: +65 66023772.
E-mail addresses: low.serena.km@alexandrahealth.com.sg (S. Low), lim.su.chi@alexandrahealth.com.sg (S.C. Lim), zhang.xiao@alexandrahealth.com.sg (X. Zhang), zhoushiyi@genomics.cn (S. Zhou), yeoh.lee.ying@alexandrahealth.com.sg (L.Y. Yeoh), liu.allen.yl@alexandrahealth.com.sg (Y.L. Liu), subramaniam.tavintharan@alexandrahealth.com.sg (S. Tavintharan), sum.chee.fang@alexandrahealth.com.sg (C.F. Sum).
1. Introduction

Diabetes Mellitus (DM) is a global health challenge. According to the International Diabetes Federation, the number of people with DM is expected to increase from 382 million in 2013 to 592 million in 2035 [1]. Approximately 60% of the world’s population with diabetes is found in Asia [2]. The myriad of complications engendered by DM therefore poses a considerable health burden. Of note, diabetic nephropathy is one of the major complications of DM; it was reported that kidney complications occurred in about 25–40% of individuals with Type 2 DM (T2DM) [3]. In Singapore, the prevalence of DM rose from 8.2% in 2004 to 11.3% in 2010 [4,5]. Given the rising prevalence of DM and the fact that diabetic nephropathy is the leading cause of new patients with end-stage kidney disease in Singapore, it is imperative to identify patients with T2 DM early for targeted management in order to prevent progression of kidney disease in DM.

There is emerging literature on the development and validation of risk models that predict progression of Chronic Kidney Disease (CKD). Such models can potentially be used for risk communications to patients to improve lifestyle and health behaviour, tailoring management for patients at different risk strata, and prognosticating patients for preparation of renal replacement therapy. The predictors included age, gender, body mass index, systolic blood pressure, serum creatinine, measure of proteinuria, urinary albumin-to-creatinine ratio (uACR), estimated glomerular filtration rate (eGFR) and even novel biomarkers [6].

However, there were some limitations of the models. Firstly, the models were predominantly derived from Caucasian populations and general population [6]. Their validity in Asians with T2DM remains unclear due to limited data on this population. Secondly, the predicted outcomes were occurrence of end-stage renal disease (ESRD) and doubling of serum creatinine which were the extreme ends of the kidney disease spectrum [6]. Given the variable outcomes of CKD such as progression to renal replacement therapy, stability and death, there is a need to develop a prediction model that aims to arrest progression at the intermediate stage prior to the occurrence of ESRD.

To address these gaps, we aimed to develop and validate a predictive model for progression of CKD in Singapore, a multi-ethnic society in Southeast Asia. The presence of this ‘intermediate’ outcome offers an opportunity to identify patients at risk early so as to prevent or slow down CKD progression to end-stage.

2. Materials and methods

2.1. Study population

This was a prospective cohort study involving patients with T2DM who were enrolled at a Diabetes Centre in a regional hospital in Singapore in 2002–2014. Subjects were excluded if they were less than 21 years old, pregnant, have active infections, active cancer and autoimmune disease, involvement of other suspected causes of renal diseases (e.g. urinary tract infection, polycystic kidney disease, haematuria or history of glomerulonephritis). For the purpose of analysis, subjects were also excluded if they had fewer than three measurements of estimated glomerular filtration (eGFR), shorter than two years of follow-up, and eGFR < 15 ml/min/1.75 m² at baseline. There were 1582 patients who were eligible for this study. Ethics approval was obtained from National Healthcare Group Domain Specific Review Board in Singapore. All patients who participated had given informed written consent.

2.2. Data collection

Demographics and clinical characteristics were obtained by trained nurses from patients’ case records and standard questionnaire administered to the patients. Blood pressure (BP) was measured using a standard automated sphygmomanometer in the sitting position after resting for at least 5 min (HEM-C7011-C1, OMRON Corp., Kyoto, Japan). The sphygmomanometer has been validated according to European Society of Hypertension (ESH) Protocol and Association for the Advancement of Medical Instrumentation (AAMI) Protocol [7]. The hospital laboratory accredited by the Royal College of the American Pathologists carried out measurements of blood and urine samples. Serum creatinine (intra-assay coefficient of variation (CV) was 0.6–1.1%; inter-assay CV was 1.1–1.4%), low density lipoprotein cholesterol (LDL-cholesterol) (intra-assay CV was 0.7–1.2%; inter-assay CV was 1.9–2.5%) and triglycerides (TG) (intra-assay CV was 0.7–1.1%; inter-assay CV was 1.6–2.0%) were quantitated using enzymatic colorimetric test (Roche cobas® c 501) [8–10], Haemoglobin A1c (HbA1c) (intra-assay CV was 1.0–1.6%; inter-assay CV was 1.4–2.0%) using Tina-quant Hemoglobin A1c Gen.3 (Roche cobas® c 501) [11], and urinary albumin (intra-assay CV was 0.7–1.6%; inter-assay CV was 1.2–2.8%) using immunoturbidimetric assay (Roche cobas® c 501) [12]. We used the Modification of Diet in Renal Disease (MDRD) formula to calculate the eGFR rate [13].

