의학석사 학위논문

# 만성신장질환에 대한 정적 및 동적 예측 모델

**Static and dynamic predictive model for chronic kidney disease progression: KNOW-CKD study**

2022년 8월

서울대학교 대학원

융합전공 혁신의과학 의학과 예방의학 전공

안 우 주

# 만성신장질환에 대한 정적 및 동적 예측 모델

지도교수 박수경

이 논문을 안우주 석사 학위논문으로 제출함

2022년 04월

서울대학교 대학원

융합전공 혁신의과학 의학과 예방의학

안 우 주

안우주의 석사 학위논문을 인준함

2022년 7월

위 원 장 오국환 (인)

부 위 원 장 박수경 (인)

위 원 이중엽 (인)

# Static and dynamic predictive model for chronic kidney disease progression: KNOW-CKD study

**by**

**Wooju An**

**A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Philosophy in Preventive Medicine at the Seoul National University College of Medicine**

**July 2022**

**Approved by Thesis Committee:**

**Professor _____Chairman**

**Professor _____Vice Chairman**

**Professor _____**

## Abstract

### Static and dynamic predictive model for chronic kidney disease progression: KNOW-CKD study



Wooju An

Integrated Major in Department of Preventive Medicine

Seoul National University

College of Medicine



**Introduction**: Chronic kidney disease (CKD) is recognized as a major public health problem worldwide. The estimated global prevalence of CKD is 13.4% (11.7-15.1%), and it is estimated that 4.902 to 7.083 million patients with end stage renal disease (ESRD) need kidney replacement treatment. The number of patients with CKD increased by 8% annually from 2016 to 2020 in Korea. In addition, the incidence and prevalence of ESRD patients are increasing. In longitudinal observational study, CKD patients had nonlinear or long-term non-progress rather than steady eGFR reduction over time. Therefore, we need a dynamic model that uses latest-available-measurement rather than a static model that uses data from a baseline visit. The objective of this study is to compare our data with a dynamic and static predictive model using a conventional statistical model and machine learning methods.

**Methods**: Data were collected from the Korean Cohort Study for Outcomes in Patients With Chronic Kidney Disease (KNOW-CKD). We analyzed data for 1,585 patients with CKD, excluding patients with autosomal dominant polycystic kidney disease (ADPKD). Missing values were imputed by multiple imputation. The primary outcome was a composite of the first occurrence of a 50% decline in eGFR from the baseline value or the onset of ESRD during follow-up. We made a static which is using baseline values and dynamic prediction model using the latest available measurement values. Cox proportional hazards, random survival forest, gradient boosting machine, and elastic net were used in predictive model. We split dataset into training and validation set by four methods which are split-sample,

temporal follow-up, temporal registration, and geographical validation. The model's performance was evaluated using time-dependent area under curve (AUC).

**Results**: Among 1,585 subjects, mean age was 54.9 ± 12.1 years and 37.2% were women. Mean urine albumin/creatinine ratio was 1.1 ± 1.5 g/g and the mean baseline estimated glomerular filtration rate (eGFR) was 49.3 ± 28.7 mL/min/1.73$m^2$. The mean systolic blood pressure was 127.7 ± 16.6 mmHg. Subjects with diabetes mellitus (DM) comprised 27.2% of the study subjects. From baseline to 3-years follow up, the hazard ratios (HRs) of eGFR and DM decreased and C statistics increased. Time-dependent AUC showed the similar tendency in cox and machine learning methods. Furthermore, it demonstrated a good level of discrimination capability in internal validations by split-sample, temporal follow-up, temporal registration, and geography methods.

**Conclusions**: In this study, we compared between static and dynamic models, finding that dynamic model showed stable and high time-dependent AUC values in all predictive models. Considering eGFR, which is difficult to predict, dynamic models that predict short-term events may be better when developing prediction models of CKD progression.

---

**Keywords**: Chronic kidney disease, predictive model, internal validation, machine learning, random survival forest, gradient boosting machine, elastic net

**Student Number**: 2020-25590

## Contents

# List of tables

# List of figures

# 1. Introduction

## 1.1. Epidemiology of chronic kidney disease

Chronic kidney disease (CKD) is defined as an abnormality of the kidney structure or function that lasts more than three months, which has a specific health effect. This is defined as a glomerular filtration rate (GFR) less than 60 mL/min/1.73$m^2$ or one or more markers of kidney dysfunction including albuminuria . The main causes of CKD are diabetes mellitus, hypertension, glomerulonephritis, and polycystic kidney disease (PKD). PKD is a common hereditary renal disease (1).

CKD is recognized as a major public health problem worldwide (2). The estimated global prevalence of CKD is 13.4% (11.7-15.1%), and it is estimated that 4.902 to 7.083 million patients with end stage renal disease (ESRD) need kidney replacement treatment (2). As a result of a cross-sectional sample of the Korea National Health and Nutrition Examination Survey (KNHANES) from 2011 to 2012, the prevalence of CKD was estimated to be 7.9% in adults over the age of 19 in Korea (3). According to data from the Health Insurance Review & Assessment Service, the number of patients with CKD increased by 8% annually from 2016 to 2020 in Korea (Figure 1) (4). In addition, the incidence and prevalence of ESRD patients are increasing. The prevalence of ESRD in 2019 has doubled over the past 10 years (5). ESRD corresponds to the fifth stage of CKD in which kidney function cannot be restored, and treatment of ESRD imposes significant social direct and indirect costs (6). Early diagnosis is important because the progression process of CKD occurs gradually over a long period of time, and in many cases progresses without any symptoms. By classifying patients to diagnose CKD progression early through a prediction model, early treatment can slow the progression and prevent complications.

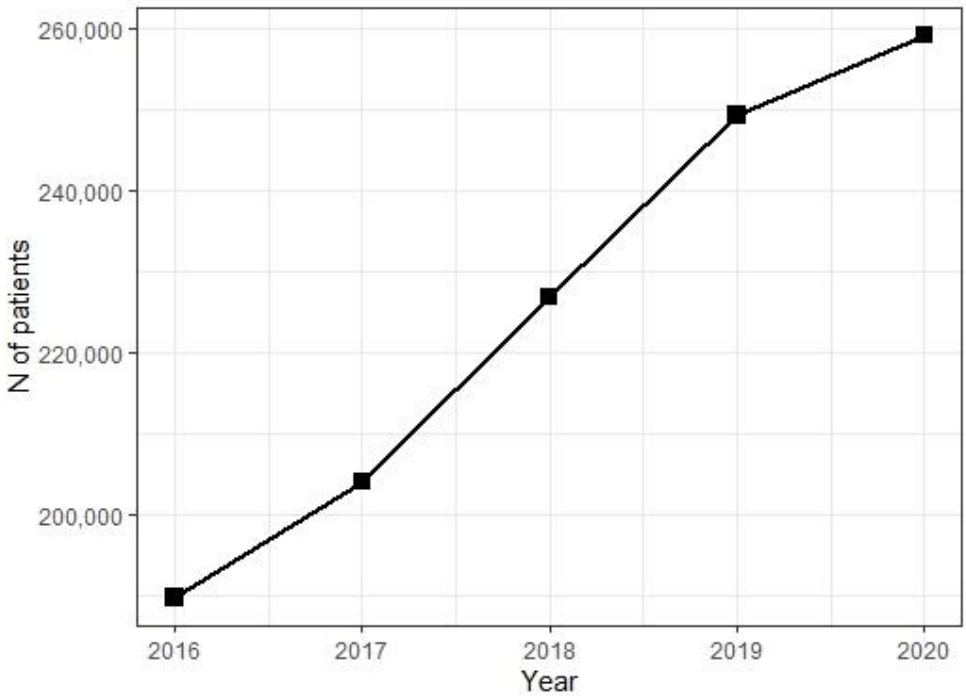



Figure 1. Number of patients with chronic kidney disease in Korea, 2016-2020.

1.2. Prediction model

These days, many studies have made prediction models for development or progression of CKD (7-13). Up to 2012, 26 articles were reported on 30 CKD occurrence and 17 CKD progression prediction models (14). The split sample approach, which is a widely used method of dividing data at an appropriate rate, is used to evaluate the performance of the model. Several studies used a split-sample approach that randomly selects a portion of the data to derive a predictive model and to evaluate the model by selecting the remaining portions (7-10). These studies can show unfair and biased evaluations of the predictive models (15). Therefore, it is necessary to evaluate the model by dividing the data in various ways. When evaluating the performance of the model, there are studies that present a standard ROC curve analysis that assumes for an individual as fixed over the whole study period (11-13). In the survival analysis, it is necessary to construct a time-dependent

ROC curve at several time points using additional information on the disease onset time of each individual (16). In a longitudinal observational study, CKD patients had nonlinear or long-term non-progress rather than a steady eGFR decline over time (17, 18). Therefore, we need a dynamic model that uses the latest-available-measurements rather than a static model that uses data from a baseline visit. In the case of an annual follow-up prospective cohort, progression of CKD should be predicted through short-term predictions.

### 1.3. Objectives

The objective of this study was to compare our data with a dynamic and static predictive model using a conventional statistical model and machine learning methods. We further analyzed predictive models using machine learning methods and a conventional static model. To evaluate these models, we divided the dataset by four different methods and used time-dependent area under curve (AUC).

## 2. Methods

### 2.1 Data source and study population

The Korean Cohort Study for Outcomes in Patients With Chronic Kidney Disease (KNOW-CKD) is a nationwide prospective cohort study involving 9 general tertiary-care hospitals in Korea. The KNOW-CKD was a multi-center prospective cohort study with a total of 2,341 patients enrolled from 2011 to 2015 in Korea voluntarily providing informed consent, excluding 103 patients without data for isotope dilution mass spectrometry (IDMS)-calibrated creatinine. Furthermore, KNOW-CKD enrolled Korean patients with CKD aged 20 to 75 years and with CKD stages 1 to 5 who were in non-dialysis-dependent. The detailed design and method of the study have previously been described elsewhere (19). The eGFR was determined using the

CKD-Epidemiology Collaboration equation (20). The exclusion criteria were as follows: (1) inability or unwillingness to provide written consent, (2) previous chronic dialysis or organ transplantation, (3) heart failure (New York Heart Association class 3 or 4) or cirrhosis (Child-Pugh class 2 or 3), (4) past or current history of malignancy, (5) pregnancy, or (6) a single kidney due to trauma or kidney donation. Among 2,238 patients, we excluded 59 without available data on renal outcome. Autosomal dominant polycystic kidney disease (ADPKD) is one of the main causes of ESRD (21). Because it is caused by mutations either in PKD1 or PKD2, subjects in the PKD subgroup have different collective characteristics from hypertensive nephropathy (HTN), diabetic nephropathy (DN) and glomerulonephritis (GN). Therefore, we also excluded 594 patients who had PKD or no genomic DNA available. We finally analyzed 1,585 patients in this study (Figure 2).

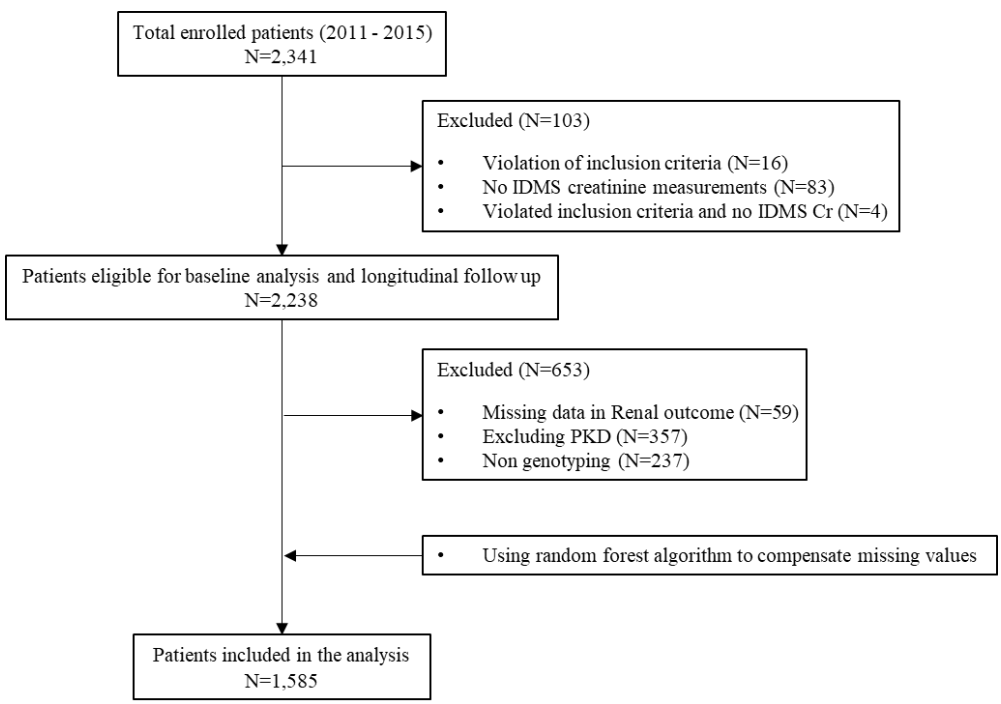



Figure 2. Flow chart of the study participants. Abbreviations: PKD, polycystic kidney disease.

