BMJ Open

Individualised risk prediction model for new-onset, progression and regression of chronic kidney disease in a retrospective cohort of patients with type 2 diabetes under primary care in Hong Kong

Lin Yang [1] , Tsun Kit Chu,[2] Jinxiao Lian,[3] Cheuk Wai Lo,[2] Shi Zhao,[1,4,5] Daihai He [6] , Jing Qin,[1] Jun Liang[2]

For numbered affiliations see end of article.

Correspondence to
Dr Lin Yang;
l.yang@polyu.edu.hk

ABSTRACT

Objectives This study is aimed to develop and validate a prediction model for multistate transitions across different stages of chronic kidney disease (CKD) in patients with type 2 diabetes mellitus under primary care.

Setting We retrieved the anonymised electronic health records of a population-based retrospective cohort in Hong Kong.

Participants A total of 26 197 patients were included in the analysis.

Primary and secondary outcome measures The new-onset, progression and regression of CKD were defined by the transitions of four stages that were classified by combining glomerular filtration rate and urine albumin-to-creatinine ratio. We applied a multiscale multistate Poisson regression model to estimate the rates of the stage transitions by integrating the baseline demographic characteristics, routine laboratory test results and clinical data from electronic health records.

Results During the mean follow-up time of 1.8 years, there were 2632 patients newly diagnosed with CKD, 1746 progressed to the next stage and 1971 regressed into an earlier stage. The models achieved the best performance in predicting the new-onset and progression with the predictors of sex, age, body mass index, systolic blood pressure, diastolic blood pressure, serum creatinine, haemoglobin A1c, total cholesterol, low-density lipoprotein, high-density lipoprotein, triglycerides and drug prescriptions.

Conclusions This study demonstrated that individual risks of new-onset and progression of CKD can be predicted from the routine physical and laboratory test results. The individualised prediction curves developed from this study could potentially be applied to routine clinical practices, to facilitate clinical decision making, risk communications with patients and early interventions.

Strengths and limitations of this study

► Early predictions for chronic kidney disease (CKD) progression and timely intervention are critical for clinical management of patients with diabetes.

► We successfully developed a multiscale multistate Poisson regression model that achieved satisfactory performance in predicting the new-onset and progression of CKDs.

► The model incorporates the predictors of demographic characteristics, routine laboratory test results and clinical data from electronic health records.

► The individualised prediction curves could potentially be applied to facilitate clinical decision making, risk communications with patients and early interventions of CKD progression.

► The cohort has a relatively short follow-up period and the retrospective study design might suffer from report bias and selection bias.

International Diabetes Federation.[1] In Hong Kong, it was estimated that 10.9% of adults aged 20–79 years had diabetes.[1] Diabetes is associated with a heavy disease burden and tremendous economic costs.[2] A local study reported that the annual direct medical costs of patients with diabetes with newly diagnosed complications were much higher than the cost of US$1984 of uncomplicated cases.[3] Chronic kidney disease (CKD) is one of the most common complications of type 2 diabetes. Globally, it was estimated that diabetes attributed to 12%–55% of end-stage renal disease (ESRD).[1] In Hong Kong, a prospective cohort of the Hong Kong Diabetes Registry reported that 10% of patients with type 2 diabetes developed into CKD, which was defined as an estimated

INTRODUCTION

Globally, in 2017, there were 425 million adults with diabetes according to the

glomerular filtration rate (eGFR) $<60\,mL/min/1.73\,m^2$, after 7 years of follow-up.[4]

Early diagnosis and treatment of CKD in patients with type 2 diabetes are critical for reducing its associated heavy disease burden. Therefore, developing an effective and efficient prediction model for CKD progression is important in terms of prioritising healthcare resources to high-risk populations. The Kidney Disease Improving Global Outcomes (KDIGO) 2012 Guideline lists several predictors for CKD progression based on previous epidemiological studies, including age, sex, ethnicity, obesity, CKD aetiology, GFR, albuminuria, blood pressure, glycaemic control, dyslipidaemia and complications of cardiovascular diseases (CVDs).[5] Specifically for people with type 2 diabetes, several prediction models have been reported in the literature, but most were Cox proportional hazards regression models on all-cause mortality and incidence of CVD.[6–9] Previous studies have investigated the risk factors that significantly predicted ESRD in Chinese populations[10 11], but none focused on CKD progression at earlier stages nor on CKD regression.

The Cox proportional hazards model is one of the most widely adopted classical modelling approaches to predict diabetes complications. However, it only allows two states (no event and event) and linear predictors, hence, it is unable to assess the progression probability across different disease stages. Some studies adopted more complex Cox models to address these limitations, such as the multistate Markov model and the Cox model with time-dependent covariates.[12 13] But these models include only a single time scale, which usually does not allow for the simultaneous estimation on the forward and backward transitions of multiple disease stages (ie, progression and regression).

A multiscale multistate Poisson regression (MSMS) model has recently been applied to estimate the risks of transition between different stages of disease progression or the incidences of multiple intermediate states and endpoints such as different hospitalisation episodes.[14] This model has the advantage of allowing multiple events in the follow-up period and thereby is capable of investigating dynamic transition across different stages of disease progression and regression. In other words, forward time scales (eg, time since enrolment into the cohort) and backward time scales (time since entry to the intermediate state) are simultaneously entered into the model as covariates. Another advantage of the MSMS model lies in its flexibility of allowing non-linear dynamic transitions by incorporating spline smooth functions in each time scale with full parametric estimation. Here we conducted a population-based retrospective cohort study, with the aim to develop and validate a prediction model for the new-onset, progression and regression of CKD in patients with type 2 diabetes under primary care in Hong Kong.

