Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biostatistics

CEB data warehouse

Real World Data in Ramathibodi Hospital

Wednesday 28th September 2022

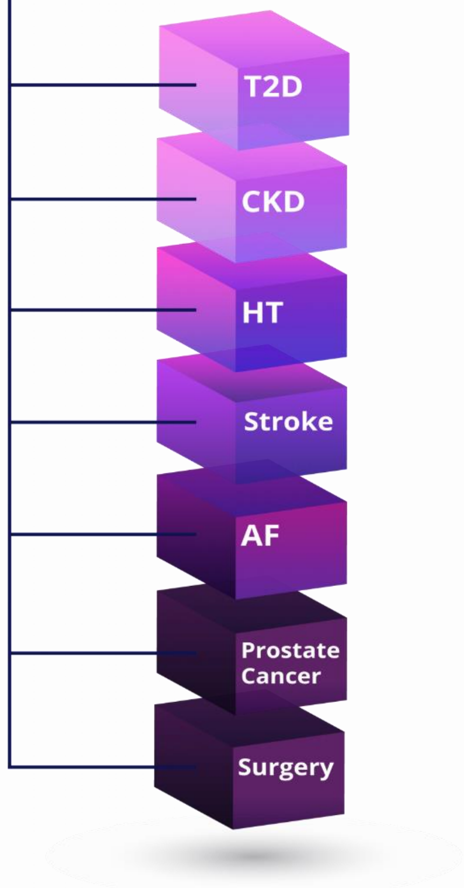
CEB's Team

Wisdom of the Land

Outlines

- CEB data warehouse
- Development of cohort datasets
- Research projects

CEB data warehouse

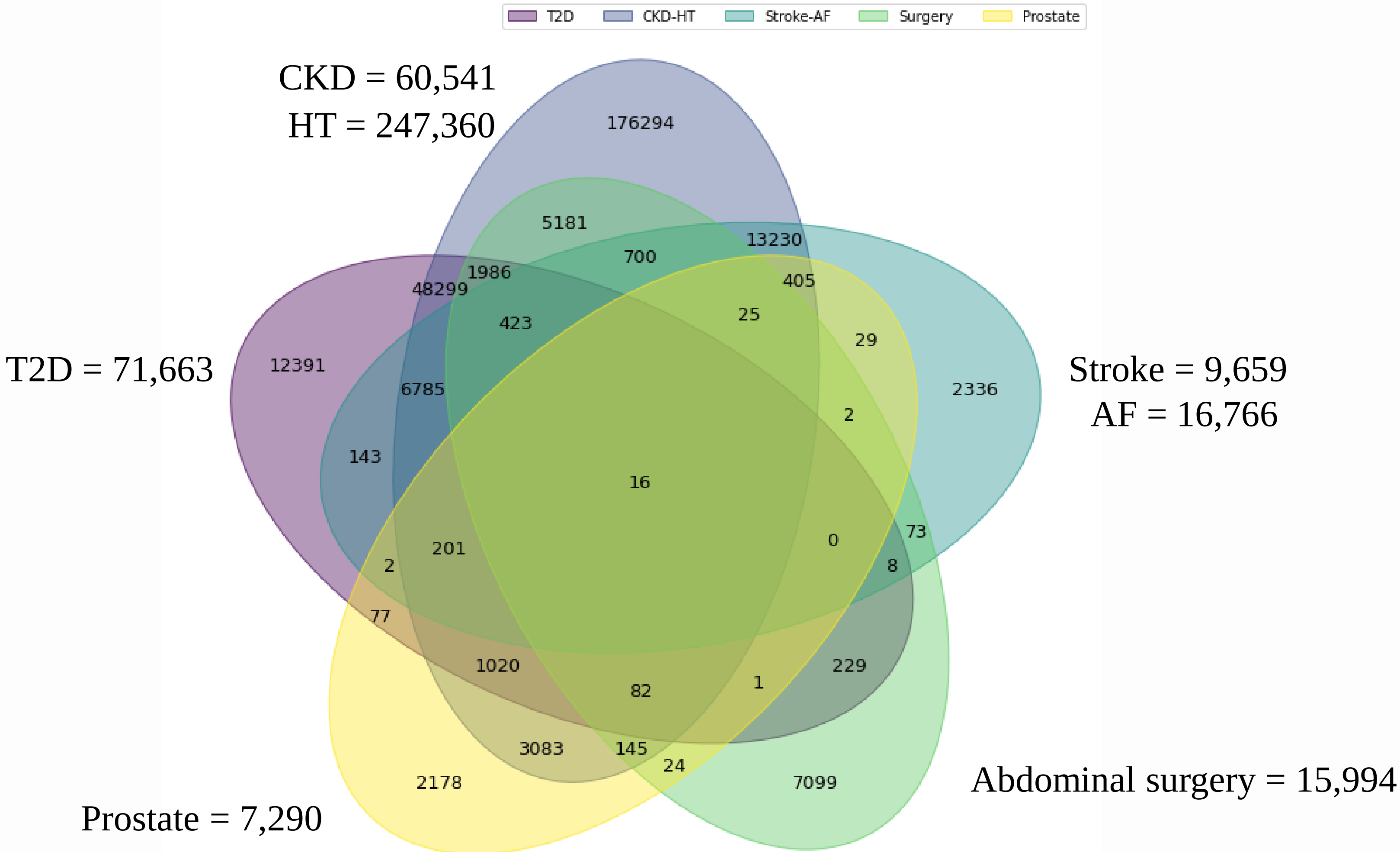


- Data warehouse for non-communicable diseases (NCDs) from Ramathibodi Hospital information system during 2010-2020 focusing on