2.2 Measurements and selection of variables

Demographic details were collected from all participants, including age and sex. The eGFR was estimated using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation using creatinine(20). Resting blood pressure was measured with electronic sphygmomanometers and cuffs of appropriate size three times for the average blood pressure. Diabetes (DM) was defined by self-reporting or use of hypoglycemic medications. The body mass index (BMI) was calculated by dividing the initial body weight (kg) by height squared ($m^2$). Urine albumin/creatinine ratio (UACR) was calculated as the urine albumin concentration divided by the urine creatinine concentration (g/g). Through backward variable selection, which is a statistical variable selection method, and conference with clinicians, our study constructed a model by putting clinically important variables in order.

2.3 Outcome definition

The primary outcome was a composite of the first occurrence of a 50% decline in eGFR from the baseline value or the onset of ESRD during follow-up. ESRD was defined as the initiation of dialysis or kidney transplantation. The study protocol indicated a creatinine measurement at 0, 6, and 12 months and then yearly thereafter. However, patients with CKD stage $\geq$ 3 were under close observation and had been followed up at 1- to 3-month intervals by all participating centers. Regardless of the study protocol, patients who reached the end points were reported by each center.

2.4 Development and evaluation of a predictive model

In this study, four predictive models were established to predict the progression of CKD. Cox proportional hazards (CPH) model is conventional statistical technique to explore the correlation between the survival time and covariates. Gradient boosting machine (GBM) is an additive algorithm, and a boosting technique is

applied to construct weaker classifiers into multiple iterations to get an improved and stronger model. The main idea of boosting is to add new models to the ensemble sequentially. In essence, boosting attacks the bias-variance-tradeoff by starting with a weak model and sequentially boosts its performance by continuing to build new trees, where each new tree in the sequence tries to fix where the previous one made the biggest mistakes (22). Random survival forest (RSF) is one of the ensemble machine learning methods with randomly produces independent decision trees, each of which receives a random subset of samples and randomly selects a subset of variables at each split in the tree for prediction. The average of the prediction of each individual tree is the result in the final prediction from the random survival forest algorithm (23). Elastic net is a type of normalized linear regression that adds constraints on linear regression coefficients to prevent the model from overfitting. Elastic net is a model compromised of the Ridge and Lasso regression.

We evaluated the model performance of the linear CPH model in the training set using the concordance index. The concordance index, or C-index, is a global measure of discriminative power of a survival model. We also used time-dependent receiver operating characteristic (ROC) curves, and the AUC values for discrimination accuracy to evaluate the performance at the specific prediction time points, of 1, 2, and 4 years, and the AUC was calculated by bootstrap. Each of the AUCs was compared using the paired t-test. The 3-year AUC was excluded because the missing value obtained in the 2-year period was more than 90%. Time-dependent AUC was performed using the risksetROC R package (24). Cumulative/dynamic (C/D), incident/dynamic (I/D), and incident/static (I/S) were proposed to estimate the time-dependent sensitivity and specificity for the censored event-times (25). The incident/dynamic defined individuals with events at time point t as the case group and individuals without events at time point t as the control group. In our study, it was appropriate to use an incident/dynamic because the aim of this study was for early treatment by identifying patients at high risk of CKD progression in the near

future (26). The standard ROC curve analysis defines the disease state of an individual after assuming that it is fixed for the entire study period (16). However, research periods generally require long follow-up, and during this period, the individual has a different time for the onset of the disease. Time-dependent ROC curve analysis enables us to configure and compare ROC curves at multiple time points using information about the onset time by observing the disease status of an individual at each time point (16).

### 2.5 Statistical analysis

The baseline characteristics of the study participants are described using the means with standard deviations (SD), and frequency is described using percentages for categorical variables. Student's t-test for continuous variables and Chi-squared test for categorical variables were used to verify if there were significant differences.

We made a static model using demographic and laboratory data from the baseline to the hospital and a dynamic model using the latest available measurement values. In the dynamic model, we included sex and diabetes mellitus as time-independent variables using the baseline value and other variables using the latest available measurement values.

We split the dataset into a training set and a validation set by four methods. Split-sample method divided the dataset into 70% and 30% subsets using the caret R package (27). We considered creating balanced splits of the data, which means that random sampling occurs within each class and preserves the overall class distribution of the data. Temporal follow-up method defined those who were followed up until the eighth year as the training set and those who did not develop a renal outcome before the eighth year and were followed up until the ninth year as the validation set. Temporal registration method divided the dataset into those who were registered at 1, 2, and 3 years and at 4 and 5 years in the same institution, which was split into 70% and 30%. Geography method divided the dataset into hospitals in the

metropolitan area and province, which was split into 80% and 20% (28). The methods of splitting the dataset in this study are summarized in Figure 3.

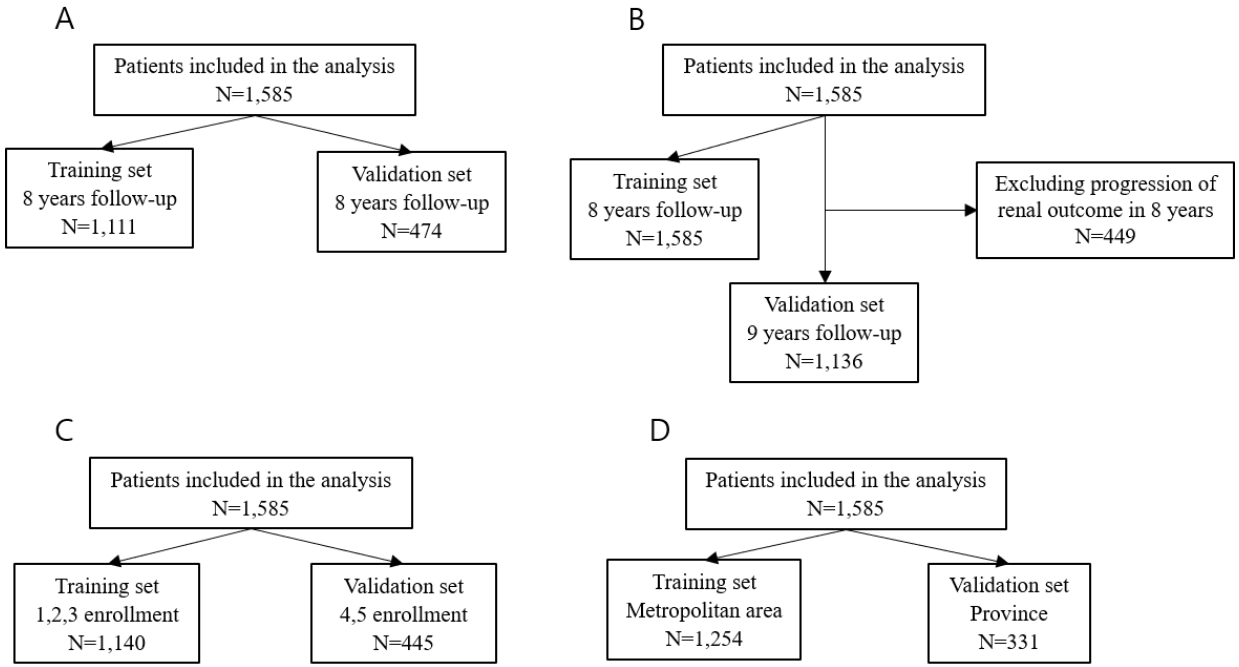



Figure 3. Flow chart of different methods for constructing predictive models. We split the dataset into a training and a validation set by four methods. (A) Split-sample: divided the dataset into 70% and 30%. (B) Temporal follow-up: divided into those who followed up until the eighth year as training set and those who followed up until the ninth year among those who have not developed renal outcome until the eighth year as validation set. (C) Temporal registration: divided the dataset into those who were registered at 1, 2, 3 years and at 4, 5 years in the same institution. (D) Geography: divided the dataset into hospitals in metropolitan area and province.

In this study, four predictive models were trained to predict the progression of CKD. CPH model, GBM, RSF, and elastic net were trained using the survival, gbm, RandomForestSRC, and glmnet R packages, respectively (29-31).

We used a grid search to investigate the combination of hyperparameters and defined the optimal hyperparameter with the best performance on a given dataset (Supplementary Table 1-3). Multiple imputation was performed using the MissForest package in R (32). MissForest is another machine learning-based data imputation algorithm that operates on the Random Forest algorithm. By averaging over many unpruned classifications or regression trees, random forest intrinsically constitutes a multiple imputation scheme.

The proportional hazard assumption of the Cox proportional hazard model was tested using the Schoenfeld residuals test (33) (Supplementary Table 4-6). eGFR violated the proportional hazards assumption. To solve that problem, we applied non-parametric models based on RSF, GBM, and elastic net as an alternative approach to the Cox proportional hazards models when the PH assumption was violated.

Data management and statistical analyses were performed using the R (version 4.1.2) statistical software.

## 3. Results

### 3.1 Baseline characteristics of the study population

Table 1 shows the baseline characteristics of the study subjects comparing before and after imputation. Among 1,585 subjects, mean age was 54.9 $\pm$ 12.1 years and 37.2% were women. Mean UACR was 1.1 $\pm$ 1.5 g/g and the mean baseline eGFR was 49.3 $\pm$ 28.7 mL/min/1.73$m^2$. The mean SBP was 127.7 $\pm$ 16.6 mmHg. Subjects with DM comprised 27.2% of the study subjects. The proportion of BMI higher than 30 kg/$m^2$ was 6.9%. Age between renal event and non-renal event group was not statistically significant (54.7 vs 55.0, Student's t-test $p$=0.69). UACR was

higher in renal event group than that in non-renal event group (1.8g/g vs 0.7g/g, Student's t-test $p<0.01$). eGFR was lower in renal event group than that in non-renal event group (30.1mL/min/1.73m$^2$ vs 58.2 mL/min/1.73m$^2$, Student's t-test $p<0.01$). SBP was higher in renal event group than that in non-renal event group (131.8mmHg vs 125.7 mmHg, Student's t-test $p<0.01$). Sex between renal and non-renal event groups was not statistically significant (39% vs 36.3%, Chi-squared test $p=0.32$). Renal event group was more likely to have DM than non-renal event group (43.8% vs 19.4%, Chi-squared test $p<0.01$). Number of subjects with BMI of 30 kg/m$^2$ or above was not different between renal and non-renal event groups.

There was no difference in the values before and after imputation.

Table 1. Baseline demographic and clinical characteristics of KNOW-CKD subjects before and after multiple imputation

| | **Before imputation** | | | | **After imputation** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Total (n=1,585)** | **Renal[1] event group (n=505)** | **Non-renal event group (n=1,080)** | ***p*** | **Total (n=1,585)** | **Renal[1] event group (n=505)** | **Non-renal event group (n=1,080)** | ***p*** |
| | **Mean (SD)** | **Mean (SD)** | **Mean (SD)** | | **Mean (SD)** | **Mean (SD)** | **Mean (SD)** | |
| Age (years) | 54.9 (12.1) | 54.7 (11.8) | 55.0 (12.3) | 0.69 | 54.9 (12.1) | 54.7 (11.8) | 55.0 (12.3) | 0.69 |
| UACR (g/g) | 1.1 (1.5) | 1.8 (1.9) | 0.7 (1.1) | <0.01 | 1.1 (1.5) | 1.8 (1.9) | 0.7 (1.1) | <0.01 |
| eGFR ($mL/min/1.73m^2$) | 49.3 (28.7) | 30.1 (19.2) | 58.2 (28.0) | <0.01 | 49.3 (28.7) | 30.1 (19.2) | 58.2 (28.0) | <0.01 |
| SBP (mmHg) | 127.7 (16.6) | 131.8 (17.8) | 125.7 (15.7) | <0.01 | 127.7 (16.6) | 131.8 (17.8) | 125.7 (15.7) | <0.01 |
| | **N (%)** | **N (%)** | **N (%)** | | **N (%)** | **N (%)** | **N (%)** | |
| Female | 589 (37.2) | 197 (39.0) | 392 (36.3) | 0.32 | 589 (37.2) | 197 (39.0) | 392 (36.3) | 0.32 |
| DM | 431 (27.2) | 221 (43.8) | 210 (19.4) | <0.01 | 431 (27.2) | 221 (43.8) | 210 (19.4) | <0.01 |
| BMI ($kg/m^2$) (≥30) | 109 (6.9) | 36 (7.1) | 73 (6.8) | 0.90 | 110 (6.9) | 37 (7.3) | 73 (6.8) | 0.76 |

Note: Values for categorical variables are given as number (percentage); values for continuous variable, as mean ± standard deviation or median [interquartile range]. eGFR was calculated using the CKD-EPI creatinine equation. The variables were compared in the following ways: female, male as reference; DM, patients without DM as reference; BMI, <30 as reference.