METHODS
Data sources
We obtained individual data of adults who were aged over 18 years at enrolment into the Risk Assessment and Management Programme for Patients with Diabetes Mellitus (RAMP-DM) from July 2014 to June 2017, in the general and specialty outpatient clinics managed by the New Territory West Cluster of the Hospital Authority in Hong Kong Special Administrative Region, China. This cluster served a population of 1.1 million in 2017, and more details about the data source can be found in Yang et al.[15] The RAMP-DM Programme aims to structurally screen for diabetic complications among patients diagnosed with diabetes type 1 or type 2 under primary care. All patients are eligible to join this programme without extra costs and on a voluntary basis. Doctors and nurses from different hospitals and clinics were regularly trained and followed the same protocol of RAMP-DM. In Hong Kong, more than 90% of people with diabetes have been enrolled in the RAMP-DM since 2014.[16] A local study also demonstrated that the patients who were not enrolled in the RAMP-DM were not significantly different from those who enrolled, in terms of sociodemographic and clinical characteristics.[16]

We also retrieved the longitudinal data of physical examinations and laboratory test results during scheduled clinical visits, including incidence of diabetic complications, blood pressure, eGFR (calculated from urine creatinine), urine albumin-to-creatinine ratio (UACR), total cholesterol, serum levels of creatinine, high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglycerides and serum haemoglobin A1c (HbA1c), together with annual prescriptions of angiotensin-converting-enzyme inhibitors (ACEI) and angiotensin II receptor blockers (ARB), from the Clinical Data Analysis and Reporting System from 1 January 2014 to 31 December 2017, by matching their unique patient reference numbers. All adults aged above 18 years with clinical diagnosed type 2 diabetes with or without complications, according to the International Classification of Disease Ninth Revision codes (250.00, 250.02, 250.10, 250.12, 250.20, 250.22, 250.30, 250.32, 250.40, 250.42, 250.50, 250.52, 250.60, 250.62, 250.70, 250.72, 250.80, 250.82, 250.90, 250.92), were included in the analysis. The participants who already had ESRD at enrolment, without HbA1c tests taken during the follow-up, and/or having incomplete baseline data were excluded from data analysis.

Patient and public involvement statement
Patients were not involved in the recruitment to and conduct of the study, since all data were retrospectively retrieved from electronic medical records.

Definition of CKD new-onset, progression and regression
According to the KDIGO 2012 Clinical Practice Guideline for Evaluation and Management of Chronic Kidney Diseases,[5] CKD can be classified into four risk stages for CKD outcomes based on eGFR and UACR levels: stage 0,

low risk; stage 1, moderately increased risk; stage 2, high risk; and stage 3, very high risk. Here we defined three categories of stage transitions as the event: (1) *new-onset of CKD*, a transition from stage 0 to 1; (2) *CKD progression*, a forward transition to the next advanced stage (ie, stage 1 to 2 and 2 to 3); and (3) *CKD regression*, a backward transition to the next stage (ie, stage 1 to 0, 2 to 1 and 3 to 2). Cases with the transition over two stages with the intermediate stage recorded were counted as two separate events, and those without were excluded from the study. For example, two events were included in the model if the subject was classified into stage 0 at the beginning and progressed into stage 1 and then stage 2 in the follow-up period. If a subject jumped from stage 0 to 2 without a record of stage 1, then his/her data were not included in the model.

Data matching

The eGFR and UACR test results taken on the same date of CKD outcome events (new-onset, progression and regression) were matched by the unique patient numbers. If these tests were conducted on different dates, then those taken closest to the event dates and within 180 days were matched to the CKD event dates with the unique patient numbers. Baseline sociodemographic data were collected at the initial visits to the clinics, including age, sex, education level, body mass index (BMI) and the presence of diabetic complications. Longitudinal data of blood pressure, lipid profile, serum creatinine, eGFR, UACR and HbA1c levels tested during the clinical visits, together with any ACEI/ARB prescriptions within the year (yes or no), were also matched for each participant by the dates of follow-up period between recruitment dates and event dates, or between recruitment dates and censoring dates if no events occurred by the end of the follow-up period.

MSMS model

The technical details of MSMS models can be found in Iacobelli and Carstensen,[14] Carstensen and Plummer.[17] We used calendar time as the basic time scale, and the risk of transition was calculated for the period from the enrolment time to the laboratory test dates with stage change. The incidence rate of the new-onset of CKD (or progression and regression) was estimated from the following formula:

$$E[\mu(t) \mid \alpha] = \mu_0(t - \tau) h_0(\alpha) \exp \sum_i \beta_i x_i + \beta_0 \qquad (1)$$

where t is the time since enrolment, $E[\mu(t)|\alpha]$ is the expected crude incidence rate of CKD new-onset (or progression and regression) from the diagnosis of the previous stage to time t, which is assumed to follow a Poisson distribution; τ is the time when last laboratory tests are conducted for CKD ($t > \tau \geq 0$); α is the most recent HbA1c levels prior to the progression events; x_is are the covariates of age, sex, BMI, systolic blood pressure (SBP), diastolic blood pressure (DBP), serum creatinine, total

cholesterol, HDL, LDL, triglycerides that were tested at enrolment, and annual drug prescriptions of ACEI and ARB; $\mu_0(t - \tau)$ is the natural splines function of the period $(t - \tau)$ during which the transition occurs; $h_0(\alpha)$ denotes the natural splines of HbA1c levels (α). Here we assume non-linear effects of time and HbA1c level, because the incidence rate of CKD new-onset (or progression and regression) could probably increase over time in a non-linear pattern. Unlike the model proposed by Iacobelli and Carstensen,[14] we only included one time scale $t - \tau$ in this model to simplify the model structures because there were very limited numbers of participants who experienced both the regression and progression events in the follow-up period.

