



Mahidol University

Faculty of Medicine Ramathibodi Hospital

Department of Clinical Epidemiology and Biostatistics

Association between Ambient PM 2.5 and Dementia in Thailand

HTUN TEZA

Data Science and Clinical Informatics Division

Department of Clinical Epidemiology and Biostatistics

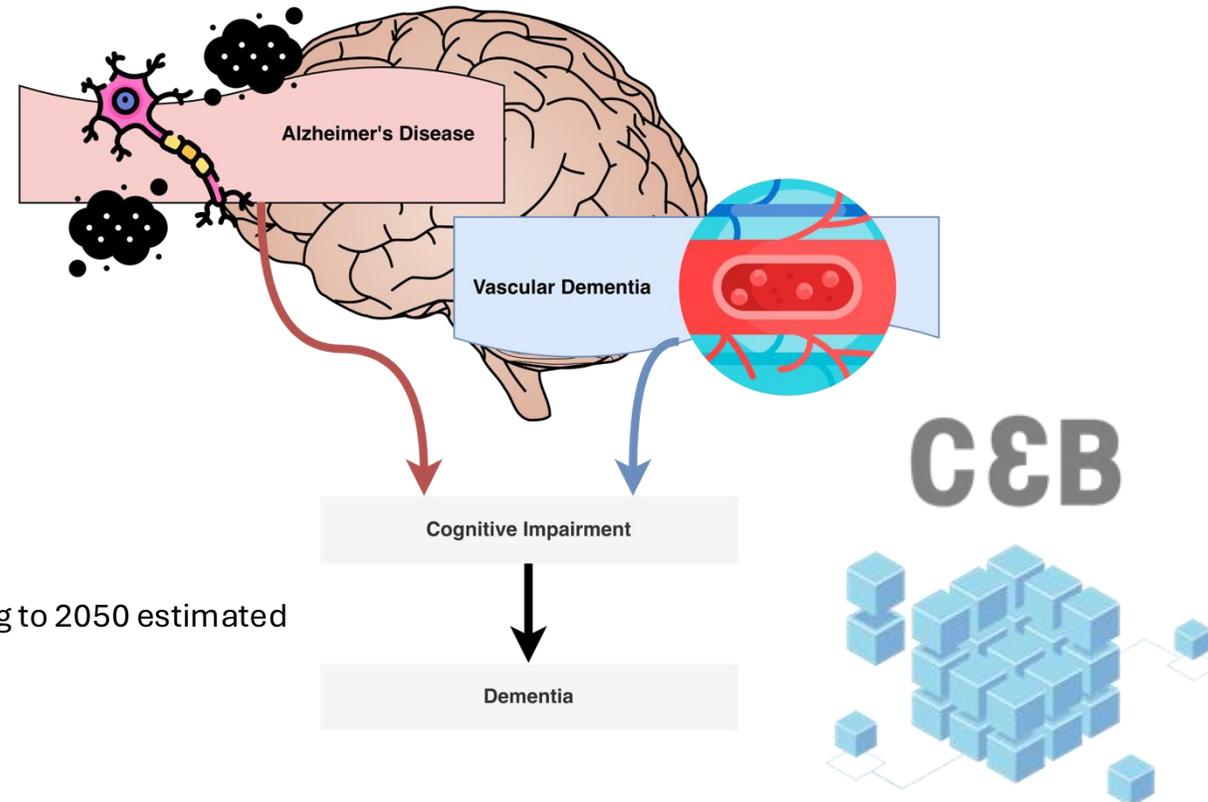




Dementia

Major Neurocognitive Disorder

- Clinical syndrome, caused by diseases that progressively destroy nerve cells
- Deterioration in cognitive functions
- Commonly, Neurodegenerative or Vascular pathology
- HICs → stabilizing or declining age-specific incidence
- LMICs → rapid increases in absolute case numbers
- Asia → worsening cardiovascular risk profiles
→ rising dementia prevalence.
- Japan → approximately 23–38% from the 1980s to the 2000s
- China → a relative annual incidence increase of around 2.9% extending to 2050 estimated

