DEVELOPMENT OF PREDICTION MODEL FOR CHRONIC KIDNEY DISEASE IN TYPE 2 DIABETICS IN THAILAND

Miss Wilailuck Tuntayothin

A Dissertation Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in Social and Administrative Pharmacy

Department of Social and Administrative Pharmacy

FACULTY OF PHARMACEUTICAL SCIENCES

Chulalongkorn University

Academic Year 2019

การพัฒนาตัวแบบการทำนายโรคไตเรื้อรังในผู้ป่วยเบาหวานชนิดที่ 2 ในประเทศไทย

น.ส.วิไลลักษณ์ ตันตะโยธิน

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาเภสัชศาสตร์สังคมและบริหาร ภาควิชาเภสัชศาสตร์สังคมและบริหาร
คณะเภสัชศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2562

Thesis Title	DEVELOPMENT OF PREDICTION MODEL FOR CHRONIC KIDNEY DISEASE IN TYPE 2 DIABETICS IN THAILAND
By	Miss Wilailuck Tuntayothin
Field of Study	Social and Administrative Pharmacy
Thesis Advisor	Assistant Professor RUNGPETCH SAKULBUMRUNGSIL, Ph.D.
Thesis Co Advisor	Professor STEPHEN JOHN KERR, Ph.D.

Accepted by the FACULTY OF PHARMACEUTICAL SCIENCES, Chulalongkorn University in Partial Fulfillment of the Requirement for the Doctor of Philosophy

..... Dean of the FACULTY OF PHARMACEUTICAL SCIENCES

(Assistant Professor RUNGPETCH SAKULBUMRUNGSIL, Ph.D.)

DISSERTATION COMMITTEE

..... Chairman

(Associate Professor Police Lieutenant PUREE ANANTACHOTI, Ph.D.)

..... Thesis Advisor

(Assistant Professor RUNGPETCH SAKULBUMRUNGSIL, Ph.D.)

..... Thesis Co-Advisor

(Professor STEPHEN JOHN KERR, Ph.D.)

..... Examiner

(Associate Professor Police Captain TANATTHA KITTISOPEE, Ph.D.)

..... Examiner

(Associate Professor SOMRATAI VADCHARAVIVAD, PharmD, BCP)

..... External Examiner

(Assistant Professor Mayuree Tangkiatkumjai, Ph.D.)

วิไลลักษณ์ ตันตะโยธิน : การพัฒนาตัวแบบการทำนายโรคไตเรื้อรังในผู้ป่วยเบาหวานชนิดที่ 2 ในประเทศไทย. (DEVELOPMENT OF PREDICTION MODEL FOR CHRONIC KIDNEY DISEASE IN TYPE 2 DIABETICS IN THAILAND) อ.ที่ปรึกษาหลัก : ผศ. ภญ. ดร.รุ่งเพ็ชร สกุลบำรุงศิลป์, อ.ที่ปรึกษาร่วม : ศ. ดร.สตีเฟ่น จอห์น เคอร์

ไตวายเรื้อรังเป็นภาวะแทรกซ้อนของโรคเบาหวาน ถ้าขาดการดูแลที่เหมาะสมจะนำไปสู่โรคไตวายระยะสุดท้ายได้ เครื่องมือที่สามารถทำนายการเกิดโรคไตในผู้ป่วยเบาหวานจะช่วยลดอุบัติการได้ อย่างไรก็ตามพบว่าเครื่องมือส่วนใหญ่ที่มีในปัจจุบันพัฒนาขึ้นเพื่อใช้ทำนายการเกิดโรคไตระยะสุดท้าย การศึกษานี้จึงมีวัตถุประสงค์เพื่อพัฒนาแบบจำลองการทำนายโรคไตวายเรื้อรังระยะ 3 ในผู้ป่วยเบาหวานชนิดที่ 2 สำหรับคนไทย การวิจัยนี้เป็นการศึกษาแบบ retrospective cohort study เก็บข้อมูลผู้ป่วยเบาหวานชนิดที่ 2 จากฐานข้อมูลอิเล็กทรอนิกส์ของโรงพยาบาลย้อนหลัง 10 ปี ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2551 ถึง วันที่ 31 ธันวาคม พ.ศ. 2561 โดยตัวแปรผลลัพธ์การรักษาที่ใช้ศึกษา คือ การเกิดโรคไตวายเรื้อรัง ซึ่งวัดจากอัตราการกรองของไตน้อยกว่า 60 มิลลิลิตรต่อนาทีต่อ 1.73 ตารางเมตร ข้อมูลผู้ป่วยทั้งหมด 2,178 ราย ที่ใช้ในการวิเคราะห์ ถูกแบ่งเป็น 2 กลุ่ม ได้แก่ กลุ่มสำหรับสร้างแบบทำนาย 1,525 คน (ร้อยละ 70) และกลุ่มทดสอบความเที่ยงตรงแบบทำนาย 653 คน (ร้อยละ 30) การศึกษานี้ใช้ Cox proportional hazard regression สำหรับการสร้างแบบทำนาย Harrell's C-statistic เพื่อทดสอบความสามารถการทำนาย และ Hosmer-Lemeshow X^2 test และ survival probability curves ในการทดสอบความแม่นยำของแบบทำนาย ผลการศึกษาพบว่า ค่ามัธยฐานของระยะการติดตามผู้ป่วย คือ 1.29 ปี โดยมีค่าพิสัย ระหว่างควอไทล์ 0.5-2.5 ปี และเป็นโรคไตเรื้อรัง 385 คน (ร้อยละ 17.68) การวิเคราะห์ทางสถิติสรุปว่า ปัจจัยการทำนายโรคไตเรื้อรัง ได้แก่ อายุ เพศหญิง ค่าอัตราส่วนอัลบูมินต่อครีเอทินิน อัตราการกรองไต และค่าน้ำตาลสะสม จาก 5 ปัจจัยนี้แบบทำนายโรคไตเรื้อรังระยะที่ 3 ได้ถูกพัฒนาเป็น 2 รูปแบบ ได้แก่ แบบทำนายที่ 1 หรือ แบบทำนายเต็มรูปแบบ และ แบบทำนายที่ 2 หรือ แบบทำนายไม่ซับซ้อน ทั้งสองแบบทำนายมีค่าความสามารถการทำนาย C-statistic 0.890 และ 0.812 ตามลำดับ และพบว่าแบบทำนายทั้งสองมีความแม่นยำในการทำนาย การวิจัยนี้สรุปว่า แบบจำลองการทำนายโรคไตเรื้อรังในผู้ป่วยเบาหวานชนิดที่ 2 มีความแม่นยำในการทำนาย ซึ่งจะเป็นประโยชน์ต่อแพทย์ในการตัดสินใจแนวทางการรักษาโรคเบาหวานเพื่อป้องกันภาวะแทรกซ้อนโรคไตจากเบาหวาน อีกทั้งช่วยให้ผู้ป่วยตระหนักรู้การดูแลสุขภาพเพื่อป้องกันภาวะไตวายระยะสุดท้าย

สาขาวิชา	เภสัชศาสตร์สังคมและบริหาร	ลายมือชื่อนิสิต .....
ปีการศึกษา	2562	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....
		ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

5776552733 : MAJOR SOCIAL AND ADMINISTRATIVE PHARMACY

KEYWORD: Chronic kidney disease, type 2 diabetics, prediction model, Thailand

Wilailuck Tuntayothin : DEVELOPMENT OF PREDICTION MODEL FOR CHRONIC KIDNEY DISEASE IN TYPE 2 DIABETICS IN THAILAND. Advisor: Asst. Prof. RUNGPETCH SAKULBUMRUNGSIL, Ph.D. Co-advisor: Prof. STEPHEN JOHN KERR, Ph.D.

Diabetes mellitus (DM) is a major contributing factor that leads to end stage renal disease (ESRD) with unsuitable clinical management. Chornic kidney disease (CKD) prediction model could prevent the progression of CKD to ESRD. However, current CKD prediction models in patients with type 2 DM were developed to predict ESRD. This study aimed to develop prediction model for CKD stage 3 in patients with type 2 DM in Thailand. This was a 10-year retrospective cohort study obtaining data of patients with type 2 DM from electronic database of Taksin hospital during 1 January 2008 to 31 December 2017. The outcome variable was the present of CKD stage 3 which was defined as estimated glomerular filtration rate (eGFR) <60 mL/min/1.73m^2. The total patients of 2,178 were randomly assigned into training dataset for developing model (N=1,525) and validation dataset for model validation (N=653). The study used Cox proportional hazard regression for model development. For model performance, while model discrimination was conducted using Harrell's C-statistic, model calibration was evaluated by Hosmer-Lemeshow chi-square test and survival probability curve. The results showed that the median follow-up time was 1.29 years (interquartile range, 0.5- 2.5 years) and 385 patients or 17.68% with CKD stage 3. Data analysis identified five CKD stage 3 predictors including age, female, urinary albumin to creatinine ratio, baseline eGFR, and Hemoglobin A1c. Based on these five CKD predictors, two CKD prediction models were developed, model 1 using laboratory testing data and model 2 using simplified or proxy data. Both models demonstrated good discrimination with C-statistic of 0.890 and 0.812, respectively, and accurate prediction. These two CKD prediction models are recommended for health providers to use as an input for decision making on clinical management which could prevent diabetic kidney disease and for raising patients' awareness on health prevention.

Field of Study:	Social and Administrative Pharmacy	Student's Signature .....
Academic Year:	2019	Advisor's Signature .....
		Co-advisor's Signature .....

ACKNOWLEDGEMENTS

.

It is always a pleasure to remind the fine people in Chulalongkorn University for their sincere guidance I received to uphold my attitude, as well as practical skill in the Social and Administrative Pharmacy program.

First of all, I would like to thank to the 90th Anniversary of Chulalongkorn University Scholarship for supporting this research.

Secondly, I would like to thank Assistant Professor Rungpetch Sakulbumrungsil, Ph.D., my advisor, for giving me invaluable guidance on my research and work. My great appreciation is conveyed to Professor Steven John. Kerr., Ph.D., my co-adviser, and Suwasin Udomkarnjananun, M.D., for guiding me the great knowledge of biostatistics and prediction model, and Assistant Professor Mayuree Tangkiatkumjai, Ph.D. and members of the dissertation committee for helping me complete this dissertation.

I would like to express my thankfulness to Chanchana Boonyakrai M.D., the nursing staffs of diabetes and metabolic center, and the staff of statistics and academic department at Taksin hospital for their assistance in data collection.

My deep thankfulness is expressed to Sirirat Tunpichart, Ph.D., my boss, for giving me many great opportunities to her invaluable guidance on my life and my research.

I would like to thank to my parent for their love, kindness, great spiritual support. Especially thank my dad for introducing me dhamma to guide me a way of life.

My many thanks are expressed to my colleagues, my students, my seniors, for kindness and great support.

Wilailuck Tuntayothin

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LISTS OF ABBREVIATIONS

ACC/AHA : **American College of Cardiology/American Heart Association**

ACCORD : **Action to Control Cardiovascular Risk in Diabetes**

ACE : **Angiotensin-converting enzyme**

ACR : **Albumin to Creatinine Ratio**

AIC : **Akaike Information Criterion**

ASCVD : **Arteriosclerotic Cardiovascular Disease**

AUSDiab : **Australian Diabetes, Obesity, and Lifestyle**

BMI : **Body Mass Index**

CKD : **Chronic Kidney Disease**

CKD-EPI : **Chronic Kidney Disease Epidemiology Collaboration**

CVD : **Cardiovascular Disease**

DM : **Diabetes mellitus**

DN : **Diabetic nephropathy**

DKD : **Diabetic kidney disease**

DR : **Diabetic retinopathy**

EGAT : **Electric Generating Authoring of Thailand**

eGFR : **estimated glomerular filtration rate**

ESRD : **End stage renal disease**

HbA1c : **Hemoglobin A1C**

HDL : **High density lipoprotein cholesterol**

HN : **Hospital number**

HR : **Hazard ratio**

ICD : **International Classification of Diseases**

IDF : **International Diabetes Federation**

JDCS : **Japan Diabetes Complications Study**

KDIGO : **Kidney Disease Improving Global Outcome**

KDOQI : **Kidney Disease Outcomes Quality Initiative**

LD : **Listwise deletion**

LDL : **Low density lipoprotein cholesterol**

MAR : **Missing data at random**

MCAR : **Missing completely at random**

MI : **Multiple imputation**

MDRD : **Modification of Diet in Renal Disease**

MICE : **Multiple Imputation by Chain Equation**

MNAR : **Missing not at random**

NSAIDs : **Nonsteroidal anti-inflammatory drugs**

OR : **Odd ratio**

PD : **Pairwise deletion**

PI : **Prognostic index**

RAAS : **Renin–angiotensin–aldosterone system**

ROC : **Receive operating characteristic**

RR : **Relative risk**

RRR : **Relative risk ratio**

RRT : **Renal replacement therapy**

SBP : **Systolic blood pressure**

Scr : **Serum creatinine**

SD : **Standard deviation**

SRR : **Summary relative risk**

TC : **Total cholesterol**

TG : **Triglyceride**

Thai SEEK : **Thai Screening and Early Evaluation of Kidney Disease**

TODAY : **Treatment Options for Type 2 Diabetes in Adolescents and Youth**

UKPDS : **United Kingdom Prospective Diabetes Study**

CHAPTER I

INTRODUCTION

BACKGROUND AND RATIONALE

Diabetes mellitus (DM) is a non-communication disease with an increasing rate throughout the world. In 2015, the prevalence of DM is 415 million people in worldwide. It can be predicted that 1 in 11 adults have diabetes (1, 2). International Diabetes Federation [IDF] (2017) had showed prevalence of adult patients with DM (age of 20-79 years) that had risen to 8.8% (425 million people), and it will be estimated to be 9.9% (629 million people) in 2045. Moreover, the prevalence diabetes by country found that China, India, and United State have the highest diabetes adult in the world high with 114.4, 72.9, 30.2 million, respectively (3). Similarly, a systematic review of Nanditha et al. (2016) about prevalence of DM in Asia and the Pacific, showed that almost 60 % of diabetes patients in worldwide were in Asia with variation in each country in ranged 3% to 47.3. Almost 50% of diabetes patients were from China, and India% (4).

Thailand has increasing diabetes similar to other countries. Prevalence of adult diabetes in Thailand was ranked in the top ten in Asia(5). The report issued by the Thai National Health Examination Survey in Thai adult population with age of 20 years old or more exhibited that the age-adjusted prevalence of DM increased from 7.7% to 9.9% within 10 years(6). In 2015,

prevalence of adult patient with DM (aged of 20-79) was 4.2 million. As result, Thailand was ranked as a country with high and medium prevalence(3). And prevalence was estimated to be 5.3 million in the next 20 years (7).

DM can cause many complications. Diabetic nephropathy (DN) is also the one of the major complications of DM. From the Thailand Diabetes Registry Project (2006), diabetes nephropathy or DN is the highest prevalence than all complications of diabetes (e.g. diabetic retinopathy, stroke, coronary artery disease) with 62.9% and 45.7% among both long-DM (>15 years) and short-DM (< 15 years) groups, respectively(8). Diabetic nephropathy, a progressive kidney damage, can contribute to chronic kidney disease (CKD) leading to end stage renal disease (ESRD) which requires renal replacement therapies (RRT).

CKD is one of increasing diabetic complication. Approximately 20 - 40 % of diabetes patients can probably develop CKD and later develop ESRD(9). In Thailand, there are a few researches studying prevalence of CKD in small groups of DM patients. A retrospective cohort study from Siriraj Hospital Mahidol University (2006) showed the prevalence of CKD stage 3 to 5, among 722 patients with type 2 DM exhibited in 235 patients (48.2%). But prevalence of only CKD stage 3 was 37.3% (10). A cross-sectional, multi-center study of Vejakama et al. (2015) among 6 primary health care units in Udonthani, exhibited prevalence of CKD stage 3 to 5 were 25.38% and 27.09% using MDRD formula and Cockcroft-Gault formula, respectively (11). A retrospective-cohort study of Kittipanyaworakun (2013) among 322 patients with type 2 DM from Saraburi hospital showed the prevalence of CKD stage 3 measured by Thai eGFR formula was

27.33%(12). However, a large retrospective cohort study of Vejakama (2015) among 15,032 diabetic patients collected from 12 Ubon Ratchathani hospitals showed that prevalence of CKD including stage G3a, G3b, and G4 were 46.6, 27.6, and 12.1, respectively(13). Regarding to the mentioned Thai studies, prevalence of CKD (stage 3 to 5) among patients with DM are in range of 12.10-48.20%. But the prevalence of CKD stage 3 in diabetic patients is in range of 25.38-46.60%.

Uncontrolled DM patients with CKD can progress fast to ESRD and mortality leading economic burden. There are high risk association between Diabetes with kidney failure and death with high hazard ratio (HR), 1.49 *($p< 0.001$)*, and HR 1.06 (*$p=0.027$*), respectively (13). Moreover, costs of care for patients with diabetic nephropathy were extraordinarily high when reaching ESRD. Renal replacement therapies (RRT) which includes kidney transplantation and kidney dialysis, are recommended management when ESRD are reached. These high costs of care lead economic burden for Thai government. In 2015, Thailand Renal Replacement Therapy Registry Report showed that 24,514 (38.57%) of ESRD patients were diabetes patients. Mean cost of one hemodialysis patient for health provider's perspective were around 1,865 baht per one session (US dollar 61.69)(14). Annual estimated cost of hemodialysis per person is 179,040 baht (US dollar 5,635.51). Moreover, the study performed by Chatterjee and colleagues (2008) showed that the median cost of complication, including among diabetes patients was significantly different ($479.93) comparing with non-complication ($115.12)(15).

For Thai healthcare system, most of diabetes patients are under treatment of endocrinologist or general physicians which are not specialist to make decisions for individual suitable clinical management in order to slow renal progression. Moreover, due to long waiting times in hospitals, some patients elect to refill their medications in community pharmacies for convenience. This leads CKD risk progression of lost-follow up DM patients cannot be monitored or assessed. As results, numbers of CKD patients (including CKD stage 3, 4, and 5) are still increasing. Therefore, individual risk CKD stage 3 detection should be required for DM patients to prevent previously mentioned adverse events.

Prediction model is a developed tool to estimate for probability presences of a disease prognostic. CKD prediction models have been developed. Results of systematic review of 26 articles during 1 January 1980 to 20 June 2012 showed that most of CKD prediction model were developed for general population (16). Even history of diabetes was one of the key predictors in these CKD predictions, some important diabetic biomarkers, i.e., hemoglobin A1c, might be not analyzed as predictors. There are some CKD prediction models that had been developed for diabetic patients, but outcome of prediction was ESRD(17-22). However, the study Low et al. (2017) had developed the CKD prediction model for diabetics patients (23). Furthermore, the study of Nelson et al. (2019) had developed a 5-year risk prediction model of an incident CKD from 15 multinational cohorts' studies among 781,6277 diabetes. Even these prediction model was developed from the large Asian population of diabetes, recruited Asian diabetes which were from Malaysia, Singapore, China, India and Philippines might have different socioeconomic

status and lifestyle from Thai population with type 2 DM (24). In Thailand, three CKD prediction models were developed. The studies of Thakkinstian (2011) and Saranburut (2017) developed CKD prediction models in Thai populations which have only 11.9% and 7.8% of DM patients, respectively, so their studies may be less specific for predicting CKD in diabetic patients. There is only one study of Kittipanyaworakun was developed CKD prediction model in 322 type DM patients(12). However, limitations of this study were obtained in terms of small sample size, lacking of validation method and eGFR calculation by using Thai eGFR formula which is not used in practice. Furthermore, most CKD prediction models require some laboratory parameters such as urinary albumin creatinine ratio that are only available in hospitals. Primary care settings such as primary clinics, community pharmacies to which DM patients also visit may not access these prediction models.

To establish prediction model of CKD stage 3 endpoint for diabetic patients which can be used for every level of health care setting including tertiary, secondary and primary health care settings. Therefore, this study was conducted to fulfill these gaps.

OBJECTIVES

1. To identify the risk predictors for chronic kidney disease stage 3 in type 2 diabetics.
2. To develop and validate the prediction model for chronic kidney disease stage 3 in type 2 diabetics.

RESEARCH QUESTIONS

1. What are the association between chronic kidney disease stage 3 in type 2 diabetics and the risk predictors?
2. Does the predictive model have good performances of validity for predicting chronic kidney disease progression in type 2 diabetics?

HYPOTHESIS

1. The risk predictors have associations with chronic kidney disease stage 3 among type 2 diabetes patients in Thailand.
2. The developed prediction model for chronic kidney disease stage 3 in type 2 diabetics has a goodness of fit.
3. The predictive model has good performances of validity for predicting chronic kidney disease stage 3 progression in type 2 diabetics.

CONCEPTUAL FRAMEWORK

This framework **(Figure 1)** explains that the risk factors listed on the left (diabetes patients' risk factors) are associated CKD stage 3 in patients with type 2 DM. With the significant associations, the prediction model can be developed.

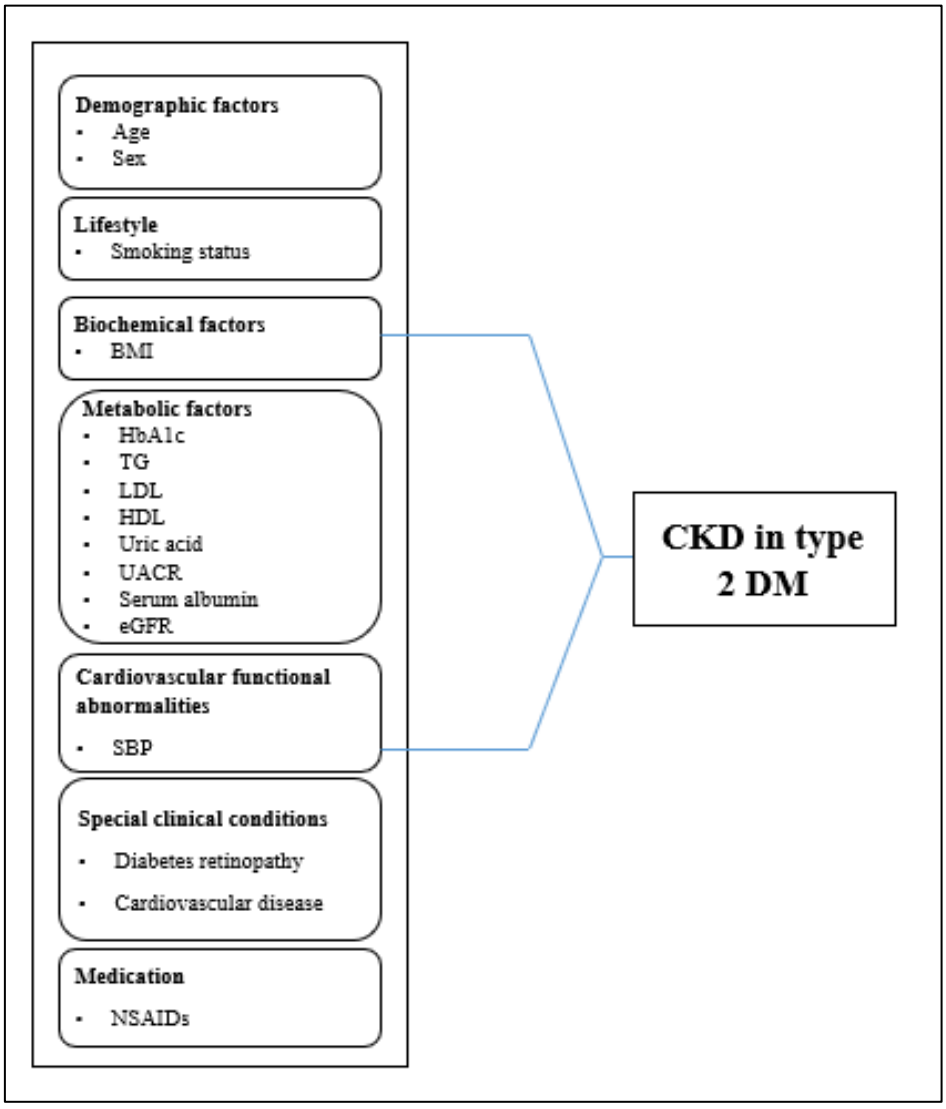


Figure 1: Conceptual framework

Definitions of constructs and variables

Diabetes patients' risk factors refer to factors that have association with CKD in patients with type 2 DM.

These CKD risk factors are categorized into 7 groups; demographic factors, lifestyle, biochemical factors, metabolic factors, cardiovascular functional abnormalities, special clinical conditions, and medication.

Demographic factors refer to socioeconomic characteristics of a population including increased age, sex.

Lifestyle refers to a way of life or style of living. Smoking is lifestyle that might affect renal function.

- **Smoking status** refers to the history of smoking. Smoking is divided into 2 categories; currently smoking (the patients who are smoking) and never-smoker (the patients who never smoke).

Biochemical factor includes body mass index (BMI). BMI is referred as body fatness. BMI can be calculated from the ratio of a person's weight in kilogram and height in squared meters.

Metabolic factors are factors that have associations to diabetes mellitus (DM) and cardiovascular disease (CVD). These metabolic factors include lipid profile, blood sugar, serum albumin, and albuminuria, estimated glomerular filtration rate (eGFR). Lipid

profile includes triglyceride (TG), high-density lipoprotein (HDL) cholesterol, and low-density lipoprotein (LDL) cholesterol. These metabolic factors can be obtained by 8 hour fasting blood test. For urinary albumin to creatinine ratio (UACR) was estimated from urine collecting.

Hemoglobin A1C (HbA1c) refers to a form of hemoglobin that is measured primarily to identify the three-month average plasma glucose concentration.

Triglyceride (TG) refers to the most common type of fat in the body. TG is from food and being produced by the body. Elevated TG is a risk factor for atherosclerosis.

Low-density lipoprotein cholesterol (LDL) refers to the bad cholesterol that contributes to fatty buildups in arteries. LDL cholesterol can be another indicator for cardiovascular disease (CVD).

High-density lipoprotein cholesterol (HDL) refers to the good cholesterol that carry LDL away from arteries and back to liver, where LDL will be broken down. Higher level of HDL means to prevent the risk of heart disease.

Uric acid is a natural waste product from digestion of food that contain purines. Hyperuricemia can lead to a gout disease that causes painful joint because of the accumulation of urate crystals.

Urinary albumin to creatinine ratio (UACR) referred to the amount of albumin in urine which indicates to a sign of kidney disease. UACR of 30-300 mg/g is defined as microalbuminuria and UACR >300 mg/g is defined as macroalbuminuria.

Serum albumin is a crystallizable albumin or mixture of albumins that constitutes the protein in blood serum and serves to maintain the osmotic pressure of the blood.

Estimated glomerular filtration rate (eGFR) indicates to level of kidney function and stage of kidney disease. eGFR was calculated from Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) formula **(Table 4).**

Cardiovascular functional abnormality refers to the risk factor related to heart disease.

- **Systolic blood pressure (SBP)** refers to amount of arteries' pressure during heart contraction. SBP is the one of component of blood pressure that is an indicator of the heart function. Blood pressure can be measured after resting at least 5 minutes by a nurse using Omron HEM 7120 Automatic Blood pressure.

The special clinical condition which refers to the retinopathy and cardiovascular disease

- **Diabetic retinopathy (DR)** refers to a diabetic complication which damage the retina of the eyes. Diabetic retinopathy can lead the vision impairment. Diabetic retinopathy includes macular edema, background retinopathy, and proliferative

diabetic retinopathy. The previous retinopathy is referred to the history of any type of diabetes retinopathy in the diabetes patients.

- **Cardiovascular disease (CVD)** is a disease that involves blood vessels and heart. Cardiovascular disease is stroke, angina pectoris, and myocardial infarction.

Medication use refers to medicine that type 2 diabetes patients were prescribed during time of study.

- **Nonsteroidal anti-inflammatory drugs (NSAIDs)** is a group of medicines with anti-inflammatory, antipyretic effects and analgesic. Examples of NSAIDs include ibuprofen, naproxen, aspirin, e.g. NSAIDs exposure is defined as total prescription days of taking NSAIDs within one year.

CKD in type 2 DM refers to incident CKD in type 2 diabetics. Chronic kidney disease stage 3 will be first diagnosed based on Kidney Disease Improving Global Outcome (KDIGO) 2012 Clinical Practice Guideline with eGFR less than 60 mL/min/1.73m^2.

CHAPTER II

LITERATURE REVIEW

This chapter describes about the literature review that related this study. The topics are included the risk factors for CKD in type 2 DM, the prediction model developing and validation, the developed CKD prediction models in diabetic patients, and multiple imputation method.

The chapter is reviewed the related main topics in order to design the appropriated methodology for this study.

I. Risk factors of chronic kidney disease in patients with type 2 DM

II. Development and validation of clinical prediction model

III. The developed CKD prediction model in diabetes patients

IV. The developed CKD prediction model in Thailand

V. Multiple Imputation by Chain Equation (MICE)

I. RISK FACTORS OF CHRONIC KIDNEY DISEASE IN PATIENTS WITH TYPE 2 DM

Even many literature reviews mentioned about CKD risk factors, including albuminuria and eGFR decline rate, or diabetic nephropathy among diabetic patients, these mentioned risk factors also had associated to develop to chronic kidney disease.