2.3. Definition of outcome

The outcome was CKD progression. This was defined as eGFR decline which involved worsening of eGFR categories (stage 1, ≥ 90 ml/min/1.73 m²; stage 2, 60–89 ml/min/1.73 m²; stage 3a, 45–59 ml/min/1.73m²; stage 3b, 30–44 ml/min/1.73 m²; stage 4, 15–29 ml/min/1.73 m²; and stage 5, <15 ml/min/1.73 m²), coupled with a 25% or more reduction in eGFR from baseline according to the 2013 Kidney Disease: Improving Global Outcomes (KDIGO) Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease KDIGO [14]. Patients with improvement in eGFR, same eGFR category, or less than 25% reduction in eGFR from baseline were deemed as having no CKD progression.

2.4. Statistical analysis

The dataset was randomly split into two datasets – training dataset for development of the prediction model (70%) and test dataset for validation of the model (30%). Categorical
variables were presented as number (percentage), and continuous variables as means (standard deviation) or median (interquartile range) as appropriate. Differences in demographic and clinical characteristics stratified by the two groups and progression were examined by Chi-Square test for categorical variables and student-t test or Mann-Whitney test for continuous variables.

In the training dataset, we first tested for univariate associations between potential variables and CKD progression using logistic regression. These variables, which were proposed as risk factors for CKD progression based on literature [6] and biological plausibility, included age, gender, ethnicity, smoking, duration of DM, body mass index, HbA$_{1c}$, systolic blood pressure, eGFR, uACR, LDL-cholesterol, TG and renin-angiotensin antagonist. Those with \( p < 0.10 \) were selected for backward stepwise multivariable logistic regression (\( p < 0.10 \) for entry and \( p < 0.05 \) for stay). The choice of backward selection was premised on the need to refine selection from a modest-sized group of potential variables, which were first identified in the univariate logistic regression, by eliminating a few variables. The candidate baseline variables for inclusion in the final model in the training dataset were age, HbA$_{1c}$, systolic blood pressure (SBP), uACR and eGFR at baseline.

2.4.1. Model validation
A series of methods was employed to validate the model. These included the following.

2.4.1.1. Discrimination. This was assessed from the area under the receiving operating characteristics curve (AUC; C statistic). Discrimination is deemed perfect, good, moderate and poor if the corresponding AUC is 1, >0.8, 0.6–0.8 and <0.6 [15,16].

2.4.1.2. Calibration (or goodness-of-fit). This was assessed by the Hosmer-Lemeshow \( C \) test. This involved dividing the cohort into quintiles of predicted risks of progression and comparing with actual outcomes. Calibration is poor if there is significant difference between the predicted and observed risks (\( p < 0.05 \)) [15,16].

2.4.1.3. Summary of classification. Sensitivity refers to the proportion of patients with CKD progression correctly classified as such by the model. Specificity refers to the proportion of patients without CKD progression correctly classified as such by the model. Positive predictive value refers to the proportion of patients identified by the model with positive result as truly having CKD progression. Negative predictive value

<table>
<thead>
<tr>
<th>Table 1 – Baseline characteristics of patients at baseline in the training and test data sets.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>Age (years)</td>
</tr>
<tr>
<td>Gender (%)</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Ethnicity (%)</td>
</tr>
<tr>
<td>Chinese</td>
</tr>
<tr>
<td>Malay</td>
</tr>
<tr>
<td>Indian</td>
</tr>
<tr>
<td>Duration of DM (years)</td>
</tr>
<tr>
<td>BMI (kg/m$^2$)</td>
</tr>
<tr>
<td>Smoking (%)</td>
</tr>
<tr>
<td>Non-smoker</td>
</tr>
<tr>
<td>Ex-smoker</td>
</tr>
<tr>
<td>Current smoker</td>
</tr>
<tr>
<td>HbA$_{1c}$ (%) (mmol/mol)</td>
</tr>
<tr>
<td>SBP (mmHg)</td>
</tr>
<tr>
<td>Urinary ACR (mg/g)</td>
</tr>
<tr>
<td>eGFR (ml/min/1.73 m$^2$)</td>
</tr>
<tr>
<td>LDL-cholesterol (mmol/l)</td>
</tr>
<tr>
<td>TG (mmol/l)</td>
</tr>
<tr>
<td>RAS-antagonist (%)</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>Follow-up period (years)</td>
</tr>
<tr>
<td>Progression (%)</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
</tr>
</tbody>
</table>