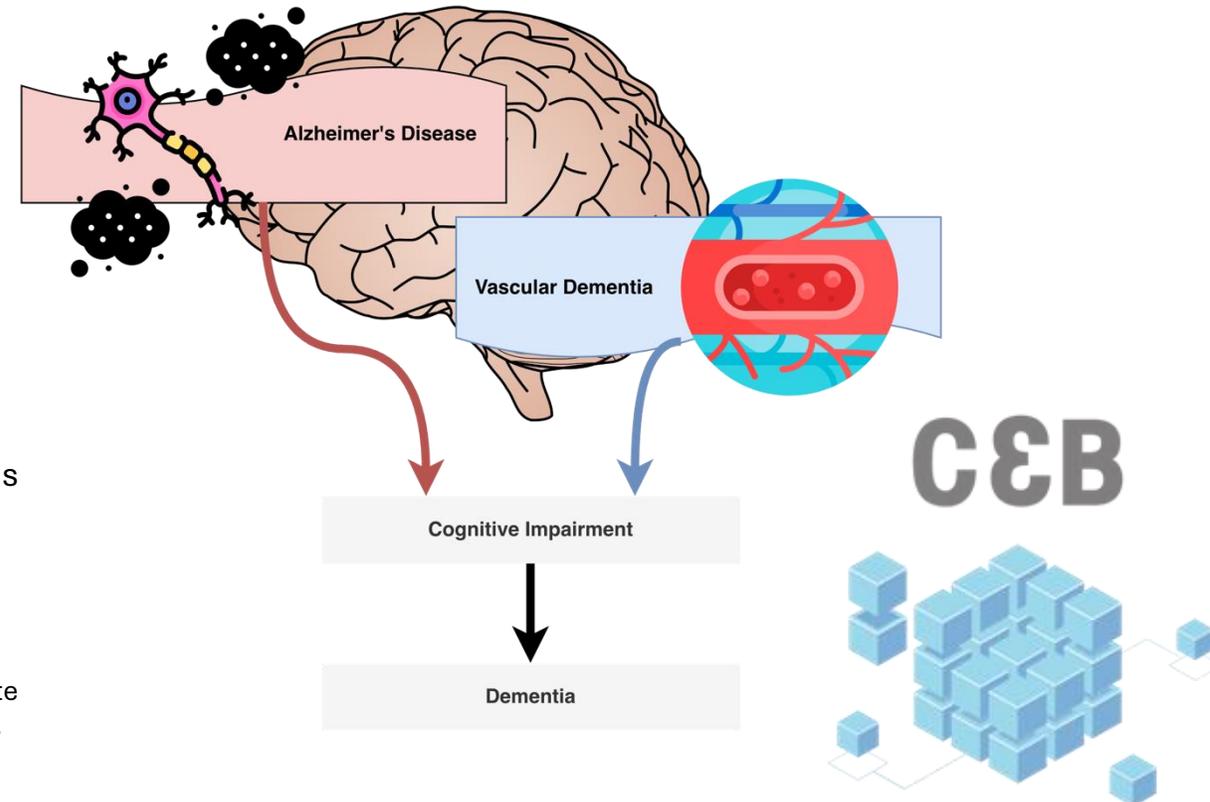




Dementia

Major Neurocognitive Disorder

- Higher long-term PM2.5 is linked to increased all-cause dementia risk,
- Pooled hazard ratios (HRs) around 1.40 (95% CI 1.23, 1.60) per 10 $\mu\text{g}/\text{m}^3$ increase reported
- PM2.5 can reach the brain via the olfactory nerve and bloodstream,
- It promotes neuroinflammation, oxidative stress, blood–brain barrier disruption, amyloid- β deposition, and tau pathology.
- PM2.5 also worsens cardiovascular disease and stroke, which themselves elevate dementia risk





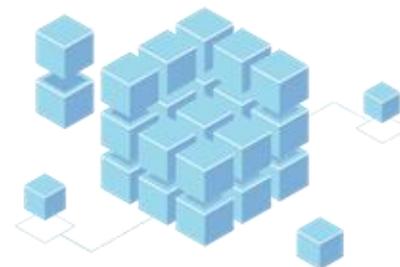
Real World Data

Dementia develops over many years

- Real-world evidence relies on data sources capable of long-term follow-up.