https://www.rama.mahidol.ac.th/ceb/CEBdatawarehouse

Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biostatistics

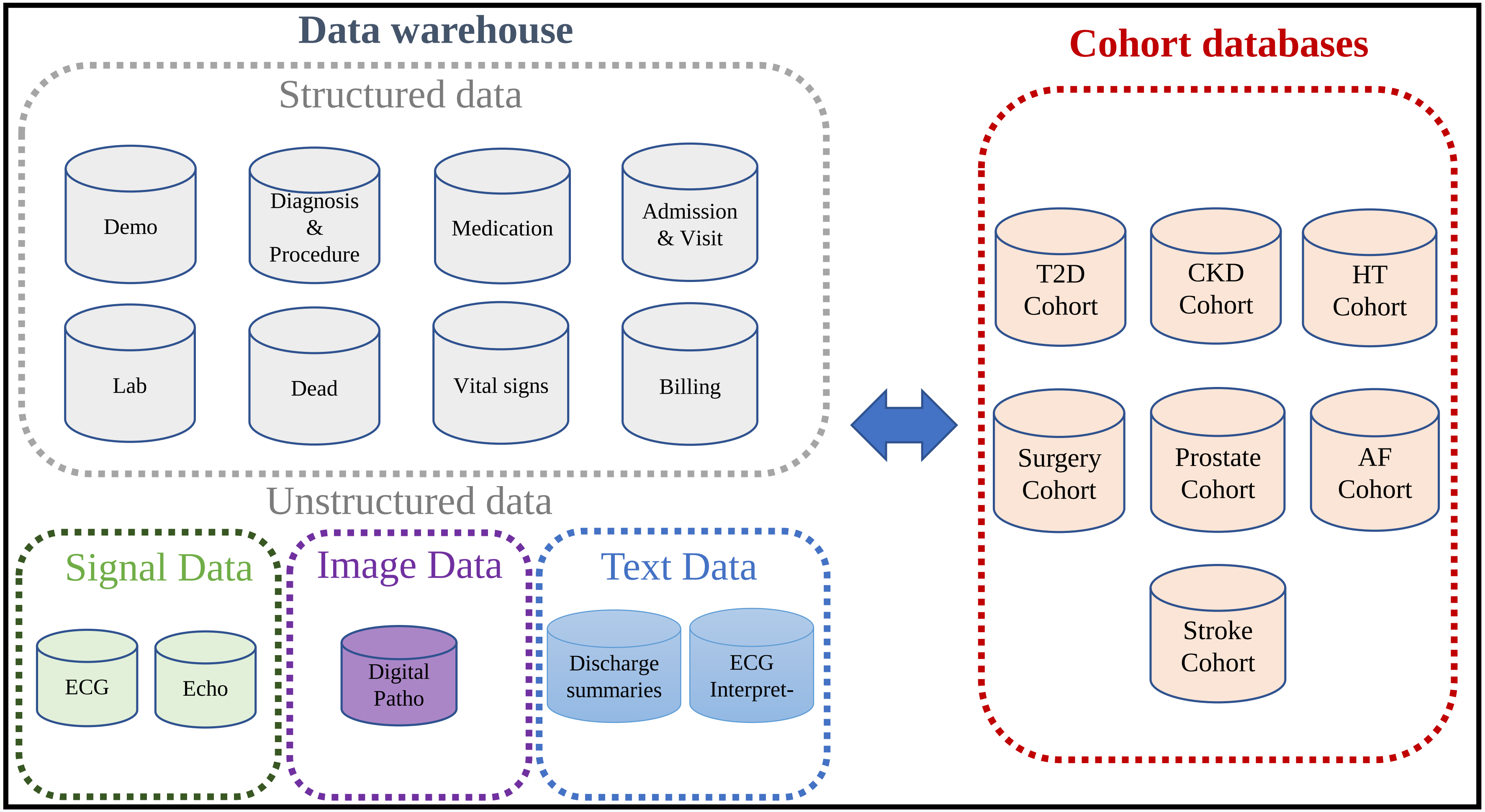
All = 282,467 subjects



Wisdom of the Land

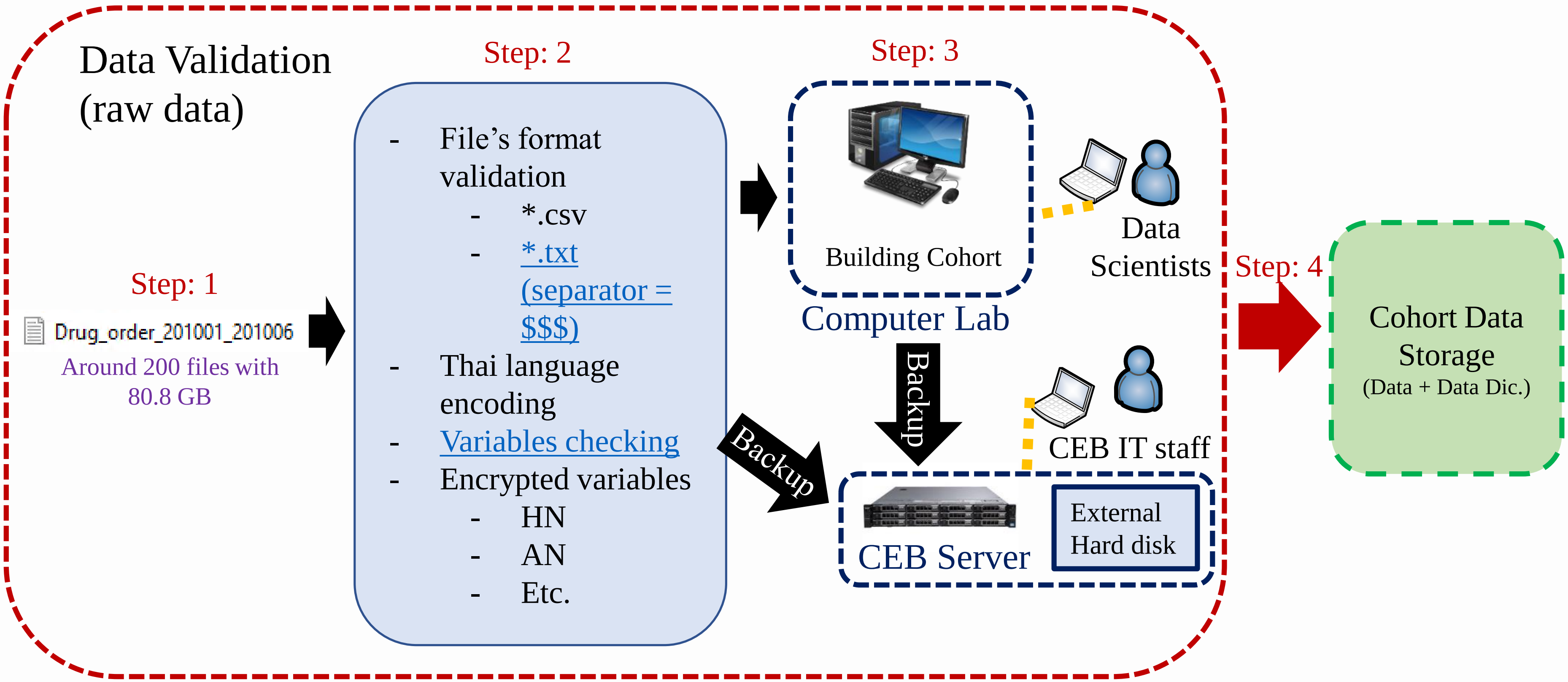
Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biostatistics

CEB - Data warehouse



Demographic	**admission and visit**	**Vital sign**	**Outcome and complication (diagnosis, operation, dead)**		**Medications**			**Laboratory**	
gender	insur	BW	Dead	DN	ANTI_ARRHYTHMIC	dpp4	asa	CBC_Hct	Lipid_Cholesterol
age	los	HIGH	CKD	HN	ANTI_COAG	glp1	other_antipl	CBC_Hb	Lipid_HDL
nationality		BMI	T2DM	ON	ANTI_DM	insulin	p2y12	Chem_glucose	Lipid_LDL
address		DBP	HT	CUN	ANTI_HT	metformin	pde	Chem_HbA1C	Lipid_Triglyceride
occupation		SBP	DLP	LN	ANTI_LIPID	sglt2	bisphosphanate	Elyte_Calcium	Lipid_vpkd
marital_status		HR	Obese	PKD	ANTI_PL	sus	calcium	Elyte_Carbondioxide	Renal_Albumin
residence		RESP	CVD	AN	ANTI_PTH	tzd	vitd	Elyte_Chloride	Renal_BUN
		SPO2	PVD	CPN	ANTI_URIC	acei	cox1	Elyte_Phosphorus	Renal_Urine_creatinine
		TEMP	Liverdis	MM	BONE	alphablocker	cox2	Elyte_Potassium	Renal_Serum_creatinine
			Pulmodis	PE	ESA	arb	phos_binder_alu	Elyte_Sodium	Renal_Cystatin_C
			CA	trauma_neph	IRON	bb	phos_binder_ca	Iron_FSH	Renal_PTH
			AIDS	TBKUB	KETOSTERIL	ccb	phos_binder_lanthanum	Iron_Serum_iron	Renal_Uric_acid
			SLE	Isch_neph	NSAIDS	diuretic	phos_binder_sevelamer	Iron_TIBC	Renal_Urine_Protein
			Gout	IGAN	PHOS_BINDER	nitrate		LFT_ALT	Renal_eGFR
			Stone	FSGS	SODAMINT	vasodilators		LFT_AST	Viral_Anti_HCV
			Dementia	CresGN	ANTI_CKDMBD	cholestyramine		LFT_Alkaline_phosphatase	Viral_Anti_HBc
			Fracture	RPGN	doac	ezetimibe		LFT_Direct_bilirubin	Viral_Anti_HBs
			Handicap	Glomdis	heparin	fibrate		LFT_GGT	Viral_Anti_HIV
			Pregnancy	HIVAN	warfarin	statinhydro		LFT_Total_Protein	Viral_HBsAg
			Sleepdis	Unknown	agis	statinlipo		LFT_Total_bilirubin	
			Gumdis						