I.1 Age

In our knowledge, the eGFR gradually drops every year. The rate of eGFR decline is in range of 0.75 to 1 mL per min per 1.73 m^2 per year among adult with age of 40 years or more (25). And an eGFR decline will increase after the age 50–60 years. Cellular and organ senescence in renal function lead to a low level of eGFR in the age more than 65 years (26). The study of Cosmo et al. (2016) on 27,029 type 2 DM participants for 4 years revealed that age was the significant variable associated to the onset of eGFR decline with the relative risk (RR), 1.37 ($P< 0.001$)(27). Furthermore, the study of Moriya et al. (2017) examined the risk factor of eGFR decline in 2,033 DM patients. The results showed that advance age (+ 10 years) was associated significantly to the rapid eGFR decline with odd ratio (OR), 1.46 (95%CI 1.12–1.91). It was discussed that the older patients might have either vascular or tubular changes that contributed to a rapid GFR decline (28).

I.2 Gender

The effect of gender on diabetic renal disease is controversy. Some studies exhibited that male gender was a risk factor for diabetic kidney disease, some studies showed that women was a higher risk of developing the kidney disorder. However, possible mechanisms for the protective effect of female gender on chronic kidney disease, i.e., gender differences in kidney anatomy, effects of sex hormones have been discussed.

The study of Cosmo et al. (2016) showed that male sex had significant association with the rapid eGFR decline and albuminuria with the relative risk ratio(RRR), 0.767(95%CI 0.68–0.86), and 1.355 (95%CI 1.22–1.50) , respectively(27). Another study of Retnakaran et.al. (2006) showed the same result that male sex had risk association of microalbuminuria and macroalbuminuria with HR 1.52(95%CI 1.10–2.10), and 1.47 (95%CI1.06–2.02), respectively (29).

I.3 Cigarette smoking

Cigarette smoking is associated to microalbuminuria, macroalbuminuria and renal failure in both type 1 and type 2 DM.

In the U.K. Prospective Diabetes Study [UKPDS] (2006) on 5,102 patients with type 2 DM. The results showed that smoking had the risk association of microalbuminuria significantly with hazard ratio (HR) 1.60 (95%CI 1.26–2.05). In multivariate model, HR of smoking to microalbuminuria was 1.2 (95%CI 1.01–1.42) significantly (29). In the cross-sectional study of Cosmo et al. (2006) on 158 currently smoking and 158 non-smokers with type 2 DM. The adjusted OR of low eGFR was 2.20 (95%CI 1.14–4.26) significantly in currently smoking comparing with the non-smoker (30). Moreover, results from the meta-analysis included 19 observation studies with type 1 and 2 DM showed significant risks of diabetic nephropathy (DN) in both type 1 and 2 DM who were ex-smokers comparing non-smokers with summary relative risk (SRR) of 1.31 (95%CI 1.06-1.62), and 1.44 (95%CI 1.24-1.67), respectively. Similarly, ex-smokers had significantly associated to macroalbuminuria in both type 1 and 2 DM with SRR of

1.27 (95%CI 1.10-1.48), and 1.72 (95%CI 1.04-2.84), respectively. Even mechanism of smoking effect to renal dysfunction is still unknown, but possible explained pathway was carboxyhemoglobin, prothrombotic factors, and platelet activation were increased from smoking. These factors cause oxidative stress, glomerulosclerosis, and tubular atrophy (31).

I.4 Body mass index (BMI)

Obesity-related glomerular disease is related to high flow of renal plasma (32). The eGFR declines in type 2 DM will be vary by BMI. The obese diabetic patients would have eGFR decline higher than non-obese diabetic patients (normal weight or overweight) (33). The results of Cosmo's cohort study (2006) among 27,029 patients with type 2 DM, showed that BMI more than 30 kg/m^2 were associated to rapid eGFR and albuminuria with relative risk ratio (RRR) 1.33 (95%CI 1.09–1.63) comparing with BMI less than $27kg/m^2$. And BMI 27-30 kg/m^2 were associated with albuminuria with RRR 1.36 (95%CI 1.09–1.7) comparing with BMI less than 27 kg/m^2(30).

I.5 Hemoglobin A1c (HbA1c)

Hemoglobin (HbA1C) is widely used to determine the long-term glycemic control. Therefore, HbA1c has been widely adopted for the diagnoses of type 2 DM. Inadequate glycemic controls are associated with poor complication in DM including microvascular and macrovascular outcomes. Many studies show that glycemic controlling could slow renal function decline.

A study of Dodhia et al. (2016), 70 patients with type 2 DM were enrolled in the study for 6 months. The diabetic patients were divided into 2 groups; group 1 was patients with serum

creatinine (Scr) < 1.2 mg/dl, and group 2 was patients with Scr 1.2 mg/dl or more. The results were found that the mean HbA1c of the first group was 7.97% which was lower than the mean of the latter group by 9% ($p<0.05$). It concluded that uncontrolled glycemic could lead to progression of DN(34).

The Action to Control Cardiovascular Risk in Diabetes (ACCORD) (35) study (2008) conducted on random 10,251 diabetes patients that are classified into 2 groups: the group 1 was the intensive therapy (HbA1c < 6%) and group 2 was the standard therapy (HbA1c 7-7.9%) with 3.5 year follow up. The results showed that macroalbuminuria was significantly decreased in the intensive therapy group with HR, 0.69 (95%CI 0.55– 0.85)(36). The finding of Sheen et al.'s study (2013) on 577 DM patients showed that the higher glycated hemoglobin (HbA1c) at baseline 8.5±2.1% was significantly associated with rapid eGFR comparing the group with HbA1c at baseline 7.9±1.8%. The association between higher HbA1c baseline and the rapid eGFR decline was significant with the adjusted odd ratio (OR) 1.014(95%CI 1.00–1.03)(37). The results of Yokoyama et al.'s cohort study (2009) on 729 DM patients showed the multiple logistic repression analysis that eGFR $\geq$7% and 6.0 to 6.9% had significantly associated to rapid eGFR, with the odd ratio (OR) 2.93 (95%CI 1.76 - 4.87), and 1.42 (95%CI 0.89 -2.27) comparing baseline HbA1c more than 6%, respectively(38).

I.6 Albuminuria

In general, levels of albuminuria can predict loss of renal function. The more albumin in the urine, the larger the progressive renal function loss. Albuminuria is classified into 2 types;

microalbuminuria (UACR 30-300 mg/g) and macroalbuminuria (UACR > 300 mg/g)(39). Regarding the International Developing Education on Microalbuminuria for Awareness of renal and cardiovascular risk in Diabetes (DEMAND) (40) study (2006), the prevalence of microalbuminuria and macroalbuminuria are 39% and 10%, respectively (40). In Vigil et al.'s study (2015), the prospective cohort study on the stage II CKD patients, the results showed that high proteinuria was associated with the rapid kidney decline significantly with odd ratio (OR), 1.817(CI 95%, 1.21–2.72) (41). Moreover, the findings of prospective cohort study of Babazono et al. (2009) among 5,449 Japanese type 1 and 2 DM patients showed that within the normal range of UACR (<30 mg/g), UACR 10 mg/g or more in female, and 5 mg/g or more in male were significantly related to the rate of eGFR decline comparing with UACR less than 5 mg/g (42).

I.7 Uric acid

Uric acid is a chemical end product from breaking down of purine compounds. Purine compounds are from two pathways, including diets and body synthesis. Uric acid is mostly production of liver. However, uric acid can be produced in intestine and kidney. Hyperuricemia is defined as levels of uric acid higher than 6 mg/dl in female or 7 mg/dl in male. Chronic high levels of uric acid cause uric acid depositing in part of joints and soft tissues which can lead to inflammation of arthritis and tophi called gout. In Jalal's study (2011) explain about relation between uric acid and diabetic nephropathy that there was still consensus because of complicated relationship about uric acid and renal function. Because uric acid level was rised in patients with

diabetic nephropathy(43). Even complicated relationship was obtained, but there were many studies showing uric acid predicting for diabetic nephropathy developing.

The cohort study among type 2 diabetics from the database of the Italian Association of Clinical Diabetologists network (2016) was conducted. The study examined the correlation between serum uric acid and the onset of CKD. The results showed the significant risk association to CKD incident in the groups of patients with level of uric acid 4.4, 5.1 and 7.1 mg/dL comparing with the group of patient with low uric acid of 3.5mg/dL(27). Similarly, a prospective cohort study of Hovind et al (2009) investigated uric acid as a predictor of DN (defined as persistent either microalbuminuria or macroalbuminuria) among 263 type 1 diabetics. The result of multiple Cox-regression analysis showed that uric acid could predict microalbumin with a hazard ratio of 2.37 (95%CI 1.04-5.37) (44).

Study of Jalal et at. (2010), a prospective cohort study recruiting 324 type 1 diabetics, evaluated predicting of uric acid for albuminuria, including microalbuminuria and macroalbuminuria with six-year follow up. The results showed that baseline of uric acid level was predictive of albuminuria; adjusted odd ratio for the development of albuminuria was 1.18 (95%CI 1.2-2.7) for every 1 mg/dL increase in uric acid level (*P=0.005*)(45). An observational study of Spencer et al. (1986) among obese type 2 DM using the Treatment Options for Type 2 Diabetes in Adolescents and Youth (TODAY) Study to evaluate association between uric acid and diabetic kidney disease (urinary albumin excretion ≥30 mg per gram). The average follow-up times were 5.7 years. Multiple Cox proportional hazard regression model exhibited that

increasing uric acid was predictive for diabetic kidney disease; hazard ratio for the development of diabetic kidney disease was 1.24 (95%CI 1.03-1.48) (*P=0.022*) (46). The mechanism is to cause renal dysfunction by placing intraluminal crystals in the collecting duct of the nephron in a manner reminiscent of gouty arthropathy(47) .

I.8 Serum albumin

Serum albumin is an abundant protein in blood. Serum albumin has roles in maintaining the oncotic pressure between blood vessels and body tissues. Low serum albumin or hypoalbuminemia has associated to diabetic nephropathy (48). The mechanism of hypoalbuminemia leading to kidney disease progression may be explained that the level of serum albumin affects the degree of proteinuria. Another possibility is low serum albumin can reflect an inflammation, which leads to the kidney disease progression (49). The result of Leehey's observational study (2005) that recruited 343 diabetes patients showed the strong association of age-adjusted low initial serum albumin (<35 mg/dl) to eGFR decline with F ratio 14.5 ($p < 0.001$) (50). Similarly, a retrospective cohort study of Zhang et al. (2019) conducted in 188 diabetes patients to find out correlation between serum albumin and ESRD. Diabetes patients were divided into 4 groups; 1) normal group (serum albumin≥35 g/L), 2) mild group (serum albumin 30-35 g/L), 3) moderate group (serum albumin 25-30 g/L), and 4) severe group (serum albumin <25 g/L). The results showed that every severity group of albumin levels, including mild, moderate and severe group, comparing the normal group had significant association of end stage

renal disease progression with hazard ratio of 2.09 (95%CI 0.67-6.56), 6.20 (95%CI 1.95-19.76), and 7.37 (95%CI 1.24-43.83), respectively (51).

I.9 Hyperlipidemia

Diabetes patients were found that most of them had dyslipidemia, such as a rising in very low-density lipoprotein (V-LDL) cholesterol and low-density lipoprotein (LDL) cholesterol and a decreasing in high-density lipoprotein (HDL) cholesterol. Dyslipidemia is shown risk association to development and progression of DN in many studies.

According to the studies in the Kidney Disease Outcomes Quality Initiative (KDOQI) guideline 2012, there are some studies revealing the inclusive correlation the high levels of low density lipoprotein (LDL) cholesterol , total cholesterol (TC), triglycerides (TG), and low levels of high density lipoprotein (HDL) cholesterol and rate of kidney progression (39). However, the results of Cosmo's study (2016) showed that relative risk ratio (RRR) for TG with each 10 mg/dL increasing by 1.02, *(P<0.001)* and LDL-c by 10 mg/dl (0.97, *P=0.004*) were related significantly to the onset of eGFR reduction (27). Another prospective observational study of Retnakaran et al. (2006) analyzed risk factors for renal dysfunction (defined as level of creatinine clearance less than 60 mL/min/1.73m^2 or double of plasma creatinine) and albuminuria among 5,102 UKPDS participants. Results were revealed that the level of LDL was a significant risk factor for macroalbuminuria; hazard ratio (HR) for macroalbuminuria occurring was 1.17 (95%CI 1.02-1.33) for every 1 mmol/L increase in LDL levels (*P= 0.022*). However, HDL had risk association to double of plasma creatinine with HR of 2.78 (95%CI 1.01–7.68) (29). Moreover, a

retrospective multi center cohort (2016) studied about triglycerides and the levels of HDL for predicting diabetic kidney disease (DKD) in Italian 47,177 type 2 DM patients with baseline of LDL $\leq$130 mg/dL. All patients were followed up 4 years. DKD was defined as presenting of either low eGFR (less than 60 mL/min/1.73m^2) or the reduction of eGFR more than 30% and/or albuminuria. The investigators found that patients with TG $\geq$ 150 mg/dL had 1.26 times for low eGFR risk (95%CI 1.11–1.42), 1.29 times of an eGFR reduction <30% risk from baseline (95%CI 1.12–1.48), and 1.19 times of albuminuria risk (95%CI 1.09–1.31) comparing patient with TG <150 mg/dL. For HDL cholesterol, it was found that patients with low HDL cholesterol (defined as less than 40 mg/dL in male and less than 50 mg/dL in female) had 1.27 times for low eGFR risk (95%CI 1.12–1.44), 1.28 times of an eGFR reduction <30% risk from baseline (95%CI 1.11–1.47), and 1.24 times of albuminuria risk (95%CI 1.13–1.36) comparing patient without low HDL cholesterol (52).

I.10 Systolic blood pressure

High blood pressure can contribute to CKD progression. Blood vessels in the kidneys can be damaged by high blood pressure. This condition can lead to reduce the abilities of renal function in term of removing fluids and waste products from the bold. As the result, it can lead to kidney failure.

In Chiang's observational study (2014) among 2,144 CKD patients to examine the association between systolic blood pressure (SBP) and clinical outcomes in CKD patients. All

were followed for 2 years (2009-2010), or death. It was found that SBP more than 140 mmHg was significantly associated to the renal outcome and rapid decline of renal function in DKD patients with the adjusted HR, 2.60 (95%CI 1.29–5.26) comparing with SBP range 111-120 mmHg (53).

In meta-analysis of 84 non-randomized and randomized trial (2013) in CKD patients. It had shown that deceasing 10 mmHg of mean arterial pressure could improve the rate of eGFR declining of 0.18 ml/min/1.73m^2. In a multivariate analysis, the finding showed that 10-mmHg reduction in SBP decreased proteinuria significantly with the regression coefficient -0.14 (95%CI 0.22 to -0.06)(54).

In Another study of Retnakaran et al. (2006) on diabetes patients showed the same result that there were significantly association between SBP and microalbuminuria and macroalbuminuria with HR 1.15 (95%CI 1.11–1.20), and 1.15 (95%CI 1.07–1.24), respectively (29).

I.11 Diabetic retinopathy

Diabetic retinopathy (DR) is a microvascular complication of DM and leads blindness. There are many studies showing prediction for diabetic kidney disease of DR.

The fifth Korea National Health and Nutrition Examination Survey (KNHANES) of Lee et al. (2014) with an adjusted OR of 1.9 (95%CI 1.04-4.26) comparing patient without DR (55).

The Japan Diabetes Complications Study (JDCS) (2017), a prospective study of patients with type 2 DM, examined the risk factors for rapid eGFR decline (defined as eGFR declines of

≥3 ml/min/1.73m^2) in 1,470 DM patients with 8 years of follow-up. Patients were separated into 4 groups based on levels of eGFR: G1 (eGFR≥120), G2 (90 ≤ eGFR <120), G3 (60 ≤ eGFR <90), G4 (eGFR <60). The results of multiple logistic regression implied that DR was the significant risk factor for the rapid eGFR decline with HR 2.24 (95%CI 1.54–3.26)(28). In Zhang et al.'s study , a cross-sectional study among 250 patients with type 2 DM, evaluate the relationship between DR and DN. Prediction models for renal outcome were generated into 3 models based on covariates, including model 1 (adjusted by age, cigarette smoking, hypertension, gender, duration of diabetics), model 2 (model 1 + hemoglobin A1C, serum creatinine, and hematuria), and model 3 (model 2 + other renal pathological finding). In an analysis of multivariate Cox proportional hazard regression showed that diabetic retinopathy (DR) was an independent risk factor for predicting of ESRD progression in model 1,2 and 3 with adjusted HR of 1.93 (95%CI 1.08-3.45), 2.65(95%CI 1.27-5.53). The study concluded that DR may be predictor for renal progression in type 2 diabetics (56). Similarly, a Taiwan multicenter cohort study was conducted to identify risk factor of diabetic retinopathy to chronic kidney disease. Data of 4,050 diabetes patients from the Epidemiology and Risk Factors Surveillance of the CKD project (2008–2013) and the National Health Insurance Research Database (NHIRD) (2001–2013) were observed. The results showed that type 2 diabetics with diabetic retinopathy had risk association to develop for CKD stage 3a to 5 with an odd ratio of 1.66 (95%CI 1.36-2.02) comparing with diabetes patients with diabetic retinopathy(57). Even the exact association

between vasoconstriction of retinal arterioles and renal dysfunction was unclear, but retinopathy might be correlated with more advanced glomerular lesions (57).

I.12 Cardiovascular disease

The UKPDS study (2006) had examined the risk factors for the renal dysfunction. This study that recruited 5,102 type 2 DM patients with 15 –year follow up. The incidences of microalbuminuria, macroalbuminuria, and the reduced creatinine clearance were the renal outcomes. In the univariate analysis has shown that the previous cardiovascular disease (defined as history of myocardial infarction, angina pectoris, or transient ischemic attack) has the association with microalbuminuria, macroalbuminuria, and the reduced creatinine clearance with HR, 1.58 (95%CI 1.31–1.90), 1.64 (95%CI 1.18–2.28), and 1.71 (95%CI 1.51–1.93), respectively (29). Moreover, another prospective cohort study of Vigil et al. (2015) on the stage II CKD patients, the results showed that previous cardiovascular disease (defined as history of heart failure, stroke, or acute myocardial infarction) was associated with the rapid kidney decline significantly with odd ratio (OR), 1.90 (CI 95% 1.03-3.52)(41).

I.13 Non-steroidal anti-inflammatory drugs

Non-steroidal anti-inflammatory drugs (NSAIDs) are cyclooxygenase function inhibitor, and NSAIDs have the ability of prostaglandin production reducing. NSAIDs can change hemodynamics in the kidney function which leads to the acute renal failure. Taking NSAIDs is associated significantly to chronic kidney disease among patients with type 2 DM. In short-term

uses, NSAIDs also cause renal effects including sodium retention, the alteration of glomerular filtration rate, and elevation of blood pressure (58, 59).

The finding of Tsai et al.'s large retrospective cohort study (2015) among Taiwanese 48,715 type 2 diabetes exhibited that taking NSAIDs for a period of 1-89 days were associated risk of CKD with the adjusted HR, 1.28 (95%CI 1.20–1.35) comparing with not taking NSAIDs. And also, taking NSAIDs more than 90 days were associated a higher risk of chronic kidney disease with the adjusted HR 1.37 (95%CI 1.26–1.49) comparing with not taking NSAIDs. The investigators recommended to evaluate risk and benefit of NSAIDs dispensing among diabetics (60).

I.14 Duration of diabetics

The findings of many studies about the correlation between duration of diabetes and eGFR decline is still controversy. In Rossing et al.'s prospective observational study (2004), 227 diabetes participants were observed Caucasian patients with type 2 DM until diabetic nephropathy developed. Median of time follow-up was 6.5 (interquartile range [IQR] 3-17) years. The result exhibited that duration of diabetics was not associated to eGFR decline(61). Cosmo et al.'s cohort study (2016), the result also showed the same way. Five years of diabetes was not related to the rapid eGFR decline(27).

In contrast, the results of Zoppini's cohort study (2012) revealed that diabetic duration $\geq$15 years had the mean annual rapid eGFR -1.0±0.1. And duration of diabetes <15 years had the mean annual rapid eGFR -0.7±0.1. As the result, it exhibited that longer duration of diabetes had

associated with the rapid eGFR significantly ($p<0.05$) comparing with the duration of diabetes less than 15 years(62). Similarly, a retrospective cohort study using The Australian Diabetes, Obesity, and Lifestyle (AUSDiab) study (2004) evaluated risk factors for albuminuria (defined as either microalbuminuria or macroalbuminuria) among 11,247 adult diabetes. The findings showed that duration of diabetics was a predictor for microalbuminuria in newly diagnosed diabetics and known diabetics with odd ratio of 1.34 (95%CI 1.06-1.70) by increasing every 10 years of durations of diabetics. When analysis in only known diabetics, duration of diabetics also predicted for albuminuria with odd ratio of 1.38 (95%CI 1.05-1.81) by increasing every 10 years of durations of diabetics(63).

I.15 Other risk factors

Other reviewed CKD risk factors for type 2 diabetics include oxidative stress, subclinical inflammation, genetic background, ethnicity, and glomerular hyperfiltration.

For the oxidative stress, Hinokio et al.'s study (2002), a prospective longitudinal study, showed that increased oxidative stress leads pathogenesis of DN. It was found that rising 8-oxo-7,8-dihydro-2'-deoxyguanosine (8-oxodG), a marker of oxidative stress tested by urinary, was predictive for incident DN among type 2 diabetics(64).

Diabetes patients with proinflammatory cytokines and chemokines, including interleukin (IL) 6, IL 18, high-sensitivity C-reactive protein, monocyte chemoattractant protein-1 or adhesion molecules, are higher risk to contribute to nephropathy and severe kidney disease. The studies exhibited the relationship between microalbuminuria and endothelial dysfunction and low-grade

inflammation in both type diabetes (type 1 and type 2 DM). An observational study (2002) determining risk of endothelial dysfunction and chronic inflammation for microalbuminuria among 328 type 2 diabetics. The univariate analysis showed that Von Willebrand factor, Tissue-type plasminogen activator, soluble E-selectin, C-reactive protein showed relation to development of urinary albumin excretion with coefficient of 0.08 (95%CI 0.03 - 0.13), 0.08 (95%CI 0.02 - 0.13), and 0.08 (95%CI 0.04 - 0.12), respectively. However, the investigators described that these endothelial dysfunction markers had relationship with albuminuria, but the mechanism was unclear (65). Similarly, a randomized-controlled trial of Persson et al. (2008) using the Irbesartan in Patients with Type 2 diabetes and Microalbuminuria (IRMA 2) study to evaluate whether biomarkers of endothelial dysfunction and inflammation can predict for progression of diabetic nephropathy (defined as onset of persistent albuminuria). The finding indicated that endothelial dysfunction (i.e., von Willebrand Factor (vWf), soluble vascular cell adhesion molecule-1 (sVCAM-1), soluble intercellular cell adhesion molecule-1 (sICAM-1), sE-selectin) was predictive for the development for diabetic nephropathy in type 2 diabetics and microalbuminuria with hazard ratio of 3.2 (95%CI 1.56-5.65)(66).

Genetic factors had association for development of DN. Type 1 and type 2 diabetics who had a sibling having DN have higher risk of DN than those who have a sibling without diabetic nephropathy(67-69). Angiotensin-converting enzyme (ACE) gene polymorphism which is a gene component of the renin–angiotensin–aldosterone system (RAAS). RAAS is a system related in the pathogenesis of progressive renal disease. ACE was showed risk association for diabetic

nephropathy developing in some studies (70, 71). Some studies contributed contrast results(72, 73). However, a meta-analysis of Wang et al. (2012) was performed with 26,580 subjects recruiting 63 published studies to evaluate the risk of ACE insertion/deletion (I/D) polymorphism for diabetic nephropathy developing. The finding showed that all ACE I/D polymorphism were associated as the risk of diabetic nephropathy in all genetic models. When subgroup analysis was conducted, results were found that Asian with type 2 DM exhibited critical associations for all genetic models. The investigators explained that ACE I/D polymorphism could contribute for the development of diabetic nephropathy (74). African-American and Pima Indians were reported with higher risk of diabetic nephropathy than Caucasians. The possible explaining was socioeconomic and genetics factors in these ethnic group contributed different glycemic and blood pressure controlling (75, 76).

Glomerular hyperfiltration showed relationship between glomerular hyperfiltration and development of DN in both types of diabetics (77, 78). A cohort study (2015) using the coronary artery calcification in type 1 diabetics evaluated whether renal hyperfiltration (defined as level of eGFR 120 mL/min/1.73m^2) have associated to rapid eGFR decline. Finding showed that renal hyperfiltration were predictive for rapid GFR decline with OR of 5.00 (95%CI 3.03-8.25)(78). Similarly, A cohort study of Remuzzi et al. (2006) among 600 type 2 DM patients without macroalbuminuria to determine the association between hyperfiltration and nephropathy (defined as macroalbuminuria). The finding showed that patients with renal hyperfiltration had risk of macroalbuminuria progression with HR of 2.16 (95%CI 1.13-4.14). The possible mechanisms of

nephropathy from renal hyperfiltration was due to increased glomerular pressure contributing to mesangial expansion and thicken glomerular basement membrane (79, 80).

II. DEVELOPMENT AND VALIDATION OF CLINICAL PREDICTION MODEL

Clinical prediction model is a developed tool to estimate probabilities of prognosis outcome from multiple parameters. Clinical prediction model is applied for screening risk individuals for a disease, predicting future of disease. Moreover, clinical prediction model can assist health provider to make a decision for disease management and supplying health education for patients(81). Even though many reports and books suggest methods for developing clinical prediction models, there is no standard of guideline(82, 83). Finally, five summarized steps are obtained.

II.1 Preparing data for creating clinical prediction model

To obtain a good and accurate prediction model, multivariate variables using extensive dataset are used. Components, including target outcome target patients, target user, should be raised to clarify prediction model. If a researcher would liked to establish a prediction model for general population, prediction model should be simple with categorical questions (e.g., yes/no choices).

II.2 Dataset selection

Aim of study leads dataset selection. Cross-sectional study usually is used to develop clinical prediction model for screening undiagnosed disease (i.e., diabetes mellitus, chronic kidney disease). Example of clinical prediction model in cross-sectional study is Korean Diabetes Scores(84). In contrast, longitudinal or cohort studies are used for predicting incidence of disease (e.g., cardiovascular disease, cancer) to prevent these diseases. Example of clinical prediction model in cohort study is the American College of Cardiology (ACC)/American Heart Association (AHA) Arteriosclerotic Cardiovascular Disease (ASCVD) risk equation (85).

II.3 Handling variables

Collected data are contained categorical or continuous variables. For continuous variables, the skewed distributions can be provided. However, the skewness can violate of the statistical assumption which leads to the statistical analysis yielding the invalid results (86). For the Cox-proportional hazard regression, the proportionality assumption is the relative hazard of groups of interest is constant over time.

Logarithm transformation or squared methods are used to solve the skewness problem (86). However, logarithm transformation is usually applied to reduce the skewness of each variable as possible in order to raise the validity of the statistical analyses. In this study, the skewness of each variable was tested by using the skewness and kurtosis test for normality based on D'Agostino's X^2 test, the histogram and the distributional diagnostic plots (87). For the skewness and kurtosis test, if the p-value of a variable is more than 0.05, the variable has no

skewed. For histogram, the distribution should be normal distribution or bell-shape curve as possible. For the distributions of the diagnostic plots, the distribution should be closed to a straight line meaning to have closely normal distribution. Even a logarithm transformed variable does not provide the good normal distribution, the less skewed distribution of data is more suitable for statistical analysis.

Missing data is a common problem that can be occurred. There are many methods to handle with missing data, i.e., deletion of uncompleted data, replacing with means, multiple imputation, etc. Handling of missing data was described in part of "**Multiple Imputation by Chain Equation (MICE)**".

II.4 Model development

II.4.1 Identifying risk predictors

The candidates of the risk predictors are reviewed from the literature reviews and the expert's experience. The database of these risk predictors should be from either cohort study or nested case-controlled study. Univariate and multiple variate analysis are used to estimate the association between the risk predictors and a chronic disease. For univariate analysis, predictor variables with p-value < 0.1 were considered to be included in the multivariate analysis. For multivariable analysis, predictor variables were estimated by backward elimination approach for building model. Improved performances of model selections in multivariate Cox-proportional hazard regression model were estimated by using goodness-of-fit of the Akaike Information

Criterion (AIC). AIC is an indicator of model fitting to prevent over fitting of the model. Lower AIC indicates better model(88).

II.4.2 Generating risk equation

After model generation, risk equation of disease endpoint are generated. Absolute CKD risk was generated based on the Cox model at three years, and subtracting the mean values each CKD predictor (89). Applying Cox models in this way allows another group to recalibrate the equation to their own cohort by replacing the mean values of the predictors in our cohort, and our three-year survival estimate with those of their own cohort. Predicted CKD risks were calculated from risk scores following this equation:

Risk score = $(\sum \beta_i \times X_i - \beta_i \times \bar{X}_i)$

Predicted CKD risks = $1 - S_0\,(t)^{\exp\,(Risk\ score)}$

S_0 is baseline survival function at time t, β_i is the regression coefficient for X_i, and the $\bar{X}_i$ is mean level of X_i.

II.5 Model validation

When a prediction model is developed, the prediction model should be tested for the validity of performance. There are three types of validation methods, including the apparent validation (using own sample), the internal validation (using own population) and the external validation (using other population). In this study, the internal validation method was applied.

Internal validation can be divided into 3 types based on techniques, including split-method, cross-validation and bootstrapping methods (90, 91).