Electronic Health Records	Claims	Trials Data	Population Cohorts
Clinical Practice Research Datalink (CPRD), UK	Medicare, US	Systolic Blood Pressure Intervention Trial (SPRINT), US	Rotterdam Elderly Study, Netherlands
Secure Anonymized Information Linkage (SAIL) Databank, UK		Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), US	Leiden 85+ Study , Netherlands
Epic-based provider networks , US			Kungsholmen Project (SNAC-K), Sweden
Veterans Health Administration records , US			Newcastle 85+ Study, UK

C&B





Real World Data

Geographical Gap

- Majority of studies are done in high-income countries (HICs) in
 - North America and Europe, or
 - specific East Asian jurisdictions like Taiwan.
- These settings feature mature longitudinal infrastructures and higher hypertension control rates.
- Thailand → Middle-Income Country (MIC) landscape with worsening cardiovascular risk profiles
- A Southeast Asian or Thai dataset therefore offers critical geographic diversity



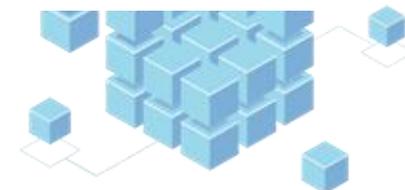


Claims Data

National Health Security Office

- Visits Data : Diagnoses and Procedures conducted
 : One record per visits
- Prescription Data : Medication prescribed for the visits
 : Multiple data entries per visit
- Claims data have unbalanced visits frequencies.
- Sick people visits more frequently while less in good health.
- The visits will be regularized into annual data.

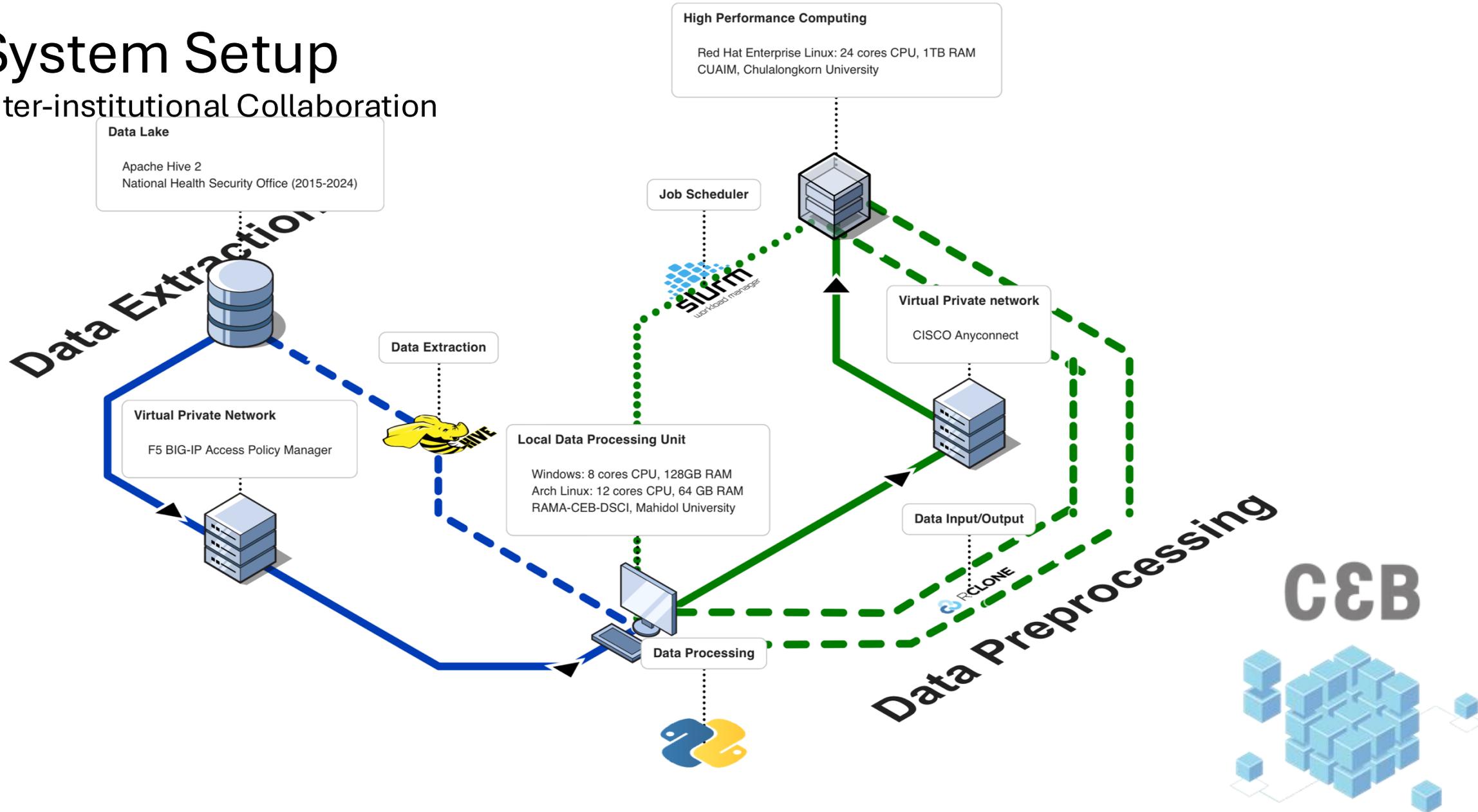
Fiscal Year	Visits (Millions)	Prescriptions (Millions)
2559	123.5	232.0
2560	128.4	234.2
2561	122.0	230.5
2562	137.7	241.3
2563	126.7	231.9
2564	121.9	219.4
2565	125.4	227.1
2566	107.0	232.9
2567	106.2	248.1
IPD	148.3	687.0