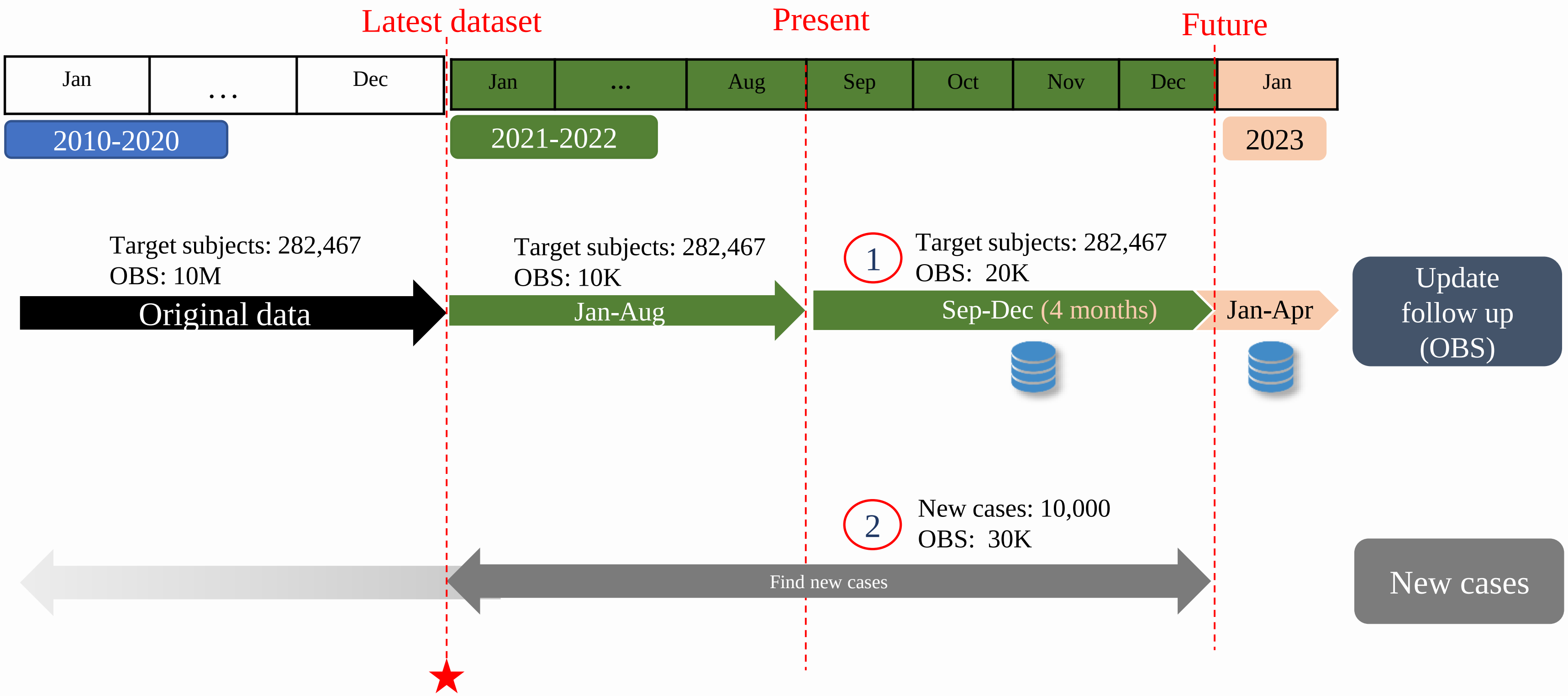
Backup raw data and cohort data



Remarks:
CEB IT staff will upload the backup data from the external hard disk to CEB-server weekly.

UPDATING DATA WAREHOUSE

Update 1. follow up and 2. new cases



Process of update new cases

DS-team

Newly diagnose patients

Loosely inclusion criteria
and **exclusion criteria**
(e.g. FPG ≥ 126 once)

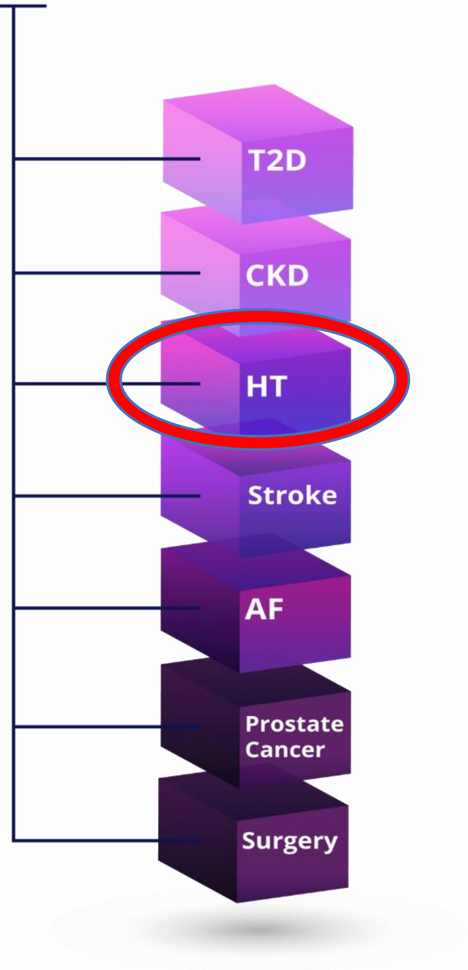
BI-team

Definite criteria
(e.g. FPG ≥ 126 twice consecutively,
HbA1c, Anti-DM drug)

Cohort

Development of cohort datasets

CEB data warehouse



The example of HT cohort database

The development of HT cohort

- Subject identification
- Data linkage
- Data lumping

Subject identification

- International Classification of Diseases (ICD)
  - a system of diagnostic codes for classifying diseases
  - a wide variety of sign, symptoms, abnormal findings and cause of condition
- In Ramathibodi Hospital Information system,
  - diagnosis of conditions are coded in tenth revision (ICD-10), and
  - ninth revision (ICD-9) for procedures, such as surgery
- To determine the inclusion criteria, an advisory panel is formed
- For Hypertension (HT) Data Warehouse project, ICD-10 codes
  - I10, I11, I12, I13 and I15.

Subject identification

- Issue 1: In UK and Thailand, clinical coders or medical coding officers are employed to
  - analyze *clinical statements* by the clinicians and
  - assign standard codes to the information system;
- examination and classifications are conducted under two different personnel at two different settings, increasing the chance of miscoding;
  - 31.0-42.0% in UK [a]
  - 62.1%-92.7% in Thailand [b]

a. Tsopra R, Peckham D, Beirne P, Rodger K, Callister M, White H, et al. The impact of three discharge coding methods on the accuracy of diagnostic coding and hospital reimbursement for inpatient medical care. Int J Med Inform. 2018;115:35-42.
b. Sukanya C. Validity of Principal Diagnoses in Discharge Summaries and ICD-10 Coding Assessments Based on National Health Data of Thailand. Healthc Inform Res. 2017;23(4):293-303.