Bootstrap method is randomly sampling with replacement from an original dataset for use in obtaining statistical inference.

Split-method is randomly split into 2 datasets; a training data for developing model, and a validation dataset to test model performance. Percentage ratio of splitting are mostly used of 70:30 or 80:20(92).

Cross-validation method is a performing of consecutively model test by randomly drawn parts of original sample. Model is developed and estimated by 50% of original sample and tested on the 50% of independent part and vice versa. Average performance of model is calculated over 2 time repeating. Other fractions are used for cross-validation, i.e., 10% cross-validation which was 10 times for repeating model testing. In this study, split-method with ratio of 70:30 was applied for training and validation datasets (**Figure 2**), respectively.

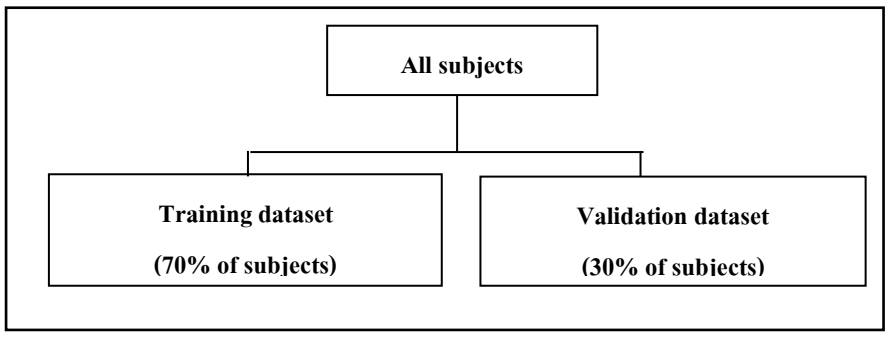


Figure 2: Schematic representation of split method

This step is to validate the performance of the developed prediction model. The approaches of validation are following.

II.5.1 Discrimination

Discrimination refers to ability of distinguishing patients with events out from those without events. C-statistic or concordance index is applied to evaluate for discriminative performance of Cox model.

The Harrell's C-statistic is a rank parameter which is a comparison between predictions with the ranks of actual event times, and not directly with the binary event status. The Harrell's C-statistic has a range of scale from 0.5 to 1.0. If C-statistic is 0.5 meaning no discriminative ability, and closed to 1 meaning perfect discriminative ability (93, 94).

II.5.2 Calibration

Calibration refers to another measurement of the predictive accuracy of the prediction model. Calibration is a measure of degree of consistency between predicted probability produced by prediction model and actual observed probability (95). There are several measures for calibration performance, such as calibration plot, calibration slope, calibration-in-the-Large(96). However, we will explain calibration measures that are applied in this study, including Hosmer-Lemeshow X^2 test, and survival probability curve for risk groups.

Hosmer-Lemeshow X^2 test is a goodness of fit test which is usually applied for binary outcomes(97). However, Hosmer-Lemeshow X^2 test is also modified to use in survival studies

(98). However, Hosmer-Lemeshow X^2 test is a comparison between predicted probabilities by prediction model and observed events divided by decile of predicted probabilities(82). A X^2 or Chi-square statistic is an estimator for goodness of fit. For the good prediction model, X^2statistic is usually non-significant (p-value >0.05) which means to the similarity between the probabilities of predicted events and observed events. In the other hands, if the probabilities of predicted events are not similar to observed events or significance is obtained in X^2 statistic, the accuracy of prediction for prediction model is poor. Hosmer-Lemeshow X^2 test can be displayed in term of either calibration graph or table. Hosmer-Lemeshow X^2 test has a easy concept, however, a difficult interpretation is obtained in small sample size (low power)(82).

Calibration of survival probabilities from Kaplan-Meier method or population-average survival curve is a calibration method by comparing predicted survival probabilities with observed curve (Kaplan-Meier curve) in risk groups. Risk groups are divided based on categorized prognostic index (PI) which are linear predictors from a cox model. Risk groups are usually divided into three to five groups. The result of this calibration is showed in term of a graph comparison (96). This calibration method is easy to understand and interpret. This method is an alternative choice on current practice. However, there are two limitations of this method (99): 1) Results of calibration is depended on risk group dividing. Even equally selection of risk groups is possible but a sensitivity analysis should be required. 2) According graphical display, there is no statistical value to determine whether calibration has significantly worse.

III. THE DEVELOPED CKD PREDICTION MODEL IN DIABETIC PATIENTS

We searched a journal from Embase and PubMed MEDLINE databases. For PubMed MEDILNE database, mesh terms were applied with "Renal Insufficiency, Chronic "AND "Decision Support Techniques". In addition, we searched manually from the reference of each recruited studies. The studies related CKD prediction models were divided into 2 groups, including CKD prediction model for CKD progression and CKD developing.

The studies which developed prediction model for renal outcome among diabetics are summaries in **Table 1.**

Table 1: The studies related CKD prediction models among DM patients

Study	Design/ Population (country)	Number of subjects/ outcomes	Age (years)	Time follow up (Years)	Model	Renal outcomes	Risk factors	Method of validation	Discrimination	Calibration
Keane et al. 2006(22)	A prospective cohort study/ type 2 diabetes and nephropathy progression (the RENAAL train from 28 countries)	1,513/341	NR	3.4	Cox	ESRD (renal replacement therapy or long-term dialysis)	Log UACR, increased serum creatinine, serum albumin, and decreased hemoglobin level	NR	None	NR
Yang et al. 2006(19)	A prospective cohort/ type 2 diabetics without ESRD (Hong Kong China)	4,438/159	Mean: 60 (SD: 20)	Median 2.9 (IQR 1.6–4.1	Cox	ESRD (defined as among death due to diabetes with renal manifestations or renal failure, hospitalization)	**Model 1:** duration of diabetes, SBP, log10 total cholesterol: HDL-cholesterol ratio, retinopathy,	**Internal validation**: split method	**Model 1:** C-statistic of 0.883	None

Table 1: The studies related CKD prediction models among DM patients (continue)

Study	Design/ Population (country)	Number of subjects/ outcomes	Age (years)	Time follow up (Years)	types of model	Renal outcomes	Risk factors	Method of validation	Discrimination	Calibration
Yang et al. 2006(19)	A prospective cohort/ type 2 diabetics without ESRD developing (Hong Kong China)	4,438/159	Mean: 60 (SD: 20)	Median 2.9 (IQR 1.6–4.1	Cox	ESRD (defined as among death due to diabetes with renal manifestations or renal failure, hospitalization due to renal failure or eGFR < 15 mL/min/1.73m^2)	**Model 2:** eGFR, hematocrit, Log10, UACR	**Internal validation**: split method	**Model 2:** C-statistic of 0.967	HL X^2 test *(P<0.1)*
							Model 3: eGFR, hematocrit, Log10 UACR, adjusted number of months using	**Internal validation**: split method	**Model 3:** C-statistic of 0.977	None

Table 1: The studies related CKD prediction models among DM patients (continue)

Study	Design/ Population (country)	Number of subjects/ outcomes	Age (years)	Time follow up (Years)	Model	Renal outcomes	Risk factors	Method of validation	Discrimination	Calibration
Desai et al. 2011(17)	A prospective cohort study nested within a randomized clinical trial/ type 2 diabetes with CKD (eGFR 20-60mL/min/1.73m^2) and anemia Developing (The TREAT study from 24 countries)	995/222 (ESRD), 407 (ESRD or death)	Mean: 68	3.5	Cox	ESRD (the initiation of dialysis therapy) or death	**Model 1:** UACR, eGFR, black race, hemoglobin level, male sex, and history of Heart failure (HF)	**Internal validation:** apparent	**Model 1:** C-statistic: 0.844, ESRD outcomes; 0.752. ESRD or death outcomes	NR
							Model 2: UACR, eGFR, black race, hemoglobin level, male sex, and history of HF, TnT and NT-pro-BNP	**Internal validation:** apparent	**Model 2:** C-statistic: 0.848, ESRD outcomes; 0.757. ESRD or death outcomes	NR

Table 1: The studies related CKD prediction models among DM patients (continue)

Study	Design/ Population (country)	Number of subjects/ outcomes	Age (years)	Time follow up (Years)	Model	Renal outcomes	Risk factors	Method of validation	Discrimination	Calibration
Jardine et al. 2012(20)	An observational analysis of a RCT/ type 2 diabetes Developing (20 countries from Asia, Australia, Europe, and North America)	11,140/ 129 with double Scr, 59 ESRD or death, 2,715 developed to albuminuria	Mean: 66 (SD: 6)	4.8	Cox	Major kidney related events (double Scr, renal replacement therapy, or renal death)	Male gender, eGFR, UACR, SBP, HbA1c, diabetic retinopathy, and age at educational attainment	**Internal validation:** Bootstrap	C-statistic: 0.847	HL X^2 test: *1.5 (*p = 0.9*)*
						Onset of albuminuria	Ethnicity, eGFR, UACR, SBP, BP-lowering treatment, HbA1c, retinopathy, waist circumference	**Internal validation:** Bootstrap	C-statistic: 0.647	HL X^2 test: 16.5(*P=0.6).*

Table 1: The studies related CKD prediction models among DM patients (continue)

Study	Design/ Population (country)	Number of subjects/ outcomes	Age (years)	Time follow up (Years)	Model	Renal outcomes	Risk factors	Method of validation	Discrimination	Calibration
Elly et al. 2013(100)	Cohort study/ type 2 diabetics without ESRD (New Zealand)	25,736/ 637	Mean: 62	Median: 7.3 (IQR 5.9-8.6)	Cox	Fatal or nonfatal ESRD (peritoneal dialysis or hemodialysis for ESRD, renal transplantation, or death from ESRD)	sex, ethnicity, age, diabetes duration, albuminuria, Scr, SBP, HbA1c, smoking status, and previous CVD	External validation	C-statistics: 0.89–0.92	HL X^2 test (display in graph)
Dunkler et al. 2015 (101)	Retrospective cohort/ type 2 diabetes with non-albuminuria or microalbuminuria (ONTARGET study from 40 countries)	6,766/694 (microalbuminria), 312 (macroalbuminuira), 105 (double Scr), and 62 (ESRD)	61-71	5.5	Cox	Alive with CKD (defined as among new microalbuminuria, macroalbuminuria or doubling of Scr, or ESRD)	**For clinical model**; glucose control, diabetes duration, number of prescribed antihypertensive drugs, previous vascular events, or vascular comorbidities.	Internal validation: bootstrap External validation	C-statistic: 0.69 (external validation dataset)	Calibration-in-the-large: 0.01 Calibration slope: 1.01,
						Alive with CKD (defined as among new microalbuminuria, macroalbuminuria or doubling of Scr, or ESRD)	**For laboratory model**: baseline albuminuria, eGFR, sex, and age	Internal validation: bootstrap External validation	C-statistic: 0.68 (external validation dataset)	Calibration-in-the-large: 0.01 Calibration slope: 1.01

Table 1: The studies related CKD prediction models among DM patients (continue)

Study	Design/ Population (country)	Number of subjects/ outcomes	Age (years)	Time follow up (Years)	Model	Renal outcomes	Risk factors	Method of validation	Discrimination	Calibration
Lin et al. 2017 (18)	Retrospective cohort study/ type 2 diabetes (China)	24,104/ 1,215	30-84	8.3	Cox	ESRD	Age, gender, age of diabetes onset, combined status of BP and anti-hypertensive medication use, Scr, variation of HbA1c, variable of SBP, diabetic nephropathy, albuminuria, anti-diabetic medication, and combined statues of dyslipidemia and anti-hyperlipemia medication	Bootstrap	C-statistic: 0.8-0.92 (validation dataset)	HL X^2 test ($p > 0.05$). Calibration slope of 0.92. Calibration-in-large of 0.009 (validation dataset)
Low et al.2017 (23)	Prospective cohort study/ type 2 diabetes (Singapore)	1,582/ 679	Mean: 57.3 (SD: 11.6)	Median: 5.5 (IQR: 4.3–7.0)	Logistic	CKD (defined as 25% or more reduction of eGFR from baseline, stage	SBP, HbA1c, eGFR, and UACR	Internal validation: Split method (70:30)	C-statistic; 0.83 (validation dataset)	HL X^2 test: 1.36($P=0.928$), validation dataset.

Table 1: The studies related CKD prediction models among DM patients (continue)

Study	Design/ Population (country)	Number of subjects/ outcomes	Age (years)	Time follow up (Years)	Model	Renal outcomes	Risk factors	Method of validation	Discrimination	Calibration
Wan et al.2017 (21)	Retrospective cohort study/type 2 diabetes without ESRD (China)	149,333/ 21,426	18-79	Median:5 (IQR 0.04-6.04	Cox	ESRD (ICD-9-CM or eGFR <15 ml/min/1.73m^2)	**Female model:** Age, anti-hypertensive medication, anti-diabetic medication, HbA1c, BP, UACR, eGFR	Internal validation: bootstrap	C-statistic: 0.862 (0.845,0.880)	Calibration plot (similar between observed event and predicted event))
							Female model: Age, anti-hypertensive medication, anti-diabetic medication, HbA1c, BP, UACR, eGFR	**external validation** (ADVANCE cohort and New Zealand cohort)	C-statistic:0.844 (0.822,0.865), **ADVANCE cohort;**0.861 (0.842,0.880), **New Zealand cohort**	Calibration plot (not similar between observed event and predicted event)
							Male model Age, anti-hypertensive medication, anti-diabetic medication, HbA1c,	**Internal validation:** bootstrap	C-statistic: 0.866 (0.849,0.882)	Calibration plot (similar between observed event and
							Male model Age, anti-hypertensive medication, anti-diabetic medication, HbA1c, BP, UACR, eGFR	**external validation** (ADVANCE cohort and New Zealand cohort)	C-statistic:0.858 (0.840,0.876), **ADVANCE cohort;** 0.863 (0.846,0.880), **New Zealand cohort**	Calibration plot (not similar between observed event and predicted event)

Table 1: The studies related CKD prediction models among DM patients (continue)

Study	Design/ Population (country)	Number of subjects/ outcomes	Age (years)	Time follow up (Years)	Model	Renal outcomes	Risk factors	Method of validation	Discrimination	Calibration
Geba et al.2019 (102)	Retrospective cohort study/type 2 diabetes (NR)	58,428/ 2,208	Mean: 61	Mean: 6.8	Cox	ESRD	BUN, CKD, albumin, pulse pressure, HbA1c, anemia, eGFR, and preexisting DM	NR	C-statistic: 0.706	NR
Nelson RG. et al. 2019(24)	15 multinational cohort studies/ diabetes (Multiple*)	781,627/ 313,646	Mean: 62	Mean: 3.9	Cox	CKD (defined as eGFR < 60 mL/min/1.73 m^2)	Age, women, black race, level of eGFR, CVD, HT, BMI, UACR, HbA1c, DM medication	**9 external validation** cohorts from OLDW	C-statistic: 0.801 (0.75, 0.819) **External validation cohorts:** 0.756(0.75, 0.758)	Calibration slope: 0.88-1.05 External validation cohorts: Calibration slope: 0.82- 1.02

NR, not report; logistic, logistic regression; ESRD, end stage renal disease; log 10, logarithm to base 10 ; SD, standard deviation; SBP, systolic blood pressure; HDL-C, high density lipid-cholesterol; eGFR, estimated glomerular filtration Ratio; UACR, urinary albumin to creatinine ratio; HF, Heart failure; TnT, troponin T; NT-pro-BNP, N-terminal pro–brain natriuretic peptide; Scr, serum creatinine; BUN , blood urea nitrogen; HbA1c, hemoglobin A1C; CVD, cardiovascular disease; DM, diabetes mellites; BP, blood pressure; HT, hypertension; ADVANCE study, the Action in Diabetes and Vascular Disease: Preterax and Diamicron-MR Controlled Evaluation study; New Zealand cohort, the New Zealand Diabetes Cohort Study; ODWL, OptumLabs Data Warehouse; ONTARGET, ONgoing Telmisartan Alone and in combination with Ramipril Global Endpoint Trial; TREAT, the Trial to Reduce Cardiovascular Events with Aranesp Therapy; RENAAL, Reduction of Endpoints in NIDDM with the Angiotensin II Antagonist Losartan.

*Australia, Canada, China, Czech Republic, Estonia, France, Germany, Hungary, India, Ireland, Italy, Lithuania, Malaysia, Netherlands, New Zealand, Philippines, Poland, Russia, Slovakia, Norway, United States, Sweden, Singapore, Israel.

From 11 studies (**Table 1)**, most of CKD prediction models are developed to predict ESRD among patients with type 2 DM. Only two established prediction model from Low's and Nelson's studies were developed for CKD. Every study was cohort studies with years follow-up of studies were in range of 2.9-8.3 years. UACR, age and HbA1c are the risk predictors appearing in most prediction models. External and internal validation were applied for validation methods. Internal validations, i.e., bootstrapping and split method, were approaches to evaluate model performances. For discrimination performance, the C-statistics which were discrimination measures were in ranges of 0.68-0.9 which has moderate to high power of predictions. In addition, calibration measures, i.e., Hosmer-Lemeshow X^2 test, calibration slope, calibration-in-large, were applied to evaluate accuracy of prediction models.

IV. THE DEVELOPED CKD PREDICTION MODEL IN THAILAND

There are three Thai studies that developed for CKD prediction model showed in **Table 2.** Both Thai studies were conducted in Thai population studies, including the population-based Thai Screening and Early Evaluation of Kidney Disease (SEEK) study and employees of the Electric Generating Authoring of Thailand (EGAT), for developing CKD model. Population of Thai SEEK study which were randomly sampling from four regions of Thailand, including Northern, Northeastern, Central and Southern) and Bangkok (metropolitan), were diabetes patients of 11.92% (434 from 3,459). Similarly, Employees of EGAT which were volunteers for

health survey were diabetes patients of 7.8% (244 from 3,186). Population were general population not to be specifically diabetes patients. Both CKD models were developed to predict CKD outcome (defined as eGFR <60 mL/min/1.73m^2). In the study of Thankkintian et al. CKD model was established to predict CKD occurring in present, but CKD model of Saranburat et al.'s study was developed to predict CKD occurring in next 10 years. Both studies had moderate to high of discriminations in range of 0.72-0.79, and good calibration were obtained. Even two CKD prediction models have good performance of CKD stage 3 prediction, their models might not be specific to patient with type 2 DM regarding small proportion of patients with type 2 DM.

Moreover, the study of Kittipanyaworakun had developed CKD prediction model in 322 type 2 diabetes from Saraburi Hospital. Even the CKD prediction model has sensitivity, specificity and overall accuracy of 72.72, 92.31 and 86.96%, respectively. However, validation method for this study was not obtained, and eGFR calculation by using Thai eGFR equation as follows:$375.5\ x\ {Cr_{Enz}}^{-0.848}\ x\ age^{-0.364}\ x\ 0.712\ (if\ female)$, where Cr_{Enz} is serum creatinine which is measured by using enzymatic method. Even Thai eGFR equation for calculating eGFR stage 3 in Thai population has higher sensitivity (85.1%) and specificity (82.8%) than CKD-EPI equation (sensitivity of 59.8% and specificity of 87.6%), Thai eGFR equation is not used for calculating eGFR in practice. Furthermore, 322 patients with type 2 DM was quite small sample sizes that might be questionable for generalization.

Table 2 : Thai CKD prediction models

Study	Design/ Population	Number of subjects/ outcomes	Age (years)	Time follow up (Years)	Types of model	Renal outcomes	Risk factors	Method of validation	Discrimination	Calibration
Thankkinstian et al.2011 (103)	Cross-sectional study/ community based (Thai SEEK study)	3,459/ 606	Mean: 45.2 (SD 0.79)	NR	Logistic	CKD (eGFR < 60 ml/min/ 1.73 m^2)	Age, DM, hypertension, and history of kidney stones	bootstrap	C-statistic: 0.77	Bias observed versus predicted values: 0.045
Kittipanyaworakun et al. 2013 (12)	Retrospective cohort study/ patients with type 2 DM	322/88	Mean: 59.1(SD8.7)	mean: 5.2 (SD 0.3)	Logistic	CKD (eGFR < 60 ml/min/ 1.73 m^2)	Age, duration of DM, eGFR, increased UAE, and SBP	No validation	**NOTE:** Sensitivity 72.72% , sensitivity 92.31%, and overall accuracy 86.96%	
Saranburat et al. 2017(104)	Cohort study/ Employees of EGAT	3,186/ 271	Mean: 51 (SD: 7.4	NR	Logistic	Decreased eGFR (eGFR < 60 ml/min/ 1.73 m^2)	**Model 1:** Age, sex, SBP, DM, and waist circumference.	bootstrap	C-statistic: 0.72	HL X^2 test :9.02 (P= 0.34),
							Model 2: Age sex, SBP, DM, GFR.	bootstrap	C-statistic: 0.79	HL X^2 test: 10.87 (P=0.21),

NR, not report; HL X^2 test, Hosmer-Lemeshow X^2 test; Logistic, logistic regression; SD, standard deviation; SBP, systolic blood pressure; eGFR, estimated glomerular filtration ratio; UAE, urinary albumin excretion; DM, diabetes mellites; Thai SEEK study, Thai Screening and Early Evaluation of Kidney Disease (SEEK) study; EGAT, Electric Generating Authority of Thailand.

According to previous mentions, most prediction models in diabetes patients are for ESRD. Only two study from Singapore was related for CKD developing among type 2 diabetes **(Table 2)**. Even one Thai prediction models are predictive for CKD developing, there are some limitations in this study in terms of small sample size, lacking of validation method and eGFR calculation by using Thai eGFR formula which is not used in practice. Therefore, there is no suitable CKD prediction model for Thai diabetes patients.

V. MULTIPLE IMPUTATIONS BY CHAIN EQUATION (MICE)

This subsection was described about types of the missing data, multiple imputations by chained equation (MICE) method and how to check the imputed model.

Missing data is a problem that can occur in research study. The problem of missing data lead to the reduction of statistic power, the biased estimation of parameters, reduced representativeness of samples, and invalid of conclusion (105). To understand further about the imputation method, understanding about types of the missing data is required.

V.1 Types of missing data

Types of missing data are categorized based on mechanisms of missing data: missing data at random (MAR), missing not at random (MNAR) and missing completely at random (MCAR). MAR is the condition that missing is related to the observed response not the missing values. For example, collecting HbA1c from DM patients depends on the requirements of each

physician. In this case, the observed responses are the physicians' requirements not the DM patients' willingness because the authority of laboratory test is under a physician's recommendation. Therefore, most research studies with missing of laboratory results including our study were assumed as MAR. MCAR is the probability of missing not depending neither the observed response or the missing values, so the pattern of data missing is completely random. For example, the weight equipment is broken, so in that day nurse cannot record the weight in each patient. The last one is MNAR which is referred to the probabilities of data missing is related to the missing values itself. For example, if the present weight recording is depended on the ex-weight recording. So, if the ex-weight is not recorded, it will lead to the present weight unrecorded (105, 106).

V.2 Multiple Imputations by Chain Equation (MICE)

There are many methods for handling missing data, such as the listwise deletion (LD), the pairwise deletion (PD), mean substitution. The LD method is an approach in handling missing data by omitting cases with missing data and analyze only complete cases. The PD method is an approach to missing data by eliminating missing data when the variables having missing values for analysis. If these is no missing data to concern in analyzing variables, the existing values can be used for statistic testing. So, PD preserved information than LD. However, these techniques lead the bias estimations, underestimated errors and inefficient estimations(105-107). Mean substitution is an approach of the single imputation to missing data by substituting the missing

values with the mean of remaining values. However, mean substitution leads true values rather than the imputed values which causes overly precise results and leads to incorrect conclusion.

Multiple imputation (MI) is another method to handle the missing data. MI especially multiple imputation by chained equation or MICE, a flexible technique, can be applied for continuous and binary variables. MICE is a method under MAR condition. MICE is a technique to missing data by imputing missing values many times with establishing many imputed datasets. The imputed values are based on two parts: the individually observed values and the relations between the observed variables and variables with missing data in other subjects. The imputed values of each variable with missing data are generated into *m* dataset. And the estimates of *m* datasets are pooled to be the single estimate based on the Rubin's rule(107). To understand more about MICE method, we give the example for the missing values for HbA1c and UACR. Other variables for analysis are eGFR and age. The steps of MICE are these followings **(Figure 3)**(108):

Step 1: Replacing the missing values in each variable by the mean of each variable called "place holder" in the dataset. For example, the means of HbA1c and UACR variables are 7 kg and 300 meters, respectively. So, the missing values of "HbA1c" and "UACR" variables (place holders) are replaced with the value of 7 and 300, respectively.

Step 2: The "place holder" of HbA1c is set back to be missing value again. But the "place holder" of UACR is still replaced with the mean of height by 300.

Step 3: The observed values of HbA1c which are non-missing values are regressed with other variables, including eGFR, age, gender and UACR to find out the missing values of HbA1c in the imputed model. The variables of the imputed model may not include all variables in the datasets. The observed variables are the analytical variables for generating a prediction model. In this case we give eGFR and age to be the observed variables. In this example, "HbA1c" variable performs as the dependent variable, and other variables perform as the independent variables.

Step 4: The missing values of HbA1c are replaced with by the linear regression model. When the "HbA1c" variable is used as the independent variable for the "UACR" variable in the linear regression model, the imputed values of HbA1c will be used. If the missing values is from the categorical variable, the logistic regression will be conducted. In case of the logarithm transformed variables, the imputation should be conducted by using the logarithm transformation.

Step 5: For the missing values of "UACR" variable, step 2 to 4 are repeated. When the missing values of the "UACR" variable are completely imputed, it has been finished as one imputed dataset. The numbers of the imputed datasets are depended on the most percentage of the missing data(109). For example, if the missing of HbA1c and UACR are 20 and 10, respectively. The 20 imputed datasets should be generated. For the survival analysis, time to event should be concerned to be the observed variable.

Figure 3: The multiple imputation by chained equation (MICE) in one imputed dataset.

V.3 How to check the imputed model

After imputing the missing values of each variable in *m* imputed datasets, the validity of the imputed values should be evaluated. The comparisons of the distribution between the observed and the imputed data are evaluated in several method, such as distributions of estimates, kernel density plots, scatter plots of linear prediction. Even there are several methods, there is no the gold standard (110).The simplest method is comparing the distributions of means and standard deviation (SD) between the observed and imputed data.

CHAPTER III

METHODOLOGY

This chapter describes details of the study methodology which includes a research design, the population and sample group, steps and instruments used in intervention, and data analysis.

I. Research design

II. Population and sample

III. Data description

IV. Experimental procedure

V. Statistical analysis

VI. Software

VII. Ethical consideration

I. RESEARCH DESIGN

A retrospective cohort study was conducted by using DM patients' data from the Diabetes Centre Clinic, Taksin hospital. The data were collected data from 1 January 2008 to 31 December 2017.

II. POPULATION AND SAMPLE

II.1 Study population and samples

The data of 2,178 patients with type 2 DM were recruited from the E-Phis program, Taksin hospital.

II.1.1 Area selection

The data of type 2 diabetes patients who were diagnosed by physician from the Diabetes Centre Clinic, Taksin hospital were applied in this study.

II.1.2 Patient selections

The eligible patients were selected based on profiles of patients. The criteria are these followings.

The inclusion criteria:

1) Adult type 2 diabetes patients with age $\geq$ 18 years old, and
2) Type 2 DM patients with the preserved eGFR ($\geq$ 60 mL/min/1.73m^2).

The exclusion criteria:

1) Patients who had pregnancy because these factors could confound the level of the interested laboratory. Pregnancy was collected by the International Classification of Diseases, 10th Revision (ICD-10) code O240-O249, or
2) Patients who had either the history of cancer or autoimmune diseases (including systemic lupus erythematosus, rheumatoid arthritis) because these diseases could

confound patients' kidney functions. Cancer were collected by ICD-20 code of C000-C97 and D000-D09. ICD-10 codes of systemic lupus erythematosus and rheumatoid arthritis are M320-M329 and M050-M069, respectively, or

3) Patients who had the history of other renal diseases or previous renal function disorders, including glomerular diseases (ICD-10 code: N00-N08), renal tubulo-interstitial disease (ICD-10 code: N10-N16), renal failure (ICD-10 code: N17-N19), urolithiasis (ICD-10 code: N20-N23), other disorders of kidney and ureter (ICD-10 code: N25-N29), polycystic kidney (Q610-Q619), and hematuria (R31). Because these conditions could confound patients' kidney functions.

4) Patients who had eGFR measurements less than 2 times, or

5) Patients who had follow-up time $\leq$1 year.