System Setup

Inter-institutional Collaboration

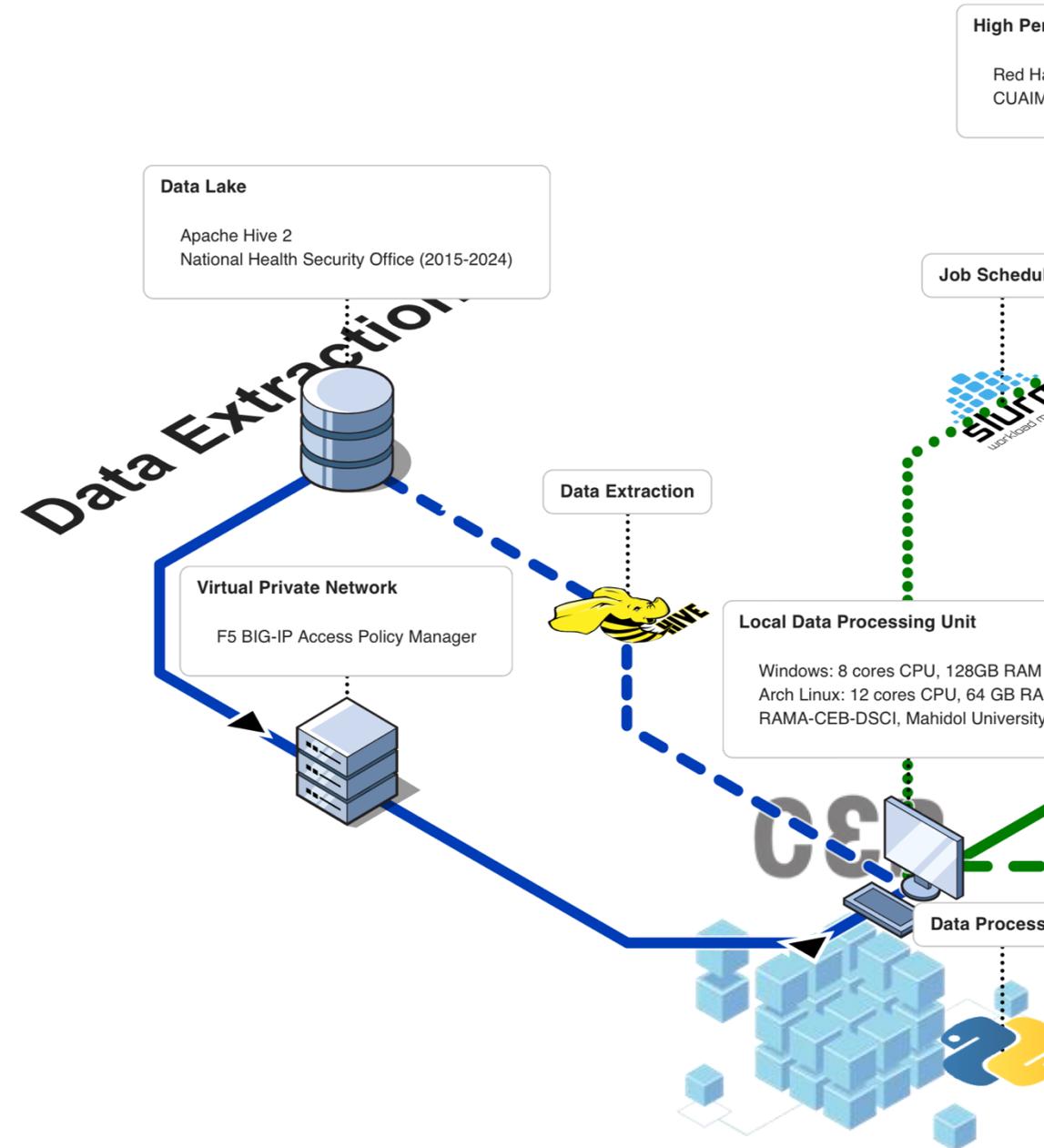