We first randomly extracted 90% of the cohort data as a training dataset to fit the MSMS model by ninefold cross-validation.[18] The remaining 10% of data were used as the test dataset to evaluate the prediction accuracy of the best-fit model for internal validation. For each outcome, we built four different models: model 1 included sex, age, BMI and HbA1c as predictors; model 2, blood pressure and total cholesterol were added to model 1; model 3, serum creatinine, HDL, LDL and triglycerides were added to model 2; model 4, annual prescription of ACEI and ARB were added to model 3. These models with different combinations of predictors were compared using the area under the receiver operating characteristic curve (AUC) and mean absolute error.[19 20] The Nagelkerke's scaled R^2 was adopted to measure the goodness-of-fit of model.[21]

Individual incidence rates of CKD new-onset, progression and regression over time since enrolment could be predicted from the final MSMS model by inputting the observed data of each subject. It is of note that the time between laboratory tests and clinical visits was not consistent, as they were determined by clinical indications and hospital manpower. Hence, to adjust for irregularly scheduled clinical visits and tests, we calculated an adjusted incidence rate by adding weights of the probability density function of clinical visits and test frequency, with the following equation $\nu(t) = \frac{\mu(t)}{g(t|\kappa)}$, where $v(t)$ is the adjusted CKD incidence rate; $g(t|\kappa)$ denotes the Gaussian kernel density function of clinical visits and tests. An example of this adjustment is shown in the online supplementary appendix 1.

All the data analyses were conducted in R (V.3.4.1) and the codes can be found online (https://github.com/yanglinpolyu/DMforecastmodel).

RESULTS

Summary statistics

A flowchart for data collection and analysis is shown in figure 1. There were 40 781 people with diabetes who were enrolled in the RAMP-DM from July 2014 to June 2017. Of them, 39 652 people with type 2 diabetes were considered eligible for this study. We excluded 8078 patients with incomplete eGFR and UACR data, 5002 with only
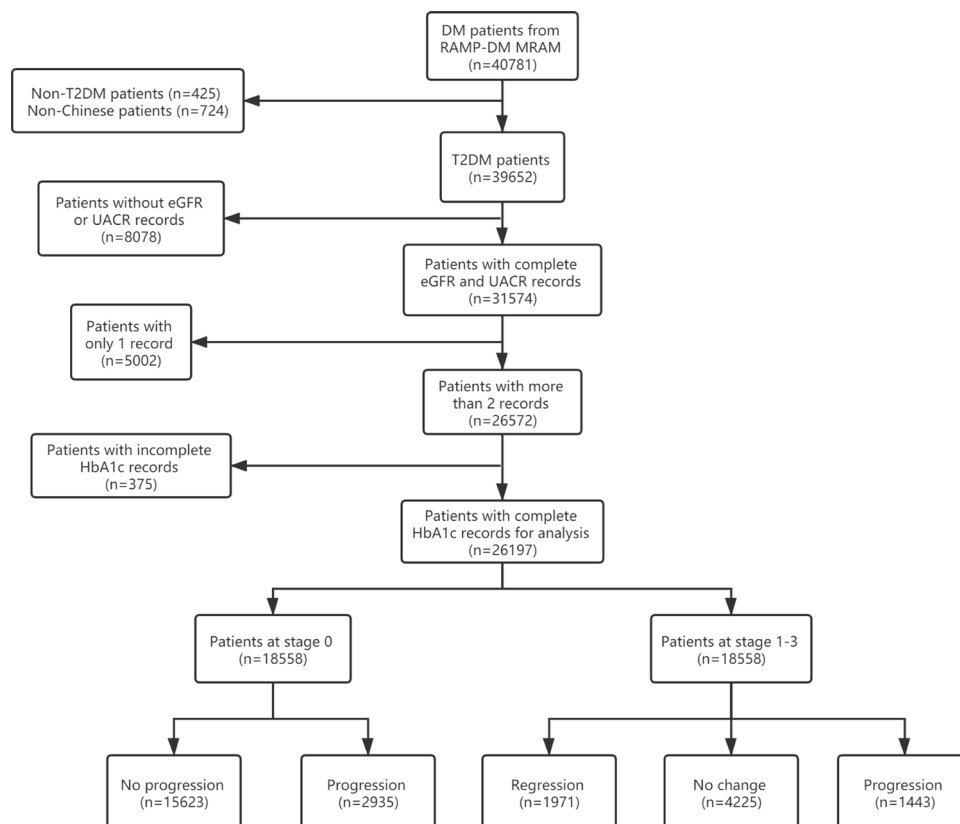
Figure 1 Flowchart of data clean and analysis procedure for CKD progression. CKD, chronic kidney disease; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; HbA1c, haemoglobin A1c; MRAM, Metabolic Risk Assessment Module; RAMP-DM, Risk Assessment and Management Programme for Patients with Diabetes Mellitus; T2DM, type 2 diabetes mellitus; UACR, urine albumin-to-creatinine ratio.

one record of eGFR and UACR and 375 with incomplete HbA1c record. In the final sample of 26 197 patients, 48.8% were men, the mean age at enrolment was 61.5 years and 70.8% did not have CKD (stage 0). At the end of the follow-up period, there were 2632 cases of CKD new-onset, 1746 progression and 1971 regression (table 1). Compared with the patients without change, those with the new-onset or progression or regression of CKD were older, more likely to be female individuals, with lower education level, higher BMI and having more complications (table 2). The mean (range) of the follow-up period was 1.83 (0.07–4.04), 1.77 (0.13–3.84), 1.43 (0.08–3.72)

and 1.89 (0.07–4.04) years, for all patients, patients with CKD new-onset/progression, with regression and without change, respectively.