1. Composite of the first occurrence of a 50% decline in eGFR from the baseline value or the onset of ESRD during follow-up. ESRD was defined as the initiation of dialysis or kidney transplantation

Abbreviations: eGFR, estimated glomerular filtration rate; UACR, urine albumin-to-creatinine ratio; DM, diabetes mellitus; BMI, body mass index; SBP, systolic blood pressure.

Table 2 shows the results of CPH analysis using training sets by four methods to divide the dataset of baseline. The hazard ratios (95% CI) of eGFR in multiple model were 0.94 (0.93-0.94) for each increase of 1 mL/min/1.73$m^2$ in split-sample, 0.94 (0.93-0.94) for each increase of 1 mL/min/1.73$m^2$ in temporal follow-up, 0.94 (0.93-0.95) for each increase of 1 mL/min/1.73$m^2$ in temporal registration, 0.94 (0.94-0.95) for each increase of 1 mL/min/1.73$m^2$ in geography. The hazard ratios (95% CI) of UACR in multiple model were 1.49 (1.40-1.57) in split-sample, 1.53 (1.46-1.61) in temporal follow-up, 1.54 (1.46-1.63) in temporal registration, 1.55 (1.46-1.63) in geography. The hazard ratios (95% CI) of SBP in multiple model were 1.11 (1.04-1.19) for each increase of 10 mmHg in split-sample, 1.09 (1.03-1.15) for each increase of 10 mmHg in temporal follow-up, 1.08 (1.01-1.15) for each increase of 10 mmHg in temporal registration, 1.07 (1.006-1.14) for each increase of 10 mmHg in geography. The hazard ratios (95% CI) of DM in multiple model were 1.58 (1.24-2.01) in split-sample, 1.78 (1.44-2.19) in temporal follow-up, 1.82 (1.46-2.26) in temporal registration, 1.54 (1.22-1.94) in geography.

Table 3 shows the results of CPH analysis using training sets by four methods to divide the data of 1-year follow up. The hazard ratios (95% CI) of eGFR in multiple model were 0.92 (0.91-0.93) for each increase of 1 mL/min/1.73$m^2$ in split-sample, 0.92 (0.91-0.93) for each increase of 1 mL/min/1.73$m^2$ in temporal follow-up, 0.92 (0.91-0.93) for each increase of 1 mL/min/1.73$m^2$ in temporal registration, 0.92 (0.92-0.93) for each increase of 1 mL/min/1.73$m^2$ in geography. The hazard ratios (95% CI) of UACR in multiple model were 1.23 (1.18-1.28) in split-sample, 1.24 (1.20-1.29) in temporal follow-up, 1.73 (1.61-1.86) in temporal registration, 1.65 (1.53-1.77) in geography. The hazard ratios (95% CI) of SBP in multiple model were 1.14 (1.05-1.24) for each increase of 10 mmHg in split-sample, 1.11 (1.03-1.96) for each increase of 10 mmHg in temporal follow-up, 1.12 (1.04-1.22) for each increase of 10 mmHg in geography. The hazard ratios (95% CI) of DM in multiple model were 1.50 (1.17-1.94) in split-sample, 1.79 (1.43-2.24) in temporal follow-up, 1.48

(1.17-1.88) in temporal registration, 1.28 (1.003-1.63) in geography.

Table 4 shows the results of CPH analysis using training sets in four methods to divide the data of 3-year follow up. The hazard ratios (95% CI) of eGFR in multiple model were 0.92 (0.90-0.93) for each increase of 1 mL/min/1.73$m^2$ in split-sample, 0.91 (0.89-0.92) for each increase of 1 mL/min/1.73$m^2$ in temporal follow-up, 0.92 (0.91-0.93) for each increase of 1 mL/min/1.73$m^2$ in temporal registration, 0.92 (0.91-0.93) for each increase of 1 mL/min/1.73$m^2$ in geography. The hazard ratios (95% CI) of UACR in multiple model were 1.71 (1.50-1.96) in split-sample, 1.76 (1.55-2.00) in temporal follow-up, 1.81 (1.61-2.03) in temporal registration, 1.93 (1.69-2.20) in geography.

Each table showed similar results among the four methods to divide the dataset. From baseline to 3-year follow up, the HRs of eGFR and DM decreased and C statistics increased.

Table 2. Hazard ratios for chronic kidney disease progression in the training sets in the static models that use data from baseline

| | Split-sample[1] | | Temporal follow-up[2] | | Temporal registration[3] | | Geography[4] | |
|---|---|---|---|---|---|---|---|---|
| | **HR (95% CI)[5]** | **HR (95% CI)[6]** | **HR (95% CI)[5]** | **HR (95% CI)[6]** | **HR (95% CI)[5]** | **HR (95% CI)[6]** | **HR (95% CI)[5]** | **HR (95% CI)[6]** |
| eGFR (mL/min/1.73m$^2$) | 0.94 (0.93-0.94) | 0.94 (0.93-0.94) | 0.94 (0.93-0.95) | 0.94 (0.93-0.94) | 0.94 (0.93-0.95) | 0.94 (0.93-0.95) | 0.94 (0.94-0.95) | 0.94 (0.94-0.95) |
| UACR (g/g) | 1.57 (1.50-1.65) | 1.49 (1.40-1.57) | 1.62 (1.55-1.69) | 1.53 (1.46-1.61) | 1.58 (1.50-1.66) | 1.54 (1.46-1.63) | 1.67 (1.59-1.75) | 1.55 (1.46-1.63) |
| Age (years) | 1.01 (0.996-1.01) | 0.97 (0.96-0.98) | 1.00 (0.99-1.01) | 0.97 (0.96-0.98) | 1.00 (0.996-1.01) | 0.98 (0.97-0.98) | 1.00 (0.99-1.01) | 0.97 (0.96-0.98) |
| SBP (mmHg) | 1.23 (1.16-1.31) | 1.11 (1.04-1.19) | 1.27 (1.20-1.34) | 1.09 (1.03-1.15) | 1.26 (1.19-1.34) | 1.08 (1.01-1.15) | 1.25 (1.18-1.33) | 1.07 (1.006-1.14) |
| Female | 1.06 (0.86-1.31) | 0.84 (0.67-1.05) | 1.06 (0.88-1.28) | 0.85 (0.70-1.03) | 1.05 (0.86-1.29) | 0.82 (0.67-1.01) | 0.98 (0.80-1.21) | 0.84 (0.68-1.04) |
| DM | 3.00 (2.43-3.70) | 1.58 (1.24-2.01) | 3.26 (2.70-3.93) | 1.78 (1.44-2.19) | 3.16 (2.58-3.86) | 1.82 (1.46-2.26) | 2.97 (2.43-3.63) | 1.54 (1.22-1.94) |
| BMI (kg/m$^2$) (≥30) | 1.19 (0.81-1.77) | 0.83 (0.55-1.25) | 1.20 (0.84-1.71) | 0.89 (0.62-1.28) | 1.24 (0.84-1.84) | 1.01 (0.68-1.51) | 1.15 (0.79-1.67) | 0.83 (0.57-1.22) |
| C-index | | 0.87 | | 0.87 | | 0.85 | | 0.86 |

Note: The variables were compared in the following ways: female, male as reference; DM, patients without DM as reference; BMI, <30 as reference.
1. Training set which consists of 70% of all patients.
2. Training set which consists of all patients following up 8-years.
3. Training set which consists of recruited patients from 2011 until 2014.
4. Training set which consists of patients registered at hospitals in Seoul.
5. A crude analysis without adjustment.
6. All predictors were included in the model.
Abbreviations: eGFR, estimated glomerular filtration rate; UACR, urine albumin-to-creatinine ratio; DM, diabetes mellitus; BMI, body mass index; SBP, systolic blood pressure; HR, hazard ratio

Table 3. Hazard ratios for chronic kidney disease progression in the training sets in the dynamic models that use data from 1-year follow up

| | **Split-sample[1]** | | **Temporal follow-up[2]** | | **Temporal registration[3]** | | **Geography[4]** | |
|---|---|---|---|---|---|---|---|---|
| | **HR (95% CI)[5]** | **HR (95% CI)[6]** | **HR (95% CI)[5]** | **HR (95% CI)[6]** | **HR (95% CI)[5]** | **HR (95% CI)[6]** | **HR (95% CI)[5]** | **HR (95% CI)[6]** |
| eGFR (mL/min/1.73m$^2$) | 0.91 (0.90-0.92) | 0.92 (0.91-0.93) | 0.91 (0.90-0.92) | 0.92 (0.91-0.93) | 0.92 (0.91-0.93) | 0.92 (0.91-0.93) | 0.92 (0.91-0.93) | 0.92 (0.92-0.93) |
| UACR (g/g) | 1.22 (1.19-1.25) | 1.23 (1.18-1.28) | 1.24 (1.21-1.26) | 1.24 (1.20-1.29) | 1.50 (1.44-1.56) | 1.73 (1.61-1.86) | 1.51 (1.46-1.57) | 1.65 (1.53-1.77) |
| Age (years) | 1.00 (0.996-1.01) | 0.97 (0.96-0.98) | 1.00 (0.99-1.01) | 0.97 (0.96-0.98) | 1.00 (0.996-1.01) | 0.98 (0.97-0.99) | 1.00 (0.99-1.01) | 0.97 (0.96-0.98) |
| SBP (mmHg) | 1.32 (1.23-1.41) | 1.14 (1.05-1.24) | 1.35 (1.27-1.43) | 1.11 (1.03-1.96) | 1.30 (1.22-1.39) | 1.03 (0.95-1.12) | 1.40 (1.31-1.49) | 1.12 (1.04-1.22) |
| Female | 1.06 (0.86-1.31) | 0.70 (0.55-0.89) | 1.06 (0.88-1.28) | 0.77 (0.62-0.95) | 1.05 (0.86-1.29) | 0.78 (0.62-0.98) | 0.98 (0.80-1.21) | 0.78 (0.62-0.98) |
| DM | 3.00 (2.43-3.70) | 1.50 (1.17-1.94) | 3.26 (2.70-3.93) | 1.79 (1.43-2.24) | 3.16 (2.58-3.86) | 1.48 (1.17-1.88) | 2.97 (2.43-3.63) | 1.28 (1.003-1.63) |
| BMI (kg/m$^2$) (≥30) | 1.43 (0.99-2.05) | 0.74 (0.44-1.23) | 1.38 (0.998-1.92) | 0.73 (0.47-1.14) | 1.38 (0.96-1.99) | 1.01 (0.66-1.52) | 1.33 (0.95-1.88) | 1.00 (0.67-1.50) |
| C-index | | 0.89 | | 0.89 | | 0.88 | | 0.88 |

Note: The variables were compared in the following ways: female, male as reference; DM, patients without DM as reference; BMI, <30 as reference.
1. Training set which consists of 70% of all patients.
2. Training set which consists of all patients following up 8-years.
3. Training set which consists of recruited patients from 2011 until 2014.
4. Training set which consists of patients registered at hospitals in Seoul.
5. A crude analysis without adjustment.
6. All predictors were included in the model.
Abbreviations: eGFR, estimated glomerular filtration rate; UACR, urine albumin-to-creatinine ratio; DM, diabetes mellitus; BMI, body mass index; SBP, systolic blood pressure; HR, hazard ratio