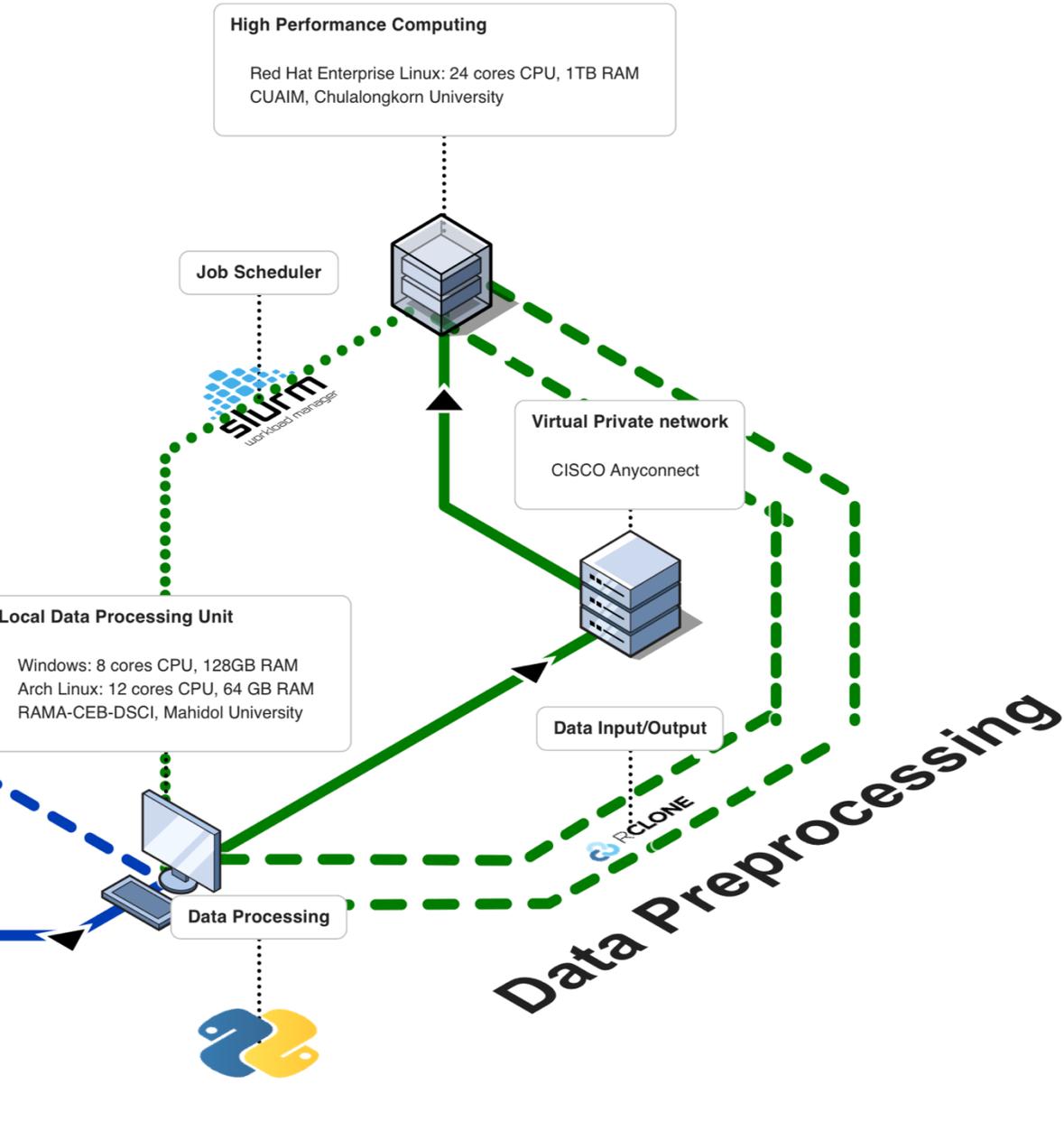


System Setup

Quantity (Data Scale)