Model performance and predictors

The MSMS model achieved better performance in predicting progression than in predicting the regression of chronic kidney disease (AUC 0.72–0.84 vs 0.50–0.71) (online supplementary appendix 2). The significant predictors for CKD new-onset included female individuals, older age, having ACEI and ARB prescriptions, high levels of BMI, SBP, total cholesterol and serum creatinine,

**Table 1** Summary of CKD progression and regression during the follow-up period

		CKD stage at the end of follow-up				New-onset/progression episodes*	Regression episodes	Person-years followed	Patients (n)
		0	1	2	3				
CKD stage at baseline	0	15 623	2632	253	50	2935	0	35 140	18 558
	1	1462	2981	859	167	1026	1462	9327	5469
	2	58	344	724	417	417	402	2518	1543
	3	5	14	88	520	0	107	1052	627
Subtotal		17 148	5971	1924	1154	4378	1971	48 038	26 197

*Any change of the status during the follow-up was counted as one episode.
CKD, chronic kidney disease.

Table 2 Descriptive statistics of cohort participants

	Total	No change	Regression	New-onset/progression
n	26 197	19 848	1971	4378
Person-years	48 038	37 478	2817	7743
Sociodemographic factors				
Sex, n (%)				
Female	12 965 (49.5)	9458 (47.7)	1138 (57.7)	2369 (54.1)
Male	13 232 (50.5)	10 390 (52.3)	833 (42.3)	2009 (45.9)
Age (years), mean±SD	62.4±10.4	61.5±10.1	64.0±11.0	65.7±10.8
Education levels, n (%)				
No formal education	3150 (12.0)	2094 (10.6)	322 (16.3)	734 (16.8)
Primary	9861 (37.6)	7320 (36.9)	779 (39.5)	1762 (40.2)
Secondary	11 734 (44.8)	9272 (46.7)	771 (39.1)	1691 (38.6)
Tertiary	1368 (5.2)	1095 (5.5)	92 (4.7)	181 (4.1)
Other	84 (0.3)	67 (0.3)	7 (0.4)	10 (0.2)
BMI, mean±SD	26.2±4.2	26.1±4.1	26.8±4.2	26.4±4.6
BMI category, n (%)				
Underweight	286 (1.1)	207 (1.0)	25 (1.3)	54 (1.2)
Normal	5180 (19.8)	4013 (20.2)	309 (15.7)	858 (19.6)
Overweight	5565 (21.3)	4333 (21.8)	379 (19.3)	853 (19.5)
Obese	15 140 (57.9)	11 280 (56.9)	1254 (63.8)	2606 (59.6)
Clinical characteristics				
Duration of diabetes (years), mean±SD	7.8±6.3	7.5±6.1	7.9±6.2	9.2±6.8
Coronary heart disease, n (%)	561 (2.1)	386 (2.1)	66 (3.3)	109 (2.5)
Stroke, n (%)	993 (3.8)	669 (3.4)	81 (4.1)	243 (5.6)
Peripheral arterial disease, n (%)				
Yes	252 (1.0)	173 (0.9)	27 (1.4)	52 (1.2)
Suspected	117 (0.4)	87 (0.4)	5 (0.3)	25 (0.6)
Diabetic retinopathy, n (%)				
No retinopathy	17 388 (66.4)	13 530 (68.2)	1259 (63.9)	2599 (59.4)
Non-sight threatening	5341 (20.4)	4014 (20.2)	409 (20.8)	918 (21.0)
Sight threatening	2651 (10.1)	1760 (8.9)	215 (10.9)	676 (15.4)
Ungradable	176 (0.7)	117 (0.6)	21 (1.1)	38 (0.9)
SBP (mm Hg), mean±SD	131.6±16.3	131.1±16.1	132.9±17.1	133.5±17.0
DBP (mm Hg), mean±SD	75.0±10.3	75.2±10.1	74.6±10.9	74.1±10.8
Laboratory tests, mean±SD				
UACR (mg/mmol)	5.6±21.4	5.5±21.6	10.9±21.0	6.0±20.0
eGFR (mL/min/1.73 m^2)	84.3±19.5	86.5±18.7	78.8±21.7	77.1±19.8
Creatinine (mmol/L)	79.2±21.3	77.7±20.5	83.5±25.4	84.1±21.7
HbA1c (%)	7.2±1.3	7.1±1.3	7.3±1.3	7.2±1.3
HbA1c (mmol/mol)	55.2±11.9	54.1±11.9	56.3±11.9	55.2±11.9
Triglycerides (mmol/L)	1.5±1.1	1.5±1.1	1.6±1.2	1.6±1.0
LDL (mmol/L)	2.3±0.7	2.3±0.7	2.4±0.7	2.3±0.7
HDL (mmol/L)	1.3±0.3	1.3±0.3	1.3±0.3	1.2±0.3
Drug prescriptions, n (%)				
ACEI	11 259 (43.0)	7992 (40.3)	991 (50.3)	2276 (52.0)
ARB	6964 (26.6)	4694 (23.6)	690 (35.0)	1580 (36.1)