Figure 4 showed the process of sample selection

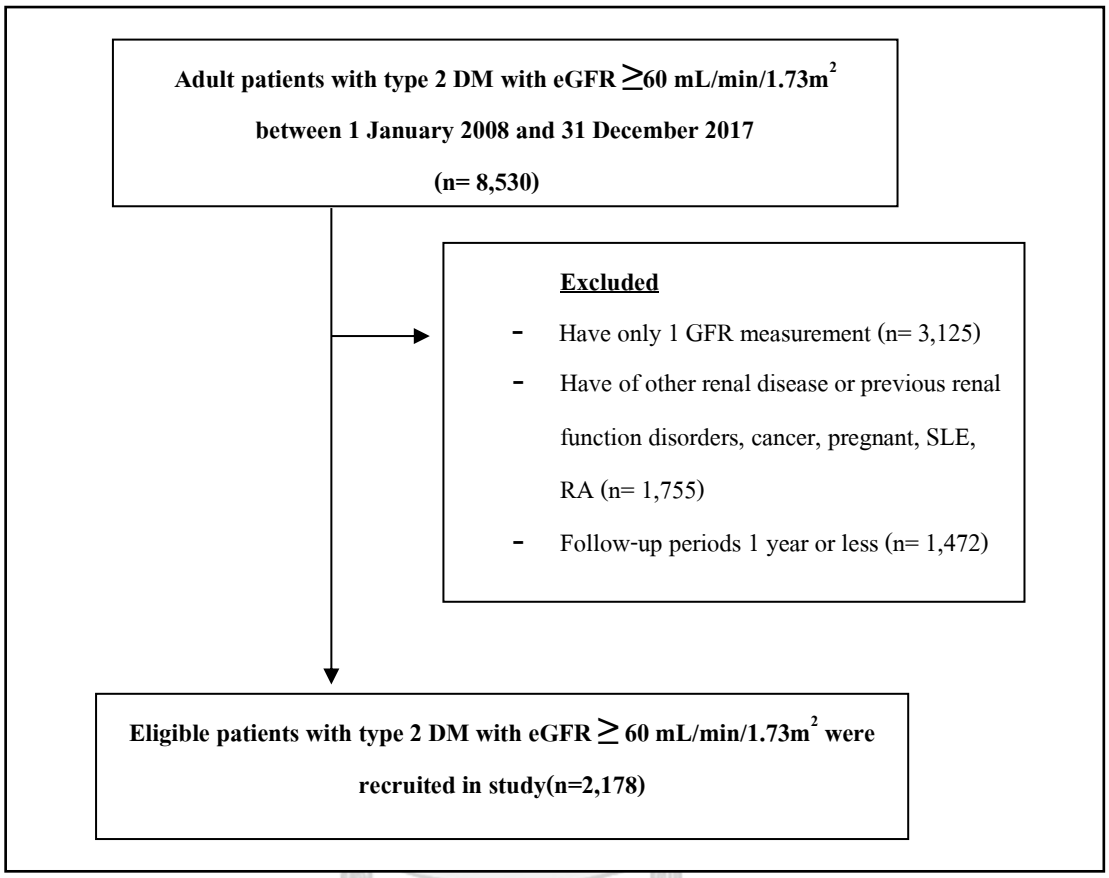


Figure 4: The flow chart of sample selection

II.2 Sample size

A rule of thumb is applied for sample size estimate to reduce the bias estimates of regression coefficient from the developed risk prediction model (111). The formula for sample size calculation is this following:

$$N = (n \times 10)/I$$

Where N is referred to the required sample size, n is referred to the numbers of variables to be analyzed and I is referred to the incidences of the combined adverse events. For this study, we assumed the prevalence of CKD among patients with type 2 DM equal to 20 % based on Thai published reports (10, 11). We assumed to use I equal to 0.20 to calculate sample sizes in this study. Candidate predictors in this study were 16. Therefore, the required sample sizes were calculated from this following:

$$N = (10 \times 16)/0.2$$

$$N = 800$$

According to a large cohort study, some researchers suggest to use the entire dataset for high power and generalization(112). Therefore, sample sizes for developing model in this study are at least 800. However, eligible subjects in this study are 1,525 (sample sizes for developing model) which are enough for conducting the study.

II.3 Split method

Split method was applied as the internal validation, 2,178 eligible subjects were randomly split into 2 datasets, including the training and validation datasets. For the training dataset, we used 70% of sample sizes to develop the prediction model. For validation dataset, we used 30% of sample sizes to validate the prediction model. As a result, 1,525 subjects were in the training dataset and 653 subjects were in the validation dataset **(Figure 5).**

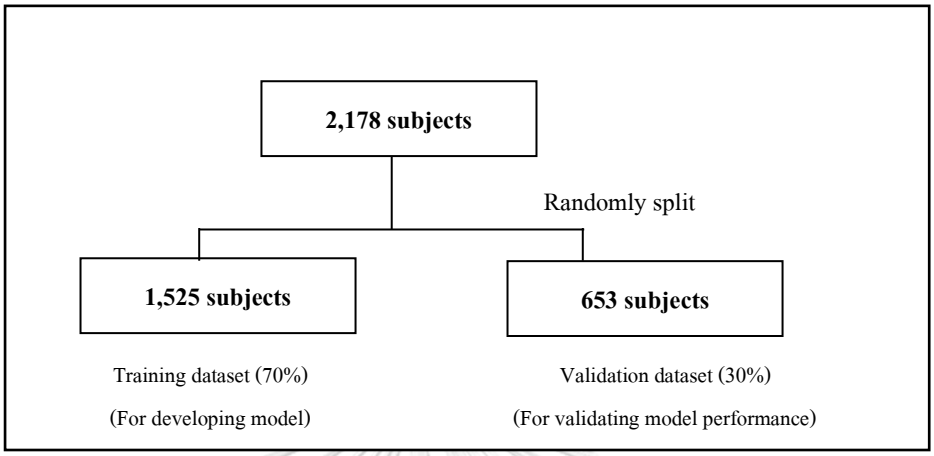


Figure 5: The training and validation datasets in the split method

III. DATA DESCRIPTIONS

The data of 2,178 patients with type 2 DM were consisted of hospital number (HN), socio-demographic factors, physical examination, diabetes related factors and biomarkers, comorbidities, medication therapies. To protect tracking personal data, patients' HN were blinded by using continuous number (i.e., 1, 2, 3).

This subsection was described about data preparation before model developing. These subsections were related to handle candidate predictors in terms of coding, candidate predictors' correlation, logarithm transformation and missing data.

III.1 Coding

The steps of handling variables were data collecting which were based on socio-demographic, physical examinations, diabetic related factors, comorbidities, medication therapy and outcome.

III.1.1 Socio-demographic factors are included age, gender, smoking status.

Age was calculated from 365-divided difference between the index date and birthdate.

Smoking status was categorized only into 2 groups, including currently smoking and non-smoker. Because smoking status were recorded only in 2017, so we cannot exactly know who were the ex-smoker. Currently smoking which was coded as 1, was defined as history of at least one time of smoking. Non-smoker was coded as 0.

Sex was categorized into 2 groups: male (coded as 1) and female (coded as 0).

III.1.2 Physical examinations were weight and height measurement. BMI was calculated by weight (kilogram) divided by squared height (meter^2).

III.1.3 Diabetic related factors and biomarkers

Blood pressure was measured by a nurse using Omron HEM 7120 Automatic Blood pressure after resting at least 5 minutes. Blood tests (i.e., hemoglobin A1c (HbA1c), triglyceride (TG), high-density lipoprotein (HDL), low-density lipoprotein (LDL), estimated glomerular filtration rate (eGFR)) were collected from blood sample after 8 hours overnight fasting. Urine tests (i.e., urinary albumin-to creatinine ratio (UACR)) were collected).

III.1.4 Comorbidities

Cardiovascular disease is defined as stroke, myocardial infarction (MI), and angina pectoris using ICD-10 codes I259, I600-I698, I200-I209, and I210-I219. Subjects who had the history of cardiovascular disease were coded as 1, and subjects without history of cardiovascular disease were coded as 0.

Diabetic retinopathy (DR) is retrieved from ICD-10 codes of E113 and H360. Subjects who had history of DR were coded as 1. Subjects who had no history of DR were coded as 0.

III.1.5 Medication therapy

History of non-steroid anti-inflammatory drugs (NSAIDs) taking were collected from the medication data. Usages of NSAIDs included oral and injectable NSAIDs. Generic names of NSAIDs were etoricoxib, celecoxib, meloxicam, piroxicam, ibuprofen, naproxen, indomethacin, diclofenac sodium, mefenamic acid, diclofenac sodium (injection), and parecoxib sodium (injection). Patients who were prescribed with NSAIDs medication were assumed with taking or using all of NSAIDs. NSAIDs exposures were calculated from the sum of prescription days of any NSAIDs within 1 year. Prescription days were the date of NSAIDs prescription plus the duration of NSAIDs supply(60). NSAIDs exposure was categorized into 3 groups; no taking NSAIDs (coded as 0), taking NSAIDs for 1-89 days (coded as 1) and taking NSAIDs $\geq$90 days (coded as 2)(60).

III.1.6 Outcome was presenting of CKD stage 3 with eGFR with less than 60 mL/min/1.73m^2. Censor is defined as lost-follow up of eGFR measurements more than 2 years. Serum creatinine and age were adopted to calculate for eGFR following the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equations (113) **(Table 3)**. CKD presenting is code by 1, and non-CKD is coded by 0.

Table 3: The CKD-EPI formula

Sex	Creatinine concentration	Formula for estimating GFR
Male	$\leq$ 0.9 mg/dL	$141x(\frac{Scr}{0.9})^{-0.411}x\ (0.993)^{Age}$
	> 0.9 mg/dL	$141x(\frac{Scr}{0.9})^{-1.209}x\ (0.993)^{Age}$
Female	$\leq$ 0.7 mg/dL	$144x(\frac{Scr}{0.7})^{-0.329}x\ (0.993)^{Age}$
	> 0.7 mg/dL	$144x(\frac{Scr}{0.7})^{-1.209}x\ (0.993)^{Age}$

Therefore, the patient's first eGFR record within the study period (1 January 2008) was defined as the index date. Baseline characteristics including socio-demographics, diabetes related factors and biomarkers, comorbidities, and medications were collected by using mean of each variable within 1 year after the index date. Follow-up ended when a patient developed CKD stage 3, and remaining patients were censored if they were lost to follow-up > 24 months or on 31 December 2017. **Table 4** showed the characteristic of sixteen CKD candidate predictors.

Table 4: The characteristic of sixteen potential CKD candidate predictors and outcome

Candidate predictors	Range	Types of data	Descriptions
Age, years	18-90	continuous	Age at recruiting study
Sex	0,1	categorical	1= male, 0 = female
Smoking status	0,1	categorical	1= currently smoke, 0 = non-smoker
BMI, kg/m^2	13.7-53	continuous	Mean of body mass index within 1 year
SBP, mmHg	85-229	continuous	Mean systolic blood pressure of a patient in 1 year
HbA1c, %	4-16.3	continuous	Mean HbA1c of a patient within 1 year
TG, mg/dL	31-3605.4	continuous	Mean TG of a patient within 1 year
LDL, mg/dL	32.3-409	continuous	Mean LDL of a patient within 1 year
HDL, mg/dL	21.8-133.5	continuous	Mean HDL of a patient within 1 year
Uric acid, md/dL	1.5-10.8	continuous	Mean uric acid of a patient within 1 year
Serum albumin, mg/dL	1.8-5.8	continuous	Mean serum albumin of a patient within 1 year
UACR, mg/g	1.3-6469.5	continuous	Mean UACR of a patient within 1 year
eGFR, $mL/min/1.73m^2$	60-179	continuous	Mean eGFR of a patient within 1 year

Candidate predictors	Range	Types of data	Descriptions
CVD	0,1	categorical	1 = CVD, 0 = non-CVD
Diabetic retinopathy	0,1	categorical	1 = retinopathy, 0 = non-retinopathy
NSAIDs exposure	0, 1, 2	categorical	2 = NSAIDs taking for ≥ 90 days, 1 = NSAIDs taking for 1-89 days, 0 = no NSAIDs taking
CKD stage 3 event	0,1	categorical	1=eGFR <60 mL/min/1.73m^2, 0 =eGFR ≥60 mL/min/1.73m^2
Year follow-up, years	0.003-7.4	continuous	Years follow up until patients were lost follow-up ≥2 years, CKD occurring.

BMI, body mass index; SBP, systolic blood pressure; HbA1c, hemoglobin A1c; TC; LDL, low-density lipoprotein cholesterol; HDL, high-density lipoprotein cholesterol; TG, triglyceride; UACR, urinary albumin to creatinine ratio; eGFR, estimated glomerular filtration rate; CVD, cardiovascular disease; NSAIDs, nonsteroidal anti-inflammatory drugs; CKD, chronic kidney disease.

III.2 Candidate predictors' correlation

According to using several variables for analysis, correlation matrix should be examined to see whether there was highly correlation between sixteen variables. Generally, if the correlation between two variables was high equal to 0.75, those variables should be removed. Nevertheless, the findings were showed that the correlation between eGFR and age was 0.703. This could be explained because age existed partially of eGFR **(Table 5)**.

Table 5:Pearson correlation matrix of sixteen variables

	Age	Gender	Smoking	BMI	SBP	eGFR	HbA1c	UUACR	Uric acid	Serum	LDL	HDL	TG	CVD	DR	NSAIDs
Age	1															
Gender	0.081	1														
Smoking	-0.018	-0.321	1													
BMI	-0.362	-0.054	0.061	1												
SBP	-.0228	-0.048	0.001	0.174	1											
eGFR	***-.0703***	-0.071	-0.104	0.200	0.116	1										
HbA1c	-0,162	-0.219	-0.114	-0.023	0.166	0.309	1									
UUACR	0.055	-0.174	0.096	-0.221	0.111	-0.085	0.252	1								
Uric acid	-0.239	0.451	-0.143	0.320	-0.125	0.047	-0.474	-0.075	1							
Serum albumin	-0.254	0.372	-0.209	0.062	-0.201	0.105	-0.238	-0.186	0.256	1						
LDL	-0.019	-0.179	-0.106	0.033	0.189	0.121	0.237	0.097	-0.274	0.010	1					
HDL	0.031	-0.241	-0.052	-0.001	0.041	-0.014	0.023	-0.072	-0.260	-0.151	0.298	1				
TG	-0.160	0.282	-0.180	0.030	0.096	0.133	0.091	0.012	0.208	0.103	-0.094	-0.200	1			
CVD	0.294	0.277	-0.028	-0.188	-0.203	-0.281	-0.245	-0.073	0.039	-0.018	-0.247	-0.086	-0.189	1		
DR	-0.004	-0.132	-0.058	-0.220	-0.173	0.211	0.147	-0.020	-0.180	-0.060	0.012	-0.017	-0.127	-0.053	1	
NSAID≥90 days	0.011	-0.191	0.062	-0.075	0.041	0.033	-0.118	-0.041	0.004	-0.075	0.198	0.108	-0.147	-0.020	-0.107	1

Correlation coefficient greater than 0.7 or smaller than -0.7 are type in italics with underline.

BMI, body mass index; SBP, systolic blood pressure; eGFR, estimated glomerular filtration rate; HbA1c, hemoglobin A1c; UACR, urinary albumin to creatinine ratio; LDL, low-density lipoprotein cholesterol;

III.3 Logarithm transformation

Generally, continuous data with skewed distribution can produce errors and violate to model assumption. Logarithmic transformation is applied to transform a highly skewed variable to be a less skewed variable as possible in order to increase validity of the associated statistical analyses(86).

In this study, the skewness and kurtosis test for normality based on D'Agostino's X^2 test, the histogram and the distributional diagnostic plots were applied for the skewness testing (87). Eleven continuous variables, including age, BMI, SBP, HbA1c, TG, LDL, HDL, uric acid, serum albumin, eGFR and UACR.

The skewness and kurtosis test showed that age had no skewness and normal distribution (*p-value = 0.395*) **(Figure 6)**. Ten continuous candidate predictors, including BMI, SBP, HbA1c, TG, LDL, HDL, uric acid, serum albumin, eGFR and UACR, had skewed distribution. Only age had normal distribution. All ten skewed continuous candidate predictors were transformed by logarithm. We checked the distributions of each logarithm transformed candidate predictor by histogram and the distributional diagnosis plot again.

However, it was found that the distribution of logarithm transformed eGFR had more skewed than eGFR without logarithm transformation **(Figure 7 and 8)** (other distributional diagnostic plots of other variables were omitted). Therefore, eGFR and age were not transformed by logarithm.

Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	joint adj chi2(2)	joint Prob>chi2
age_int	2,178	0.1727	0.9902	1.86	0.3948
BMI	2,127	0.0000	0.0000	.	0.0000
SBP	2,174	0.0000	0.0000	.	0.0000
HbA1c	1,973	0.0000	0.0000	.	0.0000
TG	2,058	0.0000	0.0000	.	.
LDL	2,011	0.0000	0.0000	.	0.0000
HDL	1,953	0.0000	0.0000	.	0.0000
uric_acid	878	0.0001	0.3173	15.07	0.0005
serumAlb	665	0.0000	0.0000	.	0.0000
base_eGFR	2,178	0.0755	0.0001	17.97	0.0001
UrineAlbumin	1,486	0.0000	0.0000	.	.

Figure 6: Skewness-Kurtosis tests for normaliy of CKD candidate predictors

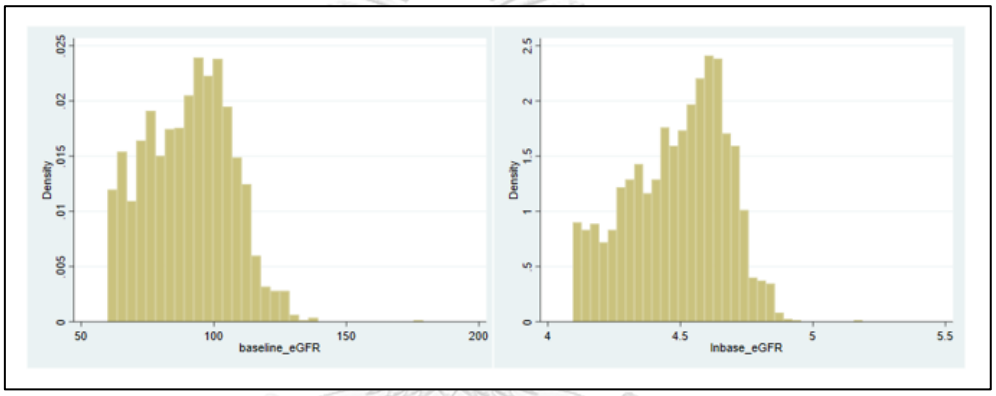


Figure 7: The histograms of eGFR (left) and logarithm transformation of eGFR (right)

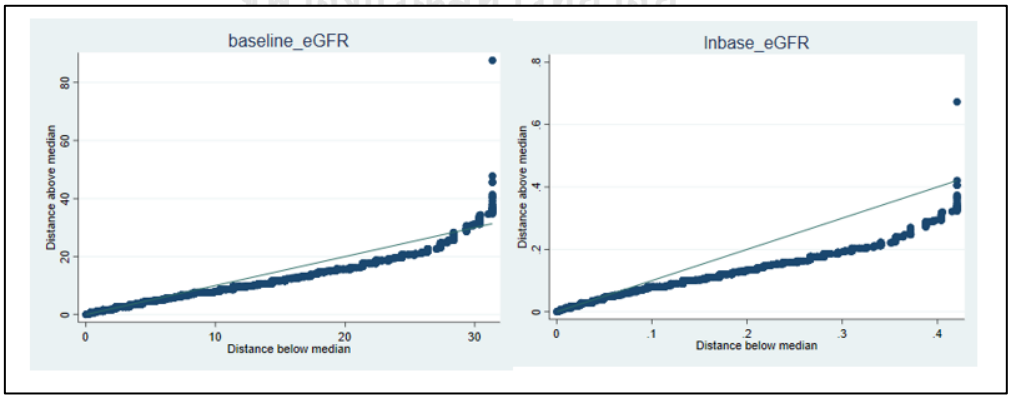


Figure 8: The distributional diagnosis plots of eGFR (left) and logarithm transformed eGFR (right)

III.4 Missing data

Missing data is a common problem in data collecting. In this study, percentage of smoking status, log transformed (log) serum albumin and log uric acid were 76.72, 69.47 and 59.69, respectively. Candidate predictors with missing data more than 50% might lead bias in model developing. Therefore, these three candidate predictors are excluded. Finally, thirteen candidate predictors (including age, sex, eGFR, log BMI, log SBP, log Hba1c, log TG, log LDL, log HDL, log UACR, history of DR, history of CVD and days of NSAIDs exposure) were left to develop for a CKD prediction model. Eight eligible variables (% missing data) include log UACR (30.89%), log HDL (9.64%), log HbA1c (8.52%), log LDL (7.28%), log TG (5.25%), days of NSAIDs exposure (3.02%), log BMI (2.69%) and log SBP (0.26%) **(Table 6)**. Because the highest percentage of missing data was 30.89%, so we imputed the missing data by mean of Multiple Imputation by Chain Equation (MICE) with 30 imputations(109). Linear regression model was obtained for continuous variables. The independent variables were CKD event, time of follow-up, age, sex, eGFR, history of diabetic retinopathy, history of cardiovascular disease. Moreover, we used history of hypertension, serum creatinine as auxiliary variables. Auxiliary variable is a variable that has relationship to a missing variable but auxiliary variable is not the interested variable for developing model. According to categorized NSAIDs exposure, we imputed the days of NSAIDs taking which was continuous variable, then we categorized into 3 groups; NSAIDs taking $\geq$ 90 days, NSAIDs taking <90 days and non-NSAIDs exposure.

In this study, split method was applied for internal validation. Eligible subjects were randomly split into 2 datasets; training and validation datasets. Therefore, MICE must be conducted in both datasets separately. Split method was mentioned in detail again in section of **EXPERIMENTAL PROCEDURE.**

Table 6: Percent of missing data of CKD candidate predictors in the training and validation datasets

Candidate predictors	Training dataset (N=1,525)		Validation dataset (N=653)	
	Missing (n)	Percent	Missing (n)	Percent
Smoking status	1,191	78.10	480	73.51
Log serum albumin	1,066	69.90	447	68.45
Log uric acid	881	57.77	419	64.17
Log UACR	471	30.89	221	33.84
Log HDL	147	9.64	78	11.94
Log HbA1c	130	8.52	75	11.49
Log LDL	111	7.28	56	8.58
Log TG	80	5.25	40	6.13
Days of NSAIDs taking	46	3.02	25	3.83
Log BMI	41	2.69	10	1.53
Log SBP	4	0.26	0	0

UACR, albumin to creatine ratio; HDL, high-density lipoprotein cholesterol; HbA1c, hemoglobin A1c; LDL, low density lipoprotein cholesterol; TG, triglyceride; NSAIDs, nonsteroidal anti-inflammatory drugs; BMI, body mass index; SBP, systolic blood pressure; log, logarithm transformation.

The imputed values distribution and patterns were obtained to check for the imputation integrity. The results showed similar distributions in terms of mean and standard deviation (SD) between the observed and imputed data in the training and validation datasets (**Table 7 and 8).** It indicates that model the unbiased imputed data of each dataset can be applied to generate for a robust model.

Table 7: The summary statistics of the imupted and observed data for seven imcomplete candidate predictors in the training dataset

Candidate predictors	Observed data			Imputed data		
	N	mean	SD	N	mean	SD
Log BMI	1,484	3.26	0.19	41	3.23	0.2
Log SBP	1,521	4.91	0.15	4	4.88	0.13
Log HbA1c	1,395	1.96	0.19	130	1.95	1.95
Log HDL	1,378	2.93	0.22	147	3.94	0.21
Log TG	1,445	4.95	0.47	80	4.89	0.49
Log LDL	1,414	4.69	0.28	111	4.69	0.28
Log UACR	1,054	131.63	532.2	471	112.37	541.22
Days of NSAIDs taking	1,479	5.19	14.6	46	4.20	14.6

N, number; SD, standard deviation; min, minimum; max, maximum; BMI, body mass index; SBP, systolic blood pressure; HbA1c, hemoglobin A1c, HDL, high-density lipoprotein cholesterol, TG, triglyceride; LDL, low-density lipoprotein cholesterol, UACR, urinary albumin to creatinine ratio; NSAIDs, nonsteroidal anti-inflammatory drugs; log, logarithm transformation

Table 8: The summary statistics of the imupted and observed data for seven imcomplete candidate predictors in the validation dataset

Candidate predictors	Observed data			Imputed data		
	N	mean	SD	N	mean	SD
Log BMI	642	3.26	0.19	10	3.21	0.19
Log HbA1c	578	1.96	0.20	75	1.96	0.2
Log HDL	575	3.94	0.21	78	3.93	0.22
Log TG	613	4.94	0.48	40	4.85	0.49
Log LDL	597	4.69	0.27	56	4.71	0.27
Log UACR	432	3.02	0.69	221	2.95	1.59
Days of NSAIDs taking	628	5.56	16.01	30	5.61	16.6

N, number; SD, standard deviation; min, minimum; max, maximum; BMI, body mass index; SBP, systolic blood pressure; HbA1c, hemoglobin A1c, HDL, high-density lipoprotein cholesterol, TG, triglyceride; LDL, low-density lipoprotein cholesterol, UACR, urinary albumin to creatinine ratio; NSAIDs, nonsteroidal anti-inflammatory drugs; log, logarithm transformation

IV. EXPERIMENTAL PROCEDURE

This experimental procedure was set up to develop and validate for CKD prediction model in type 2 diabetes. The processes of model development included univariate analysis, multivariate analysis, model simplifying and creating CKD risk equation. Model development was operated in the training dataset (70% of sample sizes). Model validation evaluated by using the validation dataset (30% of sample sizes) was consisted of discrimination and calibration

estimations. **Figure 9** showed flow chart of the process of CKD prediction model development and validation with the training and validation datasets.

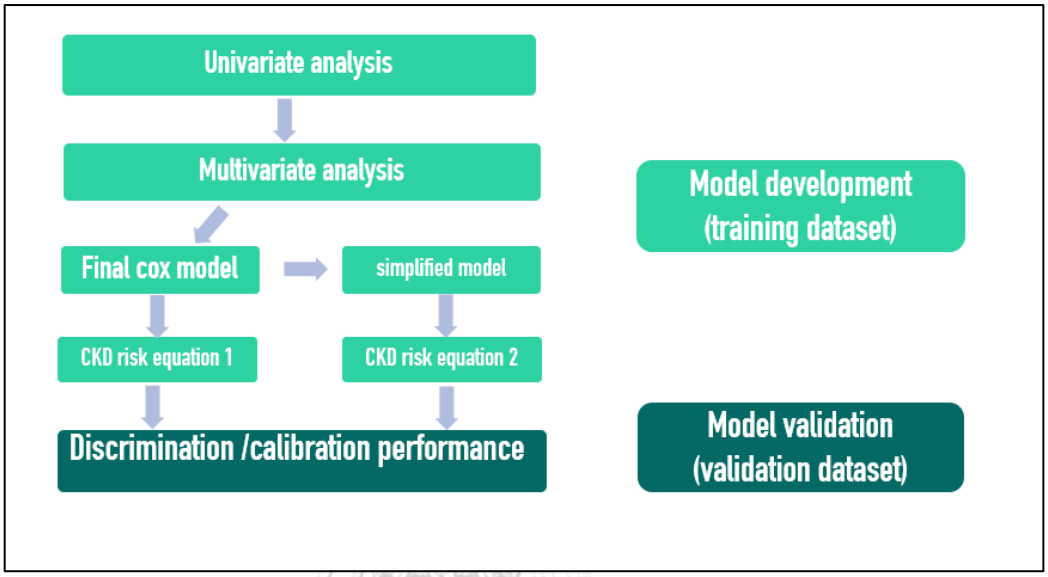


Figure 9: The summary of CKD prediction model development and validation

IV.1. Model development

This part is related with univariate analysis, multivariate analysis, model refining and CKD risk equation generating. This step used 1,525 subjects in the training dataset. **Figure 10** summarized processing of model development.

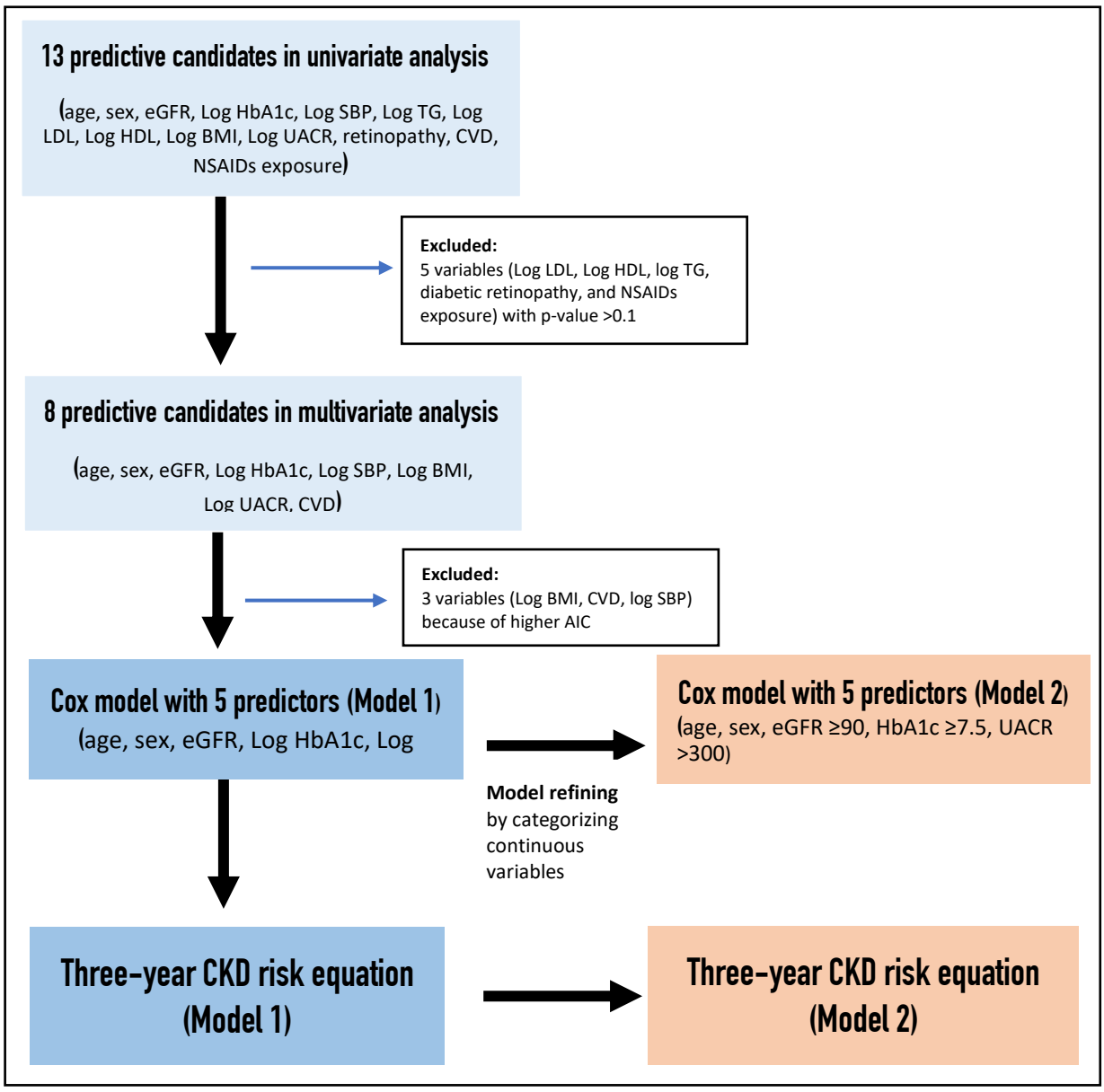


Figure 10: Summarized processing of model developing.