- National claims data at tens to hundreds of GB per file (visit + prescription layers)
- Longitudinal coverage across multiple fiscal years → multi-terabyte total volume
- Does not fit into memory or storage of standard consumer machines
- Requires distributed storage and chunk-based processing

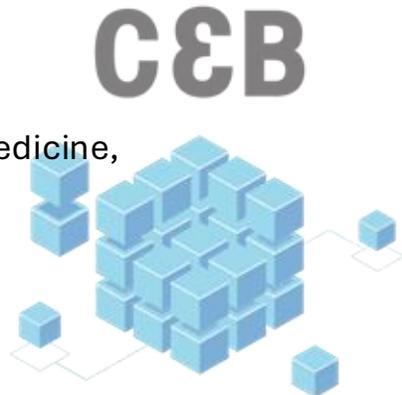




System Setup

Computation (Processing Requirements)

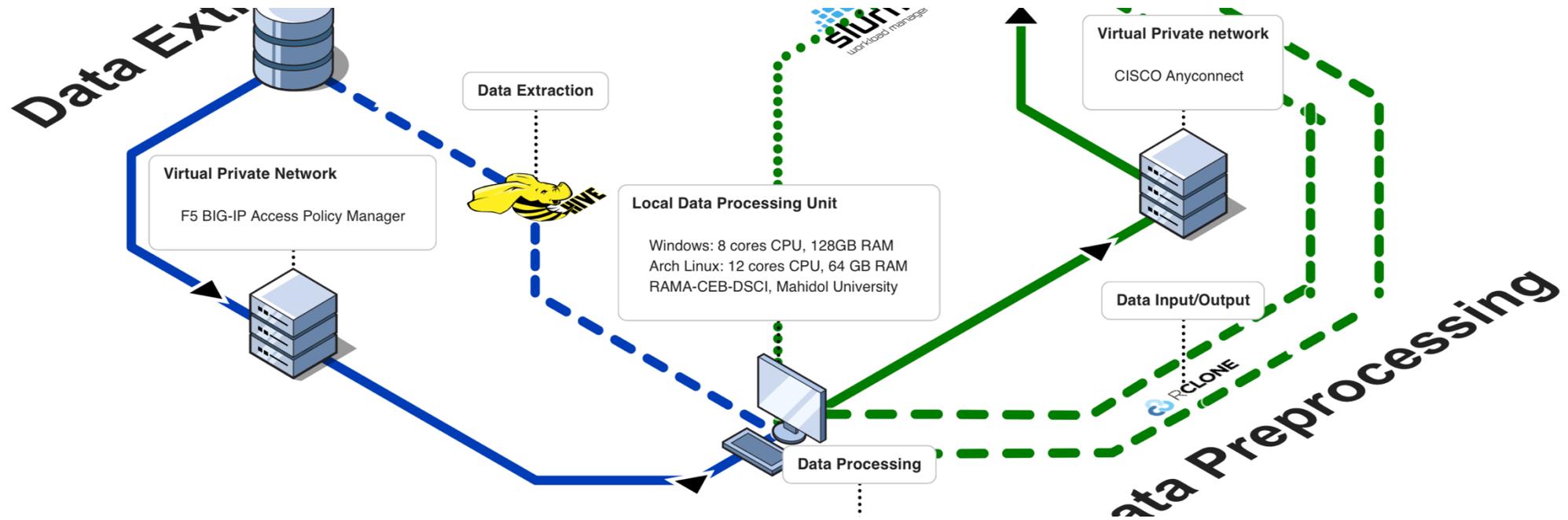
- Complex preprocessing:
 - Visit-prescription linkage (one-to-many structure)
 - Temporal aggregation and cohort construction
- High RAM + parallel compute required for:
 - Grouping / joins at national scale
 - Iterative cohort refinement
- Local machines (≤ 128 GB RAM) are insufficient for full pipeline execution
- High Performance Computing provided by Center for AI in Medicine (CU-AIM), Faculty of Medicine, Chulalongkorn University