Table 4. Hazard ratios for chronic kidney disease progression in the training sets in the dynamic models that use data from 3-years follow up

| | **Split-sample[1]** | | **Temporal follow-up[2]** | | **Temporal registration[3]** | | **Geography[4]** | |
|---|---|---|---|---|---|---|---|---|
| | **HR (95% CI)[5]** | **HR (95% CI)[6]** | **HR (95% CI)[5]** | **HR (95% CI)[6]** | **HR (95% CI)[5]** | **HR (95% CI)[6]** | **HR (95% CI)[5]** | **HR (95% CI)[6]** |
| eGFR (mL/min/1.73m$^2$) | 0.89 (0.88-0.90) | 0.92 (0.90-0.93) | 0.89 (0.88-0.90) | 0.91 (0.89-0.92) | 0.90 (0.89-0.91) | 0.92 (0.91-0.93) | 0.90 (0.89-0.91) | 0.92 (0.91-0.93) |
| UACR (g/g) | 1.88 (1.79-1.98) | 1.71 (1.50-1.96) | 1.88 (1.80-1.96) | 1.76 (1.55-2.00) | 1.89 (1.80-1.98) | 1.81 (1.61-2.03) | 1.93 (1.84-2.03) | 1.93 (1.69-2.20) |
| Age (years) | 1.00 (0.996-1.01) | 0.97 (0.96-0.99) | 1.00 (0.99-1.01) | 0.97 (0.95-0.98) | 1.00 (0.996-1.01) | 0.98 (0.96-0.99) | 1.00 (0.99-1.01) | 0.97 (0.96-0.99) |
| SBP (mmHg) | 1.29 (1.20-1.38) | 1.09 (0.97-1.22) | 1.33 (1.25-1.42) | 1.02 (0.91-1.15) | 1.32 (1.23-1.42) | 1.06 (0.96-1.18) | 1.36 (1.26-1.47) | 1.07 (0.94-1.22) |
| Female | 1.06 (0.86-1.31) | 0.75 (0.52-1.07) | 1.06 (0.88-1.28) | 0.88 (0.63-1.24) | 1.05 (0.86-1.29) | 0.84 (0.62-1.13) | 0.98 (0.80-1.21) | 0.88 (0.63-1.22) |
| DM | 3.00 (2.43-3.70) | 1.30 (0.89-1.91) | 3.26 (2.70-3.93) | 1.33 (0.92-1.92) | 3.16 (2.58-3.86) | 1.26 (0.90-1.77) | 2.97 (2.43-3.63) | 1.03 (0.72-1.49) |
| BMI (kg/m$^2$) (≥30) | 1.50 (1.07-2.11) | 0.85 (0.46-1.58) | 1.42 (1.04-1.93) | 0.64 (0.36-1.14) | 1.34 (0.95-1.90) | 0.64 (0.38-1.10) | 1.34 (0.96-1.86) | 0.58 (0.32-1.06) |
| C-index | | 0.90 | | 0.91 | | 0.90 | | 0.90 |

Note: The variables were compared in the following ways: female, male as reference; DM, patients without DM as reference; BMI, <30 as reference.
1. Training set which consists of 70% of all patients.
2. Training set which consists of all patients following up 8-years.
3. Training set which consists of recruited patients from 2011 until 2014.
4. Training set which consists of patients registered at hospitals in Seoul.
5. A crude analysis without adjustment.
6. All predictors were included in the model.
Abbreviations: eGFR, estimated glomerular filtration rate; UACR, urine albumin-to-creatinine ratio; DM, diabetes mellitus; BMI, body mass index; SBP, systolic blood pressure; HR, hazard ratio

3.2 Validation of prediction models

Performance of prediction models comparing the static and dynamic models were described in Table 5, 6 and Figure 4-7.

Most of time-dependent AUC values of the dynamic model were higher than those of the static model. The results of split-sample validation were as follows. Time-dependent AUC of CPH model were 0.814 (0.780-0.841) in 2 years, 0.805 (0.775-0.840) in 4 years from static model vs 0.843 (0.807-0.870) in 2 years, 0.892 (0.853-0.931) in 4 years from dynamic model. Time-dependent AUC of GBM were 0.825 (0.791-0.855) in 2 years, 0.819 (0.786-0.847) in 4 years from static model vs 0.874 (0.837-0.912) in 2 years, 0.866 (0.824-0.905) in 4 years from dynamic model. Time-dependent AUC of RSF were 0.792 (0.749-0.832) in 2 years, 0.760 (0.723-0.794) in 4 years from static model vs 0.839 (0.789-0.891) in 2 years, 0.809 (0.729-0.871) in 4 years from dynamic model. Time-dependent AUC of elastic net were 0.813 (0.784-0.845) in 2 years, 0.795 (0.765-0.829) in 4 years from static model vs 0.864 (0.828-0.896) in 2 years, 0.892 (0.852-0.932) in 4 years from dynamic model.

The results of temporal follow-up validation were as follows. Time-dependent AUC of CPH model were 0.771 (0.705-0.833) in 2 years, 0.759 (0.699-0.820) in 4 years from static model vs 0.807 (0.753-0.861) in 2 years, 0.871 (0.811-0.916) in 4 years from dynamic model. Time-dependent AUC of GBM were 0.806 (0.739-0.868) in 2 years, 0.779 (0.722-0.828) in 4 years from static model vs 0.843 (0.793-0.891) in 2 years, 0.875 (0.827-0.922) in 4 years from dynamic model. Time-dependent AUC of RSF were 0.793 (0.707-0.923) in 2 years, 0.659 (0.601-0.755) in 4 years from static model vs 0.851 (0.755-0.947) in 2 years, 0.775 (0.677-0.867) in 4 years from dynamic model. Time-dependent AUC of elastic net were 0.780 (0.712-0.852) in 2 years, 0.757 (0.701-0.812) in 4 years from static model vs 0.813 (0.771-0.850) in 2 years, 0.875 (0.818-0.918) in 4 years from dynamic model.

The results of temporal registration validation were as follows. Time-dependent AUC of GBM were 0.848 (0.814-0.896) in 2 years, 0.808 (0.748-0.894) in 4 years

from static model vs 0.889 (0.855-0.916) in 2 years, 0.871 (0.807-0.951) in 4 years from dynamic model. Time-dependent AUC of RSF were 0.804 (0.763-0.862) in 2 years, 0.695 (0.625-0.795) in 4 years from static model vs 0.866 (0.829-0.906) in 2 years, 0.850 (0.762-0.951) in 4 years from dynamic model.

The results of geographical validation were as follows. Time-dependent AUC of GBM were 0.862 (0.821-0.904) in 2 years, 0.807 (0.768-0.837) in 4 years from static model vs 0.915 (0.880-0.939) in 2 years, 0.862 (0.789-0.921) in 4 years from dynamic model. Time-dependent AUC of RSF were 0.832 (0.789-0.875) in 2 years, 0.719 (0.683-0.771) in 4 years from static model vs 0.905 (0.862-0.937) in 2 years, 0.812 (0.734-0.890) in 4 years from dynamic model.

This result showed the similar tendency in cox and machine learning methods. Furthermore, it demonstrated a good level of discrimination capability in all methods of splitting data.

Table 5. Comparison of time-dependent AUC between cox, GBM, RSF in predicting progression of chronic kidney disease from baseline

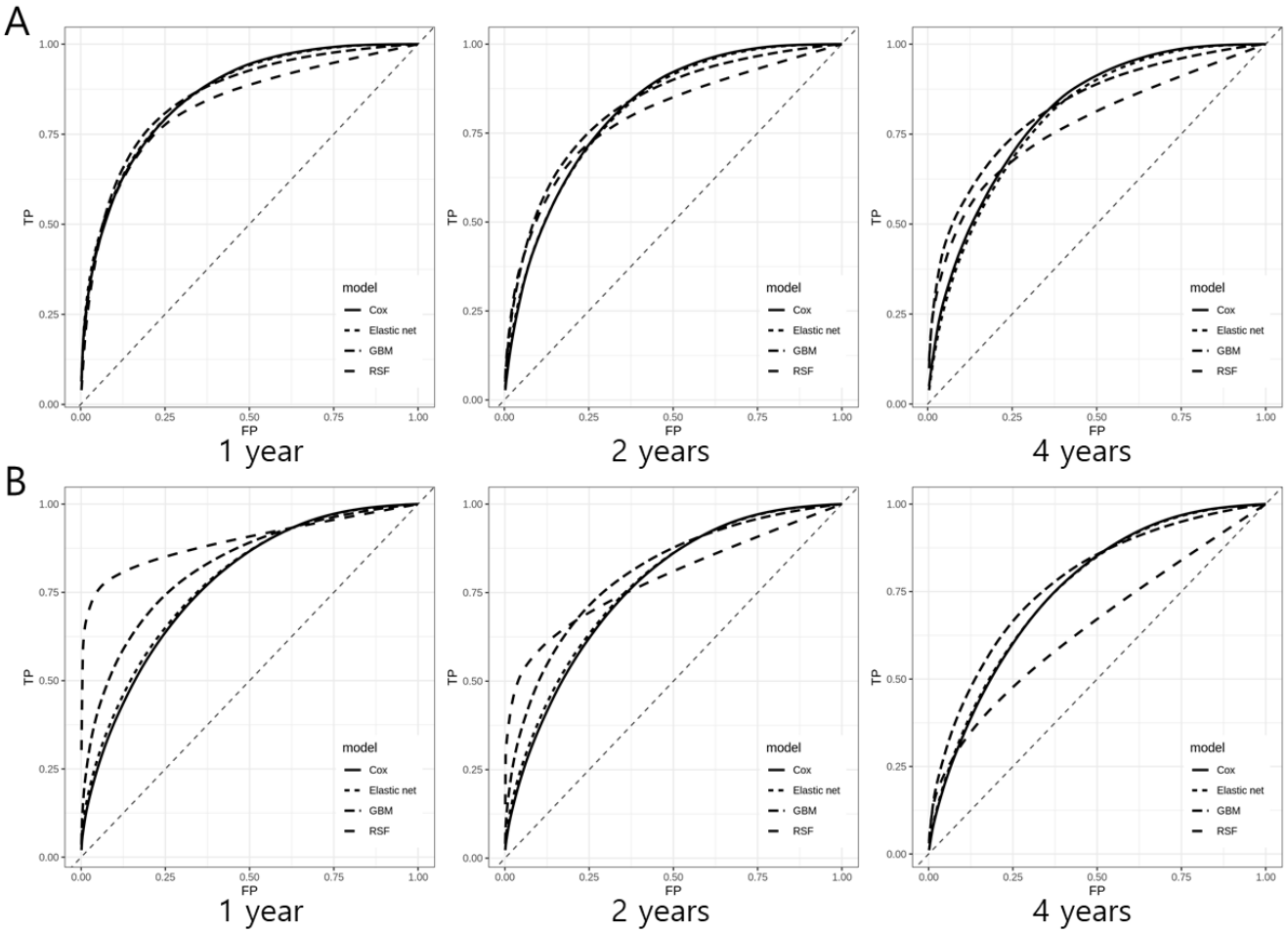
| Model | 1-yr AUC (95% CI) | 2-yr AUC (95% CI) | 4-yr AUC (95% CI) |
|---|---|---|---|
| **Split-sample validation** | | | |
| Cox | 0.854 (0.810-0.893) | 0.814 (0.780-0.841) | 0.805 (0.775-0.840) |
| GBM | 0.856 (0.805-0.899) | 0.825 (0.791-0.855) | 0.819 (0.786-0.847) |
| RSF | 0.822 (0.775-0.869) | 0.792 (0.749-0.832) | 0.760 (0.723-0.794) |
| Elastic net | 0.857 (0.811-0.896) | 0.813 (0.784-0.845) | 0.795 (0.765-0.829) |
| **Temporal follow-up validation** | | | |
| Cox | 0.781 (0.708-0.846) | 0.771 (0.705-0.833) | 0.759 (0.699-0.820) |
| GBM | 0.825 (0.755-0.891) | 0.806 (0.739-0.868) | 0.779 (0.722-0.828) |
| RSF | 0.899 (0.774-0.988) | 0.793 (0.707-0.923) | 0.659 (0.601-0.755) |
| Elastic net | 0.790 (0.721-0.859) | 0.780 (0.712-0.852) | 0.757 (0.701-0.812) |
| **Temporal registration validation** | | | |
| Cox | 0.855 (0.816-0.897) | 0.837 (0.805-0.871) | 0.825 (0.768-0.888) |
| GBM | 0.856 (0.817-0.916) | 0.848 (0.814-0.896) | 0.808 (0.748-0.894) |
| RSF | 0.850 (0.798-0.916) | 0.804 (0.763-0.862) | 0.695 (0.625-0.795) |
| Elastic net | 0.843 (0.797-0.895) | 0.820 (0.783-0.854) | 0.797 (0.733-0.877) |
| **Geographical validation** | | | |
| Cox | 0.885 (0.856-0.918) | 0.856 (0.818-0.895) | 0.798 (0.761-0.834) |
| GBM | 0.884 (0.852-0.918) | 0.862 (0.821-0.904) | 0.807 (0.768-0.837) |
| RSF | 0.880 (0.842-0.912) | 0.832 (0.789-0.875) | 0.719 (0.683-0.771) |
| Elastic net | 0.863 (0.829-0.899) | 0.839 (0.798-0.876) | 0.775 (0.739-0.813) |

Abbreviations: AUC, area under the curve; GBM, gradient boosting machine; RSF, random survival forest