VI.1.1 Univariate analysis

Thirteen predictor candidates, including age, sex, eGFR, logarithm transformed (log) HbA1c, log SBP, log TG, log LDL, log HDL, log BMI, log UACR, history of DR, history of

CVD, NSAIDs exposure, were recruited for the univariate analysis by using Cox-proportional hazard regression. Predictors candidates with p-value less than 0.1 were considered to be included in the multivariate analysis.

VI.1.2 Multivariate analysis

For the multivariable analysis, eligible predictor candidates (including age, sex, eGFR, log HbA1c, log UACR, log BMI, log SPB and history of CVD) were estimated by backward elimination approach for building a model. For improved performances of model selections in multivariate Cox-proportional hazard regression, models were estimated by using the goodness-of-fit of the Akaike Information Criterion (AIC)(88). A model which had lowest of AIC values was chosen to be a CKD prediction model.

Final cox model which had final predictors was defined as **model 1 (laboratory model)**; **Model 1** was further refined to a simpler model as **model 2 (simplified model)** which could facilitate it's use in routine heath practice. The proportion hazard assumption was assessed by testing whether the log hazard ratio function of our Cox model was constant over the time (114). The steps of multivariate analysis and assumption of CKD model were presented in **Appendix A.**

VI.1.3 CKD risk equation generating

After obtaining the cox model, CKD risk models were generated to estimate 3-year risk of CKD based on final Cox model(89). The individual CKD risk is estimates as follow:

$$\text{Predicted CKD risks} = 1 - S_0\,(t)^{\exp\,(\sum \beta_i X_i - \sum \beta_i \bar{X}_i)}$$

S_0(t) is baseline survival function at time t (e.g. t =3 years), X_i is the individual's values of the predictors obtained from the Cox model, β_i is the regression coefficient foX_ir, and the $\bar{X}_i$ is the mean of X_i.

The advantage of using this method for generating CKD risk equation is that this method can be easily recalibrated to other cohorts. A second cohort can recalibrate our three-year risk CKD equation by replacing mean values of our predictors in the equation with their own mean values, and replacing our three-year CKD-free survival probability with an estimate from follow-up of their cohort.

In this study, we used three-years CKD-free survival probability of 0.7307. Finally, the three-year CKD risk models are generated for **model 1 (laboratory model)** and **model 2 (simplified model)** in these followings:

Model 1 (laboratory model):

Predicted 3-year CKD risks

$$= 1 - 0.7307^{\exp\,[(\beta_1 Age + \beta_2 Male + \beta_3 GFR + \beta_4 \log(HbA1c) + \beta_5 \log(ACR)] - [\beta_1 mean(Age) + \beta_2 proportion(Male) + \beta_3 mean(GFR) + \beta_4 (\log(mean(HbA1c)) + \beta_5 (\log\,(mean(ACR))}$$

Model 2 (simplified model):

Predicted 3-year CKD risks

$$= 1 - 0.7307^{\exp\left[(\beta_6(if\ age \geq 50) + \beta_7 Male + \beta_8(if\ GFR \geq 90) + \beta_9(if\ HbA1c > 7.5) + \beta_{10}(if\ ACR > 300)\right] - \left[\beta_6 proportion(if\ age \geq 50) + \beta_7 proportion(Male) + \beta_8 proportion(if\ GFR \geq 90) + \beta_9 proportion(if\ HbA1c \geq 7.5) + \beta_{10} proportion(if\ ACR > 300)\right]}$$

Appendix B showed the three -year risk of CKD equation for **model 1 and model 2.**

VI.2 Model validation

The developed three-year risk of CKD equation **(model 1 and model 2)** were evaluated for the performances by means of discrimination and calibration. The validation dataset was used to estimate both models' performances.

VI.2.1 Discrimination

Discrimination is ability of distinguish CKD risk patients from non-CKD risk patients. The C-statistic has ranges from 0.5 (poor predictive ability) to 1 (perfect predictive ability) (115). In this step, the estimation of discrimination with 3-year CKD risk model had to calculate for Harrell's C by using Somers' D method(116) **(Appendix C)**.

VI.2.2 Calibration

The modified Hosmer-Lemeshow X^2 test and survival probability curves were applied to evaluate for the performance of two 3-year risk of CKD equations. The modified Hosmer-Lemeshow X^2 test was applied with decile of prognosis risks. Non-significance should be

obtained (115, 117). Moreover, The survival probability curves were comparing between the predicted survival probabilities and the observed survival probabilities (Kaplan-Meier curves) based on three risk groups (low, moderate and high risks of CKD) (96).

VI.2.2.1 The modified Hosmer-Lemeshow X^2 test is comparing between the predicted CKD probabilities and the observed events of CKD. The comparisons were classified into deciles based on the predicted CKD risk in decile risk groups. The observed events of CKD were obtained from the Kaplan-Meier estimators. In contrast, the predicted CKD probabilities were obtained from the cox model. (118). The calculations were presented in **Appendix D**.

VI.2.2.2 Calibration of survival probabilities by Kaplan-Meier method

The principle of this calibration method was adapted from the Kaplan-Meier survival estimations during the time of following-up. Graphs were the comparisons between the observed and the predicted survival probability curves. The curves of observed (Kaplan-Meier) and predicted probabilities (predicted from the fitted Cox model) were estimated individually at time 0, 1, 2, …, *t* years of the following-up (in this study used three-year follow-up) by CKD risk groups. The CKD risk groups estimated by prognostic indexes (PI) were divided into 3 groups; low, moderate and high CKD risks.

Assessing calibration of the survival probabilities by Cox model is following this formula:

$$S(t;x) = S_0\,(t)^{\exp\,(x\beta)}$$

Where $S_0(t)$ is baseline survival function, xβ is cumulative of multiplying between predictors and coefficient ($x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k$) or prognostic index (PI). $S_0(t)$ is calculated from baseline hazard function following this formula:

$$S_0(t) = \exp\{-H_0(t)\}$$

Where $H_0(t)$ is baseline cumulative hazard function from time 0 – *t*. $H_0(t)$ can be retrieved from the fitted Cox model.

And according to specificity of Cox model in both training and validation datasets, similar baseline survival functions in both datasets should be obtained (99).

Steps of predicting survival probabilities using example of prediction model 1

The CKD prediction model **(model 1)** contained predictors, including age, gender, eGFR, log UACR and log HbA1c. The outcome was time to CKD presenting. The examples were presented in the training and validation datasets(99).

Step 1: Calculate prognostic index (xβ) from the prediction model.

Calculate prognostic index (PI) from the prediction equation in the training dataset. After results of PI, individual predicted survival function is obtained.

PI = β_1Age + β_2Male + β_3eGFR+β_4(log UACR) +β_5(log HbA1c)

Where β_i is a pooled coefficient of each predictor from 30 imputed prediction equations. However, 3-year risk of CKD were generated, so we had calculated PI based on risk score of each model.

PI = ((1.180327*log(HbA1c)) + (0.2435595*log(UACR)) + (-0.1027086*(eGFR)) + (-0.2942762*(if male)) + (0.0221663*age) - ((1.180327* log(7.214512)) + (0.2435595* log(123.7432)) + (-0.1027086*90.46844) + (-0.2942762*0.4255738) + (0.0221663*55.87869))

Center the calculated PI on the training dataset mean.

PI are divided by tercile into 3 risk groups (low, moderate and high CKD risk groups).

PI are divided by tercile into 3 risk groups (low, moderate and high CKD risk groups).

Step 2: Calculated individual predicted survival probabilities

Calculate individual predicted survival probabilities at the censoring time or observed events (3-yeat follow-up) with this following formula:

$$S(t;x) = S_0\,(t)^{\exp\,(PI)}$$

After getting individual predicted survival probabilities, average of predicted survival function was calculated for each risk group. As result, predicted survival curves of 3 CKD risk groups were obtained.

Step 3: The observed survival probabilities was estimated from actual results at the censor time or observed events (3-year follow-up). As a result, the observed survival probability curves (or Kaplan-Meier curves) were obtained.

Step 4: The predicted survival probability curves and the observed survival probability curves were plotted to compare the trends. Moreover, 95% confidence interval can be presented for the observed probabilities survival to clarify how well the predicted survival curve is closed to

observed probabilities survival. If the predicted survival probabilities curve is closed to the observed survival probabilities curve, it was indicated that the prediction model had a good accuracy of CKD prediction or good calibration.

The predicted survival probability curves and the observed survival probability curves were plotted to compare the trends. Moreover, 95% confidence interval can be presented for the observed probabilities survival to clarify how well the predicted survival curve is closed to observed probabilities survival. If the predicted survival probabilities curve is closed to the observed survival probabilities curve, it was indicated that the prediction model had a good accuracy of CKD prediction or good calibration.

V. STATISTICAL ANALYSIS

The statistical analysis was used in many steps of the study's methodology, including baseline characteristic and prediction model developing.

The baseline characteristic comparing the training dataset and validation dataset were shown in **RESULT section**. In comparisons of baseline characteristics between training and validation datasets we presented continuous data as means (SD) or median (IQR), and formal comparisons were made with student-t test or Mann-Whitney Wilcoxon test as appropriate; categorical data were presented as n (%) and formal comparisons made using a chi-square test.

For model development, Cox-proportional hazard regression was applied in both univariate and multivariate analysis. P-value less than 0.05 was determined as statistically significance.

VI. SOFTWARE

All data pre-processings, including handling predictors-coding, logarithm transformation, missing data, model developing, and model validation were conducted using STATA version 16 (Statacorp, College Station, TX, USA).

VII. ETHICAL CONSIDERATION

This study had been approved by two Institutional Review Boards (IRB), including The Research Ethics Review Committee for Research Involving Human Research Participants, Health Sciences Group, Chulalongkorn University, and the IRB of the Medical Service Department, Bangkok.

CHAPTER IV

RESULTS

This section provides the results of the study, including baseline characteristic, prediction model developing and model validation. In part of the prediction model development have three parts, including univariate analysis, multivariate analysis and CKD risk equation generating. The topics are these followings:

I. Baseline characteristics

II. Model development

1. Univariate analysis

2. Multivariate analysis

3. Three-year CKD risk equation generating

III. Model performance

It starts with the results of the baseline characteristic comparing between the training dataset and the validation dataset.

I. BASELINE CHARACTERISTICS

Of the remaining eligible 2,178 patients with type 2 DM, the mean (SD) age was 55.74 (11.19). 1,525 and 653 patients were assigned to the training and validation datasets, respectively. The characteristics of CKD risk covariates of DM patients comparing between the training and validation datasets were shown in **Table 9.** The median of the follow-up time in the training and validation dataset were 1.29 (interquartile range, [IQR] 0.50-2.50) years and 1.20 (IQR, 0.49-2.41) years. The mean age was 55.74 years (range 18-90 years), 56.70% were female, 17.91% had cardiovascular disease, 3.08% had diabetic retinopathy. Mean BMI were 26.45 kg/m^2, mean HbA1c were, 7.2%. Moreover, patients had mean eGFR and urinary ACR of 90.38% and 16.15 mg/g, respectively. Most of patients (75.57%) had no NSAIDs taking from the hospital. The proportion of CKD events in the training and validation datasets were 18.16% and 16.54%, respectively. The similarities of characteristics of risk covariates were obtained in both datasets.

Table 9: Baseline characteristics of CKD risk covariate in the training and the validation datasets

Variables	Overall (*n* = 2,178)	Training dataset (*n* =1,525)	Validation dataset (*n* = 653)	*P-value*
Socio-demographic factors				
Age (Mean, SD), years	55.74(11.19)	55.88(11.30)	55.42(10.92)	0.*81*
Age group, n (%)				*0.31*
<40	159 (7.30)	117 (7.67)	42 (6.43)	
40-59	1,212 (55.65)	838 (54.95)	374 (57.27)	
60-69	560 (25.71)	387 (25.38)	173 (26.49)	
≥70	247 (11.34)	183 (12.00)	64 (9.80)	
Sex, n (%)				0.29
- Male	943 (43.30)	649(42.56)	294(45.02)	
- Female	1,235(56.70)	876(57.44)	359(54.98)	
BMI (Mean, SD), kg/m^2	26.45 (5.20)	26.45 (5.18)	26.46 (5.24)	*0.79*
BMI level, n (%)				*0.60*
< 23	548 (25.16)	374(24.52)	174(26.65)	
23.0-24.9	396 (18.18)	282 (18.49)	114(17.46)	
25.0-29.9	779 (35.77)	555(36.39)	224 (34.30)	
≥30	455 (20.89)	314(20.59)	141 (21.59)	
Diabetes-related factor and biomarkers				
SBP (Mean, SD), mmHg	136.23 (20.27)	136.50 (20.39)	135.59 (19.99)	*0.46*
Uncontrolled SBP [a], n (%)	827 (37.97)	577 (37.84)	250 (38.28)	*0.84*
HbA1c (Mean, SD), %	7.20 (1.57)	7.21 (1.53)	7.19 (1.65)	*0.34*
HbA1c level, n (%)				*0.60*
<7	1,145 (52.57)	801 (52.52)	344 (52.68)	
7.0-8.0	520 (23.88)	357 (23.41)	163 (24.96)	
>8	513 (23.55)	367 (24.07)	146 (22.36)	

Variables	Overall (*n* = 2,178)	Training dataset (*n* =1,525)	Validation dataset (*n* = 653)	*P-value*
LDL (Mean, SD), mg/dL	113.20 (31.64)	113.43 (32.57)	112.65(29.38)	*0.99*
Uncontrolled LDL [b], n (%)	1,352 (62.08)	977 (64.07)	375 (57.43)	*0.003*
HDL (Mean, SD), mg/dL	52.32(11.58)	52.24(11.65)	52.50(11.43)	*0.53*
Uncontrolled HDL[c], n (%)	567 (26.03)	415 (27.21)	152 (23.28)	*0.06*
TG (Mean, SD), mg/dL	158.59(117.35)	159.28(91.69)	157.01(162.26)	*0.19*
Uncontrolled TG[d], n (%)	928 (43.41)	669 (43.87)	259 (42.25)	*0.50*
UACR [Median, IQR], mg/g	16.15 [5.7-119]	16 [5.6-113.4]	16.8 [5.95-139.2]	*0.45*
Albuminuria level, n (%)				*0.39*
<30	1,350 (62.03)	946 (62.03)	405 (62.02)	
30-300	460 (21.12)	331 (21.70)	129 (19.75)	
>300	367 (16.85)	248 (16.26)	119 (18.22)	
eGFR (mean, SD), mL/min per 1.73m^2	90.38 (16.18)	90.47 (16.22)	90.17 (16.08)	*0.66*
Kidney stage, n (%)				*0.98*
≥90 (stage 1)	1,018 (46.74)	713 (46.75)	305 (46.71)	
<90 (stage 2)	1,160 (53.26)	812(53.25)	348 (53.29)	
Comorbidities				
CVD [e], n (%)	390 (17.91)	276(18.10)	114(17.46)	*0.72*
DR, n (%)	67(3.08)	45 (2.95)	22(3.37)	*0.61*
Medication				
NSAIDs taking, n (%)				*0.60*
No NSAIDs taking	1,671 (76.72)	1,179 (77.31)	492 (75.34)	
-1-89 days	492 (22.59)	336 (22.03)	156 (23.89)	
-≥90 days	15 (0.69)	10 (0.66)	5 (0.77)	

Variables	Overall (*n* = 2,178)	Training dataset (*n* =1,525)	Validation dataset (*n* = 653)	*P-value*
Follow-up time [median, IQR],	1.25	1.29	1.20	
years	[0.49-2.47]	[0.50-2.50]	[0.49-2.41]	*0.43*
CKD events, n (%)	385(17.68)	277(18.16)	108(16.54)	*0.36*

BMI, body mass index; SBP, systolic blood pressure; HbA1c, hemoglobin A1c; LDL, low-density lipoprotein cholesterol; HDL, high-density lipoprotein cholesterol; TG, triglyceride; UACR, urinary albumin to creatinine ratio; eGFR, estimated glomerular filtration rate; CVD, cardiovascular disease; DR, diabetic retinopathy; NSAIDs, nonsteroidal anti-inflammatory drugs; CKD, chronic kidney disease; SD, standard deviation; IQR, interquartile range.

[a] SBP $\geq$140 mmHg

[b] LDL cholesterol < 100 mg/dL

[c] HDL cholesterol $\geq$ 40 mg/dL if male, and HDL $\geq$ 50 mg/dL if female

[d] TG < 150 mg/dL

[e] Cardiovascular disease was defined as myocardial infraction, stroke, and angina pectoris

II. MODEL DEVELOPMENT

In this subsection, the results of the training datasets were separated into 3 parts: univariate analysis, multivariate analysis, and CKD risk equation generating.

II.1. Univariate analysis

The training dataset was obtained to develop CKD model. Thirteen potential prognostic variables (including age, sex, eGFR, log HbA1c, log LDL, log TG, log HDL, log UACR, log SBP, log BMI, history of CVD, history of DR and NSAIDs exposure) were assessed in univariate analysis. Log LDL, log HDL, log TG, DR and NSAIDs taking did not meet criteria for inclusion

in multivariate analysis. **Table 10** shows the hazard ratio (HR) of each candidate predictor in univariate analysis.

Table 10: Univariate Cox regression for candidate predictors

Candidate predictors	**HR**	**(**95%CI**)**	***p-value***	β
Sex (male vs. female)	0.69	0.54, 0.89)	*0.004*	-0.36
Age (years)	1.07	(1.06, 1.08)	*<0.001*	0.07
eGFR (mL/min/1.73m^2)	0.90	(0.89,0.91)	*<0.001*	-0.10
Log UACR (mg/g)	1.28	(1.19, 1.38)	<0.001	0.25
Log HbA1c (%)	1.89	(1.02, 3.50)	*0.04*	0.64
Log TG (mg/dl)	0.82	(0.63, 1.06)	*0.12*	-0.20
Log LDL (mg/dl)	0.82	(0.53, 1.26)	*0.36*	-0.20
Log HDL (mg/dl)	1.33	(0.75, 0.36)	*0.32*	0.29
Log BMI (kg/m^2)	0.45	(0.24, 0.87)	*0.02*	-0.79
Log SBP (mmHg)	3.24	(1.48, 7.12)	*0.003*	1.18
Diabetic retinopathy (yes vs.no)	1.53	(0.86, 2.72)	*0.15*	0.42
CVD (yes vs. no)	1.66	(1.27, 2.18)	*<0.001*	0.51
NSAIDs taking - $\geq$ 90 days (yes vs. non-NSAIDs)	0.90	(0.67, 1.21)	*0.49*	-0.10
- 1-89 days (yes vs. non-NSAIDs)	0.83	(0.20, 3.33)	*0.79*	-0.19

HR, hazard ratio; CI, confidence interval; β, ln(HR); HbA1c, hemoglobin A1c; eGFR, estimated glomerular filtration rate; UACR, urinary albumin to creatinine ratio; LDL, low density lipoprotein cholesterol; HDL, high density lipoprotein cholesterol; TG, triglyceride; BMI, body mass index; SBP, systolic blood pressure; CVD, cardiovascular disease NSAIDs, nonsteroidal anti-inflammatory drugs; log, logarithm transformation.

II.2 Multivariate analysis

In multivariate analysis history of CVD, history of DR and log SBP were excluded regarding greater AIC. As a result, five final predictors were selected by backward elimination: age, sex, log HbA1c, log UACR, and eGFR. The steps of multivariate analysis were presented in **Appendix A**. The final multivariate cox model was defined as **model 1.** We further refined **model 1** to be a simpler version (**model 2)** which used clinically relevant dichotomous groupings for the continuous covariates in **model 1**. HR (95%CI) from these multivariate models are presented in **Table 11.** The proportional hazard assumption was met for both models.

We established three-CKD risk equations based on each final cox model **(model 1 and model 2)** by using coefficients of each predictor. Three-year risk of CKD equation is preferable for our final CKD prediction equations shown in **Appendix B.**

Table 11*:* Multivariate Cox regression results for models 1 and 2

Predictors	CKD models in the training dataset (n= 1,525)					
	Model 1 (laboratory model)			Model 2 (simplified model)		
	HR (95%CI)	*p-value*	β	HR (95%CI)	*p-value*	β
Age (year)	1.02 (1.01, 1.04)	*0.001*	0.02			
Male sex vs female	0.75 (0.58, 0.96)	*0.025*	-0.29	0.70 (0.54,0.90)	*0.005*	-0.36
eGFR (mL/min/1.73m^2 increase)	0.90 (0.89, 0.91)	*<0.001*	-0.10			
Log HbA1c (%)	3.25 (1.65, 6.40)	*0.001*	1.18			
Log UACR (mg/g)	1.27 (1.18, 1.38)	*<0.001*	0.24			
Age $\geq$ 50 years				2.13 (1.38,3.31)	*0.001*	0.76
eGFR $\geq$ 90 (mL/min/1.73m^2)				0.09 (0.06,0.13)	*<0.001*	-2.42
HbA1c > 7.5%				1.38 (1.05,1.80)	*0.018*	0.32
UACR >300 mg/g				2.25 (1.58,3.19)	*<0.001*	0.81

HR, hazard ratio; CI, confidence interval; β, ln(HR); HbA1c, hemoglobin A1c; eGFR, estimated glomerular filtration rate; UACR, urinary albumin to creatinine ratio; log, logarithm transformation.

III. MODEL PERFORMANCE

In this part both CKD prediction models **(model 1 and model 2)** were evaluated for the discrimination and calibration performances in the validation dataset. C-statistic were applied for discrimination estimating. And the modified Hosmer-Lemeshow X^2 test and survival probabilities curve were methods for calibration.

III.1 Discrimination

Both **model 1** and **model 2** showed good discriminative performance in the training dataset, with C-statistics of 0.873 (95%CI 0.856-0.892) and 0.798 (95%CI 0.774-0.823), respectively. The C-statistics of **model 1 and model 2** in the validation dataset showed similarly high values of 0.890 (95%CI 0.870-0.911) and 0.812 (95%CI 0.781-0.842), respectively, indicating good performance of both models in distinguishing patients with type 2 DM who developed CKD from those who did not

III.2 Calibration performance

III.2.1 The modified Hosmer-Lemeshow X^2 test

Both model 1 and model 2 showed adequate calibration in the training and validation datasets comparing the observed versus events by decile of risk, with no significant difference in the Hosmer and Lemeshow X^2 $\mathrm{P-values}$. Expected and observed events, and goodness of fit statistics from **model 1** and **model 2** in the training and validation dataset are presented in **Table 12 and 13. Model 1 and model 2** and summaries are shown in **Table 14.**

Table 12: Hosmer-Lemeshow X^2 table with decile of risk groups in training and validation datasets (Model 1)

Risk group	Training dataset			Validation dataset		
	N	Predicted event	Observed event	N	Predicted event	Observed event
1	152	1.56	0	65	0.64	2
2	153	4.11	7	65	1.79	0
3	152	7.11	4	65	3.35	0
4	153	12.35	8	66	5.72	4
5	152	20.05	15	65	8.59	5
6	153	33.12	21	65	14.01	1
7	152	55.12	20	66	22.99	25
8	153	88.55	66	65	35.92	28
9	152	123.43	97	65	51.99	39
10	153	149.50	139	66	64.50	58
total	1525	494.91	378	653	209.50	163
X^2		15.15			15.82	
p-value		*0.09*			*0.11*	

Table 13: Hosmer-Lemeshow X^2 table with decile of risk groups in training and validation datasets (model 2)

Risk group		Training dataset*			Validation dataset	
	N	Predicted event	Observed event	N	Predicted event	Observed event
1	147	6.19	1	61	2.58	2
2	149	8.93	4	62	3.65	0
3	151	12.82	11	60	5.05	5
4	307	37.28	18	73	7.95	0
5	127	43.06	27	66	9.49	2
6	161	96.20	57	59	20.07	19
7	61	43.54	28	74	44.14	39
8	217	158.01	112	35	24.39	19
9	205	183.31	152	85	61.89	37
10				78	69.58	50
total	1525	589.34	410	653	248.78	174
X^2		8.43			13.87	
p-value		*0.39*			*0.13*	

*Eight degree of freedom

Table 14: Model 1 and model 2 performances summaries

Method	Training dataset		Validation dataset	
	Model 1	**Model 2**	**Model 1**	**Model 2**
Discrimination				
C-statistic (95%CI)	0.873 (0.856-0.892)	0.798 (0.774-0.823)	0.890 (0.870-0.911)	0.812 (0.781-0.842)
Calibration				
Hosmer-Lemeshow X^2 (*p-value*)	15.15 (*0.09*)	8.43 (*0.39*)	15.82 (*0.11*)	13.87 (*0.13*)

III.2.2 Survival probability curves by Kaplan-Meier method

We also estimated the performance of calibration with Kaplan-Meier method. Survival probability curves of 3-year CKD risk based on three CKD risk groups (low, moderate and high risk), were compared by plotting the predicted CKD survival probabilities and observed CKD events. The predicted CKD probabilities curves fell within the 95% CI for observed CKD probabilities in both **model 1 and model 2 (Figure 11 and 12).**

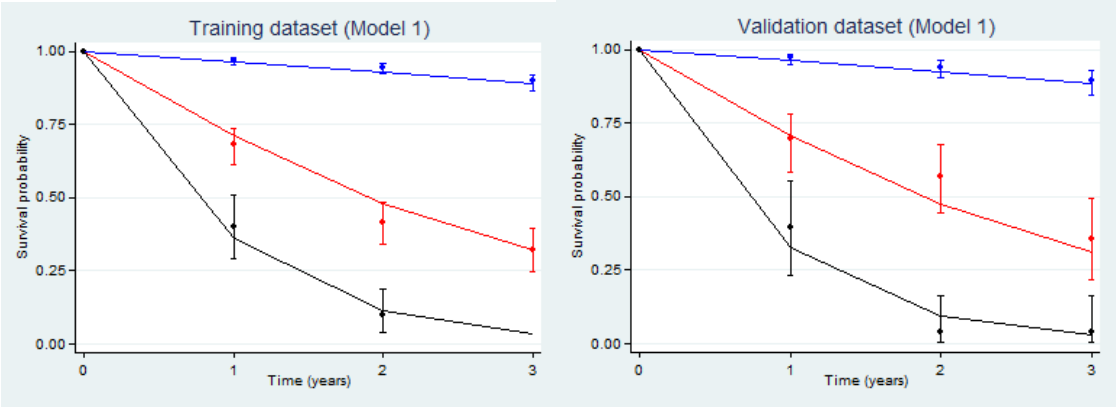


Figure 11:The observed CKD survival probabilities vs. predicted CKD probabilities for the Cox model in the training and the validation datasets for model 1.

Predicted survival probabilities are smooth lines, and the observed CKD survival probabilities from the Kaplan-Meier method with 95% confidence intervals are represented by vertical capped lines. Three prognosis groups plotted represent low risk (blue lines), moderate risk (red lines), and high risk (black lines) groups.

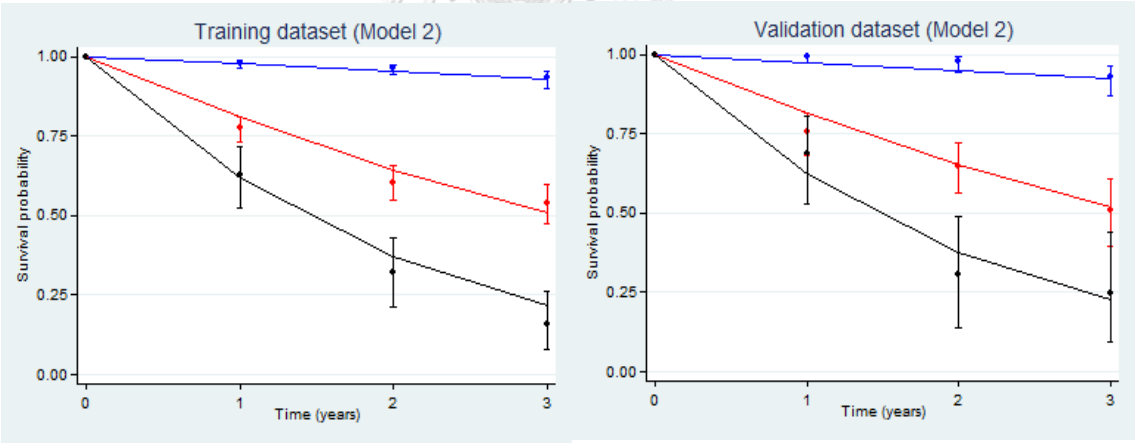


Figure 12: The observed CKD survival probabilities vs. predicted CKD probabilities for the Cox model in the training and the validation datasets for model 2.