System Setup

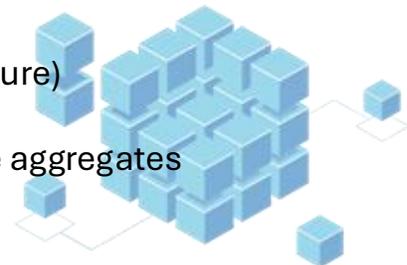
Security (Data Governance)



- Data contains person-level health records
- Strict access control enforced:
 - No raw data movement outside secure environment

- All transfers and computation via:
 - Private VPN tunnels
 - Controlled HPC access (CU-AIM infrastructure)
- Analysis outputs are restricted to non-identifiable aggregates

C&B



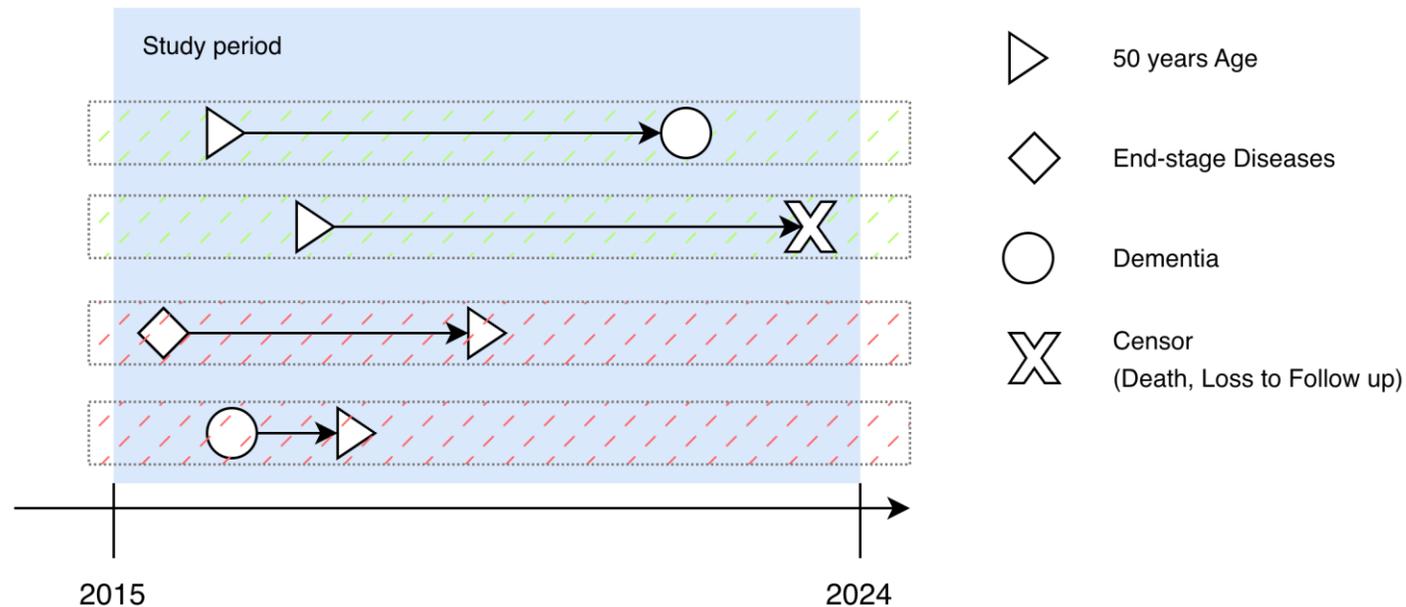


Population: Cohort Criteria

- Adults aged 50 years or older observed in NHSO database, during the study period (2010–2024).

Exclusion Criteria

- those with end-stage chronic diseases (such as end-stage kidney failure, liver failure, or terminal cancer), and
- those diagnosed with dementia before the age of 50.



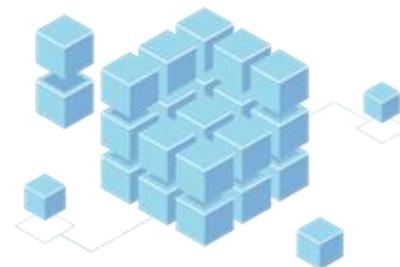


Outcome: Dementia

Criteria

Dementia	Criteria
Alzheimer's dementia	F000, F001, F002, F009, G300, G301, G308, G309
Vascular dementia	F010, F011, F012, F013, F018, F019
Unspecified dementia	Unspecified dementia: F03, Use of Donepezil, Rivastigmine, Galantamine, Memantine
Other dementia	Dementia with Lewy Bodies: G3183, Frontotemporal Dementia: G310, Multiple System Atrophy: G232, G233, Dementia in other diseases classified elsewhere: F02*, Progressive Supranuclear Palsy: G231, Corticobasal degeneration: G3185, HIV Dementia: B220
Mixed Dementia	Coexistence of two or more dementia pathologies recorded in the same individual