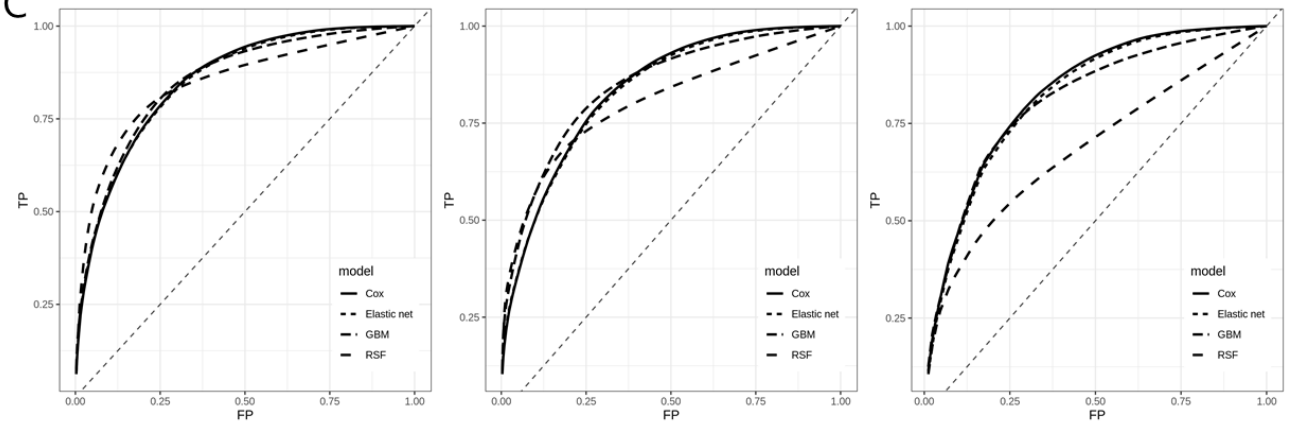
Table 6. Comparison of time-dependent AUC between cox, GBM, RSF in predicting progression of chronic kidney disease from latest available measurement

| Model | 1-yr AUC (95% CI)[1] | 2-yr AUC (95% CI)[2] | 4-yr AUC (95% CI)[3] |
|---|---|---|---|
| **Split-sample validation** | | | |
| Cox | 0.854 (0.810-0.893) | 0.843 (0.807-0.870) | 0.892 (0.853-0.931) |
| GBM | 0.857 (0.810-0.900) | 0.874 (0.837-0.912) | 0.866 (0.824-0.905) |
| RSF | 0.822 (0.775-0.869) | 0.839 (0.789-0.891) | 0.809 (0.729-0.871) |
| Elastic net | 0.857 (0.811-0.896) | 0.864 (0.828-0.896) | 0.892 (0.852-0.932) |
| **Temporal follow-up validation** | | | |
| Cox | 0.781 (0.708-0.846) | 0.807 (0.753-0.861) | 0.871 (0.811-0.916) |
| GBM | 0.814 (0.748-0.883) | 0.843 (0.793-0.891) | 0.875 (0.827-0.922) |
| RSF | 0.899 (0.774-0.988) | 0.851 (0.755-0.947) | 0.775 (0.677-0.867) |
| Elastic net | 0.790 (0.721-0.859) | 0.813 (0.771-0.850) | 0.875 (0.818-0.918) |
| **Temporal registration validation** | | | |
| Cox | 0.855 (0.816-0.897) | 0.789 (0.711-0.901) | 0.916 (0.865-0.955) |
| GBM | 0.859 (0.816-0.921) | 0.889 (0.855-0.916) | 0.871 (0.807-0.951) |
| RSF | 0.850 (0.798-0.916) | 0.866 (0.829-0.906) | 0.850 (0.762-0.951) |
| Elastic net | 0.843 (0.797-0.895) | 0.729 (0.646-0.871) | 0.904 (0.855-0.946) |
| **Geographical validation** | | | |
| Cox | 0.885 (0.856-0.918) | 0.801 (0.690-0.912) | 0.887 (0.852-0.926) |
| GBM | 0.882 (0.851-0.918) | 0.915 (0.880-0.939) | 0.862 (0.789-0.921) |
| RSF | 0.883 (0.845-0.919) | 0.905 (0.862-0.937) | 0.812 (0.734-0.890) |
| Elastic net | 0.863 (0.829-0.899) | 0.770 (0.645-0.904) | 0.882 (0.840-0.926) |

1. Predicted by baseline.
2. Predicted by 1-year follow-up data.
3. Predicted by 3-years follow-up data.
Abbreviations: AUC, area under the curve; GBM, gradient boosting machine; RSF, random survival forest


A
TP
FP
model
Cox
Elastic net
GBM
RSF
1 year
2 years
4 years
B
1 year
2 years
4 years

C


TP
FP
model
Cox
Elastic net
GBM
RSF
1 year
2 years
4 years

D

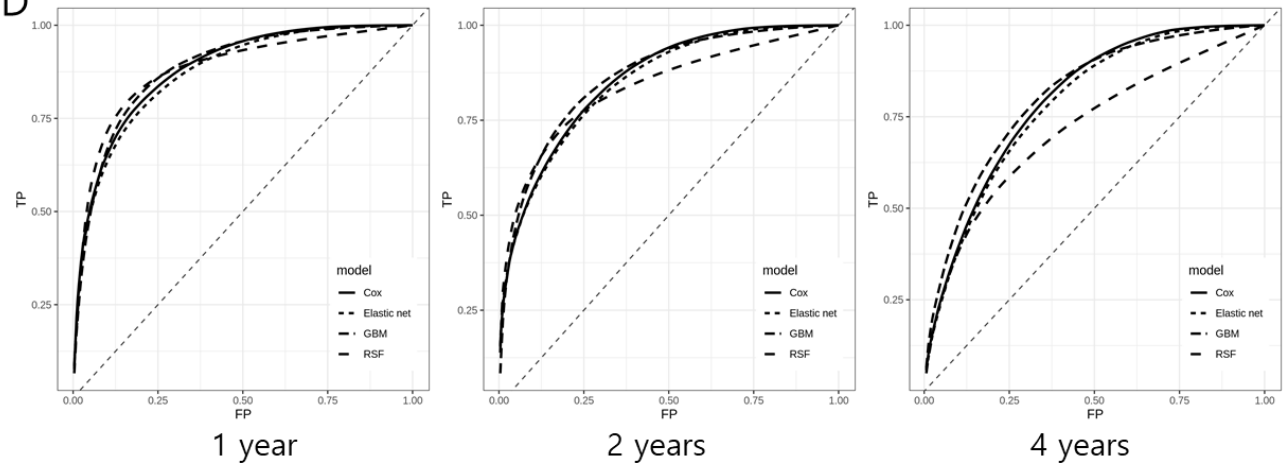
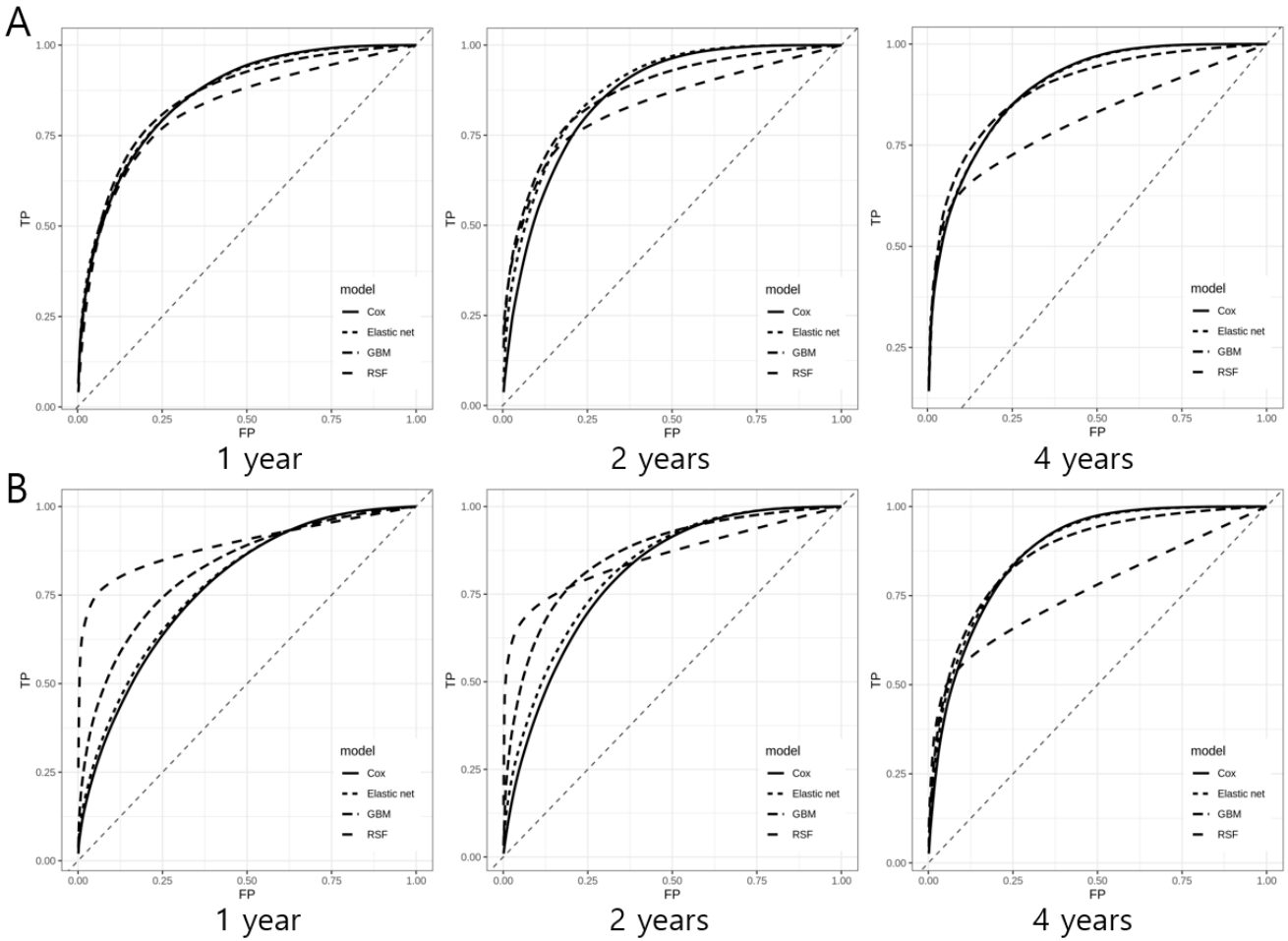
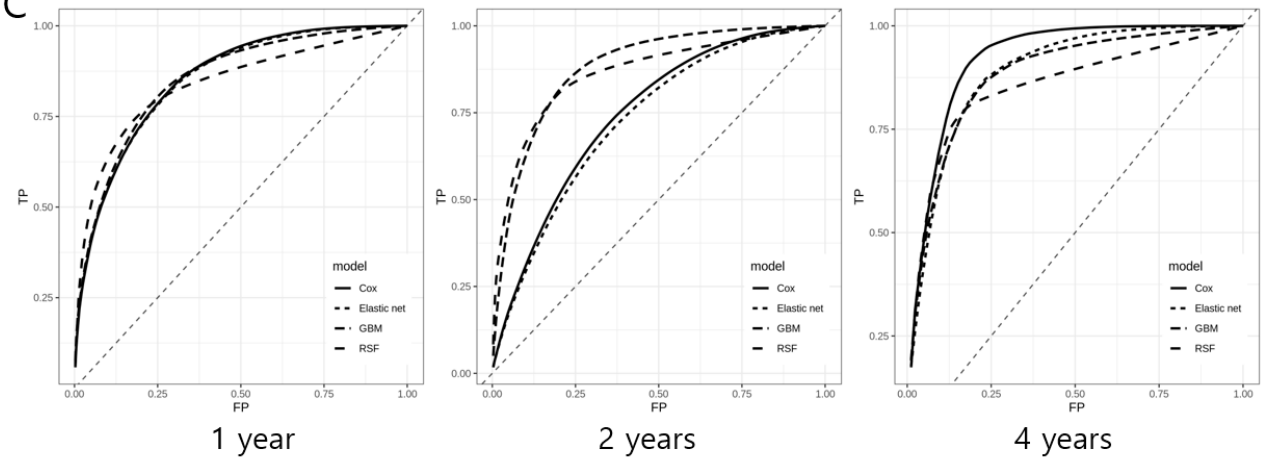

TP
FP
model
Cox
Elastic net
GBM
RSF
1 year
2 years
4 years

Figure 4. Time-dependent ROC curves for progression of chronic kidney disease static models at t=1, 2, 4 years in the (A) split-sample validation set (B) temporal follow-up validation set (C) temporal registration validation set (D) geographical validation set. The ROC curves of clinical parameters model for COX, GBM, RSF. Abbreviations: TP, true positive; FP, false positive; AUC, area under the curve; ROC, receiver operating characteristic; GBM, gradient boosting machine; RSF, random survival forest; eGFR, estimated glomerular filtration rate; UACR, urine albumin-to-creatinine ratio; DM, diabetes mellitus; BMI, body mass index; SBP, systolic blood pressure.