Subject identification

- Issue 2: Clinical statements mainly notes present illness rather than overall medical history.
- Less likely to report underlying or pre-existing conditions such as DM, CKD and HT.
- Issue 3: The criteria for diagnosis changes over time (concept drift) due to increasing prevalence and awareness
- Conditions are often observed to be underdiagnosed on the retrospective review [a]

a. Morley KI, Wallace J, Denaxas SC, Hunter RJ, Patel RS, Perel P, et al. Defining Disease Phenotypes Using National Linked Electronic Health Records: A Case Study of Atrial Fibrillation. PLOS ONE. 2014;9(11):e110900

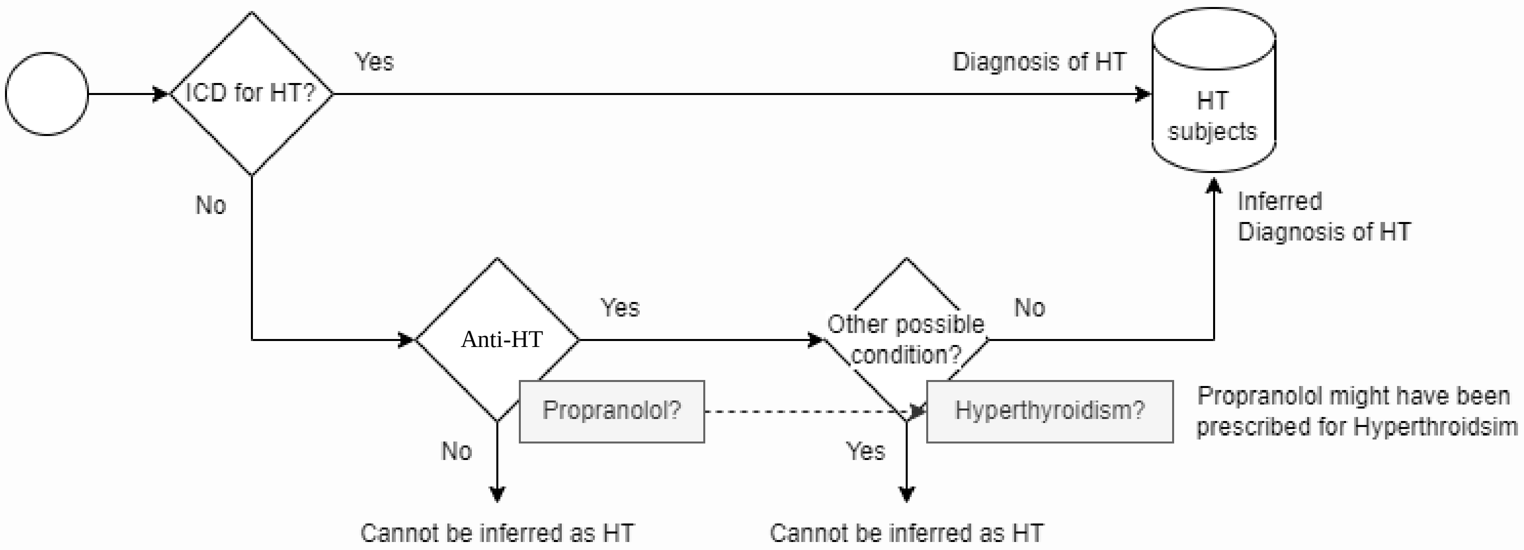
Subject identification

- Therefore, the advisory panel addresses the issues -
  - For patients without interested diagnostic code, what other criteria can be used to *reasonably infer diagnosis*?

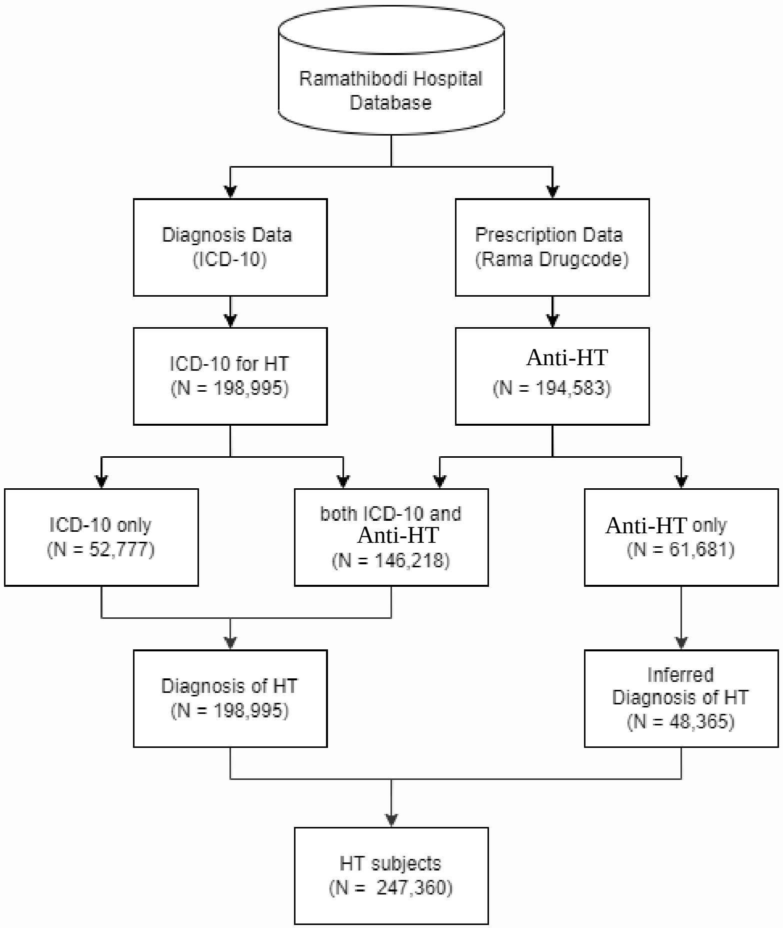
Data Warehouse	Diagnosis	Operation/Procedure	Laboratory	Medication
Type-2 DM	ICD-10		Hemoglobin A1C Fasting Blood Sugar	Anti-diabetic drugs
CKD	ICD-10	Renal Replacement Therapy	Estimated glomerular filtration rate Urine Albumin-Creatinine ratio Urine Protein-Creatinine ratio	
HT	ICD-10			Anti-hypertensive medications (Anti-HT)

Subject identification

- For HT data warehouse project,
  - ICD-10 notation of HT, or
  - At least one Anti-HT medication NOT prescribed for NON-HT conditions



Subject identification



ICD-10 codes used for Diagnosis of Hypertension (HT)

I10 Essential (primary) hypertension
I11 Hypertensive heart disease
I12 Hypertensive chronic kidney disease
I13 Hypertensive heart and chronic kidney disease
I15 Secondary hypertension

Antihypertensive Medicatio (Anti-HT) used for Diagnosis of HT

At least one drug group prescribed
Angiontensin-converting Enzyme Inhibitor,
Angiotensin II Receptor Blocker,
Calcium Channel Blocker,
Alpha Agonist,
Alpha Blocker,
Alpha Beta Blocker,
Beta Blocker,
Diuretic,
Ergot Alkaloid,
Hydralazine,
Minoxidil,
Neprilysin Inhibitor,
Renin Inhibitor,
Reserpine,
Statin