ACEI, angiotensin-converting-enzyme inhibitors; ARB, angiotensin II receptor blockers; BMI, body mass index; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; HbA1c, haemoglobin A1c; HDL, high-density lipoprotein; LDL, low-density lipoprotein; SBP, systolic blood pressure; UACR, urine albumin-to-creatinine ratio.
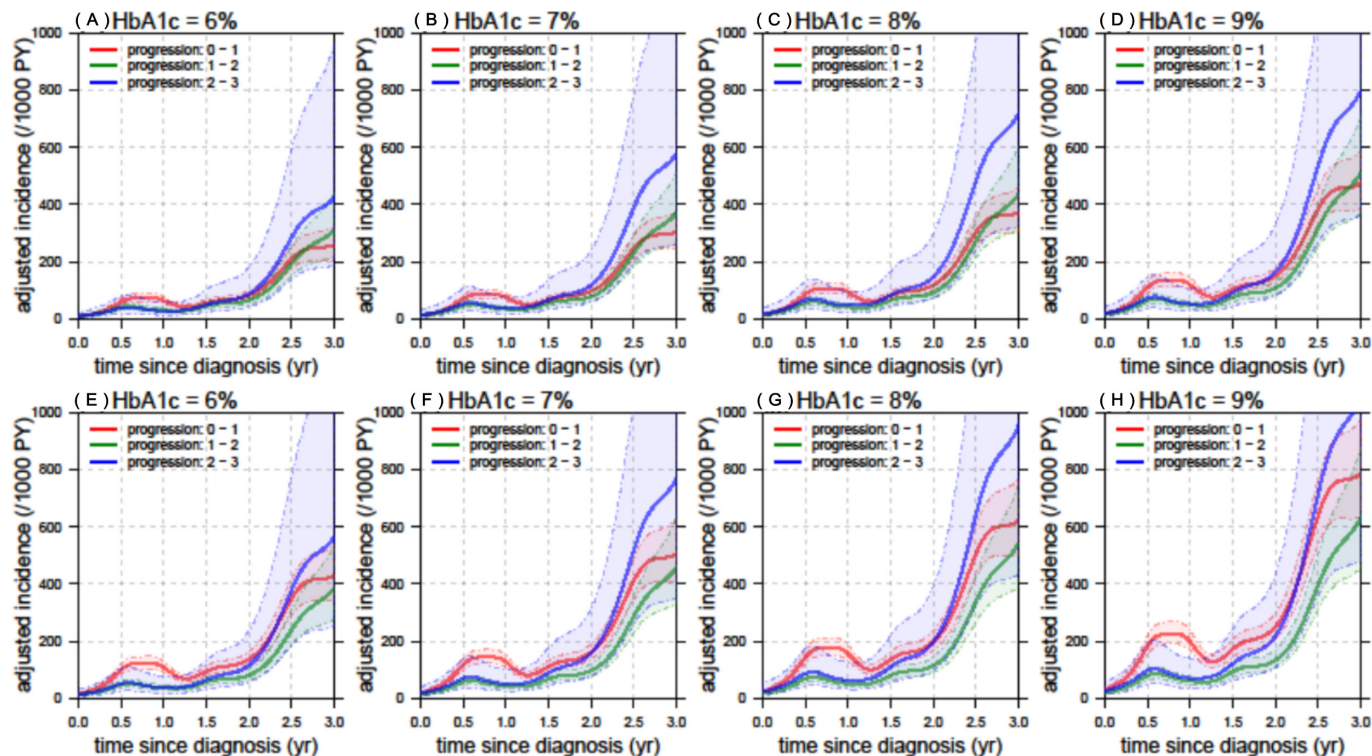
Figure 2 Adjusted incidence rates per 1000 person-years (PY) of CKD new-onset or progression over time since the diagnosis of previous stages in female patients, at the HbA1c levels ranging from 6% to 9%, for new-onset (red solid line), progression from stage 1 to 2 (green solid line) and from stage 2 to 3 (blue solid line). Upper panel (A–D) shows adjusted incidence rates in patients aged below 65 years. Lower panel (E–H) shows adjusted incidence rates in patients aged 65 years or above. The broken line and shadow area indicate the 95% CI. CKD, chronic kidney disease; HbA1c, haemoglobin A1c.

low levels of DBP, HDL and LDL (online supplementary appendix 3). Fewer significant predictors were found for those who progressed from stage 1 to 2 and from 2 to 3, and the effect estimates of significant predictors were similar to those of new-onset. The only exception is that lower BMI was associated with a higher risk of progression from stage 2 to 3.

Rates of new-onset and progression

To facilitate the comparison between groups, we plotted the prediction curves of adjusted rates per 1000 person-year by assuming the mean values of each predictor in the MSMS model: BMI=25.7, SBP=130 mm Hg, DBP=75 mm Hg, total cholesterol=4.0 mmol/L, creatinine=75 mmol/L, HDL=1.20 mmol/L, LDL=2.10 mmol/L and triglycerides=1.21 mmol/L. The mean age was 56.4 years for female individuals aged <65 years, 73.7 years for female individuals aged 65+ years, 56.1 years for male individuals aged <65 years and 72.4 years for male individuals aged 65+ years.

The temporal trend of adjusted rates of CKD new-onset and progression from the MSMS model across different levels of HbA1c are shown in figures 2 and 3, for women and men, respectively. The CKD progression rate dramatically increased in women 2 years after diagnosis, but this change was less evident in men. An exponential increasing trend of incidence rates over time was observed for all three types of CKD progression and gradually elevated when the

HbA1c level increased from 6% to 9%. The new-onset rate was high within the first year of enrolment (diagnosed as stage 0), particularly among older adults aged 65 years or over. The progression rate from stage 2 to 3 dramatically increased 2 years after the diagnosis of stage 2.

Rates of regression

The significant predictors for regression from stage 1 to 0 include male individuals, younger age, lower levels of SBP and creatinine, without ACEI or ARB prescriptions (online supplementary appendix 3). The stage 2–1 and 3–2 regression has similar predictors, including male individuals, older age, lower levels of SBP and creatinine. Higher levels of DBP and without ACEI are only significant for the stage 3–2 regression.

The regression from earlier stages shows a less evident temporal pattern in women and men (online supplementary appendix 4 and 5). The adjusted incidence rates were high during the first half year and then quickly declined to very low levels after 1 year. The regression rates were higher in the group from stage 2 to 1, followed by 1 to 0 and stage 3 to 2 in both sex groups. The regression rates of older people tend to be slightly higher than the younger group.