A
TP
FP
model
Cox
Elastic net
GBM
RSF
1 year
2 years
4 years
B
1 year
2 years
4 years

C


TP
FP
model
Cox
Elastic net
GBM
RSF
1 year
2 years
4 years

D

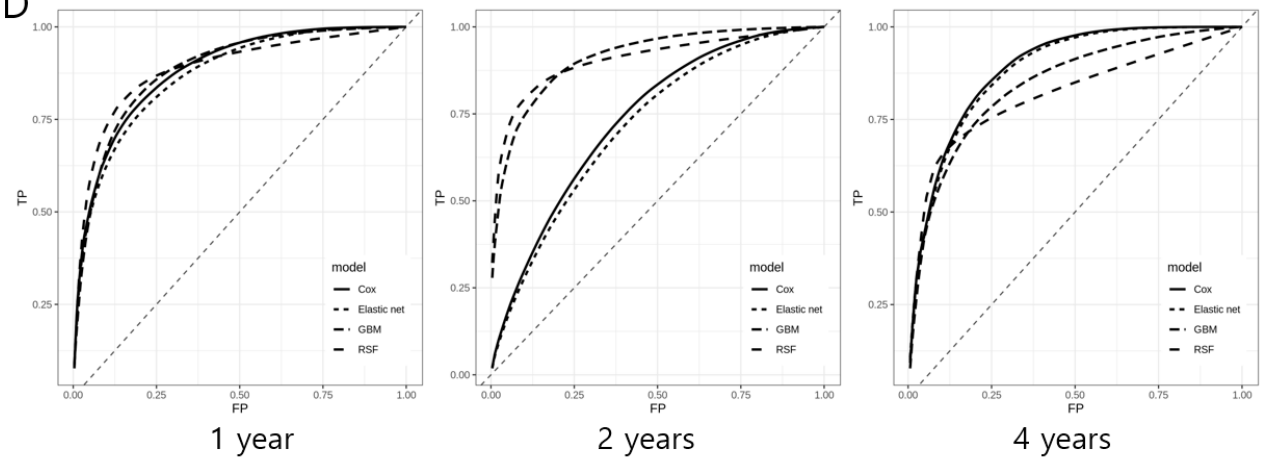

TP
FP
model
Cox
Elastic net
GBM
RSF
1 year
2 years
4 years

Figure 5. Time-dependent ROC curves for progression of chronic kidney disease dynamic models at t=1, 2, 4 years in the (A) split-sample validation set (B) temporal follow-up validation set (C) temporal registration validation set (D) geographical validation set. The ROC curves of clinical parameters model for COX, GBM, RSF. Abbreviations: TP, true positive; FP, false positive; AUC, area under the curve; ROC, receiver operating characteristic; GBM, gradient boosting machine; RSF, random survival forest; eGFR, estimated glomerular filtration rate; UACR, urine albumin-to-creatinine ratio; DM, diabetes mellitus; BMI, body mass index; SBP, systolic blood pressure.

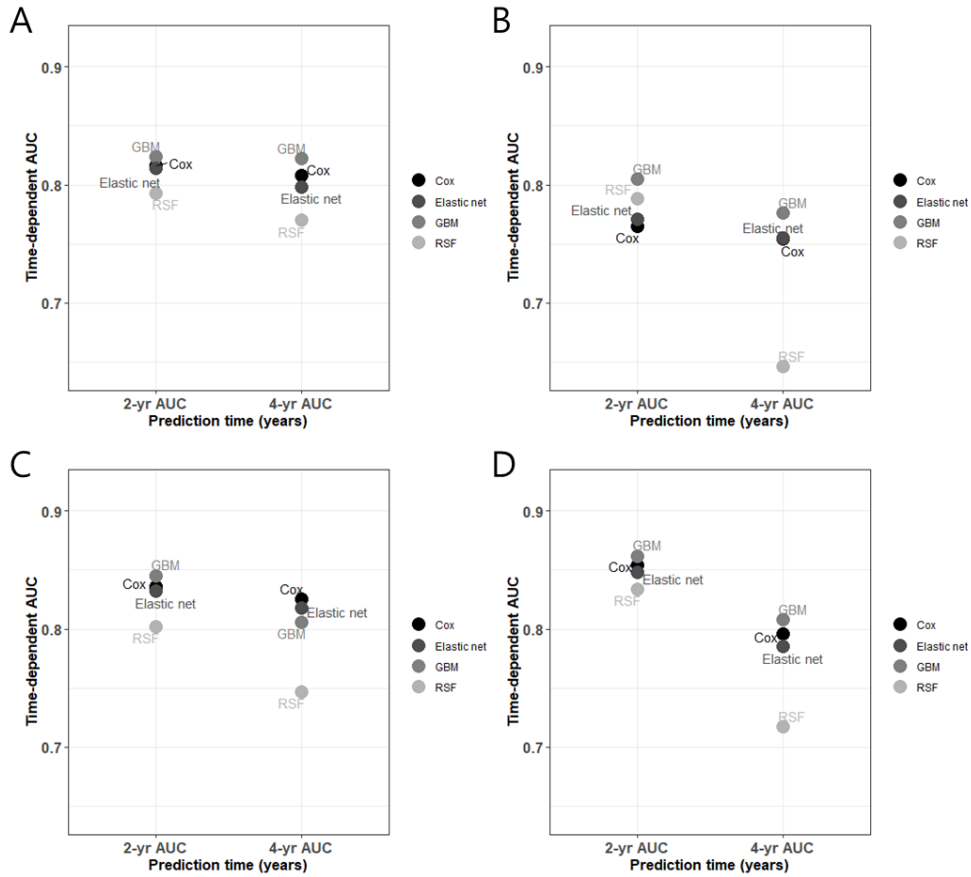



Figure 6. Dot plots summarizing AUC results of static model at t=2, 4 years in the (A) split-sample training set (B) temporal follow-up training set (C) temporal registration training set (D) geography training set. Abbreviations: GBM, gradient boosting machine; RSF, random survival forest.

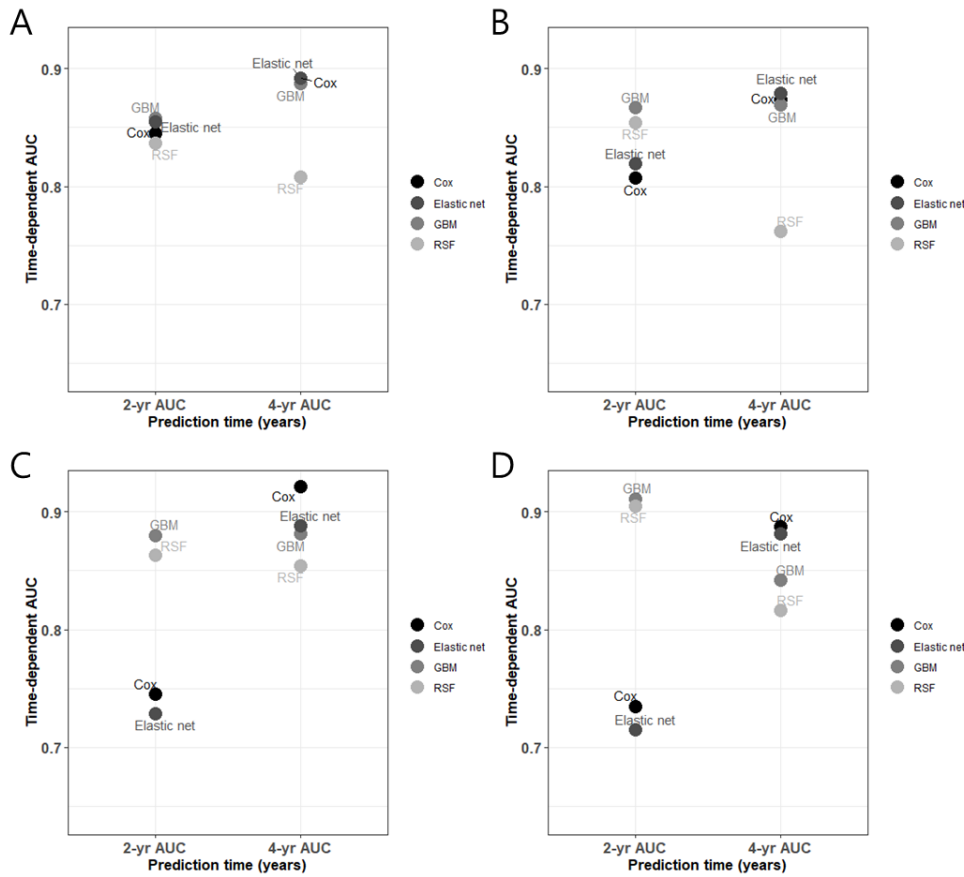



Figure 7. Dot plots summarizing AUC results of dynamic model at t=2, 4 years in the (A) split-sample training set (B) temporal follow-up training set (C) temporal registration training set (D) geography training set. Abbreviations: GBM, gradient boosting machine; RSF, random survival forest.

In order to compare the AUC values of the static and dynamic models, paired t-test was performed with the AUC values extracted from Table 5 and 6 (Table 7). The average AUC (95% CI) for 2 year for the static model was 0.818 (0.747-0.889), and the average AUC (95% CI) for 2 years for the dynamic model was 0.837 (0.705-0.969). The average AUC (95% CI) for 4 years for the static model was 0.772 (0.669-0.876), and the average AUC (95% CI) for 4 years for the dynamic model was 0.865 (0.774-0.956). The dynamic model outperformed the static model at all-year AUCs, which are significantly different.

Table 7. Comparison of time-dependent AUC between static and dynamic model in predicting progression of chronic kidney disease

| **Model** | **2-yr AUC (95% CI)[1]** | **4-yr AUC (95% CI)[2]** |
|---|---|---|
| Static[3] | 0.818 (0.747-0.889) | 0.772 (0.669-0.876) |
| Dynamic[4] | 0.837 (0.705-0.969) | 0.865 (0.774-0.956) |
| *p* | <0.01 | <0.01 |

1. Mean value of 2 years AUC for all prediction models according to the validation methods.
2. Mean value of 4 years AUC for all prediction models according to the validation methods.
3. Prediction model using demographic and laboratory data from the baseline to the hospital.
4. Prediction model using the latest available measurement values.

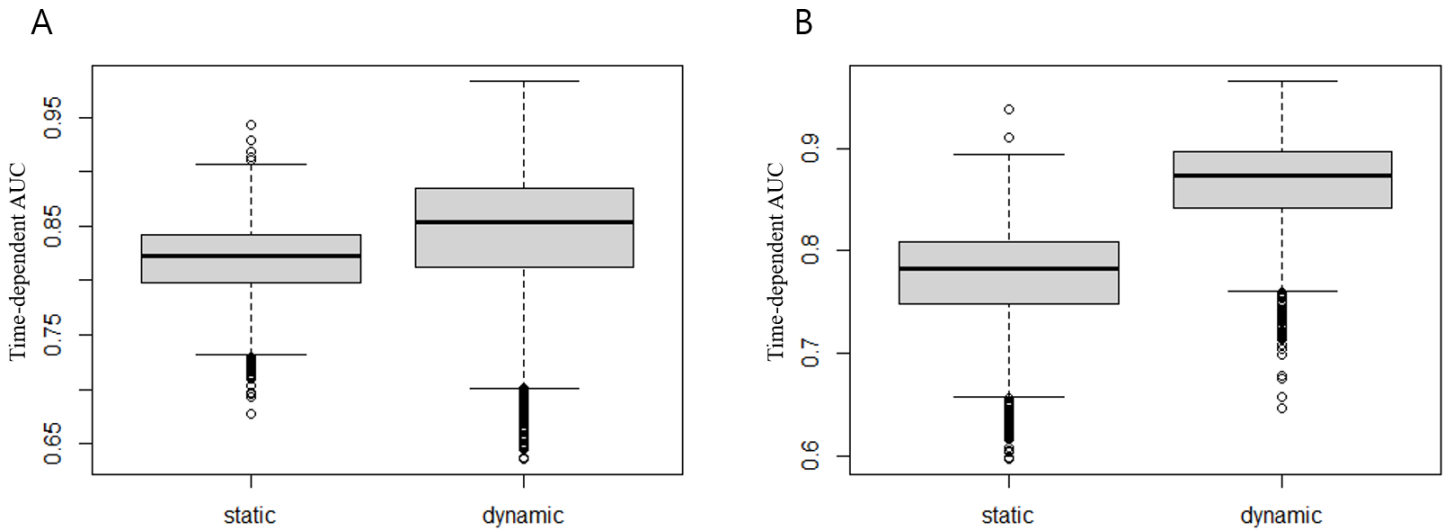



Figure 8. Box plots for comparing AUC results between static and dynamic models at (A) t=2 years (B) t=4 years.

## 4. Discussion

Our study compared the dynamic predictive model developed with the latest available measurement values with the static predictive model developed with the baseline values at the first visit to the hospital. In the case of the static model, the AUC values gradually decreased over years 1, 2, and 4, and in the case of the dynamic model, the AUC values tended to be stable or increased. Both the cox and machine learning methods generally showed high AUC values in the dynamic model. It also showed good performance in the discrimination capability for all methods of splitting the data.