Predicted survival probabilities are smooth lines, and the observed CKD survival probabilities from the Kaplan-Meier method with 95% confidence intervals are represented by vertical capped lines. Three prognosis groups plotted represent low risk (blue lines), moderate risk (red lines), and high risk (black lines) groups.

CHAPTER V

DISUSSION

This section discusses about the findings of this study. The topics are these following:

I. Study population

II. CKD prediction model development and performance

III. CKD predictors

IV. Comparing model performance with other CKD prediction models

V. Two model establishing

VI. Using the prediction models in practice

VII. Strengths and limitations

I. STUDY POPULATION

The prevalence of CKD stage 3 among patients with type 2 DM in this retrospective cohort study was 17.68% which was not in range comparing the previous Thai's studies (25.38-48%). According to the exclusion criteria of this study for developing model, all patients with type 2 DM from Taksin hospital were not recruited leading less CKD prevalence than other studies(10-13).

According to baseline characteristic of our cohort, the median of follow-up time was 1.29 year which is very short time. Because of Thailand's problematic health referral system, primary health care unit usually refer DM patients to tertiary hospital (such as Taksin hospital) with severe stage of CKD so most of referred DM patients to Taksin hospital have level of eGFR closed to 60 mL/min/1.73m^2. The evidence was confirmed that most patients (53.26%) had baseline of eGFR less than 90 mL/min/1.73m^2. Most patients had controlled HbA1c (52.57%), non-microalbuminuria (62.03%) and controlled SBP (62.03%) but DM patients also had high proportion of obesity (35.77%), uncontrolled LDL (62.08%) and uncontrolled TG (43.41%). In our knowledge, these uncontrolled LDL, TG and obesity have association to rapid eGFR decline leading CKD stage 3 incident quickly.

II. **CKD PREDICTION MODEL DEVELOPMENT AND PERFORMANCE**

In our cohort of patients with type 2 DM, we created a CKD prediction equation **(model 1)** for three-year risk of CKD endpoints from the training datasets. The performances of **model 1** had good discrimination and calibration in both training and validation datasets. Moreover, we also created a simplified three-year risk CKD equation **(model 2)** by categorizing predictors in dichotomous characteristic from **model 1**. The simplified model (**model 2)** also had good discrimination and calibration. As a result, these showings indicated that **model 1 (laboratory model)** and **model 2 (simplified model)** had accurate predictions and high power of prediction for patients with type 2 DM.

According to categorized predictors in model 2, this leads predictive performance of **model 1** better than one of **model 2.** Confirmed results showed in survival probabilities by Kaplan-Meier method **(Figure 11 and 12).** The difference of predicted survival probabilities at 3 years between **model 1 and model 2** are 0.25 or 25%.

III. CKD PREDICTORS

Our findings showed that older age, male sex, lower eGFR, higher UACR, and higher hemoglobin a1c (HbA1c) are associated to CKD stage 3 developing in patients with type 2 DM. Our finding was quite similar to a previous study interesting CKD progression on patients with type 2 DM. According to Kittipanyaworakun' retrospective cohort study among 322 patients with type 2 DM from Saraburi hospital, increased age, diabetes duration, eGFR, increased urinary albumin excretion, and increased SBP. Increased age, eGFR, and urinary albumin excretion were similar to our predictors. A prospective cohort study on 1,582 Singaporean type 2 DM patients, the prediction model for CKD progression (the reduction of $\geq$ 25% below the eGFR baseline) included higher age, higher SBP, lower eGFR, higher UACR, higher LDL and higher HbA1c (23). Three of those predictors, including lower eGFR, higher UACR and higher HbA1c were similar to our study's results. According to the study of Nelson et al. (2019), 15 multinational cohort studies among 781,627 diabetes, a 5-year CKD prediction model that included age, sex, black race, history of CVD, hypertension, lower eGFR values, higher UACR, elevate HbA1c,

types of diabetic medication presenting as risk factors showed some overlap with the parameters that were significant in our study(24).

According to the rare prediction model of CKD progression in DM patients, we compared obtained predictors with other studies with different CKD outcomes (i.e., end stage renal disease (ESRD), major kidney related events, or onset of albuminuria). We found that our predictors (including eGFR, UACR, age and HbA1c) partially overlapped in these studies.

The Action in Diabetes and Vascular Disease: Preterax and Diamicron MR Controlled Evaluation (ADVANCE) (2012) study indicated predictors for risks of major kidney related events (defined as double serum creatinine, renal replacement, or renal death) and onset of albuminuria. For major kidney related events, history of diabetic retinopathy, male gender, eGFR, UACR and HbA1c and age at the educational attainment were used as predictors(20). However, male gender showed as risk predictors in the ADVANCE study, but male gender was prevention effect in our study. For onset of albuminuria, ethnicity of Asian, SBP, BP-lowering agents, eGFR, UACR and HbA1c were predictors. Another study of Dunkler (2015) using Ongoing Telmisartan Alone and in combination with Ramipril Global Endpoint Trial (ONTARGET) and Outcome Reduction with Initial Glargine Intervention (ORIGIN) as the developing and validating model showed that baseline albuminuria, eGFR, female gender, age as predictors for alive DM patients with CKD outcome (defined as among new microalbuminuria, or macroalbuminuria, or double Scr or ESRD). Those four predictors were quite similar to our study. Furthermore, our predictors have similarities to Thai case-control study (2017) which conducted in 470 diabetic and

hypertension patients in the primary care and secondary care units, Nakhonratchasima, a province in northeastern Thailand, with the predictors of uncontrolled HbA1c $\geq$7 %, old age with 70-79 years and female gender(119).

According to different demographic, ethnicity, medication, and time of following-up, our predictors were partially similar compared to other studies. However, UACR, eGFR and HbA1c were mainly predictive for CKD outcome among patients with type 2 DM(19, 22, 23, 101, 102). These findings were strongly emphasized that the good controlled blood glucose and controlled albuminuria could prevent diabetic kidney disease.

To explain how each predictor could lead to developing CKD, the explanation was obtained in the following:

Increased age is the well-known risk leading loss of renal function. Aging leads to a decrease in glomerular filtration rate (GFR) and renal blood flow (120). Thus, elderly patients with type 2 DM will have reserved capacity of renal less than younger patients with type 2 DM.

For gender, estrogen has renoprotective effect by reducing albuminuria.(121, 122). Conversely, our results showed CKD prevention in male gender comparing female gender. After subgroup analysis, we found that the number of female patients with the age of 50 or more (76.37%) is significantly greater than that of male patients with age of 50 or more (65.02%) (X^2= 23.57, $p<0.001$). It meant that most female patients might be postmenopausal which might have low estrogen leading to decreased renoprotective effect.

Glomerular filtration rate (GFR) is a critical predictor for CKD developing in patients with type 2 DM. Higher GFR levels indicates to a better kidney function. Similarly, low initial GFR level increased risk of clinical renal outcome(123). Many cohort studies among DM patients showed similar results in an effective prevention on higher eGFR level to end stage renal disease (ESRD), onset of albuminuria and death(17-22, 100-102).

For HbA1c, the mechanisms that cause kidney damage are the following: 1) Tubulointerstitial injury by activating protein kinase C, 2) increasing of the renin-angiotensin-aldosterone system (RAS) leads to glomerular damage, and 3) generating advanced glycation end products (AGEs) leads to overproduction of mesangial cell matrix. The three mechanisms lead to nephron loss and proteinuria(124). Elevated HbA1c exhibited inducing to rapid eGFR decline, and progression of albuminuria (28, 29).

The level of albuminuria predicts renal function loss. Albumin creatinine ratio level more than 300 mg/dL or proteinuria induced tubular chemokine expression and complement activation. This leads to inflammatory cell infiltration in the interstitium and sustained fibrogenesis. Rapid eGFR decline can exhibit by extension of proteinuria(125). Theoretically, the eGFR and UACR are well-known as independent predictors of CKD progression and ESRD (126). The eGFR reflects to renal function, and UACR reflects to renal damage (18). Recognizing the independence between UACR and eGFR leads to the reclassification CKD stage by consolidating UACR and levels of eGFR(127). Even UACR can predict the onset of albuminuria among DM patients, we tried to add more CKD predictors to increase abilities of CKD prediction(20).

These predictors excepting age and sex reflect to metabolic profile on CKD progression. Moreover, UACR and HbA1c are modified factors. Thus, controlled blood sugar, controlled albuminuria, or slowly GFR decline should be emphasized to health provider for clinical management and patient education.

IV. COMPARING MODEL PERFORMANCE WITH OTHER CKD PREDICTION MODELS

We tried to compare performance with other CKD model (including 10-year risk of decreased eGFR of Saranburat's study and 5-year risk of incident CKD of Nelson's study) using their developed CKD perdiction models with our own cohort. The Thai CKD prediction model of Saranburat et al's study (2017) which studies among of amongs 3,186 employess of the Electronic Generating to evaluate 10 year risk of decreased eGFR. The first model (clinical only) which had good performance of C-statistic of 0.72 consisted of age, sex, SBP, history of DM, and waist circumference. The second model (clinical + limited laboratory tests) which had better C-statistic of 0.79. included age, sex, SBP, histdory of DM and eGFR. However, we could not evaluate the performance of this equation in our own population as one of the predictors was waist circumference which was not available in our study(104). Moreover, long time follwing up with 10 years was not able to be obatined in our cohort.

Another study is Nelson' study which developed five-year risk CKD incident equation among 781,627 DM patients from 15 cohort studies. Nelson's model has good discrimination

with C-statistic of 0.801. However, we were unable to use Nelson's eqaution to compare performance to our study because smoking status was missing in 78% of our cohort participants(24)**(Appendix E).** As a result, we could not conclude whether our CKD model is better than either Nelson's model or Saranburat's model.

V. TWO MODEL ESTABLISHING

In Thailand, we have three levels of medical care, including primary care, secondary care and tertiary care, based on performances in medical care services. Most diagnosed DM patients were treated under health care units with high abilities of DM monitoring and management, i.e., tertiary medical care units and some secondary medical care units (general hospitals, regional hospitals, or). These medical care units can monitor DM patients especially DM complications by routine laboratories test, i.e., FBS, HbA1c, eGFR, urinary ACR and so on). However, some DM patients were also treated under primary health care units (i.e., primary clinic, community pharmacy) with limitation of some laboratory test, i.e., urinary ACR. And some patients prefer to refill medicine in community pharmacy which is a primary care unit with limitation of laboratory test, i.e., HbA1c, urinary ACR. In this study we wanted to develop two CKD prediction models to assist every level heath care units to screen or early detect patients with type 2 DM.

Therefore, our first CKD prediction model **(model 1)** which contained continuous laboratory is suitable for patients with type 2 DM who had complete laboratory testing in health care units such as secondary or tertiary hospitals. **Model 2** can be used friendly to estimate CKD

risk for patients with type 2 DM in primary health care settings with limitation facilities for laboratory test especially UACR test such as primary care clinics or community pharmacies.

VI. USING 3-YEAR RISK OF CKD EQUATIONS IN PRACTICE

This subsection was separated into 2 parts; accessibility of laboratory in health care settings and application of three-year risk of CKD stage 3 equations in patients with type 2 DM.

Accessibility of laboratory in health care settings

Regarding to our predictors, age and sex can be easily obtained. HbA1c, eGFR and UACR must be obtained from blood or urine test. In these days these three laboratory tests were more available for DM patients in health care units, i.e., hospital.

The 2019 Standards of Medical Care in Diabetes of the American Diabetes Association recommend all DM patients should be monitored for diabetic nephropathy by eGFR and urinary albumin (i.e., spot urinary albumin-to creatinine ratio) at least once a year, and blood glucose level should be primarily monitored by HbA1c(128). In Thailand, the Thai Clinical Practice Guideline for Diabetes 2017 suggests testing HbA1c and UACR for every DM patient to monitor a glucose level at least once a year (129). In health policy, a HbA1c and microalbuminuria test are supported by National Health Security Office (NHSO) for health care units under the operations of NHSO, (130).

As a result, most patients with type 2 DM have trended to access these three-laboratory tests. Supporting with data of the MedResNet (2019), a survey of assessment for Diabetic and Hypertension Care of hospitals under the Ministry of Public Health and hospitals under the Bangkok Metropolitan Administration in 2018, exhibited that the average percentages of HbA1c test and microalbuminuria at least once a year were 86.5 and 61.7, respectively. A HbA1c test had a high prevalence especially among public hospitals, including regional hospitals (93.4%), hospitals under the Jurisdiction of the Ministry of Defense/Ministry of Interior (91.7%), sub-district health promotion hospitals (88.1%), general hospitals (87.8%), and community hospitals (85.4%). In contrast, private hospitals had a lower percentage of HbA1c test (69.6%). Even the average percentage of microalbuminuria or UACR test was fair, its trend was increasing every year from 36.2% (in 2010) to 61.7% (in 2018)(131).

Application of three-year risk of CKD stage 3 equations in patients with type 2 DM

Individual predicted CKD risk could be calculated from a formula in **Appendix B.** As previously mentioned, **model 1** can be applied for predicting individual 3-year CKD risks for patients with type 2 DM with laboratory test, including eGFR, UACR, and HbA1c. This model can strongly predict CKD risk at 3 years. Eligible range of each predictors for age, eGFR, UACR, and HbA1c are 18-90 years old, 179-60 mL/min/1.73 m^2, 1.3-6469.5 mg/g, and 4-16.3%, respectively.

The calculated percent of three-year risk of CKD stage 3 were divided into 5 risk groups, including very low (<5%), low to moderate (5-15%), moderate to high (16-25), very high (26-40), and extremely high (>40%). The reason for dividing five groups are we would like to provide specific recommendations for specialists not only general physicians who may take care patients with type 2 DM very high or extremely high CKD stage 3 risk in order to prevent DM complications, i.e., cardiovascular disease. We used Hosmer-Lemeshow chi-square test to calibrate the five-CKD risk groups with non-significant results in both training and validation datasets **(Appendix F).** The result of Hosmer-Lemeshow chi-square indicated that using CKD risk equation dividing into the 5 risk groups have accuracy of prediction. Supporting with results of survival probability curves of 3-year CKD risk based on five CKD risk groups (very low, low to moderate, moderate to high, vary high and extremely high), were compared by plotting the predicted CKD survival probabilities and observed CKD events. The predicted CKD probabilities curves fell within the 95% CI for observed CKD probabilities in both **model 1 and model 2.**

Recommendations in each risk group are provided in each CKD risk group based on clinical guideline **(Appendix H)(39, 128, 132),** for example, if DM patient has UACR >30 mg/g, blood pressure controlling <130/80 mmHg, avoiding nephrotoxicity agents and ACEI/ARB agent should be provided for this patient. Moreover, we provided additional recommendations for health care provider to make decisions for clinical management in each risk group. For example, if DM patient has got a very low risk (<5%), lifestyle, herb and dietary supplement education will be provided with CKD risk monitoring once a year. For DM patients with low to moderate risk

(5-15%), additional recommendation for closely monitoring in drug nephrotoxicity, i.e., NSAIDs, herb and dietary supplement monitoring, controlling for low salt and protein intaking will be provided. If DM patients have moderate to high risks (16-25%) in community pharmacist or primary health care setting, referral to specialists, i.e., nephrologist, will be provided. Regrading to high prevalence of proteinuria (20%) from our training dataset, a risk of AKI incident, in this moderate to high risk group, self-monitoring of acute kidney injury (AKI) episode, i.e., proteinuria, nocturia, oliguria, should be provided for these patients, and CKD risk evaluation should be estimated twice a year. For DM patients with high risk (26-40%), additional recommendations are avoiding drug nephrotoxicity and diabetic retinopathy should be provided. For DM patients with extremely high risk (>40%), additional recommendations are screening for metabolic complications and comorbidity due to CKD progression, i.e., electrolyte abnormalities, metabolic acidosis, anemia.

Model 1 (laboratory model)

Our first CKD prediction model (**model 1**) which contained continuous laboratory is suitable for patients with type 2 DM who had comeplete laboratory test in health care units such as secondary or tertiary hospitals, or patients with type 2 DM who can regognized their exact eGFR, HbA1c and UACR results.

If a male type 2 DM patient with 55 years old have eGFR of 95 mL/min/1.73m^2, HbA1c of 7.8% and urinary albumin creatinine ratio (UACR) of 31 mg/g, the three-year risk of CKD

stage 3 endpoint at 3 year will be calculated equal to be 13.64% which is in low to moderate risk group. According to microalbuminuria, BP controlling <130/80 mmHg, and ACEI/ARB should be provided following clinical guideline. Moreover, additional recommendation for monitoring of drug nephrotoxicity, herb and dietary supplement monitoring, low salt controlling, lifestyle education should be provided, and the patient should be monitored CKD risk by three-year risk of CKD equation once a year.

Model 2 (simplified model)

For **model 2** or simplified model can be used in this mentioned DM patient with less discriminative performance than **model 1** (0.890 vs. 0.812). We recommend health care providers to use **model 2** when they have limitation of laboratory test, i.e., lacking for urinary albumin creatinine ratio test, or have limitation of laboratory information.

When **model 1** was refined to attain **model 2**, we had categorized continuous variables, including eGFR, UACR and HbA1c, based on the clinical practice guideline. We categorized eGFR levels into 2 groups based on stage of CKD; eGFR $\geq$90 mL/min/1.73 m^2 (CKD stage 1) and eGFR <90 mL/min/1.73 m^2 (CKD stage 2)(82). For UACR, we categorized into 2 groups based on the definition of proteinuria: UACR >300 mg/g (proteinuria or macroalbuminuria) and UACR $\leq$ 300 mg/g(82). At first, we used a HbA1c level of 7 as a cut point, but the significant result was not found. Therefore, uncontrolled DM was categorized using HbA1c levels of 7.5 as a cut point instead; HbA1c >7.5% (uncontrolled DM) and HbA1c <7.5% (controlled DM) (129). According to the mean age of 55, we categorized age group by 50 years older more.

Therefore, codes of each categorized predictor are these followings; eGFR $\geq$90 mL/min/1.73 m^2 (code 1) and eGFR <90 mL/min/1.73 m^2 (code 0), UACR >300 mg/g (proteinuria or macroalbuminuria) given in code 1 and UACR $\leq$ 300 mg/g (code 0), HbA1c >7.5% (code 1) and HbA1c $\leq$ 7.5% (code 0), age $\geq$50 years old (code 1) and age < 50 years old (code 2), male (code 1) and female (code 0).

If **model 2** or simplified model is used for the previous case: a male (code 1) type 2 DM patient with 55 years old (code 1) have eGFR of 95 mL/min/1.73m^2 (code 1), HbA1c of 7.8% (code 1) and urinary albumin creatinine ratio (UACR) of 31 mg/g (code 0), the three-year risk of CKD stage 3 endpoint at 3 year will be calculated equal to be 10.49% which is in low to moderate risk group. As a result, the previous recommendations should be provided.

According to categorized UACR >300 mg/g, UACR can be estimated by using urine dipstick when UACR is not available based on the KDIGO 2012 guideline suggestion(133). Therefore, UACR >300 mg/g or proteinuria, one predictor in **model 2**, can be estimated by using an urine dipstick test which is possible to be available in every healthcare units instead of UACR estimation. If an urine dipstick test has 3+, it means that albumin in urine is more than 300 mg/g(134).

A HbA1c test can be calculated form an equation of Mekvanich's study (2014) which were generated among 1,440 DM patients (age $\geq$35 years old) with 8 hour fasting blood at Prananklao hospital **(135)**. The equation of Mekvanich's was generated from HbA1c and FBS association by using linear regression **(Appendix G)**. Even the aim of Mekvanich's study was to

generate equation for average plasma glucose for DM patient by estimating correlations between HbA1c and fasting plasma glucose level, we can use this equation to estimate for HbA1c in our cohort. First, the range of FBS test and HbA1c used to generate the equation in the study were 42-795 mg/dL and 4-17%, respectively, which are usual available range in most DM patients. Second, the 8 hour fasting blood tests at one time visiting hospital in cross-sectional study is a simple and standard technique in most healthcare settings, including community pharmacies. Third, the study conducting with Thai large DM patients which might have similar demographic data and lifestyle to other Thai DM patients. As a result, the calculated HbA1c from Mekvanich's equation can be applied for Thai patients with type 2 DM including DM patients who visit in community pharmacy. In practice, if an estimated HbA1c level from Mekvanich's equation is 8%, the HbA1c level will be defined as 1 (HbA1c >7.5%) in **model 2 (simplified model).** But if a calculated HbA1c level is 6%, it will be defined as 0 (HbA1c $\leq$7.5%) instead.

For primary clinic or primary health care units with limitation some laboratory tests, patients might obtain eGFR and HbA1c level except urinary ACR. Therefore, primary clinic can used model 2 by using protein dipstick instead of urinary ACR. For community pharmacy, a community pharmacist can use **model 1** if a patient with type 2 DM can remember his or her eGFR, HbA1c and urinary ACR level tested from other health care units. If the patient might remember roughly about eGFR levels, urinary ACR and HbA1c level, **model 2** will be a good choice**.** If the patient remembers only eGFR level, **model 2** might be used by using protein

dipstick test and HbA1c calculation for Urinary ACR and HbA1c levels. However, the last condition validity tests should be conducted. Without laboratory, neither models can be used.

VII. STREGHTS AND LIMITATIONS

Our study has a number of limitations**.** ***First***, follow-up period was quite low with the median of 1.29 (interquartile range 0.5-2.5 years) years. Lost follow up can lead to this problem (i.e., forgetting their appointments, changing their living places, or working in other provinces). As a result of the limited follow-up, we were only able to develop 3-year CKD risk models. ***Second***, as we did not conduct external validation, our CKD prediction models might not be generalizable to all Thai patients with type 2 DM. However, the Medical Research Network of the Consortium of the Thai Medical schools (MedResNet) which is a Thai survey of Assessment for Diabetic and Hypertension Care of 906 hospitals among 36,793 DM patients in 2018 **(Appendix I**), shows patients have similar mean values of HbA1c, SBP, LDL, BMI, proportions of controlled HbA1c and hypertension to those observed in our cohort. This lends some support that our CKD prediction models can be applied to Thai patients with type 2 DM.

Despite these limitations, our study has several strengths. ***First***, although many studies have assessed the progression of CKD in the general population including in Thailand, our model is the first to assess 3-year risk of CKD stage 3 specifically in Thai patients with type 2 DM. ***Second***, the predictors in both CKD risk models are laboratory results used in routine clinical practice. Moreover, the simplified model can be used in clinical practice by using dipstick

screening for proteinuria and glucose strip test for HbA1c estimating. These are simple and instantaneous laboratory test that can be easily performed in healthcare settings with limitation laboratory test. ***Third***, for eGFR calculation we used the CKD-EPI equation which performed better with less bias than the original Modification of Diet in Renal Disease (MDRD) Study equation (113). ***Forth,*** Cox regression including time to CKD event, an essential factor for predicting event prognosis, was used to develop for a better model prediction than a model using logistic regression including only event and covariate.

CHAPTER VI

CONCLUSION AND RECOMMENDATIONS

CONCLUSIONS

This study found that age, sex, eGFR, UACR and HbA1c are the significant predictors for developing chronic kidney disease. Increased age, HbA1c and UACR had the risk effects on developing CKD. In contrast, male sex and increased eGFR had a protective effect on CKD developing. These findings showed that good blood glucose and albuminuria controls can preserve type 2 DM patients' renal function.

Two CKD risk prediction models are developed for Thai patients with type 2 DM in this study. The two models which were consisted of the laboratory and the socio-demographic data had shown good and accurate predictions. They could predict the 3-year probabilities of CKD in patients with type 2 DM. **Model 1 (laboratory model)** is suitable for the health care settings with complete laboratory test, and **model 2 (simplified model)** is suitable for some primary health care settings or community pharmacies that have limited laboratory test.

These CKD risk prediction models can assist health care providers to early detect for CKD developing and promote health care providers to prevent diabetic nephropathy. Moreover, these CKD risk models are tools for supporting patients' education about diabetic nephropathy by emphasizing them to control blood glucose and albuminuria.

CKD PREDICTION MODELS AND HEALTH CARE SERVICE

CKD prediction model can support health care service in term of improved CKD screening, clinical management and patient education.

CKD is one of non-communication disease (NCD) increasing in Thailand. These days, NHSO has the health policy of the health prevention and promotion such as chronic disease screening, health education, smoking cessation, for preventing non-communication disease (NCD) in every level of health care setting. For tertiary and secondary health care setting, including hospital, using CKD prediction models not only improve clinical management but it also emphasizes health care providers to monitor CKD-related laboratories or signs of CKD progression. This leads to improve CKD screening. Based on the results of this study, the CKD prediction model with the classified recommendations for patients with type 2 DM should be included in a part of DM clinical management guideline to provide the suitable clinical management: lifestyle, medication, with specific healthcare provided for patients with type 2 DM. CKD prediction model can be a decision support for health care providers, including general physician, nephrologists, pharmacist, and nurses to decide effective clinical management for patients with type 2 DM for improve DM outcomes. For primary care setting, i.e., primary clinic, community pharmacies, where have only general physician or community pharmacists, the CKD prediction model also can support in making decision in health prevention among patients with type 2 DM. When a patient with type 2 DM has a high risk of CKD, health care providers can make a decision immediately to refer the patients with type 2 DM to suitable healthcare units (i.e.,

hospitals) for potential treatment. As a result, the incidents of CKD the hospital admission with CKD are possibly decreased and lead to decrease the health expenditure of CKD management. Moreover, the CKD prediction model can recruit the losing follow-up type 2 DM patients with high CKD risk to health care unit for suitable DM management.

With clinical predictors, the CKD prediction model can be a patient education tool to raise awareness of controlling blood glucose, proteinuria, or controlling blood pressure. As result, CKD prediction model may increase awareness of CKD among patients which is very low (1.9%) (136).

THE RECOMMENDATIONS ON IMPLEMENTATION OF USING THREE-YEAR RISK OF CKD STAGE 3 EQUATIONS FOR FURTHER RESEARCH

For further study, multi-center analysis and external validation should be conducted. Herbal and dietary supplements should be a predictor candidate. Moreover, validity test of HbA1c calculation from Mekvanich's study should be obtained **(Appendix G)**.

According to recommendations or clinical interventions based on five CKD risk groups for both models **(model 1 and model 2)**, including very low, low to moderate, moderate to high, very high, and extremely high **(Appendix H),** prospective cohort study should be conducted to evaluate effectiveness of each clinical intervention in each CKD risked DM patient.

Moreover, regarding the required laboratory parameters for predicting CKD risks in our developed models **(model 1 and model 2**), DM patients who do not possess the laboratory test in

terms of eGFR, UACR and HbA1c, cannot be used to assess CKD risks with our developed models. Therefore, a questionnaire of CKD risk assessment without laboratory parameters should be established based on our developed models, and the reliability test and the validity test should be performed. Examples of CKD questionnaires with simple questions are presented in **Appendix J**.

REFERENCES

1. Roglic G. WHO Global report on diabetes: A summary. Int J Non-Commun Dis. 2016;1.

2. Selvin E, Coresh J, Brancati FL. The burden and treatment of diabetes in elderly individuals in the u.s. Diabetes care. 2006 29(11):2415-9.

3. International Diabetes Federation. IDF Diabetes Atlas 8th Edition (2017). 2017.

4. Nanditha A, Ma RCW, Ramachandran A, Snehalatha C, Chan JCN, Chia KS, et al. Diabetes in Asia and the Pacific: Implications for the Global Epidemic. Diabetes care. 2016;39(3):472-85.

5. Chan JC, Malik V, Jia W, Kadowaki T, Yajnik CS, Yoon KH, et al. Diabetes in Asia: epidemiology, risk factors, and pathophysiology. Jama. 2009;301(20):2129-40.

6. Aekplakorn W, Chariyalertsak S, Kessomboon P, Assanangkornchai S, Taneepanichskul S, Putwatana P. Prevalence of Diabetes and Relationship with Socioeconomic Status in the Thai Population: National Health Examination Survey, 2004-2014. J Diabetes Res. 2018;2018.

7. Guariguata L. Contribute data to the 6th edition of the IDF Diabetes Atlas. Diabetes Res Clin Pract. 2013;100(2):280-1.

8. Leelawattana R, Pratipanawatr T, Bunnag P, Kosachunhanun N, Suwanwalaikorn S, Krittiyawong S, et al. Thailand diabetes registry project: prevalence of vascular complications in long-standing type 2 diabetes. J Med Assoc Thai. 2006;89 Suppl 1:S54-9.

9. Tuttle KR, Bakris GL, Bilous RW, Chiang JL, de Boer IH, Goldstein-Fuchs J, et al. Diabetic kidney disease: a report from an ADA Consensus Conference. Am J Kidney Dis. 2014;64(4):510-33.

10. Sriwijitkamol A, Moungngern Y, Vannaseang S. Assessment and prevalences of diabetic complications in 722 Thai type 2 diabetes patients. J Med Assoc Thai. 2011;94 Suppl 1:S168-74.