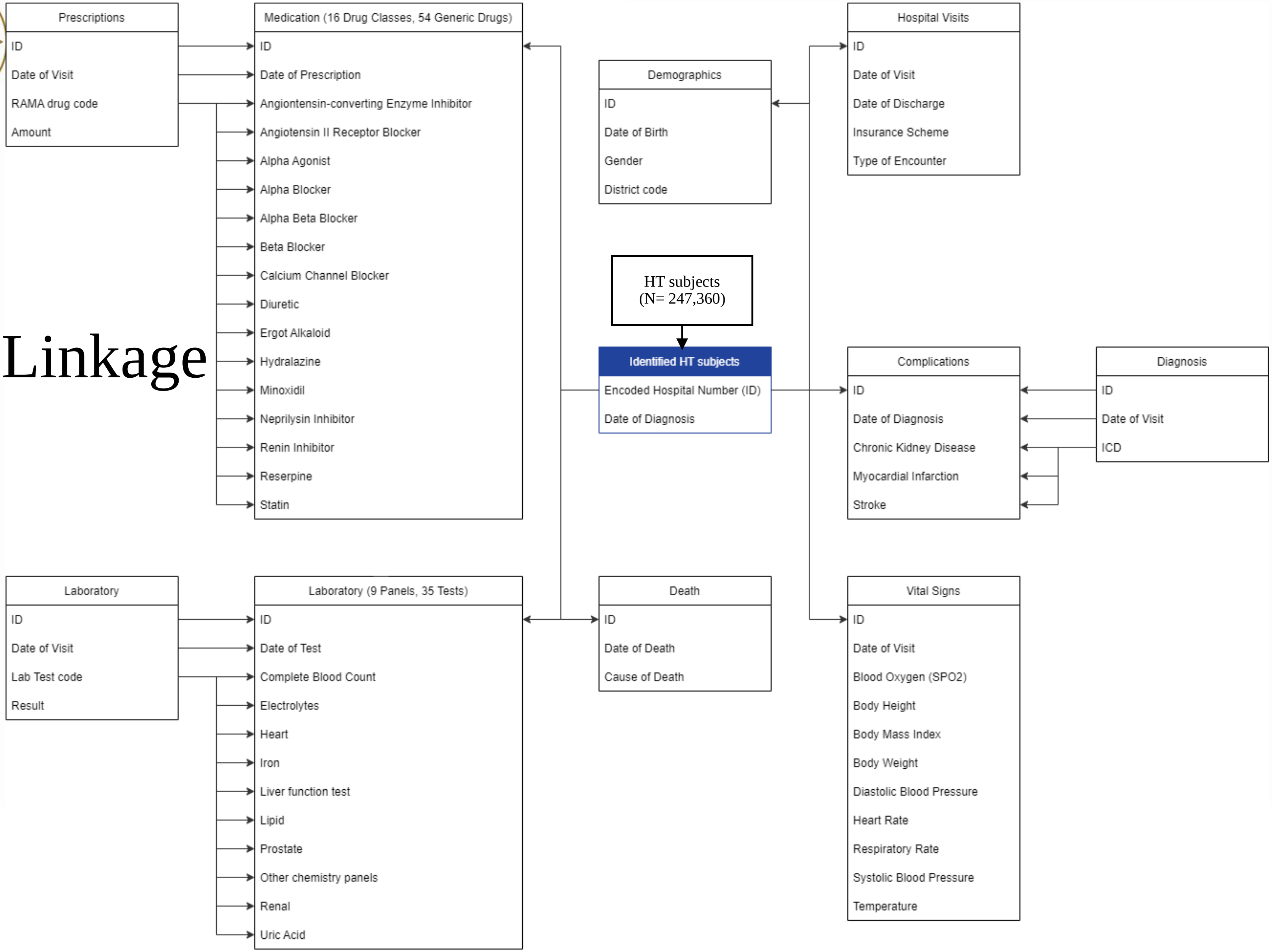
Other conditions commonly prescribed with ANTH

Angiontensin-converting Enzyme Inhibitor: Heart Failure
Angiotensin II Receptor Blocker: Heart Failure
Calcium Channel Blocker: Arrhythmia (non-Dihydropyridine)
Alpha Agonist: Hypertension in Pregnancy
Alpha Blocker: Benign Prostatic Hyperplasia
Beta Blocker: Hyperthyroidism
Neprilysin Inhibitor: Heart Failure

Data linkage

- Hospital number (HN) is a primary key (unique ID) for each patient in the system
- For patient de-identification, HN were cascade-encrypted using two Base64 hash by
  - Ramathibodi Business Intelligence Team (Extraction), and
  - Data Science and Clinical Informatics Division (Transformation, Loading).
- Encrypted HN + Date of visit is used to identify the same *visit* across multiple information systems (such as pharmacy and laboratory)

Data Linkage



Data Lumping

- Different patients have different treatment plans such as follow-up visits, so the number of observations are unbalanced between each subject.
- Not all visits undergo every selected tests, so not every data of interest is produced.
- Medical episodes can be single or multiple clinical visits, therefore a single episode can produce multiple observations.
- By linking data on encrypted HN + date of visit as the single ***episode***, it leads to a very sparse data.
- Data lumping helps reduce the sparsity of the data.
  - By preserving the record of interest and
  - Aggregating the rest of routine records for a determined time interval

Data Lumping

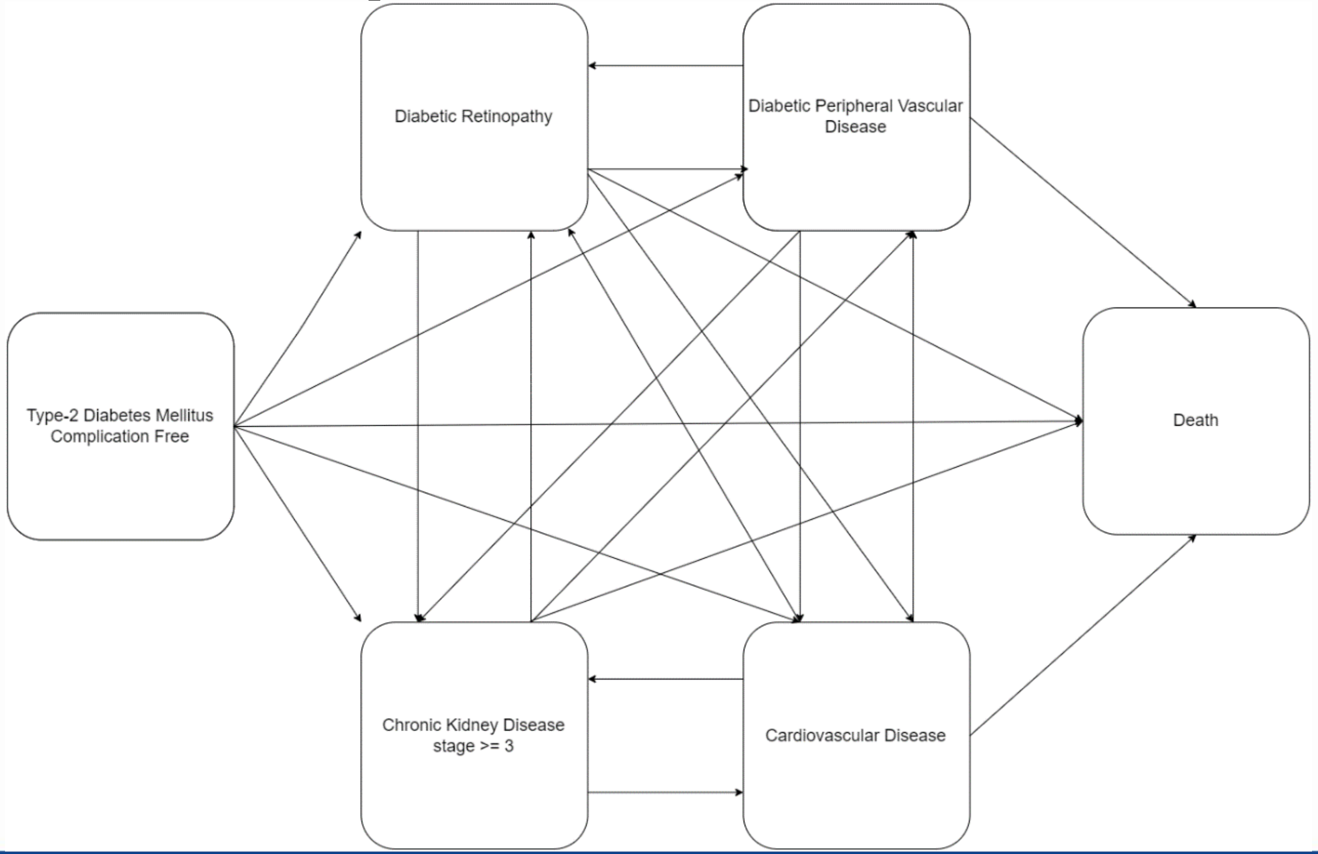
- The lumping criteria differs on each project and research question.
- For HT data warehouse project, every 180 days interval is used.
  - Categorical variables (such as comorbidities) - first record of each lump.
  - Continuous features (such as BMI) - average value
  - Except for boolean variables (such as medications) - maximum of the lump.
- For abdominal surgery data warehouse project,
  - Records from 180 days before to 90 days after the date of operation were lumped as the perioperative features.