DM, Diabetes Mellitus; BMI, Body Mass Index; HbA1c, Haemoglobin A1c; HbA1c-CV, Haemoglobin A1c Coefficient of Variation; SBP, Systolic Blood Pressure; Urinary ACR, Urinary Albumin-to-Creatinine Ratio; eGFR, estimated Glomerular Filtration Rate; LDL-Cholesterol, Low Density Lipoprotein Cholesterol; TG, Triglycerides; RAS-antagonist, Renin-angiotensin system antagonist.
refers to the proportion of patients identified by the model with negative result as truly having no CKD progression. A cut-off of 0.40 was selected as the threshold beyond which an observation has a predicted positive outcome. Using lens command, the graphs of sensitivity and specificity versus probability cut-off were produced. The cut-off point was chosen at the intersection of the sensitivity and specificity curves.

Statistical analysis was performed using STATA version 14.0 (STATA Corporation, College Station, Texas).

3. Results

During a median follow-up period of 5.5 years (IQR 4.3–7.0), 679 (42.9%) had CKD progression. Progression occurred in 467 patients (42.2%) in training dataset, and 212 patients (44.6%) in the test dataset. The patients in the training dataset and the test dataset were similar in terms of age, gender, ethnicity, body mass index (BMI), duration of diabetes, smoking status, HbA1c, LDL-Cholesterol, TG, uACR and use of renin-angiotensin system antagonists at baseline. The patients in the training dataset had a lower eGFR at baseline than those in the validation dataset ($p = 0.003$) (Table 1).

The final variables selected in the final multivariable logistic regression after backward selection were log uACR (mg/g), SBP (per 10 mmHg), HbA1c (%), eGFR (per 5 ml/min/1.73 m$^2$), LDL-Cholesterol (mmol/l) and age (per 10 years) (Table 2). The AUC of the final model was 0.80 (95% Confidence Interval (CI), 0.77–0.83). The predicted probability was not significantly different from the observed probability of CKD progression over 6 years of follow-up with Hosmer-Lemeshow $\chi^2$ of 0.65 ($p = 0.986$) (Fig. 1). Using a cut-off of $p = 0.40$, the model had a sensitivity of 71.4%, specificity of 72.2%, positive predictive value of 65.3% and negative predictive value of 77.4% (Table 3).

All the variables in the final model using the test dataset were significantly associated with progression (Table 2). The model in the test dataset shows good discrimination with AUC of 0.83 (95%CI, 0.79–0.87). The observed and predicted probabilities of CKD progression were not significantly different with Hosmer-Lemeshow $\chi^2$ of 1.36 ($p = 0.928$) (Fig. 2). Using a cut-off of $p = 0.40$, the model had a sensitivity of 75.6%, specificity of 72.3%, positive predictive value of 68.9% and negative predictive value of 78.5% (Table 3).
We have developed and validated a prediction model for CKD (Supplementary Fig. 1). These metabolic risk factors, apart from T2DM. This reflects the influence of adverse metabolic profile and age at completion of formal education, in addition to four of the variables found in our final model (eGFR, uACR, SBP and Hba1c) [17]. In the Reduction of Endpoints in NIDDM with the Angiotensin II Antagonist Losartan (RENAAL) study, uACR, serum albumin, serum creatinine and hemoglobin were identified as predictors of ESRD in patients with T2DM and nephropathy [18]. Another study included eGFR, uACR and haematocrit in their risk model for predicting ESRD in Hong Kong Chinese patients with T2DM [19].