DISCUSSION

Early predictions of complication incidence and disease progression are of great concern in the clinical
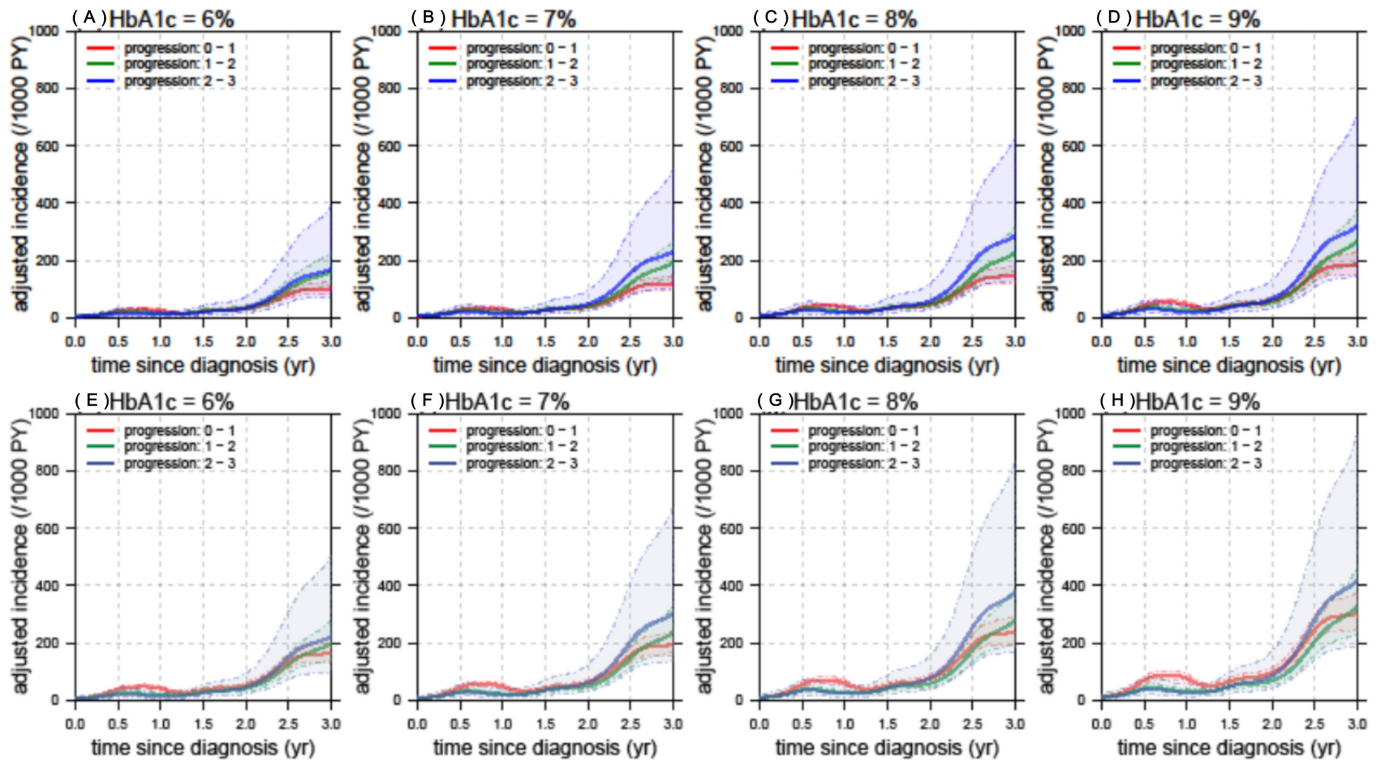
Figure 3 Adjusted incidence rates per 1000 person-years (PY) of CKD new-onset or progression over time since the diagnosis of previous stages in male patients, at the HbA1c levels ranging from 6% to 9%, for new-onset (red solid line), progression from stage 1 to 2 (green solid line) and from stage 2 to 3 (blue solid line). Upper panel (A–D) shows adjusted incidence rates in patients aged below 65 years. Lower panel (E–H) shows adjusted incidence rates in patients aged 65 years or above. The broken line and shadow area indicate the 95% CI. CKD, chronic kidney disease; HbA1c, haemoglobin A1c.

management of patients with type 2 diabetes. The MSMS model has recently been introduced to evaluate the risk of chronic diseases.[22–24] Its advantages over the classical proportional hazards (PH) model have been discussed in the literature.[14 17 25] This novel modelling approach allows the simultaneous prediction of multistate transitions, either progression or regression and thereby is capable of assessing both linear and non-linear effects of predictors. Previous studies have demonstrated that the MSMS model outperformed the PH model in both model structure and fitting performance.[14 26–28] Our study is among the first to apply this model to predict the risks of CKD progression and regression in a large sample of the Chinese population with type 2 diabetes under primary care. The advantage of the MSMS model also lies in its ability to incorporate non-linear effects. We observed rapidly increasing progression rates between different stages among female patients, but to a lesser extent among male patients (figures 2 and 3). Particularly, a turning point of the progression rates was found at 1 year postdiagnosis of the previous stage, whereas the regression rates peaked at half a year. This suggests that 1 year is a critical control window to reverse the deterioration of nephropathy among patients with type 2 diabetes. Another critical time point is 2 years postdiagnosis, beyond which the regression rates for both genders dramatically decreased, suggesting an irreversible trend of kidney function deterioration.

Based on the data availability and association with disease progression, we considered a series of predictors including demographic characteristics, clinical presentations, drug prescriptions and laboratory tests routinely conducted in these patients. Interestingly, a larger number of significant predictors were successfully identified by the MSMS model for new-onset than for progression. Nearly all these predictors significantly predicted the CKD new-onset, whereas for progression, only age, gender, BMI, SBP and serum creatinine remained significant predictors. These findings highlight the importance of early interventions for patients with type 2 diabetes when their renal functions maintain within the normal range and many risk factors are modifiable.