Research projects (T2D projects)

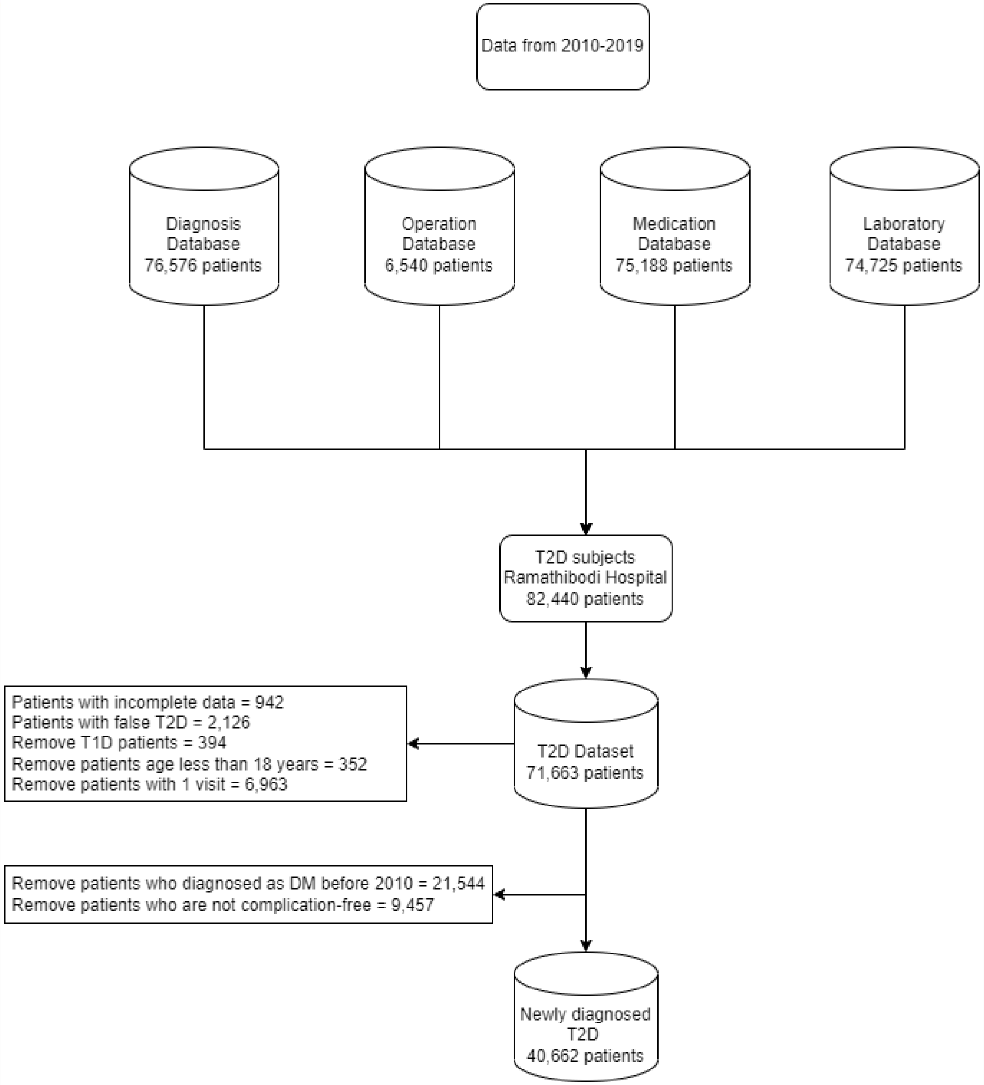
T2D team

Multistate model of T2D progression

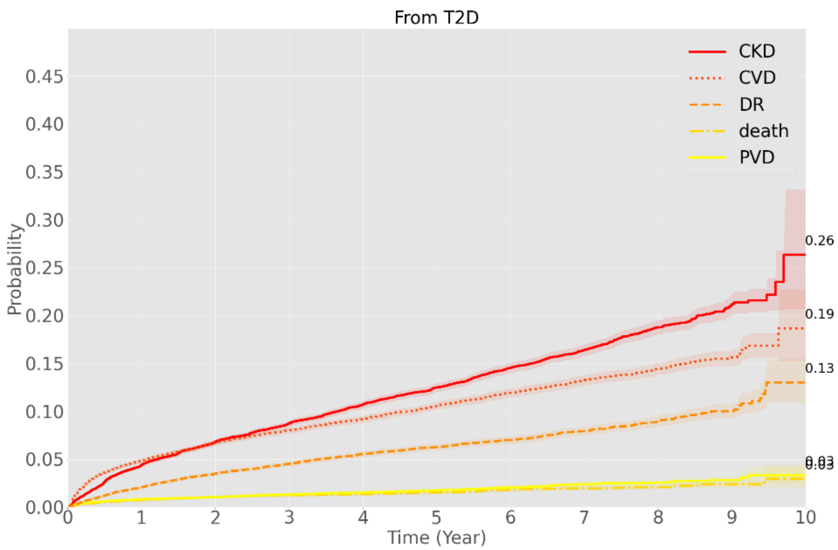
- To estimate transition probabilities

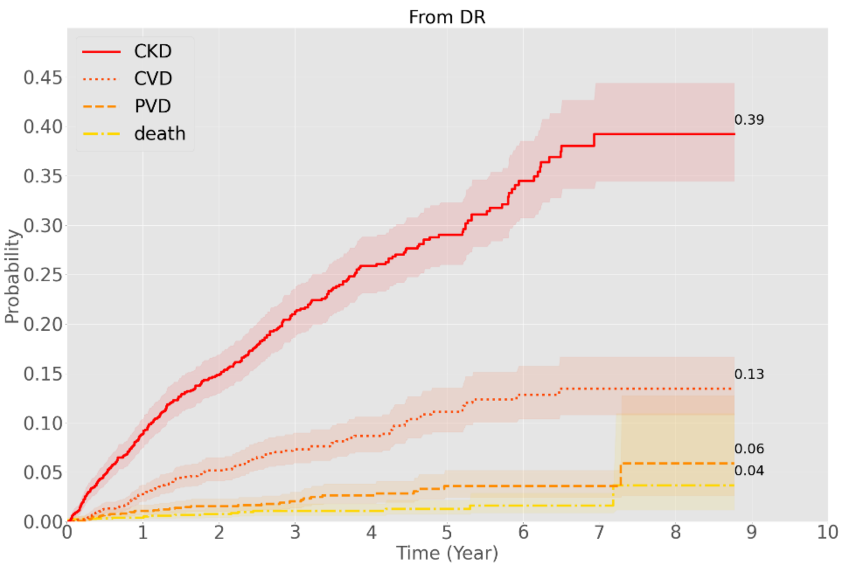


Dataflow

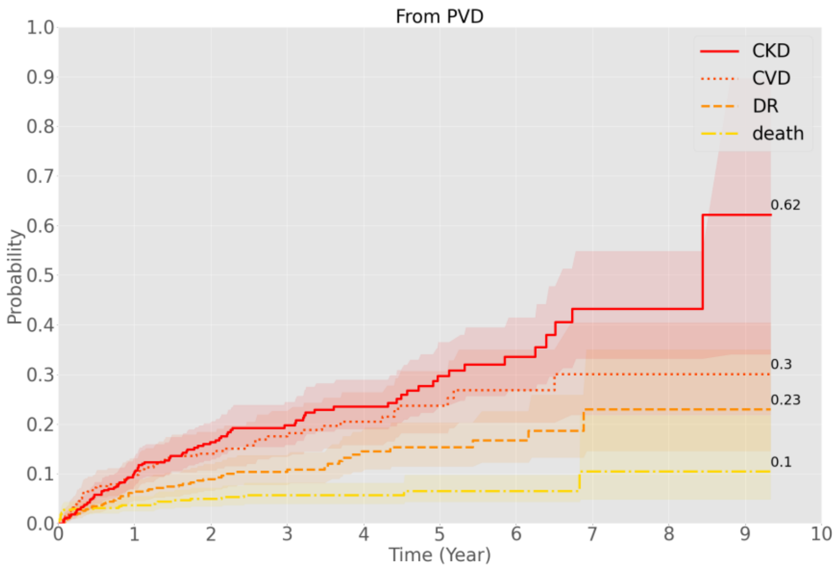

Data from 2010-2019
Diagnosis Database 76,576 patients
Operation Database 6,540 patients
Medication Database 75,188 patients
Laboratory Database 74,725 patients
T2D subjects Ramathibodi Hospital 82,440 patients
T2D Dataset 71,663 patients
Patients with incomplete data = 942