11. Narenpitak S, Narenpitak A. Prevalence of chronic kidney disease in type 2 diabetes in primary health care unit of Udon Thani province, Thailand. J Med Assoc Thai. 2008;91(10):1505-13.

12. Kittipanyaworakun S, Munprom C, Auamnoy T, Praditpornsilpa K, Vadcharavivad S. Predicting the development of renal impairment in type 2 diabetic patients with preserved kidney function. J Thai J Pharm Sci. 2013;38:275-8.

13. Vejakama P, Ingsathit A, Attia J, Thakkinstian A. Epidemiological study of chronic kidney disease progression: a large-scale population-based cohort study. Medicine (Baltimore). 2015;94(4).

14. The Nephrology Society of Thailand. Thailand renal replacement therapy year 2015. Bangkok: Nephrology Society of Thailand; 2015.

15. Chatterjee S, Riewpaiboon A, Piyauthakit P, Riewpaiboon W, Boupaijit K, Panpuwong N, et al. Cost of diabetes and its complications in Thailand: a complete picture of economic burden. Health Soc Care Community. 2011;19(3):289-98.

16. Echouffo-Tcheugui JB, Kengne AP. Risk models to predict chronic kidney disease and its progression: a systematic review. PLoS medicine. 2012;9(11).

17. Desai AS, Toto R, Jarolim P, Uno H, Eckardt KU, Kewalramani R, 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(5):717-28.

18. Lin C-C, Li C-I, Liu C-S, Lin W-Y, Lin C-H, Yang S-Y, et al. Development and validation of a risk prediction model for end-stage renal disease in patients with type 2 diabetes. Scientific Reports. 2017;7:1-13.

19. Yang XL, So WY, Kong AP, Clarke P, Ho CS, Lam CW, et al. End-stage renal disease risk equations for Hong Kong Chinese patients with type 2 diabetes: Hong Kong Diabetes Registry. Diabetologia. 2006;49(10):2299-308.

20. Jardine MJ, Hata J, Woodward M, Perkovic V, Ninomiya T, Arima H, et al. Prediction of kidney-related outcomes in patients with type 2 diabetes. Am J Kidney Dis. 2012;60(5):770-8.

21. Wan EYF, Fong DYT, Fung CSC, Yu EYT, Chin WY, Chan AKC, 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 nephrology. 2017;18(1):257.

22. Keane WF, Zhang Z, Lyle PA, Cooper ME, de Zeeuw D, Grunfeld JP, 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(4):761-7.

23. Low S, Lim SC, Zhang X, Zhou S, Yeoh LY, Liu YL, et al. Development and validation of a predictive model for Chronic Kidney Disease progression in Type 2 Diabetes Mellitus based on a 13-year study in Singapore. Diabetes Res Clin Pract. 2017;123:49-54.

24. Nelson RG, Grams ME, Ballew SH, Sang Y, Azizi F, Chadban SJ, et al. Development of

risk prediction equations for incident chronic kidney disease. Jama. 2019;322(21):2104-14.

25. Stevens LA, Coresh J, Greene T, Levey AS. Assessing kidney function-measured and estimated glomerular filtration rate. The New England journal of medicine. 2006;354(23):2473-83.

26. Glassock RJ, Winearls C. Ageing and the glomerular filtration rate: truths and consequences. Trans Am Clin Climatol Assoc. 2009;120:419-28.

27. De Cosmo S, Viazzi F, Pacilli A, Giorda C, Ceriello A, Gentile S, et al. Predictors of chronic kidney disease in type 2 diabetes: A longitudinal study from the AMD Annals initiative. Medicine (Baltimore). 2016;95(27).

28. Moriya T, Tanaka S, Sone H, Ishibashi S, Matsunaga S, Ohashi Y, et al. Patients with type 2 diabetes having higher glomerular filtration rate showed rapid renal function decline followed by impaired glomerular filtration rate: Japan Diabetes Complications Study. Journal of diabetes and its complications. 2017;31(2):473-8.

29. Retnakaran R, Cull CA, Thorne KI, Adler AI, Holman RR, Group US. Risk factors for renal dysfunction in type 2 diabetes: U.K. Prospective Diabetes Study 74. Diabetes. 2006;55(6):1832-9.

30. De Cosmo S, Lamacchia O, Rauseo A, Viti R, Gesualdo L, Pilotti A, et al. Cigarette smoking is associated with low glomerular filtration rate in male patients with type 2 diabetes. Diabetes care. 2006;29(11):2467-70.

31. Jiang N, Huang F, Zhang X. Smoking and the risk of diabetic nephropathy in patients with type 1 and type 2 diabetes: a meta-analysis of observational studies. Oncotarget. 2017;8(54):93209-18.

32. Chagnac A, Weinstein T, Korzets A, Ramadan E, Hirsch J, Gafter U. Glomerular hemodynamics in severe obesity. Am J Physiol Renal Physiol. 2000;278(5):F817-22.

33. Bentata Y, Latrech H, Abouqal R. Does body mass index influence the decline of glomerular filtration rate in diabetic type 2 patients with diabetic nephropathy in a developing country? Renal failure. 2014;36(6):838-46.

34. Dodhia SS, Barasara JD, Joshi VSJIJoMS, Health P. Glycemic control affects progression of kidney disease in patients with type 2 diabetes mellitus. 2016;5(7):1305-9.

35. Launer LJ, Miller ME, Williamson JD, Lazar RM, Gerstein HC, Murray AM, et al. Effects of intensive glucose lowering on brain structure and function in people with type 2 diabetes

(ACCORD MIND): a randomised open-label substudy. Lancet Neurol. 2011;10(11):969-77.

36. Gerstein HC, Miller ME, Byington RP, Goff DC, Jr., Bigger JT, Buse JB, et al. Effects of intensive glucose lowering in type 2 diabetes. The New England journal of medicine. 2008;358(24):2545-59.

37. Sheen Y-J, Lin J-L, Li T-C, Bau C-T, Sheu W. Peripheral Arterial Stiffness Is Independently Associated with a Rapid Decline in Estimated Glomerular Filtration Rate in Patients with Type 2 Diabetes. BioMed research international. 2013;2013.

38. Yokoyama H, Kanno S, Takahashi S, Yamada D, Itoh H, Saito K, et al. Determinants of decline in glomerular filtration rate in nonproteinuric subjects with or without diabetes and hypertension. Clin J Am Soc Nephrol. 2009;4(9):1432-40.

39. National Kidney Foundation. KDOQI Clinical Practice Guideline for Diabetes and CKD: 2012 Update. Am J Kidney Dis. 2012;60(5):850-86.

40. Parving HH, Lewis JB, Ravid M, Remuzzi G, Hunsicker LG, investigators D. Prevalence and risk factors for microalbuminuria in a referred cohort of type II diabetic patients: a global perspective. Kidney Int. 2006;69(11):2057-63.

41. Vigil A, Condés E, Cobo Jaramillo G, Gallar P, Oliet A, Rodriguez I, et al. Predictors of a Rapid Decline of Renal Function in Patients with Chronic Kidney Disease Referred to a Nephrology Outpatient Clinic: A Longitudinal Study. Advances in Nephrology. 2015;2015:1-8.

42. Babazono T, Nyumura I, Toya K, Hayashi T, Ohta M, Suzuki K, et al. Higher levels of urinary albumin excretion within the normal range predict faster decline in glomerular filtration rate in diabetic patients. Diabetes care. 2009;32(8):1518-20.

43. Jalal D, Maahs DM, Hovind P, Nakagawa T. Uric acid as a mediator of diabetic nephropathy. Seminars in nephrology. 2011;31(5):459-65.

44. Hovind P, Rossing P, Tarnow L, Johnson RJ, Parving H-H. Serum Uric Acid as a Predictor for Development of Diabetic Nephropathy in Type 1 Diabetes. Diabetes. 2009;58(7):1668-71.

45. Jalal D, Rivard C, Johnson R, Maahs D, McFann K, Rewers M, et al. Serum uric acid levels predict the development of albuminuria over 6 years in patients with type 1 diabetes: Findings from the Coronary Artery Calcification in Type 1 Diabetes study. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association. 2010;25:1865-9.

46. Bjornstad P, Laffel L, Lynch J, El Ghormli L, Weinstock RS, Tollefsen SE, et al. Elevated Serum Uric Acid Is Associated With Greater Risk for Hypertension and Diabetic Kidney Diseases in Obese Adolescents With Type 2 Diabetes: An Observational Analysis From the Treatment Options for Type 2 Diabetes in Adolescents and Youth (TODAY) Study. Diabetes care. 2019;42(6):1120-8.

47. Spencer HW, Yarger W, Robinson R. Alterations of renal function during dietary-induced hyperuricemia in the rat. Kidney International. 1976;9(6):489-500.

48. Fanali G, di Masi A, Trezza V, Marino M, Fasano M, Ascenzi P. Human serum albumin: from bench to bedside. Mol Aspects Med. 2012;33(3):209-90.

49. Navarro JF, Mora C, Maca M, Garca J. Inflammatory parameters are independently associated with urinary albumin in type 2 diabetes mellitus. Am J Kidney Dis. 2003;42(1):53-61.

50. Leehey DJ, Kramer HJ, Daoud TM, Chatha MP, Isreb MA. Progression of kidney disease in type 2 diabetes - beyond blood pressure control: an observational study. BMC nephrology. 2005;6(1):8.

51. Zhang J, Zhang R, Wang Y, Li H, Han Q, Wu Y, et al. The Level of Serum Albumin Is Associated with Renal Prognosis in Patients with Diabetic Nephropathy. J Diabetes Res. 2019;2019.

52. Russo GT, De Cosmo S, Viazzi F, Pacilli A, Ceriello A, Genovese S, et al. Plasma Triglycerides and HDL-C Levels Predict the Development of Diabetic Kidney Disease in Subjects With Type 2 Diabetes: The AMD Annals Initiative. Diabetes care. 2016;39(12):2278.

53. Chiang HP, Lee JJ, Chiu YW, Tsai JC, Hung CC, Hwang SJ, et al. Systolic blood pressure and outcomes in stage 3-4 chronic kidney disease patients: evidence from a Taiwanese cohort. Am J Hypertens. 2014;27(11):1396-407.

54. Goto A, Goto M, Noda M, Tsugane S. Incidence of type 2 diabetes in Japan: a systematic review and meta-analysis. PLoS One. 2013;8(9).

55. Lee WJ, Sobrin L, Lee M-J, Kang M, Seong M, Cho H. The Relationship Between Diabetic Retinopathy and Diabetic Nephropathy in a Population-Based Study in Korea (KNHANES V-2, 3). Investigative ophthalmology & visual science. 2014;55(10):6547-53.

56. Zhang J, Wang Y, Li L, Zhang R, Guo R, Li H, et al. Diabetic retinopathy may predict the renal outcomes of patients with diabetic nephropathy. Renal failure. 2018;40(1):243-51.

57. Lin HT, Zheng CM, Wu YC, Chang YH, Chen JT, Liang CM, et al. Diabetic Retinopathy

as a Risk Factor for Chronic Kidney Disease Progression: A Multicenter Case(-)Control Study in Taiwan. Nutrients. 2019;11(3):509.

58. Clive DM, Stoff JS. Renal syndromes associated with nonsteroidal antiinflammatory drugs. The New England journal of medicine. 1984;310(9):563-72.

59. Garella S, Matarese RA. Renal effects of prostaglandins and clinical adverse effects of nonsteroidal anti-inflammatory agents. Medicine (Baltimore). 1984;63(3):165-81.

60. Tsai HJ, Hsu YH, Huang YW, Chang YK, Liu JS, Hsu CC. Use of non-steroidal anti-inflammatory drugs and risk of chronic kidney disease in people with Type 2 diabetes mellitus, a nationwide longitudinal cohort study. Diabet Med. 2015;32(3):382-90.

61. Rossing K, Christensen, PK., Hovind, P., Tarnow, L., Rossing, P., Parving, HH. Progression of nephropathy in type 2 diabetic patients. Kidney Int. 2004;66(4):1596-605.

62. Zoppini G, Targher G, Chonchol M, Ortalda V, Negri C, Stoico V, et al. Predictors of estimated GFR decline in patients with type 2 diabetes and preserved kidney function. Clin J Am Soc Nephrol. 2012;7(3):401-8.

63. Tapp RJ, Shaw JE, Zimmet PZ, Balkau B, Chadban SJ, Tonkin AM, et al. Albuminuria is evident in the early stages of diabetes onset: results from the Australian Diabetes, Obesity, and Lifestyle Study (AusDiab). Am J Kidney Dis. 2004;44(5):792-8.

64. Hinokio Y, Suzuki S, Hirai M, Suzuki C, Suzuki M, Toyota T. Urinary excretion of 8-oxo-7, 8-dihydro-2'-deoxyguanosine as a predictor of the development of diabetic nephropathy. Diabetologia. 2002;45(6):877-82.

65. Stehouwer CD, Gall MA, Twisk JW, Knudsen E, Emeis JJ, Parving HH. Increased urinary albumin excretion, endothelial dysfunction, and chronic low-grade inflammation in type 2 diabetes: progressive, interrelated, and independently associated with risk of death. Diabetes. 2002;51(4):1157-65.

66. Persson F, Rossing P, Hovind P, Stehouwer CD, Schalkwijk CG, Tarnow L, et al. Endothelial dysfunction and inflammation predict development of diabetic nephropathy in the Irbesartan in Patients with Type 2 Diabetes and Microalbuminuria (IRMA 2) study. Scandinavian journal of clinical and laboratory investigation. 2008;68(8):731-8.

67. Seaquist ER, Goetz FC, Rich S, Barbosa J. Familial clustering of diabetic kidney disease. Evidence for genetic susceptibility to diabetic nephropathy. The New England journal of medicine.

1989;320(18):1161-5.

68. Pettitt DJ, Saad MF, Bennett PH, Nelson RG, Knowler WC. Familial predisposition to renal disease in two generations of Pima Indians with type 2 (non-insulin-dependent) diabetes mellitus. Diabetologia. 1990;33(7):438-43.

69. Satko SG, Langefeld CD, Daeihagh P, Bowden DW, Rich SS, Freedman BI. Nephropathy in siblings of African Americans with overt type 2 diabetic nephropathy. Am J Kidney Dis. 2002;40(3):489-94.

70. Jeffers BW, Estacio RO, Raynolds MV, Schrier RW. Angiotensin-converting enzyme gene polymorphism in non-insulin dependent diabetes mellitus and its relationship with diabetic nephropathy. Kidney Int. 1997;52(2):473-7.

71. Boright AP, Paterson AD, Mirea L, Bull SB, Mowjoodi A, Scherer SW, et al. Genetic variation at the ACE gene is associated with persistent microalbuminuria and severe nephropathy in type 1 diabetes: the DCCT/EDIC Genetics Study. Diabetes. 2005;54(4):1238-44.

72. Fujisawa T, Ikegami H, Shen GQ, Yamato E, Takekawa K, Nakagawa Y, et al. Angiotensin I-converting enzyme gene polymorphism is associated with myocardial infarction, but not with retinopathy or nephropathy, in NIDDM. Diabetes care. 1995;18(7):983-5.

73. Hadjadj S, Tarnow L, Forsblom C, Kazeem G, Marre M, Groop PH, et al. Association between angiotensin-converting enzyme gene polymorphisms and diabetic nephropathy: case-control, haplotype, and family-based study in three European populations. J Am Soc Nephrol. 2007;18(4):1284-91.

74. Wang F, Fang Q, Yu N, Zhao D, Zhang Y, Wang J, et al. Association between genetic polymorphism of the angiotensin-converting enzyme and diabetic nephropathy: a meta-analysis comprising 26,580 subjects. Journal of the renin-angiotensin-aldosterone system : JRAAS. 2012;13(1):161-74.

75. Valencia ME, Weil EJ, Nelson RG, Esparza J, Schulz LO, Ravussin E, et al. Impact of lifestyle on prevalence of kidney disease in Pima Indians in Mexico and the United States. Kidney International. 2005;68:S141-4.

76. Pavkov ME, Knowler WC, Hanson RL, Nelson RG. Diabetic nephropathy in American Indians, with a special emphasis on the Pima Indians. Curr Diab Rep. 2008;8(6):486-93.

77. Ruggenenti P, Porrini EL, Gaspari F, Motterlini N, Cannata A, Carrara F, et al. Glomerular

hyperfiltration and renal disease progression in type 2 diabetes. Diabetes care. 2012;35(10):2061-8.

78. Bjornstad P, Cherney DZ, Snell-Bergeon JK, Pyle L, Rewers M, Johnson RJ, et al. Rapid GFR decline is associated with renal hyperfiltration and impaired GFR in adults with Type 1 diabetes. Nephrol Dial Transplant. 2015;30(10):1706-11.

79. Remuzzi G, Benigni A, Remuzzi A. Mechanisms of progression and regression of renal lesions of chronic nephropathies and diabetes. The Journal of clinical investigation. 2006;116(2):288-96.

80. Premaratne E, Verma S, Ekinci EI, Theverkalam G, Jerums G, MacIsaac RJ. The impact of hyperfiltration on the diabetic kidney. Diabetes Metab. 2015;41(1):5-17.

81. Lee Y-H, Bang H, Kim DJ. How to Establish Clinical Prediction Models. Endocrinology and Metabolism. 2016;31(1):38-44.

82. Steyerberg EW. Clinical prediction models : a practical approach to development, validation, and updating. New York: Springer; 2009.

83. Oh SM, Stefani KM, Kim HC. Development and application of chronic disease risk prediction models. Yonsei Med J. 2014;55(4):853-60.

84. Lee YH, Bang H, Kim HC, Kim HM, Park SW, Kim DJ. A simple screening score for diabetes for the Korean population: development, validation, and comparison with other scores. Diabetes care. 2012;35(8):1723-30.

85. Goff DC, Jr., Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation. 2014;129(25 Suppl 2):S49-73.

86. Feng C, Wang H, Lu N, Chen T, He H, Lu Y, et al. Log-transformation and its implications for data analysis. Shanghai Arch Psychiatry. 2014;26(2):105-9.

87. D'Agostino RB, Belanger A, D'Agostino RB. A Suggestion for Using Powerful and Informative Tests of Normality. The American Statistician. 1990;44(4):316-21.

88. Bozdogan HJP. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. 1987;52(3):345-70.

89. Sullivan LM, Massaro JM, D'Agostino RB, Sr. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. Stat Med. 2004;23(10):1631-60.

90. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. Journal of clinical epidemiology. 2001;54(8):774-81.

91. Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. Journal of clinical epidemiology. 2016;69:245-7.

92. McNeill D, Davenport TH. Analytics in Healthcare and the Life Sciences: Strategies, Implementation Methods, and Best Practices: Pearson Education; 2014.

93. Harrell FE, Jr., Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. Jama. 1982;247(18):2543-6.

94. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Stat Med. 2004;23(13):2109-23.

95. Hilden J, Habbema JD, Bjerregaard B. The measurement of performance in probabilistic diagnosis. II. Trustworthiness of the exact values of the diagnostic probabilities. Methods of information in medicine. 1978;17(4):227-37.

96. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol. 2013;13(1):33.

97. Hosmer DW, Lemesbow S. Goodness of fit tests for the multiple logistic regression model. Communications in Statistics - Theory and Methods. 1980;9(10):1043-69.

98. Friis-Moller N, Ryom L, Smith C, Weber R, Reiss P, Dabis F, et al. An updated prediction model of the global risk of cardiovascular disease in HIV-positive persons: The Data-collection on Adverse Effects of Anti-HIV Drugs (D:A:D) study. European journal of preventive cardiology. 2016;23(2):214-23.

99. Royston P. Tools for Checking Calibration of a Cox Model in External Validation: Prediction of Population-Averaged Survival Curves Based on Risk Groups. Stata Journal. 2015;15(1):275-91.

100. Elley CR, Robinson T, Moyes SA, Kenealy T, Collins J, Robinson E, et al. Derivation and validation of a renal risk score for people with type 2 diabetes. Diabetes care. 2013;36(10):3113-20.

101. Dunkler D, Gao P, Lee SF, Heinze G, Clase CM, Tobe S, et al. Risk Prediction for Early CKD in Type 2 Diabetes. Clin J Am Soc Nephrol. 2015;10(8):1371-9.

102. Geba D, Cordova JM, Shetty S, Williams B. 2234-PUB: A Risk Prediction Model for End-

Stage Renal Disease in Patients with Type 2 Diabetes Mellitus. Diabetes. 2019;68(Supplement 1):2234-PUB.

103. Thakkinstian A, Ingsathit A, Chaiprasert A, Rattanasiri S, Sangthawan P, Gojaseni P, et al. A simplified clinical prediction score of chronic kidney disease: a cross-sectional-survey study. BMC nephrology. 2011;12(1):45.

104. Saranburut K, Vathesatogkit P, Thongmung N, Chittamma A, Vanavanan S, Tangstheanphan T, et al. Risk scores to predict decreased glomerular filtration rate at 10 years in an Asian general population. BMC nephrology. 2017;18(1):240.

105. Kang H. The prevention and handling of the missing data. Korean J Anesthesiol. 2013;64(5):402-6.

106. Dong Y, Peng C-YJ. Principled missing data methods for researchers. Springerplus. 2013;2(1):222.

107. Rubin D. Multiple imputation for nonresponse in surveys. 1987 New York. NY Wiley.

108. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res. 2011;20(1):40-9.

109. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Statistics in Medicine. 2011;30(4):377-99.

110. Nguyen CD, Carlin JB, Lee KJ. Model checking in multiple imputation: an overview and case study. Emerging themes in epidemiology. 2017;14(1):8.

111. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. Journal of clinical epidemiology. 1996;49(12):1373-9.

112. Han K, Song K, Choi BW. How to Develop, Validate, and Compare Clinical Prediction Models Involving Radiological Parameters: Study Design and Statistical Methods. Korean journal of radiology. 2016;17(3):339-50.

113. 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.

114. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika. 1994;81(3):515-26.

115. Pencina MJ, D'Agostino RB, Sr. Evaluating Discrimination of Risk Prediction Models: The

C Statistic. Jama. 2015;314(10):1063-4.

116. Newson R. Comparing the Predictive Powers of Survival Models Using Harrell's C or Somers' D. Stata Journal. 2010;10(3):339-58.

117. Rahman MS, Ambler G, Choodari-Oskooei B, Omar RZ. Review and evaluation of performance measures for survival prediction models in external validation settings. BMC Medical Research Methodology. 2017;17(1):60.

118. D'Agostino RB, Sr., Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation. 2008;117(6):743-53.

119. Wongprakob N, Piyabundikul L. Predictive Factors of Chronic Kidney Disease in Patients with Diabetes Mellitus and Hypertension. The Journal of Baromarajonani College of Nusing,Nakhonratchasima. 2017;23(2):94-106.

120. Weinstein JR, Anderson S. The aging kidney: physiological changes. Adv Chronic Kidney Dis. 2010;17(4):302-7.

121. Neugarten J. Gender and the Progression of Renal Disease. Journal of the American Society of Nephrology. 2002;13(11):2807.

122. Agarwal M, Selvan V, Freedman BI, Liu Y, Wagenknecht LE. The relationship between albuminuria and hormone therapy in postmenopausal women. Am J Kidney Dis. 2005;45(6):1019-25.

123. Norris KC, Greene T, Kopple J, Lea J, Lewis J, Lipkowitz M, et al. Baseline predictors of renal disease progression in the African American Study of Hypertension and Kidney Disease. J Am Soc Nephrol. 2006;17(10):2928-36.

124. Cade WT. Diabetes-related microvascular and macrovascular diseases in the physical therapy setting. Phys Ther. 2008;88(11):1322-35.

125. Cravedi P, Remuzzi G. Pathophysiology of proteinuria and its value as an outcome measure in chronic kidney disease. British journal of clinical pharmacology. 2013;76(4):516-23.

126. Gansevoort RT, Matsushita K, van der Velde M, Astor BC, Woodward M, Levey AS, et al. Lower estimated GFR and higher albuminuria are associated with adverse kidney outcomes. A collaborative meta-analysis of general and high-risk population cohorts. Kidney international. 2011;80(1):93-104.

127. Levey AS, de Jong PE, Coresh J, El Nahas M, Astor BC, Matsushita K, et al. The definition, classification, and prognosis of chronic kidney disease: a KDIGO Controversies Conference report. Kidney Int. 2011;80(1):17-28.

128. American Diabetes Association. 11. Microvascular complications and foot care: Standards of Medical Care in Diabetes—2019. Diabetes care. 2019;42(Supplement 1):S124-38.

129. Diabetes Association of Thailand. Clinical practice guideline for diabetes. Bangkok: Romyen Media.; 2017.

130. Fund Management Manual of the National Health Security Office 2020 [Internet]. 2019 [cited 23 November 2019]. Available from: https://www.nhso.go.th/frontend/index.aspx.

131. The survey of assessment for Diabetic and Hypertension care of hospitals under the Ministry of Public Health and hospitals under the Bangkok Metropolitan Administration in 2018 [internet]. [Internet]. 2018 [cited 17 November 2019]. Available from: https://dmht.thaimedresnet.org/document/DMHT_slide_20180911_final2059.pdf.

132. Diabetes Association of Thailand.the Endocrine Society of Thailand. Department of Medical Services. National Health Security Office. Clinical practice guideline for diabetes. 2017.

133. Kidney disease: Improving global outcomes (KDIGO) CKD work group. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int Suppl. 2013;3(1):1-150.

134. Satirapoj B. Proteinuria: A Comprehensive Approach to Diagnosis. Royal Thai Army Medical Journal. 2011;64(3):155-64.

135. Mekvanich N. The Relationship between Fasting Plasma Glucose and Hemoglobin A1c for Creating Average Plasma Glucose Chart in Diabetic Patients of Pranangklao Hospital.

136. Ingsathit A, Thakkinstian A, Chaiprasert A, Sangthawan P, Gojaseni P, Kiattisunthorn K, et al. Prevalence and risk factors of chronic kidney disease in the Thai adult population: Thai SEEK study. Nephrol Dial Transplant. 2010;25(5):1567-75.

APPENDICES

APPENDIX A

MULTIVARIATE ANALYSIS

Table 15: Multivariable hazard ratio for the selection of CKD prediction model

Candidate predictors	Model A HR	Model B HR	Model C HR
Age, year	1.02*	1.02*	1.02*
eGFR, mL/min/1.73m^2	0.90*	0.90	0.90*
Log UACR, mg/dL	1.27*	1.27*	1.27*
Log HbA1c, %	3.32*	3.31*	3.25*
Log BMI, kg/m2	0.83		
CVD, n (yes vs. no)	1.09		
Log SBP, mmHg	1.85	1.78	
Sex, n (male vs. female)	0.74*	0.75*	0.75*
C-statistic	0.874	0.874	0.874
AIC (range)	3102.72	3099.42	3099.22

HR, hazard ratio; HbA1c, hemoglobin A1c; eGFR, estimated glomerular filtration rate; UACR, urinary albumin to creatinine ratio; SBP, systolic blood pressure; CVD, cardiovascular disease; log, logarithm transformation; AIC, akaike information criterion.

*variable have significance with p-value less than 0.05.

	rho	chi2	df	Prob>chi2
lnHbA1c	0.00267	0.00	1	0.9636
lnurinAlb	0.11582	4.41	1	0.0358
base_eGFR	0.10454	3.42	1	0.0644
gender	-0.03911	0.44	1	0.5089
age_int	0.08446	1.91	1	0.1670
global test		9.54	5	0.0894

Figure 13: Test of proportional hazard assumption for model 1

	rho	chi2	df	Prob>chi2
high_HbA1c75	0.02558	0.18	1	0.6726
proteinuria	0.13227	5.04	1	0.0247
eGFRstage	0.10680	3.13	1	0.0769
gender	0.00865	0.02	1	0.8861
age50	0.04674	0.60	1	0.4387
global test		8.51	5	0.1304

Figure 14: Test of proportional hazard assumption for model 2

APPENDIX B

THREE-YEAR RISK OF CKD EQUATIONS

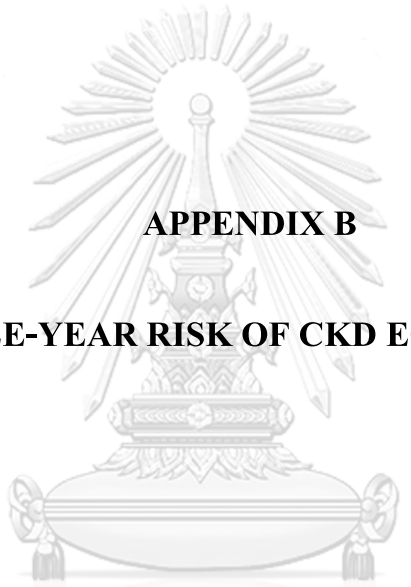


Table 16*:* Prediction equations to apply to individual patients for the 3-year risk of CKD

Model 1 **(laboratory model)**	1- 0.7307^exp((1.180327*log(HbA1c)) + (0.2435595*log(UACR)) + (-0.1027086*(eGFR)) + (-0.2942762*(if male)) + (0.0221663*age) - ((1.180327* log(7.214512)) + (0.2435595* log(123.7432)) + (-0.1027086*90.46844) + (-0.2942762*0.4255738) + (0.0221663*55.87869))
Model 2 **(simplified model)**	1-0.7307^exp((0.3207118*(if HbA1c >7.5)) + (0.8099867*(if UACR >300 mg/g)) + (-2.423474*(if eGFR ≥90)) + (-0.3603594*(if male)) + (0.7576155*(if age ≥50)) - ((0.3207118*0.3325027) + (0.8099867*0.160612) + (-2.423474*0.532459) + (-0.3603594*0.4255738) + (0.7576155*0.7154098))

HbA1c, Hemoglobin A1c; eGFR, estimated Glomerular Filtration Rate; UACR, Urinary albumin to creatinine ratio; log, logarithm transformation.