A few prediction models for progression of CKD have been developed in the literature[13 29 30] and some were specifically for patients with diabetes.[31–33] This study is among the first to define CKD progression (or regression) as a forward (or backward) transition between stages of proteinuria and deterioration of eGFR. In this study, we found that female individuals, older age, ACEI and ARB prescriptions, higher BMI, SBP, total cholesterol and creatinine, lower DBP, HDL and LDL were associated with a higher risk of CKD new-onset and progression. However, Tangri et al reported that younger age and male individuals had a faster progression to renal failure in their PH model for patients with CKD of various aetiologies.[34] This discrepancy might be due to the different

outcome variables and study populations, since our cohort included only patients with type 2 diabetes and the outcomes were a progression from earlier stages. The predictors of our model are similar to an early prediction model that was developed using the PH model for CVDs of patients with type 2 diabetes in Hong Kong.[35] This CVD model found that older age, longer duration of type 2 diabetes and higher HbA1c predicted a higher risk of CVD incidence.

This study is among the first to investigate the predictors of CKD regression in patients with type 2 diabetes under primary care, to our best knowledge. We found that younger age had a faster regression from stage 1 to 0, but slower from stage 2 to 1 or 3 to 2. This suggests that, once younger adults enter the advanced stages, it is less likely to reverse the progression.

There is ample evidence in the literature to demonstrate that both ARB and ACEI could slow down the progression of CKD. However, we found that patients with type 2 diabetes who were prescribed with ARB and ACEI had a dramatically faster progression rate of CKD and a lower regression rate. In our data, these drugs were significant only in the transition of stages 0 to 2 and only ACEI prescription predicts a lower rate of regression from 3 to 2. We speculate that patients who are prescribed with ACEI and ARB more likely had a high-risk profile than did those without these prescriptions. Future research would be useful to investigate whether the association found in our study was truly due to the drugs themselves or confounding.

Previous studies including a few from Hong Kong have reported that men with type 2 diabetes had a higher risk of CKD incidence than women.[4] On the contrary, we found that women had a faster rate of CKD progression and men had a higher rate of regression in our cohort. Interestingly, we also found that female patients with different drug usage showed distinct temporal patterns of CKD progression, whereas fewer discrepancies were observed in the male groups. This gender heterogeneity has been well recognised in the development of diabetic complications, but not in drug efficacy. Further studies are warranted to explore this important gender heterogeneity.

The results of this study have important clinical implications. Few previous studies have considered the prediction model for progression of early stages, however, a timely intervention on these early stages in the primary care level could have a significant impact to prevent or slow down the further progression to the end-stage renal failure. This would potentially reduce the disease burden and economic burden on the individual levels as well as on the secondary care of the healthcare system. Our prediction models were developed from the routine clinical data of electronic health records and thereby it could be automatically integrated into the existing clinical information system and easily adopted by clinicians to visualise the CKD progression risks in primary care settings. The graphic presentation of individualised

CKD risk prediction could be a useful tool to facilitate risk communication and health education for disease management of diabetes.

Interestingly, we found a small peak of CKD progression risks within half to 1 year since enrolment, even after we adjusted for check-up frequency (figures 2 and 3). It is of note that these participants had different lengths of years with type 2 diabetes and some already had developed complications when first enrolled into the RAMP-DM Programme. We speculate that these patients progressed faster than others, but it was actually due to their delayed laboratory tests and clinical visits. Furthermore, although we adjusted the intervals of follow-up checks in the model by calculating adjusted incidence rates, it is difficult to completely eliminate the impact of inconsistent intervals of clinical visits and laboratory tests. Therefore, the peaks within 1 year need cautious interpretations. Ideally, the prediction risk curve should be plotted over the time since diagnosis. However, this could be very difficult to achieve, as many patients did not have timely laboratory tests for diagnosis. In fact, many prediction models also used the year since enrolment as a time indicator, including a few local studies.[8 35]

There are a few limitations to this study. First, although we had a large sample size, the follow-up period was relatively short. CKD stages were defined by one test result due to the relatively short study period. Therefore, the predicted rates beyond 2 years tend to have wide confidence intervals. Second, a large number of patients in RAMP-DM were excluded due to incomplete eGFR, UACR and HbA1c data. It is unclear to us whether some were not scheduled to take tests due to mild conditions or others skipped clinical visits for some reason. Therefore, some of our model estimates might slightly overestimate or underestimate the true risks. Future studies with longer follow-up periods are warranted to obtain more accurate predictions. Last but not least, the study population were Chinese patients under primary care in a highly developed economic region. Future studies are warranted to apply our model to other populations as external validation and calibration, so that the models can eventually become a useful tool in clinical practice.

CONCLUSIONS

The MSMS model achieved satisfactory performance in predicting the progression of CKD in patients with type 2 diabetes under primary care. The prediction model developed from this study could be applied to build an online calculator for individual risks of CKD progression. This will greatly facilitate clinical decision making of individualised intervention plan and treatment target to slow the progression of CKD in Chinese adults with type 2 diabetes, based on biomarkers of glycaemic control, cardiovascular and renal function. The online calculator will also improve the risk communications of doctors and nurses with patients.

Author affiliations
[1]School of Nursing, Hong Kong Polytechnic University, Hong Kong, Hong Kong
[2]Department of Family Medicine and Primary Healthcare, New Territory West Cluster, Hospital Authority, Hong Kong, Hong Kong
[3]School of Optometry, Hong Kong Polytechnic University, Hong Kong, Hong Kong
[4]JC School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, China
[5]Shenzhen Research Institute of Chinese University of Hong Kong, Shenzhen, China
[6]Department of Applied Mathematics, Hong Kong Polytechnic University, Kowloon, Hong Kong

ORCID iDs
Lin Yang http://orcid.org/0000-0002-5964-3233
Daihai He http://orcid.org/0000-0003-3253-654X