### 4.1 Comparison with previous studies

In a recent study, a CKD progression prediction model was constructed using machine learning methods based on logistic regression. Although the predictive model was constructed using Elastic net, Lasso, Ridge, XGboost, SVM, random forest, ANN, and k-NN methods, the model was constructed with a small number of samples, and survival data analysis over time was not performed (11). Furthermore, although CKD progression prediction models are designed to be optimal for development samples by proceeding only with internal validation, their accuracy may be reduced due to overfitting when testing on similar individuals (34, 35).

### 4.2 Strengths and limitations

Our work has the following limitations. First, the majority of time dependent AUC values tended to be stable, gradually decreasing, or increasing, but the AUC values sharply dropped in random survival forest and elastic net. The discriminative power seems to be low due to a lack of samples. Second, the cox model provides clinically understandable results on covariate effects, while machine learning is close to “black box” models. Although our study analyzed prediction models with machine learning,

there is a limitation in providing meaningful interpretations due to the characteristics of black box models. Technology development for machine learning is necessary to ensure the reliability of these machine learning models. Third, when comparing the performance of machine learning and the cox model in a recent paper, the machine learning-based model performed as well as the cox regression (36). In our study, the result of the machine learning showed a performance similar to that of the cox model. Fourth, the temporal follow-up method may have excluded those who had an outcome until an eight-year follow-up, excluding those who had a rapid progression of CKD. The temporal follow-up method of our study seems to have caused a selective survival bias, also called Neiman's fallacy. Fifth, because of the proportional hazard assumption was violated, we estimated biased results using the cox proportional models.

Our study has several strengths. First, the performance of the model was evaluated by dividing the data by four methods, and the performance of the validation showed a good level of discrimination capability. Second, in the field of cancer, machine learning or deep learning papers using survival analysis are emerging (36, 37), but they are insufficient in the field of CKD. Our study constructed a machine learning study considering the survival time, making them useful analytical tools for predicting CKD progression. Third, time-dependent AUC values for evaluating the performance of the prediction model that varies over time were presented. Basically, the epidemiological concept focuses on a cumulative event to define the case, but our study evaluated the performance during various time with time-dependent AUCs.

## 5. Conclusion

In this study, we compared static and dynamic models finding that the dynamic model showed stable and high time-dependent AUC values in all predictive models. We confirmed that the eGFR and UACR are important impacts in RSF and GBM, and UACR and DM are important impacts in elastic net on the predictability of the

model. Considering the eGFR, which is difficult to predict, dynamic models that predict short-term events may be better when developing prediction models of CKD progression. Further studies are warranted to compare the static and dynamic predictive models through the cox model satisfying the proportional hazard assumption.

Supplementary Table 1. Using grid search parameter optimization for static model that use baseline data

| Algorithm | Grid search script | Validation | Optimal parameter |
|---|---|---|---|
| Random survival forest | tune(formula,<br>data=DB_train,<br>mtryStart=ncol(DB_train)/2,<br>nodesizeTry=c(1:9, seq(1,10,by=1)),<br>ntreeTry=5000) | Split-sample validation | nodesieze = 1<br>mtry = 4 |
| | | Temporal follow-up validation | nodesieze = 3<br>mtry = 5 |
| | | Temporal registration validation | nodesieze = 4<br>mtry = 4 |
| | | Geographical validation | nodesieze = 2<br>mtry = 4 |
| Gradient boosting machine | expand.grid(<br>shirnkage= c(.01,.05, .1, .3),<br>n.minobsinnode= c(5,10,15),<br>RMSE =0) | Split-sample validation | Shrinkage = 0.01<br>n.minobsinnode = 5 |
| | | Temporal follow-up validation | Shrinkage = 0.01<br>n.minobsinnode = 5 |
| | | Temporal registration validation | Shrinkage = 0.01<br>n.minobsinnode = 5 |
| | | Geographical validation | Shrinkage = 0.01<br>n.minobsinnode = 15 |
| Elastic net | cv.glmnet(x_train, y_train, family = 'cox', type.measure = 'C') | Split-sample validation | lambda.min = 0.037 |
| | | Temporal follow-up validation | lambda.min = 0.028 |
| | | Temporal registration validation | lambda.min = 0.035 |
| | | Geographical validation | lambda.min = 0.023 |

Supplementary Table 2. Using grid search parameter optimization for dynamic models that use data from 1-year follow up

| Algorithm | Grid search script | Validation | Optimal parameter |
|---|---|---|---|
| Random survival forest | tune(formula,<br>data=DB_train,<br>mtryStart=ncol(DB_train)/2,<br>nodesizeTry=c(1:9, seq(1,10,by=1)),<br>ntreeTry=5000) | Split-sample validation | nodesieze = 2<br>mtry = 5 |
| | | Temporal follow-up validation | nodesieze = 6<br>mtry = 5 |
| | | Temporal registration validation | nodesieze = 3<br>mtry = 7 |
| | | Geographical validation | nodesieze = 3<br>mtry = 4 |
| Gradient boosting machine | expand.grid(<br>shirnkage= c(.01,.05, .1, .3),<br>n.minobsinnode= c(5,10,15),<br>RMSE =0) | Split-sample validation | Shrinkage = 0.01<br>n.minobsinnode = 15 |
| | | Temporal follow-up validation | Shrinkage = 0.3<br>n.minobsinnode = 5 |
| | | Temporal registration validation | Shrinkage = 0.01<br>n.minobsinnode = 15 |
| | | Geographical validation | Shrinkage = 0.01<br>n.minobsinnode = 15 |
| Elastic net | cv.glmnet(x_train, y_train, family = 'cox', type.measure = 'C') | Split-sample validation | lambda.min = 0.029 |
| | | Temporal follow-up validation | lambda.min = 0.130 |
| | | Temporal registration validation | lambda.min = 0.021 |
| | | Geographical validation | lambda.min = 0.017 |

Supplementary Table 3. Using grid search parameter optimization for dynamic models that use data from 3-year follow up

| Algorithm | Grid search script | Validation | Optimal parameter |
|---|---|---|---|
| Random survival forest | tune(formula,<br>data=DB_train,<br>mtryStart=ncol(DB_train)/2,<br>nodesizeTry=c(1:9, seq(1,10,by=1)),<br>ntreeTry=5000) | Split-sample validation | nodesieze = 7<br>mtry = 5 |
| | | Temporal follow-up validation | nodesieze = 9<br>mtry = 4 |
| | | Temporal registration validation | nodesieze = 3<br>mtry = 4 |
| | | Geographical validation | nodesieze = 4<br>mtry = 5 |
| Gradient boosting machine | expand.grid(<br>shirnkage= c(.01,.05, .1, .3),<br>n.minobsinnode= c(5,10,15),<br>RMSE =0) | Split-sample validation | Shrinkage = 0.3<br>n.minobsinnode = 5 |
| | | Temporal follow-up validation | Shrinkage = 0.01<br>n.minobsinnode = 15 |
| | | Temporal registration validation | Shrinkage = 0.01<br>n.minobsinnode = 5 |
| | | Geographical validation | Shrinkage = 0.3<br>n.minobsinnode = 5 |
| Elastic net | cv.glmnet(x_train, y_train, family = 'cox', type.measure = 'C') | Split-sample validation | lambda.min = 0.005 |
| | | Temporal follow-up validation | lambda.min = 0.029 |
| | | Temporal registration validation | lambda.min = 0.059 |
| | | Geographical validation | lambda.min = 0.026 |

Supplementary Table 4. Test for the proportionality in the cox model that use baseline data

| | Split-sample[1] | | Temporal follow-up[2] | | Temporal registration[3] | | Geography[4] | |
|---|---|---|---|---|---|---|---|---|
| | **Statistic** | **p-value** | **Statistic** | **p-value** | **Statistic** | **p-value** | **Statistic** | **p-value** |
| eGFR ($mL/min/1.73m^2$) | 25.69 | <0.01 | 14.81 | <0.01 | 26.94 | <0.01 | 17.75 | <0.01 |
| UACR (g/g) | 0.24 | 0.63 | 0.58 | 0.44 | 0.01 | 0.90 | 0.11 | 0.74 |
| Age (years) | 0.52 | 0.47 | 0.08 | 0.77 | 0.01 | 0.91 | 0.26 | 0.61 |
| SBP (mmHg) | 7.01 | 0.01 | 8.00 | <0.01 | 3.97 | 0.05 | 4.45 | 0.03 |
| Female | 0.17 | 0.68 | 0.58 | 0.45 | 0.33 | 0.57 | 0.02 | 0.89 |
| DM | <0.01 | 0.97 | <0.01 | 0.96 | 0.45 | 0.50 | 0.36 | 0.55 |
| BMI ($kg/m^2$) (≥30) | 0.04 | 0.84 | 0.01 | 0.91 | 0.07 | 0.79 | 0.83 | 0.36 |

Note: The variables were compared in the following ways: female, male as reference; DM, patients without DM as reference; BMI, <30 as reference.
1. Training set which consists of 70% of all patients.
2. Training set which consists of all patients following up 8-years.
3. Training set which consists of recruited patients from 2011 until 2014.
4. Training set which consists of patients registered at hospitals in Seoul.
Tests based on the scaled Schoenfeld residuals.
Abbreviations: eGFR, estimated glomerular filtration rate; UACR, urine albumin-to-creatinine ratio; DM, diabetes mellitus; BMI, body mass index; SBP, systolic blood pressure; HR, hazard ratio

Supplementary Table 5. Test for the proportionality in the cox model in that use data from 1-year follow up

| | Split-sample[1] | | Temporal follow-up[2] | | Temporal registration[3] | | Geography[4] | |
|---|---|---|---|---|---|---|---|---|
| | **Statistic** | **p-value** | **Statistic** | **p-value** | **Statistic** | **p-value** | **Statistic** | **p-value** |
| eGFR ($mL/min/1.73m^2$) | 50.9 | <0.01 | 60.21 | <0.01 | 38.69 | <0.01 | 36.04 | <0.01 |
| UACR (g/g) | 17.78 | <0.01 | 51.13 | <0.01 | 0.74 | 0.39 | 1.92 | 0.17 |
| Age (years) | 0.04 | 0.83 | 0.48 | 0.49 | 0.18 | 0.67 | 0.01 | 0.92 |
| SBP (mmHg) | 1.24 | 0.27 | 3.49 | 0.06 | 1.04 | 0.31 | 1.72 | 0.19 |
| Female | 0.14 | 0.71 | 0.50 | 0.48 | 0.21 | 0.65 | 0.24 | 0.63 |
| DM | 0.53 | 0.47 | 1.78 | 0.18 | 0.02 | 0.90 | 0.51 | 0.47 |
| BMI ($kg/m^2$) (≥30) | 5.3 | 0.02 | 8.19 | <0.01 | 1.48 | 0.22 | 3.16 | 0.08 |

Note: The variables were compared in the following ways: female, male as reference; DM, patients without DM as reference; BMI, <30 as reference.
1. Training set which consists of 70% of all patients.
2. Training set which consists of all patients following up 8-years.
3. Training set which consists of recruited patients from 2011 until 2014.
4. Training set which consists of patients registered at hospitals in Seoul.
Tests based on the scaled Schoenfeld residuals.
Abbreviations: eGFR, estimated glomerular filtration rate; UACR, urine albumin-to-creatinine ratio; DM, diabetes mellitus; BMI, body mass index; SBP, systolic blood pressure; HR, hazard ratio

Supplementary Table 6. Test for the proportionality in the cox model that use data from 3-year follow up

| | **Split-sample[1]** | | **Temporal follow-up[2]** | | **Temporal registration[3]** | | **Geography[4]** | |
|---|---|---|---|---|---|---|---|---|
| | **Statistic** | **p-value** | **Statistic** | **p-value** | **Statistic** | **p-value** | **Statistic** | **p-value** |
| eGFR ($mL/min/1.73m^2$) | 19.62 | <0.01 | 20.21 | <0.01 | 27.24 | <0.01 | 26.67 | <0.01 |
| UACR (g/g) | 3.2 | 0.07 | 0.07 | 0.79 | 1.94 | 0.16 | 5.78 | 0.02 |
| Age (years) | 0.07 | 0.79 | 4.12 | 0.04 | 0.05 | 0.82 | 0.98 | 0.32 |
| SBP (mmHg) | 3.91 | 0.05 | 0.73 | 0.39 | 1.82 | 0.18 | 2.29 | 0.13 |
| Female | 1.36 | 0.24 | 0.88 | 0.35 | 1.06 | 0.30 | 1.56 | 0.21 |
| DM | 0.75 | 0.39 | 0.01 | 0.92 | 0.03 | 0.86 | 0.82 | 0.36 |
| BMI ($kg/m^2$) (≥30) | 1.27 | 0.26 | 1.58 | 0.21 | 1.07 | 0.30 | 1.13 | 0.29 |