Where 0.7307 is the 3-year CKD-free survival probability in our cohort.

APPENDIX C

THE ESTIMATION OF DISCRIMINATION FOR THREE-YEAR RISK OF CKD EQUATIONS

Model 1 (laboratory model)

1. Calculate predicted CKD risk from 3-year risk of CKD equation (model 1), we use the commands

gen score = ((1.180327*log(HbA1c)) + (0.2435595*log(UACR)) + (-0.1027086*(eGFR)) + (-0.2942762*(if male)) + (0.0221663*age) - ((1.180327* log(7.214512)) + (0.2435595* log(123.7432)) + (-0.1027086*90.46844) + (-0.2942762*0.4255738) + (0.0221663*55.87869))

gen predictive_risk = 1-0.7307^exp(score)

2. Replace value of predicted risk to be negative value.

replace predictive_risk = -predictive_risk

3. Generate censorship indicator variable to a *somersd* censorship indicators variable

gen censind = 1-_d if _st==1

where, _st is created by *stset* command; 1 in observations with right-censored lifetimes (where _d [event of CKD] is 0); and 0 in observation with uncensored lifetimes (where _d is 1).

4. Calculate C-statistic for the training dataset.

somersd _t predictive_risk if _st ==1 & sample==1, tr(c) tdist cenind(censind)

The C-statistics of three-year risk CKD model were 0.873 and 0.890 in the training and validation dataset, respectively (**Figure 15 and 16).**

Symmetric 95% CI for Harrell's c

_t	Coef.	Jackknife Std. Err.	t	P>|t|	[95% Conf.	Interval]
predictive_risk	.8731701	.009416	92.73	0.000	.8547004	.8916398

Figure 15: C-statistic with 95%CI of three-year risk CKD model (model 1) in the training dataset

Symmetric 95% CI for Harrell's c

_t	Coef.	Jackknife Std. Err.	t	P>|t|	[95% Conf.	Interval]
predictive_risk	.8904952	.0105918	84.07	0.000	.8696971	.9112933

Figure 16: C-statistic with 95%CI of three-year risk CKD model (model 1) in the training dataset

Model 2 (simplified model)

1. Calculate predicted CKD risk from 3-year risk of CKD equation (model 1), we use the commands

gen score_2 = ((0.3207118*(if HbA1c >7.5)) + (0.8099867*(if UACR >300)) + (-2.423474*(if eGFR ≥90)) + (-0.3603594*(if male)) + (0.7576155* (if age ≥50)) - ((0.3207118*0.3325027) + (0.8099867*0.160612) + (-2.423474*0.532459) + (-0.3603594*0.4255738) + (0.7576155*0.7154098))

gen predictive_risk = 1-0.7307^exp(score_2)

2. Replace value of predicted risk to be negative value.

replace predictive_risk = -predictive_risk

3. Generate censorship indicator variable to a *somersd* censorship indicators variable

gen censind = 1-_d if _st==1

where, _st is created by *stset* command; 1 in observations with right-censored lifetimes (where _d [event of CKD] is 0); and 0 in observation with uncensored lifetimes (where _d is 1).

4. Calculate C-statistic for the training dataset.

somersd _t predictive_risk if _st ==1 & sample==1, tr(c) tdist cenind(censind)

The C-statistics of three-year risk CKD model were 0.798 and 0.812 in the training and validation dataset, respectively (**Figure 17 and 18).**

Symmetric 95% CI for Harrell's c

_t	Coef.	Jackknife Std. Err.	t	P>|t|	[95% Conf.	Interval]
predictive_risk	.7984826	.0123536	64.64	0.000	.7742508	.8227144

Figure 17: C-statistic with 95%CI of three-year risk CKD model (model 2) in the training dataset

Symmetric 95% CI for Harrell's c

_t	Coef.	Jackknife Std. Err.	t	P>|t|	[95% Conf.	Interval]
predictive_risk	.8118328	.0152945	53.08	0.000	.7818004	.8418652

Figure 18: C-statistic with 95%CI of three-year risk CKD model (model 2) in the validation dataset

APPENDIX D

THE ESTIMATION OF MODIFED HOSMER-LEMESHOW X^2 FOR THREE-YEAR RISK OF CKD EQUATIONS

Model 1 (laboratory model) in the training dataset

1. Calculate predicted CKD risk from 3-year risk of CKD equation (model 1), we use the commands

gen score = ((1.180327*lnHbA1c) + (0.2435595*lnurinAlb) + (-0.1027086*base_eGFR) + (-0.2942762*gender) + (0.0221663*age_int)) - ((1.180327*log(7.214512)) + (0.2435595*log(123.7432)) + (-0.1027086*90.46844) + (-0.2942762*0.4255738) + (0.0221663*55.87869))

gen predictive_risk = 1-0.7307^exp(score)

2. Risk group (***full10***) were divided into decile based on the calculated predicted risk from 1.

egen full10=cut(predictive_risk), group(10)

sort full10

by full10: egen mean_fullrisk = mean(predictive_risk)

3. Calculate mean of predicted hazard function in each risk group.

table full10, c(mean predictive_risk n predictive_risk)

4. Predicted CKD events were calculated by multiplying between number of patients of each risk group and the predicted hazard function in each decile of risk group **(Table 17).**

Table 17: Calculation of the predicted CKD risk

Risk group	Predicted hazard function	N	Predicted CKD events
1	0.0098969	65	0.64
2	0.0275004	65	1.79
3	0.0515167	65	3.35
4	0.0866657	66	5.72
5	0.1320810	65	8.59
6	0.2154872	65	14.01
7	0.3482845	66	22.99
8	0.5526576	65	35.92
9	0.7998934	65	51.99
10	0.9773067	66	64.50

5. Calculate observed 3-year hazard function by using the command

sts list, fail at(0 2 3) by(full10)

6. Calculate observed CKD events by multiply between ***N*** in each risk group and the observed hazard functions **(Table 18)**.

Table 18: The calculation of the predicted and observed CKD events in the validation dataset (model 1)

Risk group	Predicted hazard function	N	Predicted CKD events	Observed hazard functions	Observed CKD events
1	0.0098969	65	0.64	0.0303	2
2	0.0275004	65	1.79	0	0
3	0.0515167	65	3.35	0	0
4	0.0866657	66	5.72	0.0625	4
5	0.1320810	65	8.59	0.0795	5
6	0.2154872	65	14.01	0.0222	1
7	0.3482845	66	22.99	0.3797	25
8	0.5526576	65	35.92	0.4299	28
9	0.7998934	65	51.99	0.5950	39
10	0.9773067	66	64.50	0.8842	58

7. Calculate chi-square (X^2)values between the predicted CKD events and the observed CKD events in each risk group following this formula;

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$X^2 = \frac{(2-0.64)^2}{0.64} + \frac{(0-1.79)^2}{1.79} + \frac{(0-3.35)^2}{3.35} + \frac{(4-5.72)^2}{5.72}$$
$$+ \frac{(5-8.59)^2}{8.59} + \frac{(1-14.01)^2}{14.01} + \frac{(25-22.99)^2}{22.99} +$$
$$\frac{(28-35.52)^2}{35.52} + \frac{(39-51.99)^2}{51.99} + \frac{(58-64.5)^2}{64.5}$$

After this step, X^2 and p-value were obtained for **model 1**.

The calculation of chi-square for **model 1** in the training dataset and for **model 2** in both the training and validation datasets were performed similarly from step 1-7. **Table 19-21** showed

the predicted and observed hazard function of model 1 and model 2 in the training and validation datasets base on the decile of risk groups.

Table 19: **The calculation of the and predicted and observed CKD events in the training dataset (model 1)**

Risk group	Predicted hazard function	*N*	Predicted CKD events	Observed hazard functions	Observed CKD events
1	0.01	152	1.56	0.00	0
2	0.03	153	4.11	0.05	7
3	0.05	152	7.11	0.03	4
4	0.08	153	12.35	0.05	8
5	0.13	152	20.05	0.10	15
6	0.22	153	33.12	0.14	21
7	0.36	152	55.12	0.13	20
8	0.58	153	88.55	0.43	66
9	0.81	152	123.43	0.64	97
10	0.98	153	149.50	0.91	139

Table 20: The calculation of the and predicted and observed CKD events in the training dataset (model 2)*

Risk group	Predicted hazard function	N	Predicted CKD events	Observed hazard functions	Observed CKD events
1	0.042103	147	6.19	0.0095	1
2	0.059961	149	8.93	0.0246	4
3	0.084894	151	12.82	0.0717	11
4	0.121428	307	37.28	0.0602	18
5	0.33908	127	43.06	0.2122	27
6	0.597493	161	96.20	0.3513	57
7	0.713745	61	43.54	0.462	28
8	0.728162	217	158.01	0.5181	112
9	0.894209	205	183.31	0.7399	152

***Eight degree of freedom**

Table 21: The calculation of the and predicted and observed CKD events in the validation dataset (model 2)

Risk group	Predicted hazard function*)*	N	Predicted CKD events	Observed hazard functions	Observed CKD events
1	0.042366	61	2.58	0.0333	2
2	0.058811	62	3.65	0	0
3	0.08413	60	5.05	0.0833	5
4	0.108925	73	7.95	0	0
5	0.143727	66	9.49	0.0345	2
6	0.340093	59	20.07	0.3207	19
7	0.596538	74	44.14	0.5297	39
8	0.696946	35	24.39	0.5565	19
9	0.728082	85	61.89	0.4342	37
10	0.9773067	78	69.58	0.6425	50

APPENDIX E

FIVE-YEAR RISK OF INCIDENT CHRNIC KIDNEY DISEASE FOR DM PATIENTS (THE STUDY OF NELSON ET AL.)

Table 22*:* Prediction equation for the 5-year absolute risk of incident CKD (eGFR <60 ml/min/m^2)

Incident CKD	1 – exp (-5^0.9766551 × exp [-2.647004 + 0.1351572 × (age/5 – 11) + 0.1381975 × (if female) + 0.0920208 × (if black) + 0.3546697 × (15 – min(eGFR, 90)/5) – 0.1525133 × max(0, eGFR-90)/5 + 0.1870637 × (if has history of CVD) + 0.0619679 × (HbA1c –7) + 0.1078296 × (if insulin use) – 0.150944 × (if no DM medication use) + 0.023959 × (HbA1c –7) × (if insulin use) +0.0398424 × (HbA1c –7) × (if no DM medication use) – 0.00084 × (if ever smoking) + 0.3653268 × (if hypertensive) + 0.050306 × (BMI/5-5.4) + 0.3737905 × (log10ACR – 1)])

CKD, chronic kidney disease; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate; CVD, cardiovascular disease; HbA1c, hemoglobin A1c; BMI, body mass index; ACR, urine albumin to creatinine ratio

APPENDIX F

CALIBRATION RESULTS FOR FIVE-CKD RISK GROUPS IN TRAINING AND VALIDATION DATASETS

Table 23 Result of Hosmer-Lemeshow chi-square test based on five CKD risk groups in the training and validation datasets (model 1)

Risk group	Training dataset			Validation dataset		
	N	Predicted event	Observed event	N	Predicted event	Observed event
Very low	407	9.99	12	158	2.67	2
Low to moderate	327	30.71	22	160	15.19	9
Moderate to high	148	28.78	17	61	12.06	1
Very high	140	44.62	23	65	20.70	18
Extremely high	503	380.80	325	209	157.88	135
Total	1,525	494.91	398	653	209.5	166
X^2		5.76			8.25	
p-value		0.22			0.08	

Table 24 Result of Hosmer-Lemeshow chi-square test based on five CKD risk groups in the training and validation datasets (model 2)

Risk group	Training dataset			Validation dataset		
	N	Predicted event	Observed event	N	Predicted event	Observed event
Very low	92	3.40	0	37	1.37	2
Low to moderate	649	59.81	29	268	24.61	6
Moderate to high	52	10.08	14	35	6.80	5
Very high	49	16.10	12	23	7.60	3
Extremely high	683	499.94	354	290	208.39	155
Total	1,525	589.34	409	653	248.78	171
X^2		7.95			7.39	
p-value		*0.09*			*0.12*	

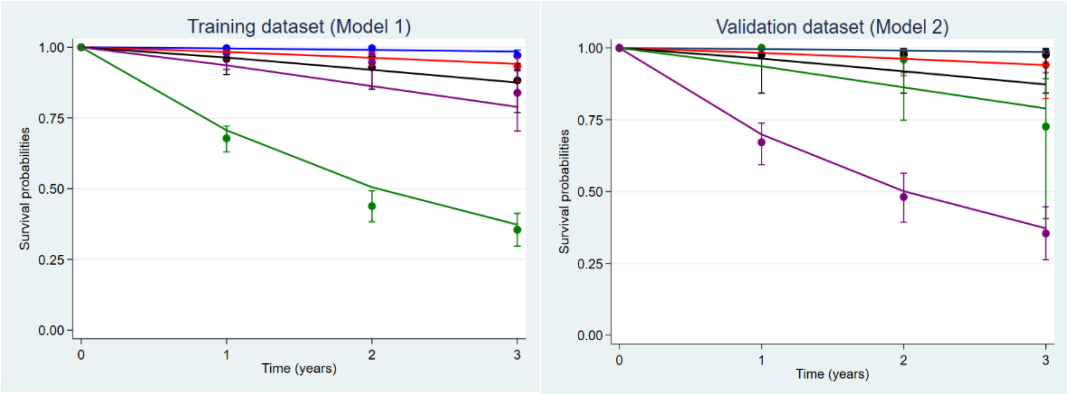


Figure 19: The observed CKD survival probabilities vs. predicted CKD probabilities for the Cox model in the training and the validation datasets for model 1.

Predicted survival probabilities are smooth lines, and the observed CKD survival probabilities from the Kaplan-Meier method with 95% confidence intervals are represented by vertical capped lines. Three prognosis groups plotted represent very low risk (blue lines), low to moderate risk (red lines), moderate to high risk (black lines), very high risk (purple lines), and extremely high risk (green line) groups.

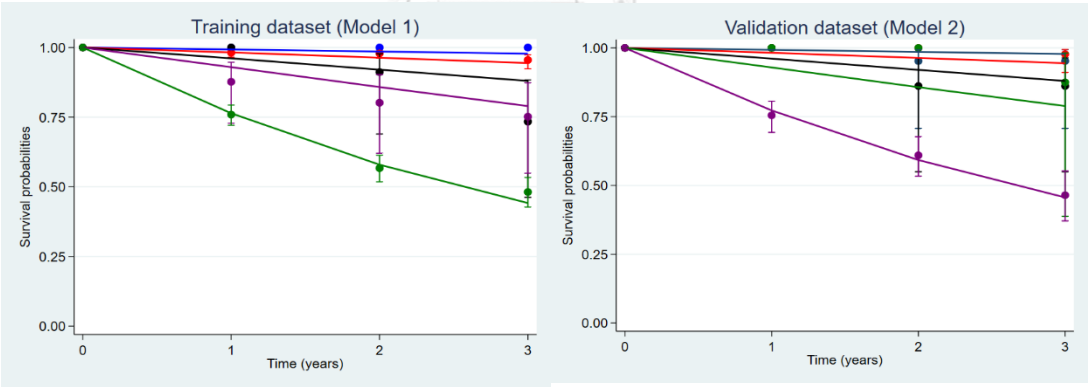


Figure 20: The observed CKD survival probabilities vs. predicted CKD probabilities for the Cox model in the training and the validation datasets for model 2.

Predicted survival probabilities are smooth lines, and the observed CKD survival probabilities from the Kaplan-Meier method with 95% confidence intervals are represented by vertical capped lines. Three prognosis groups plotted represent very low risk (blue lines), low to moderate risk (red lines), moderate to high risk (black lines), very high risk (purple lines), and extremely high risk (green line) groups.

APPENDIX G

EQUATION TO ESTIMATE THE EXPECTED HbA1c

(MEKVANICH'S STUDY)

Table 25*:* Equation to estimate the expected HbA1c

Expected values of HbA1c*	$\frac{FBG + 26.68}{24.119}$

HbA1c, hemoglobin A1c; FBS, fasting blood glucose

***age $\geq$ 35 years old and range of FBG is 42-795 mg/dL**

APPENDIX H

APPLICATIONS AND RECOMMENDATIONS OF THREE-YEAR RISK OF CKD EQUATIONS

Gender (1= male, 0= female)	1
Age (18-90 years)	40
eGFR (CKD-EPI fornula), 1.73mL,	70
urinary ACR, mg/g (>=1)	302
HbA1c, %	7.6

result	
3-year CKD risk (%)	82.12
severity	extremely high
CKD staging*	G2
Need for specialist refer?	refer
frequency of evaluation	twice per year

Recommendation	
blood pressure target, mmHg*	<130/80
Anithypertensive agent*	consider the use of ACEI/ARB if albuminuria occuring
Glycermic control target	A1c target is typically 7% but may not be apropriate for all DM type 2 patients
Lipid target*	lipid target is considered to follow relevant guideline , i.e., KDIGO guideline. Check lipids once to establish as baseline and once after therapy initiated.
Smoking	Encorage smoking cessation
lifestyle	give lifestyle education, i.e., food , exercise
	emphazige healthy food controlling and exercise
	Low salt, low protein intaking
drug	education of the use of nephrotoxic drug, dietary supplement and herb.
	avoid of the use of nephrotoxic drug (i.e., NSAIDs), dietary supplement and herb, and monitor closely
other recommendation	diabetic retinopathy monitoring
	Self monitoring of acute kidney injury (AKI) episode (proteinuria, noctuaries, oliguria)
	prevention of metabolic complication and comorbidity due to CKD progression(i.e., BP, volumn overload, electrolyte abnormalities, metabolic acidosis,anemia, metabolic bone disease - Clinical laboratories should be monitored as baseline.

Figure 21: Application program for 3-year CKD risk estimation (model 1)

Gender	(1= male, 0= female)	1
Age	(1= age>=50, 0 = age <50)	0
eGFR (CKD-EPI fornula)	(1=GFR>=90, 0= GFR <90)	0
urinary ACR, mg/g	(1=ACR >300, 0 = ACR <=300)	1
HbA1c, %	(1= HbA1c >=7.5%, 0= HbA1c <7.5%)	0

result	
3-year CKD risk (%)	61.58
severity	extremely high
CKD staging*	G2
Need for specialist refer?	refer
frequency of evaluation	twice per year

Recommendation	
blood pressure target, mmHg*	<130/80
Anithypertensive agent*	consider the use of ACEI/ARB if albuminuria occuring
Glycermic control target	A1c target is typically 7% but may not be apropriate for all DM type 2 patients
Lipid target*	lipid target is considered to follow relevant guideline , i.e., KDIGO guideline. Check lipids once to establish as baseline and once after therapy initiated.
Smoking	Encorage smoking cessation
lifestyle	give lifestyle education, i.e., food , exercise
	Low salt, low protein intaking
drug	education of the use of nephrotoxic drug, dietary supplement and herb.
	avoid of the use of nephrotoxic drug (i.e., NSAIDs), dietary supplement and herb, and monitor closely
other recommendation	diabetic retinopathy monitoring
	Self monitoring of acute kidney injury (AKI) episode (proteinuria, noctuaries, oliguria)
	volumn overload, electrolyte abnormalities, metabolic acidosis,anemia, metabolic bone

Figure 22: Application program for 3-year CKD risk estimation (model 2)

Table 26: Classified recommendations in 5 risk groups of CKD

Risk group	3-year predicted CKD risk (%)	Recommendations
Very Low	<5	-Monitor risk of CKD stage 3 once a year - Lifestyle education - Education for drug nephrotoxicity, i.e., NSAIDs.
Low to moderate	5-15	- Monitor risk of CKD stage 3 once a year - Lifestyle education - Closed monitor NSAIDs, herb, or dietary supplement use - Low salt, low protein intaking
Moderate to high	16-25	- Refer to specialist to follow up renal function - Monitor risk of CKD stage 3 twice a year - Lifestyle education - Closed monitor NSAIDs, herb, or dietary supplement use - Low salt, low protein intaking - self monitoring of AKI episode (proteinuria, noctuaries, oliguria)
Very high	26-40	- Refer to specialist to follow up renal function - Monitor risk of CKD stage 3 twice a year - Lifestyle education - Avoid NSAIDs, herb, or dietary supplement use - Low salt, low protein intaking - self monitoring of AKI episode (proteinuria, noctuaries, oliguria)
Extremely high	>40	- Refer to specialist to follow up renal function - Monitor risk of CKD stage 3 twice a year - Lifestyle education

Risk group	3-year predicted CKD risk (%)	Recommendations
		- Avoid NSAIDs, herb, or dietary supplement use - Low salt, low protein intaking - self monitoring of AKI episode (proteinuria, noctuaries, oliguria) - Screening for metabolic complications and comorbidity due to CKD progression, i.e., electrolyte abnormalities, metabolic acidosis, anemia

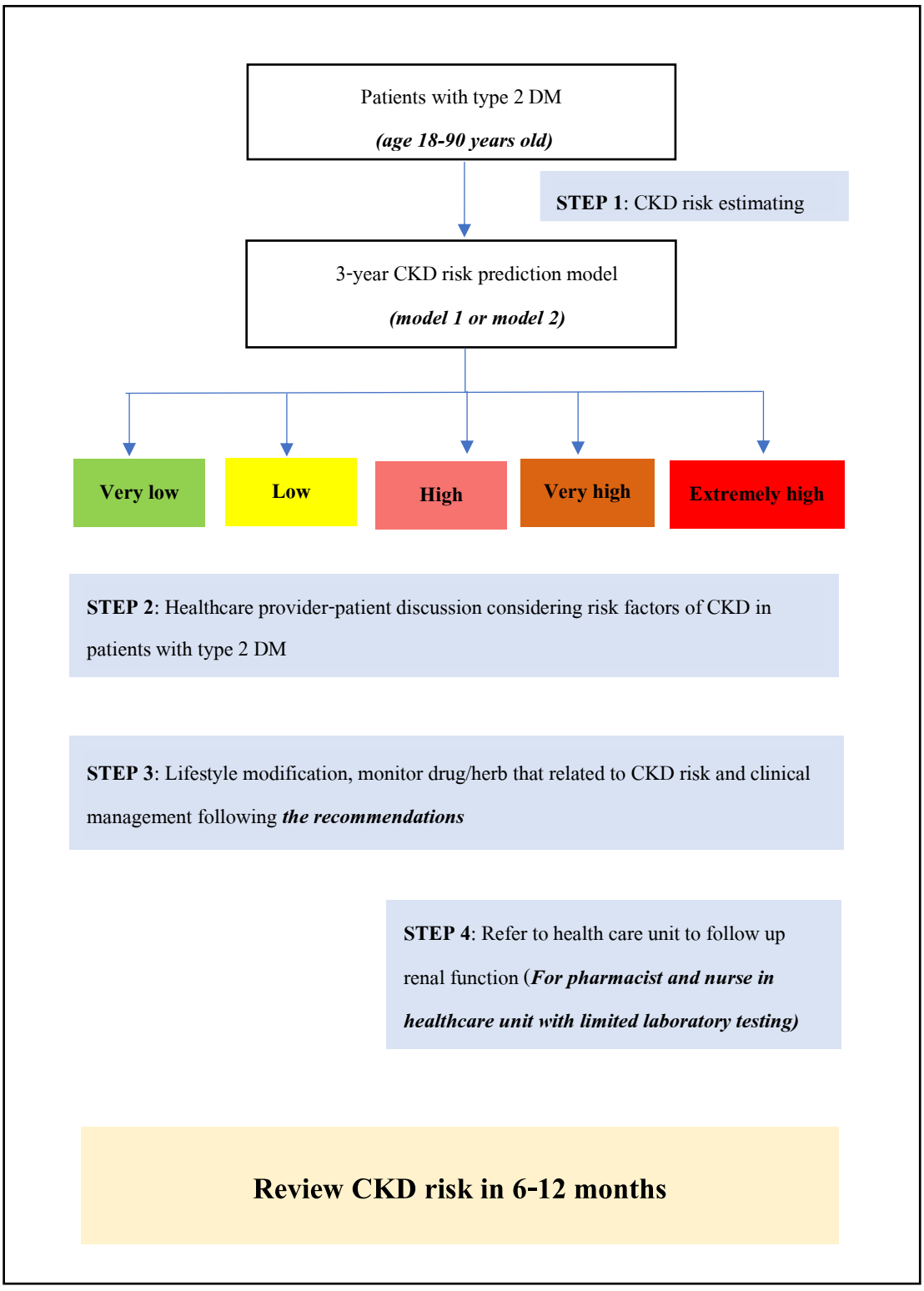


Figure 23: Flow chart of CKD risk estimating process

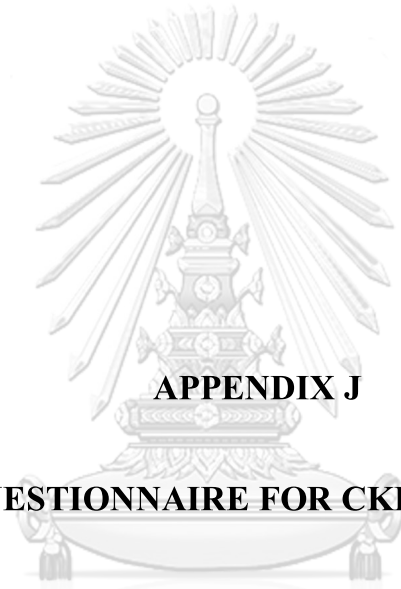
APPENDIX I

MedRestNet REPORT 2018

Table 27: Comparison of characteristics of patients with type 2 DM between the MedResNet report 2018 and this study

Laboratories	Our study (n=2,178)	MedResNet report 2018 (n= 36,793)
Mean of SBP, mmHg	136.23±20.27	133.1±15.3
Mean of DBP, mmHg	75.77±12.13	74.6 ±10.2
Mean of LDL, mg/dL	132.9±31.26	104.9±37.9
Mean of HbA1c, %	7.23	7.92
BMI, kg/m^2	26.45	24.7-26.5
Percentage of controlled FPG (70-130 mg/dL)	34.81%	37.1%
Percentage of controlled HbA1c (HbA1c <7%)	54.64%	36.5% (26.2-50.8%)
Percentage of uncontrolled HbA1c (HbA1c ≥9%)	12.16%	22.8%
Percentage of controlled LDL (LDL <100 mg/dL)	36.7%	49.2%
Percentage of comorbidity with hypertension	80.53%	78.5%
Percentage of comorbidity with Diabetes retinopathy (DR)	3.08%	5.2%
Diabetes nephropathy	4.73%	7.4%
Percentage of microalbuminuria test at least once a year	68.23%	61.7%
Percentage of HbA1c monitoring at least once a year	90.57%	86.5% (65.6%-93.9%)

NR, no report; SBP, systolic blood pressure; DBP, diastolic blood pressure; BMI, body mass index; FPG, fasting plasma glucose; HbA1c, hemoglobin A1c; LDL, low-density lipoprotein cholesterol



APPENDIX J

DRAFTED QUESTIONNAIRE FOR CKD RISK ESTIMATION

(MODEL 2)

Name……...……. Age……...years old

Gender ☐ male ☐ female

Underlining diseases ☐ Hypertension ☐ Diabetes Mellitus ☐ DLP ☐ other……

Duration of diabetics……years

Smoking status ☐ Never ☐ Currently smoking……. /day

☐ Stop smoking……years

Alcohol drinking status ☐ Never ☐ Current drinking ….mL/day

☐ Stop drinking….years

How many days did you take NSAIDs within 1 ear?.....days

Do you have history of diabetes nephropathy in family? ☐ No ☐ Yes……...

Do you take dietary supplement/ herbs? ☐ No ☐ Yes ……

Blood pressure……mmHg Pulse rate……bpm

Level of eGFR…….mL/1.73min/m^2

Questionnaires for CKD risk assessment

Abnormal urine assessment

(1) Do you have foamy urine?	(Yes)	(No)

Assessment of uncontrolled DM

(1) Increasing anti-diabetic medicine (injection or oral)	(Yes)	(No)
(2) Weight losing	(Yes)	(No)
(3) Thirsty	(Yes)	(No)
(4) Infected event such as UIT, infected wound, TB	(Yes)	(No)
(5) Polyurea	(Yes)	(No)

Figure 24: The draft of questionnaire for CKD risk estimating (Model 2*)*

VITA

NAME	Wilailuck Tuntayothin
DATE OF BIRTH	19 November 1986
PLACE OF BIRTH	Udonthani
INSTITUTIONS ATTENDED	2005-2010 Bachelor of Science in Pharmacy
HOME ADDRESS	1209 Ladproa Rd. Wang Thonglang, Bangkok 10310 Tel: 66 25384906
AWARD RECEIVED	2014 The education scholarships "The 100th Anniversary Chulalongkorn University for Doctoral Scholarship" 2016 The scholarships of SAKURA Exchange Program in Science 2017 Poster Presentation Award from The JSPS+-NRCT Follow-Up Seminar 2017 and 33rd International Annual Meeting in Pharmaceutical Science in topic of "COMMUNITY PHARMACISTS AND PUBLIC HEALTH CENTER COLLABORATION: THE MISSING CONNECTION IN PHARMACEUTICAL CARE"