REFERENCES
1 Federation ID. *IDF diabetes atlas*. 8th. Brussels, 2017.
2 Seuring T, Archangelidi O, Suhrcke M. The economic costs of type 2 diabetes: a global systematic review. *Pharmacoeconomics* 2015;33:811–31.
3 Jiao FF, Lam CLK, Wong CKH, *et al*. Patient-Level estimates of diabetic complications on direct medical cost. *Value Health* 2014;17:A340.
4 Luk AOY, Ma RCW, Lau ESH, *et al*. Risk association of HbA1c variability with chronic kidney disease and cardiovascular disease in type 2 diabetes: prospective analysis of the Hong Kong diabetes registry. *Diabetes Metab Res Rev* 2013;29:384.
5 KDIGO. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney International* 2013;3:1–150.
6 Hong Kong Diabetes Registry, Yang X, So WY, *et al*. Development and validation of an all-cause mortality risk score in type 2 diabetes. *Arch Intern Med* 2008;168:451–7.
7 Wan EYF, Fong DYT, Fung CSC, *et al*. Incidence and predictors for cardiovascular disease in Chinese patients with type 2 diabetes

mellitus - a population-based retrospective cohort study. *J Diabetes Complications* 2016;30:444–50.
8 Wan EYF, Yu EYT, Chin WY, *et al*. Ten-Year risk prediction models of complications and mortality of Chinese patients with diabetes mellitus in primary care in Hong Kong: a study protocol. *BMJ Open* 2018;8:e023070.
9 Yang X, So W-Y, Kong APS, *et al*. Development and validation of stroke risk equation for Hong Kong Chinese patients with type 2 diabetes: the Hong Kong diabetes registry. *Diabetes Care* 2007;30:65–70.
10 Wan EYF, Fong DYT, Fung CSC, *et al*. Prediction of new onset of end stage renal disease in Chinese patients with type 2 diabetes mellitus - a population-based retrospective cohort study. *BMC Nephrol* 2017;18:257–65.
11 Yang XL, So WY, Kong APS, *et al*. End-Stage renal disease risk equations for Hong Kong Chinese patients with type 2 diabetes: Hong Kong diabetes registry. *Diabetologia* 2006;49:2299–308.
12 Meira-Machado L, de Uña-Alvarez J, Cadarso-Suárez C, *et al*. Multi-State models for the analysis of time-to-event data. *Stat Methods Med Res* 2009;18:195–222.
13 Tangri N, Inker LA, Hiebert B, *et al*. A dynamic predictive model for progression of CKD. *Am J Kidney Dis* 2017;69:514–20.
14 Iacobelli S, Carstensen B. Multiple time scales in multi-state models. *Stat Med* 2013;32:5315–27.
15 Yang L, Chu TK, Lian J, *et al*. Risk factors of chronic kidney diseases in Chinese adults with type 2 diabetes. *Sci Rep* 2018;8:14686.
16 Lian J, McGhee SM, So C, *et al*. Five-Year cost-effectiveness of the patient Empowerment programme (PEP) for type 2 diabetes mellitus in primary care. *Diabetes Obes Metab* 2017;19:1312–6.
17 Carstensen B, Plummer M. Using Lexis Objects for Multi-State Models in R. *J Stat Softw* 2011;38:1–18.
18 McLachlan G, Do K-A AC. *Analyzing microarray gene expression data*. John Wiley & Sons, 2005.
19 Owen A. Empirical likelihood ratio confidence regions. *The Annals of Statistics* 1990;18:90–120.
20 Owen ARTB. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 1988;75:237–49.
21 Nagelkerke NJD. A note on a general definition of the coefficient of determination. *Biometrika* 1991;78:691–2.
22 Andrésdóttir G, Jensen ML, Carstensen B, *et al*. Improved prognosis of diabetic nephropathy in type 1 diabetes. *Kidney Int* 2015;87:417–26.
23 Carstensen B, Read SH, Friis S, *et al*. Cancer incidence in persons with type 1 diabetes: a five-country study of 9,000 cancers in type 1 diabetic individuals. *Diabetologia* 2016;59:980–8.
24 Rasmussen BSB, Yderstraede KB, Carstensen B, *et al*. Substantial reduction in the number of amputations among patients with diabetes: a cohort study over 16 years. *Diabetologia* 2016;59:121–9.
25 Plummer M, Carstensen B. Lexis : An *R* Class for Epidemiological Studies with Long-Term Follow-Up. *J Stat Softw* 2011;38.
26 Efron B. Logistic regression,survival analysis, and the Kaplan-Meier curve. *J Am Stat Assoc* 1988;83:414–25.
27 Efron B. The two-way proportional hazards model. *J Royal Statistical Soc B* 2002;64:899–909.
28 Whitehead J. Fitting COX's regression model to survival data using GLIM. *J R Stat Soc C-appl* 1980;29:268–75.
29 Echouffo-Tcheugui JB, Kengne AP. Risk models to predict chronic kidney disease and its progression: a systematic review. *PLoS Med* 2012;9:e1001344.
30 Schanstra JP, Zürbig P, Alkhalaf A, *et al*. Diagnosis and prediction of CKD progression by assessment of urinary peptides. *J Am Soc Nephrol* 2015;26:1999–2010.
31 Desai AS, Toto R, Jarolim P, *et al*. Association between cardiac biomarkers and the development of ESRD in patients with type 2 diabetes mellitus, anemia, and CKD. *Am J Kidney Dis* 2011;58:717–28.
32 Keane WF, Zhang Z, Lyle PA, *et al*. Risk scores for predicting outcomes in patients with type 2 diabetes and nephropathy: the RENAAL study. *Clin J Am Soc Nephrol* 2006;1:761–7.
33 Looker HC, Colombo M, Hess S, *et al*. Biomarkers of rapid chronic kidney disease progression in type 2 diabetes. *Kidney Int* 2015;88:888–96.
34 Tangri N, Stevens LA, Griffith J, *et al*. A predictive model for progression of chronic kidney disease to kidney failure. *JAMA* 2011;305:1553–9.
35 Wan EYF, Fong DYT, Fung CSC, *et al*. Development of a cardiovascular diseases risk prediction model and tools for Chinese patients with type 2 diabetes mellitus: a population-based retrospective cohort study. *Diabetes Obes Metab* 2018;20:309–18.