One of the strengths of our study is the choice of definition of CKD progression in line with the KDIGO guidelines. To the best of our knowledge, this is the first study using this definition as the outcome for developing prediction model. Previous studies have developed or validated models on major endpoints such initiation of chronic dialysis, kidney transplantation or doubling of serum creatinine [17–25]. Only one study defined progressive CKD as having the most renal function decline (top 20% of the total population) and eGFR < 60 ml/min/1.73 m² using follow-up [26]. In view of the lengthy course of kidney disease in DM, the use of prediction models in both early and late stages of kidney disease provides a wide window of opportunity for prevention. Our prediction tool can potentially help clinicians to identify patients early before CKD progression starts, thereby allowing intensive treatment for those at high risk. Second, our study is based on a cohort with T2DM exclusively. This reduced the heterogeneity of the study population in assessing the contribution of different risk factors to CKD progression and addressed an important unmet need of a population at extremely high risk for CKD. Our results will also add to the findings from the current limited pool of research which focused on T2DM [17–19]. The third strength is the use of easily available clinical and laboratory information which are routinely used in clinical practice. Furthermore, the prediction tool is practical and can be easily incorporated into the clinical information system. It was reported that the presence of risk prediction model and integration into clinical practice guidelines have resulted in improved compliance to treatment guidelines and promoted individual decision-making.

There are a few limitations in this study. First, our sample size was moderate in comparison with other studies. Second, our cohort came from Diabetes Centre in a regional hospital and hence was not fully representative of the entire population of individuals with T2DM in Singapore. This limits the generalizability of findings to other cohorts such as patients in primary care. Third, there is a potential problem of measurement error using MDRD equation to estimate GFR. It was, however, shown that MDRD had better or similar performance than CKD-EPI equation in patients with DM [27]. Lastly, there may be other risk factors which may influence CKD progression but not explored in this study such as genetic factors, socio-economic status, diet and exercise.

Our findings also partially overlap with those in earlier studies on patients with T2DM, albeit on different outcomes. For example, in the Action in Diabetes and Vascular Disease: Preterax and Diamicron MR Controlled Evaluation (ADVANCE) study, the prediction model on major kidney related outcomes (doubling of serum creatinine, renal replacement therapy or renal death) included gender, diabetic retinopathy and age at completion of formal education, in addition to four of the variables found in our final model (eGFR, uACR, SBP and Hba1c) [17].

Table 3 – Performance of the model in the training and test data sets.

<table>
<thead>
<tr>
<th>Table 3 – Performance of the model in the training and test data sets.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training data set</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Discrimination</td>
</tr>
<tr>
<td>AUC (95% CI)</td>
</tr>
<tr>
<td>P</td>
</tr>
<tr>
<td>Calibration</td>
</tr>
<tr>
<td>Hosmer-Lemeshow Ĉ test P</td>
</tr>
<tr>
<td>Sensitivity (%)</td>
</tr>
<tr>
<td>Specificity (%)</td>
</tr>
<tr>
<td>PPV (%)</td>
</tr>
<tr>
<td>NPV (%)</td>
</tr>
<tr>
<td>AUC, area under the receiver operating characteristics curve; PPV, positive predictive value; NPV, negative predictive value.</td>
</tr>
</tbody>
</table>

Fig. 2 – Observed versus predicted risk of CKD progression during a 6-year follow-up in the test data set.

4. Discussion

We have developed and validated a prediction model for CKD progression in patients with T2DM. The AUC or C-statistic in our model was 0.80 in the training dataset and 0.83 in the test dataset, showing good discriminatory performance. In a systematic review by Echouffo-Tcheugi on risk models to predict CKD and its progression, it was reported that the C-statistic ranged from 0.56 to 0.94, indicating modest to good performance [6]. Our model has performed relatively well in comparison with these studies. In addition, our model showed good calibration with p > 0.90.

Our study demonstrated that lower eGFR, higher uACR, higher SBP, higher Hba1c, higher LDL-Cholesterol and older age are significantly associated with CKD progression in T2DM. This reflects the influence of adverse metabolic profile on CKD progression. These metabolic risk factors, apart from age, are potentially modifiable and the importance of their control should be emphasized in patient education and clinical management. The model incorporating these risk factors can potentially be used as a simple medical calculator (Supplementary Fig. 1).
5. Conclusions

We have developed and validated a model to predict CKD progression in Asians with T2DM. The model demonstrated that eGFR, uACR, SBP, HbA1c, LDL-Cholesterol and age influenced the risk of CKD progression. The model performed well among individuals who share characteristics similar to our source-population and can be a practical tool for easy implementation in routine clinical practice.

Conflict of interest statement

The authors declare that they have no conflict of interest.

Acknowledgements

This study was supported by Alexandra Health Grant (AHPL SIGII/11001; SIG/11029 and SIG/12024).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.diabres.2016.11.008.

REFERENCES