Note: The variables were compared in the following ways: female, male as reference; DM, patients without DM as reference; BMI, <30 as reference.
1. Training set which consists of 70% of all patients.
2. Training set which consists of all patients following up 8-years.
3. Training set which consists of recruited patients from 2011 until 2014.
4. Training set which consists of patients registered at hospitals in Seoul.
Tests based on the scaled Schoenfeld residuals.
Abbreviations: eGFR, estimated glomerular filtration rate; UACR, urine albumin-to-creatinine ratio; DM, diabetes mellitus; BMI, body mass index; SBP, systolic blood pressure; HR, hazard ratio

Supplementary Table 7. Baseline demographic and clinical characteristics of KNOW-CKD subjects with no genomic DNA available before and after multiple imputation

| | **Before imputation** | | | | **After imputation** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Total (n=2,238)** | **Renal[1] group (n=720)** | **Non-renal group (n=1,518)** | ***p*** | **Total (n=2,238)** | **Renal[1] group (n=720)** | **Non-renal group (n=1,518)** | ***p*** |
| | **Mean (SD)** | **Mean (SD)** | **Mean (SD)** | | **Mean (SD)** | **Mean (SD)** | **Mean (SD)** | |
| Age (years) | 53.67 (12.24) | 54.41 (11.70) | 53.32 (12.48) | 0.05 | 53.67 (12.24) | 54.41 (11.70) | 53.32 (12.48) | 0.05 |
| UACR (g/g) | 0.90 (1.42) | 1.61 (1.86) | 0.57 (1.00) | <0.01 | 0.94 (1.44) | 1.69 (1.87) | 0.58 (0.99) | <0.01 |
| eGFR (mL/min/1.73m$^2$) | 53.07 (30.72) | 32.87 (21.30) | 62.66 (29.85) | <0.01 | 53.07 (30.72) | 32.87 (21.30) | 62.66 (29.85) | <0.01 |
| SBP (mmHg) | 127.83 (16.20) | 132.01 (17.51) | 125.87 (15.17) | <0.01 | 127.88 (16.18) | 132.08 (17.44) | 125.88 (15.16) | <0.01 |
| | **N (%)** | **N (%)** | **N (%)** | | **N (%)** | **N (%)** | **N (%)** | |
| Female | 871 (38.9) | 294 (40.8) | 577 (38.0) | 0.22 | 871 (38.9) | 294 (40.8) | 577 (38.0) | 0.22 |
| DM | 518 (23.1) | 277 (38.5) | 241 (15.9) | <0.01 | 518 (23.1) | 277 (38.5) | 241 (15.9) | <0.01 |
| BMI (kg/m$^2$) (≥30) | 146 (6.6) | 50 (7.0) | 96 (6.4) | 0.62 | 147 (6.6) | 51 (7.1) | 96 (6.3) | 0.56 |

Note: Values for categorical variables are given as number (percentage); values for continuous variable, as mean ± standard deviation or median [interquartile range]. eGFR was calculated using the CKD-EPI creatinine equation. The variables were compared in the following ways: female, male as reference; DM, patients without DM as reference; BMI, <30 as reference.

1. Composite of the first occurrence of a 50% decline in eGFR from the baseline value or the onset of ESRD during follow-up. ESRD was defined as the initiation of dialysis or kidney transplantation

Abbreviations: eGFR, estimated glomerular filtration rate; UACR, urine albumin-to-creatinine ratio; DM, diabetes mellitus; BMI, body mass index; SBP, systolic blood pressure.

## 6. Reference


1. Charles C, Ferris AH. Chronic Kidney Disease. Prim Care. 2020;47(4):585-95.
2. Lv JC, Zhang LX. Prevalence and Disease Burden of Chronic Kidney Disease. Adv Exp Med Biol. 2019;1165:3-15.
3. Ji E, Kim YS. Prevalence of chronic kidney disease defined by using CKD-EPI equation and albumin-to-creatinine ratio in the Korean adult population. Korean J Intern Med. 2016;31(6):1120-30.
4. Healthcare Bigdata Hub Frequent disease statistics of patients with chronic kidney disease from 2016 to 2020 [Available from: http://opendata.hira.or.kr/op/opc/olapMfrnIntrsIlnsInfo.do.
5. Hong YA, Ban TH, Kang CY, Hwang SD, Choi SR, Lee H, et al. Trends in epidemiologic characteristics of end-stage renal disease from 2019 Korean Renal Data System (KORDS). Kidney Res Clin Pract. 2021;40(1):52-61.
6. Wang V, Vilme H, Maciejewski ML, Boulware LE. The Economic Burden of Chronic Kidney Disease and End-Stage Renal Disease. Semin Nephrol. 2016;36(4):319-30.
7. Hippisley-Cox J, Coupland C. Predicting the risk of chronic Kidney Disease in men and women in England and Wales: prospective derivation and external validation of the QKidney Scores. BMC Fam Pract. 2010;11:49.
8. Kshirsagar AV, Bang H, Bomback AS, Vupputuri S, Shoham DA, Kern LM, et al. A simple algorithm to predict incident kidney disease. Arch Intern Med. 2008;168(22):2466-73.
9. Bang H, Vupputuri S, Shoham DA, Klemmer PJ, Falk RJ, Mazumdar M, et al. SCreening for Occult REnal Disease (SCORED): a simple prediction model for chronic kidney disease. Arch Intern Med. 2007;167(4):374-81.
10. Hemmelgarn BR, Culleton BF, Ghali WA. Derivation and validation of a clinical index for prediction of rapid progression of kidney dysfunction. QJM. 2007;100(2):87-92.
11. Xiao J, Ding R, Xu X, Guan H, Feng X, Sun T, et al. Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. J Transl Med. 2019;17(1):119.
12. Zhao J, Zhang Y, Qiu J, Zhang X, Wei F, Feng J, et al. An early prediction model for chronic kidney disease. Sci Rep. 2022;12(1):2765.
13. Chien KL, Lin HJ, Lee BC, Hsu HC, Lee YT, Chen MF. A prediction model for the risk of incident chronic kidney disease. Am J Med. 2010;123(9):836-46 e2.
14. Echouffo-Tcheugui JB, Kengne AP. Risk models to predict chronic kidney disease and its progression: a systematic review. PLoS Med. 2012;9(11):e1001344.
15. Collins GS, Omar O, Shanyinde M, Yu LM. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. J Clin Epidemiol. 2013;66(3):268-77.
16. Kamarudin AN, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. BMC

Med Res Methodol. 2017;17(1):53.
17. Li L, Astor BC, Lewis J, Hu B, Appel LJ, Lipkowitz MS, et al. Longitudinal progression trajectory of GFR among patients with CKD. Am J Kidney Dis. 2012;59(4):504-12.
18. Xie Y, Bowe B, Xian H, Balasubramanian S, Al-Aly Z. Estimated GFR Trajectories of People Entering CKD Stage 4 and Subsequent Kidney Disease Outcomes and Mortality. Am J Kidney Dis. 2016;68(2):219-28.
19. Oh KH, Park SK, Park HC, Chin HJ, Chae DW, Choi KH, et al. KNOW-CKD (KoreaN cohort study for Outcome in patients With Chronic Kidney Disease): design and methods. BMC Nephrol. 2014;15:80.
20. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009;150(9):604-12.
21. Colbert GB, Elrggal ME, Gaur L, Lerma EV. Update and review of adult polycystic kidney disease. Dis Mon. 2020;66(5):100887.
22. Boehmke B, Greenwell B. Hands-on machine learning with R: Chapman and Hall/CRC; 2019.
23. Qiu X, Gao J, Yang J, Hu J, Hu W, Kong L, et al. A Comparison Study of Machine Learning (Random Survival Forest) and Classic Statistic (Cox Proportional Hazards) for Predicting Progression in High-Grade Glioma after Proton and Carbon Ion Radiotherapy. Front Oncol. 2020;10:551420.
24. Heagerty PJ, Saha-Chaudhuri P, Saha-Chaudhuri MP. Package 'risksetROC'. 2012.
25. Heagerty PJ, Lumley T, Pepe MS. Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. Biometrics. 2000;56(2):337-44.
26. Bansal A, Heagerty PJ. A comparison of landmark methods and time-dependent ROC methods to evaluate the time-varying performance of prognostic markers for survival outcomes. Diagn Progn Res. 2019;3:14.
27. Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, et al. Package 'caret'. The R Journal. 2020;223:7.
28. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart. 2012;98(9):691-8.
29. Ridgeway G, Southworth MH, RUnit S. Package 'gbm'. Viitattu. 2013;10(2013):40.
30. Ishwaran H, Kogalur UB, Kogalur MUB. Package 'randomForestSRC'. breast. 2022;6:1.
31. Friedman J, Hastie T, Tibshirani R, Narasimhan B, Tay K, Simon N, et al. Package 'glmnet'. Journal of Statistical Software 2010a. 2022;33(1).
32. Stekhoven DJ, Buhlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28(1):112-8.
33. Grambsch PM, Therneau TM. Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. Biometrika. 1994;81(3):515-26.
34. Tangri N, Inker LA, Hiebert B, Wong J, Naimark D, Kent D, et al. A Dynamic Predictive Model for Progression of CKD. Am J Kidney Dis. 2017;69(4):514-20.

35. Moons KG, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart. 2012;98(9):683-90.
36. Moncada-Torres A, van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. Sci Rep. 2021;11(1):6968.
37. Matsuo K, Purushotham S, Jiang B, Mandelbaum RS, Takiuchi T, Liu Y, et al. Survival outcome prediction in cervical cancer: Cox models vs deep-learning model. Am J Obstet Gynecol. 2019;220(4):381 e1- e14.

# 초록

**연구 배경:** 만성 신장 질환은 전 세계적으로 주요 공중 보건 문제로 인식되고 있다. 전 세계 만성 신장 질환 유병률은 13.4%(11.7~15.1%)로 추정되며, 말기 신장 질환 환자 490만 2000명에서 708만 3000명이 신장이식 치료가 필요한 것으로 추산된다. 한국 만성 신장 질환 환자 수는 2016년부터 2020년까지 매년 8%씩 증가했다. 또한 말기 신장 질환 환자의 발생률과 유병률이 증가하고 있다. 종단 관찰 연구에서 만성 신장 질환 환자는 시간에 따른 지속적인 사구체 여과율 감소보다는 비선형적이거나 장기적인 비진행 상태를 보였다. 따라서 우리는 처음 방문했을 때의 데이터를 사용하는 정적 모델보다는 사용 가능한 최근에 측정된 데이터를 사용하는 동적 모델이 필요하다. 이 연구의 목적은 기존의 통계 모델과 기계 학습 방법을 사용하여 우리의 데이터를 동적 및 정적 예측 모델과 비교하는 것이다.

**연구 방법:** 데이터는 Korean Cohort Study for Outcomes in Patients With Chronic Kidney Disease (KNOW-CKD)에서 수집되었다. 상염색체 우성 다낭성 신장병(ADPKD) 환자를 제외한 만성 신장 질환 환자 1,585명의 데이터를 분석했다. 결측값은 multiple imputation 방법을 통해 대체되었다. 주요 결과지표는 기준 값에서 eGFR이 50% 감소하거나 추적 관찰 중에 이식이나 투석을 하게 되는 경우이다. 우리는 기준 값을 사용하는 정적 모델과 사용 가능한 최신 측정값을 사용하는 동적 모델을 만들었다. 예측 모델에는 콕스 비례 위험, random survival forest, gradient boosting machine, elastic net들이 사용되었다. 우리는 데이터 세트를 split-sample, temporal follow-up, temporal registration 및 geographical validation라는 네 가지 방법에 의해 훈련용 데이터와 검

증 데이터로 나누었다. 모델의 성능은 시간 의존적 곡선 아래 면적(AUC)을 사용하여 평가되었다.

**연구 결과:** 1,585명의 피험자 중 평균 연령은 54.9 ± 12.1세였으며 37.2%가 여성이었다. 평균 소변 알부민 크레아티닌 비율은 1.1 ± 1.5 g/g, 평균 기준선 추정 사구체 여과율은 49.3 ± 28.7 mL/min/1.73 m^2였다. 평균 수축기 혈압은 127.7 ± 16.6 mmHg이었다. 여성은 589명(37.2%)이었고 당뇨를 받은 피험자는 27.2%였다. 기준치에서 3년 추적 관찰까지 추정 사구체 여과율과 당뇨의 위험비가 감소하고 C statistics가 증가했습니다. 모델의 성능은 콕스 비례 위험과 기계 학습 방법에서 유사한 경향을 보였다. 또한 내부 검증에서 우수한 수준의 식별 능력을 보여주었다.

**결론:** 본 연구에서는 정적 모델과 동적 모델을 비교하여 동적 모델이 모든 예측 모델에서 안정적이고 높은 시간 의존성 AUC 값을 나타냈다는 것을 발견했다. 예측이 어려운 추정 사구체 여과율을 고려하면 만성 신장 질환 악화 예측 모델을 개발할 때 단기 이벤트를 예측하는 동적 모델이 더 나을 수 있다.

---

주요어: 만성신장질환, 예측 모델, 내부 검증, 머신 러닝