Patients with false T2D = 2,126
Remove T1D patients = 394
Remove patients age less than 18 years = 352
Remove patients with 1 visit = 6,963
Remove patients who diagnosed as DM before 2010 = 21,544
Remove patients who are not complication-free = 9,457
Newly diagnosed T2D 40,662 patients

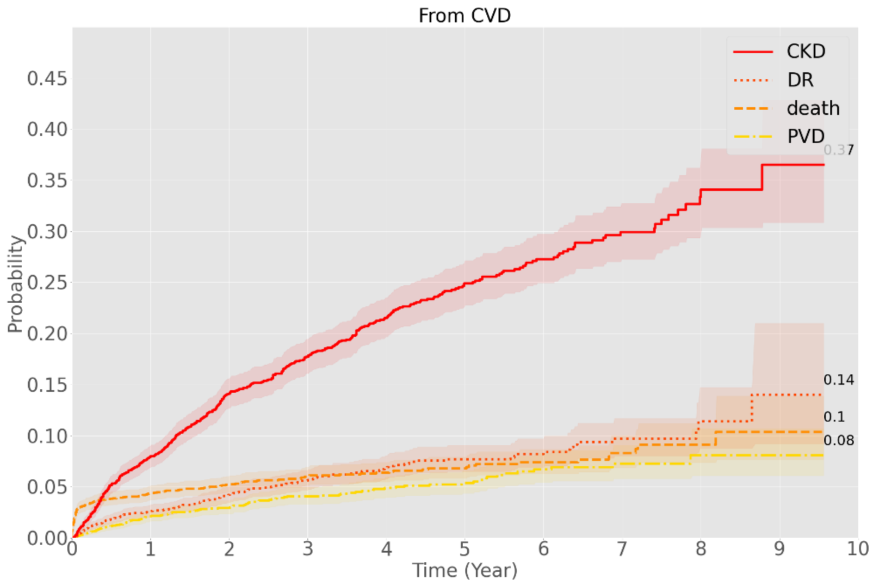
Transition Probabilities (1)





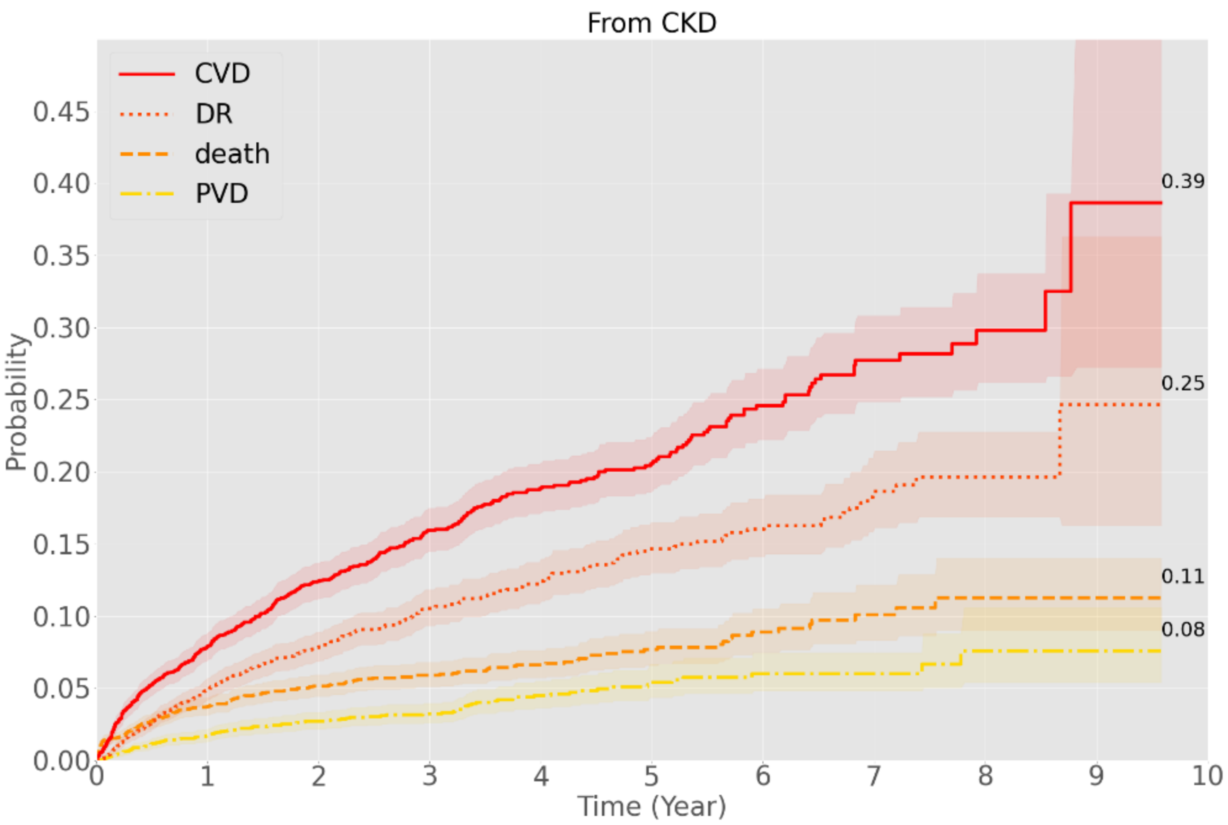
Transition Probabilities (2)





Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biostatistics

Transition Probabilities (3)



Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biostatistics

Effects of second-line antihyperglycemic drugs on the risk of chronic kidney disease: Emulating a target trial using a hospital-based cohort of Thai patients with type 2 diabetes

T2D team

Cardiovascular Diabetology: In press 23-09-2022

Wisdom of the Land

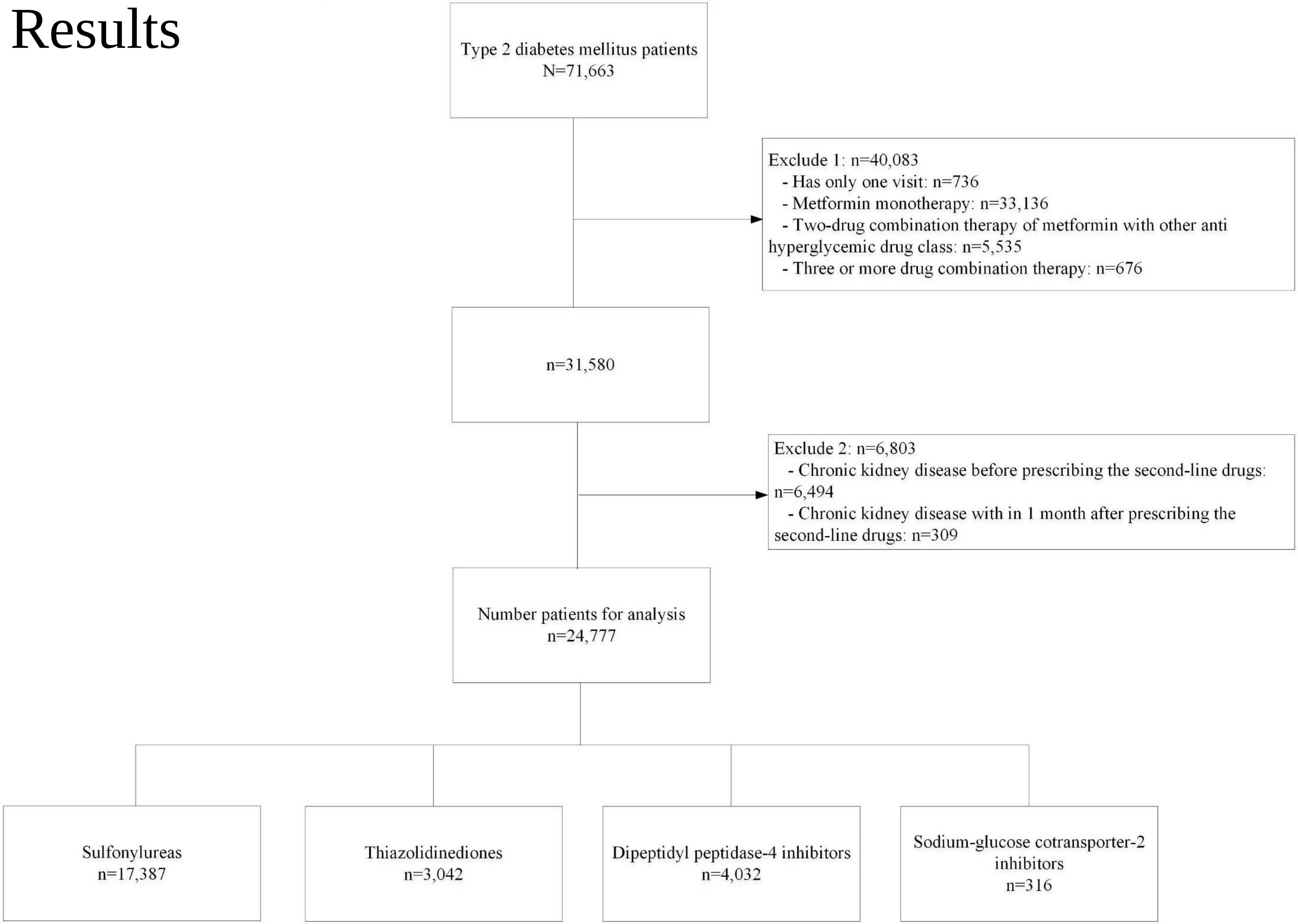
Study objective

- To assess the effectiveness of second-line antihyperglycemic drugs when added to metformin on the risk of chronic kidney disease (CKD) development in Thai patients with type 2 diabetes.

Methods

- A real-world, hospital-based, type 2 diabetes cohort was retrospectively assembled at Ramathibodi Hospital from 2010 to 2019.
- Patients who received sulfonylureas (SU), thiazolidinediones (TZD), dipeptidyl peptidase-4 inhibitors (DPP4i), or sodium-glucose cotransporter-2 inhibitors (SGLT2i), as second-line antihyperglycemic treatment were included.
- Treatment effect models with inverse probability weighting and regression adjustment (IPWRA) were used to estimate CKD risk according to treatment.

Results


Type 2 diabetes mellitus patients
N=71,663
Exclude 1: n=40,083
- Has only one visit: n=736
- Metformin monotherapy: n=33,136
- Two-drug combination therapy of metformin with other anti hyperglycemic drug class: n=5,535
- Three or more drug combination therapy: n=676
n=31,580
Exclude 2: n=6,803
- Chronic kidney disease before prescribing the second-line drugs: n=6,494
- Chronic kidney disease with in 1 month after prescribing the second-line drugs: n=309
Number patients for analysis
n=24,777
Sulfonylureas
n=17,387
Thiazolidinediones
n=3,042
Dipeptidyl peptidase-4 inhibitors
n=4,032
Sodium-glucose cotransporter-2 inhibitors
n=316

Estimation of CKD risks by intention to treat approach

Treatment	POM	Lower limit	Upper limit
SGLT2i	0.037	0.012	0.063
DPP4i	0.133	0.122	0.143
TZD	0.175	0.157	0.193
SU	0.179	0.173	0.185

Estimation of relative effects between second-line drugs: Treatment effect model with IPWRA

ATE (95% CI) →

RR (95% CI) ↓

Treatment	SU	TZD	DPP4i	SGLT2i
ITT				
SU	ref	-0.004 (-0.023, 0.014)	-0.046 (-0.059, -0.034)	-0.142 (-0.167, -0.116)
TZD	0.98 (0.87, 1.08)	ref	-0.042 (-0.063, -0.021)	-0.137 (-0.168, -0.106)
DPP4i	0.74 (0.68, 0.81)	0.76 (0.66, 0.86)	ref	-0.095 (-0.122, -0.068)
SGLT2i	0.21 (0.07, 0.35)	0.21 (0.07, 0.36)	0.28 (0.09, 0.47)	ref

Conclusions

- Our study identified 14.2%, 13.7%, and 9.5% reduced CKD risk in Thai patients with type 2 diabetes who were treated with SGLT2i compared to those treated with SU, TZD, and DPP4i, respectively, in real-world clinical data.
- Previous evidence of a reno-protective effect of SGLT2i reported in other populations is consistent with our observations in this Southeast Asian cohort.

Research on cohorts of real world data

- Treatment effectiveness
- Disease progression
- Counterfactual prediction model
- Economic evaluation
- Data analytics
  - Advance statistic models
  - Natural language processing
  - Machine learning
  - Deep learning
  - AI deployment

https://www.rama.mahidol.ac.th/ceb/CEBdatawarehouse

Data sharing for research

Data sharing

- Research team
  - Your own team + Rama co-investigators
- Acknowledgement
  - Faculty of Medicine Ramathibodi Hospital
  - Grant resource

Mahidol University
Faculty of Medicine Ramathibodi Hospital
Department of Clinical Epidemiology and Biosta










Thank you