B

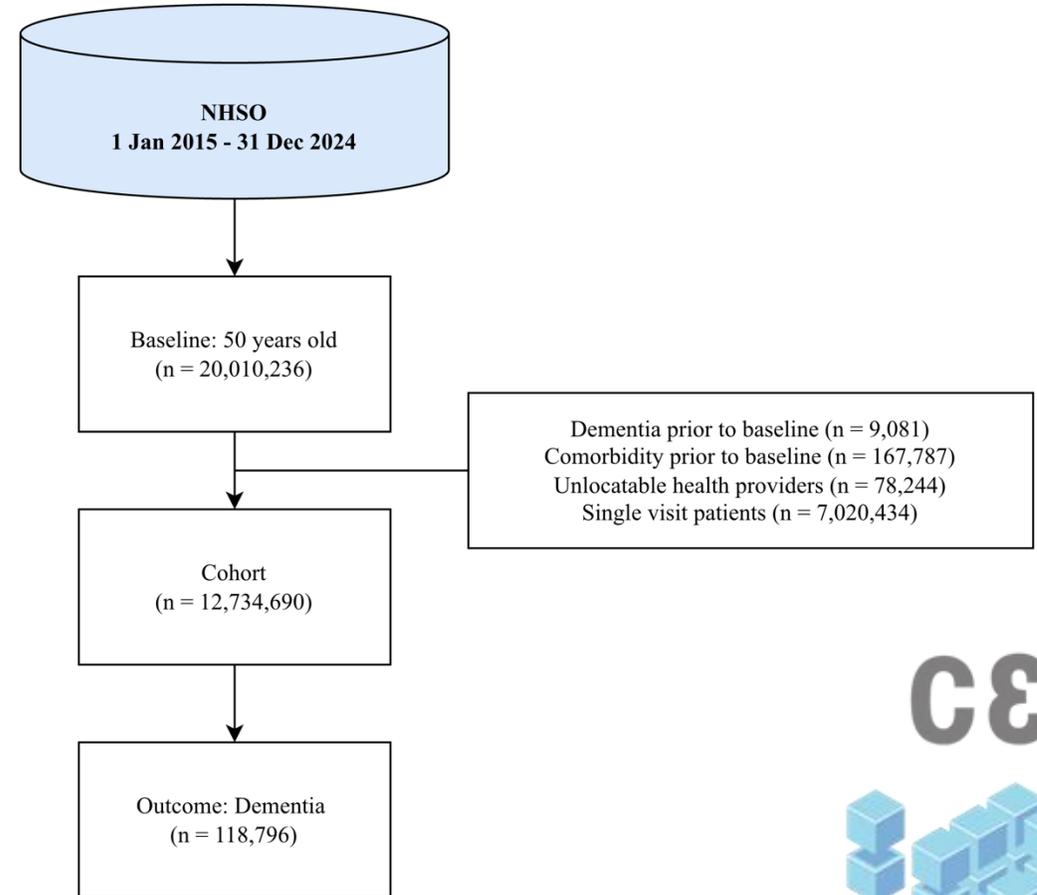




Cohort Identification

- For 12,734,690 subjects with total 74,958,678.278 person-years
- 118,796 number of subjects develop dementia.
- Incidence rate : 1.585 (1.576, 1.594) per 1000 person-year.

	Follow-up period (years)	Observations
Mean	5.89 (2.91)	5.49 (3.04)
Median	6.70	5.00
Range	0.003 – 9.94	2.00 – 12.00
IQR	3.38-8.67	3.00 – 9.00





Dementia

Criteria

Dementia	N	Cohort = 12,734,690	Dementia = 118,796
Alzheimer's disease	49,163	0.39 %	41.38 %
Vascular dementia	11,188	0.09 %	9.42 %
Unspecified dementia	28,271	0.22 %	23.8 %
Other dementia	499	0.004 %	0.43 %
Mixed Dementia	29,675	0.23 %	24.98 %

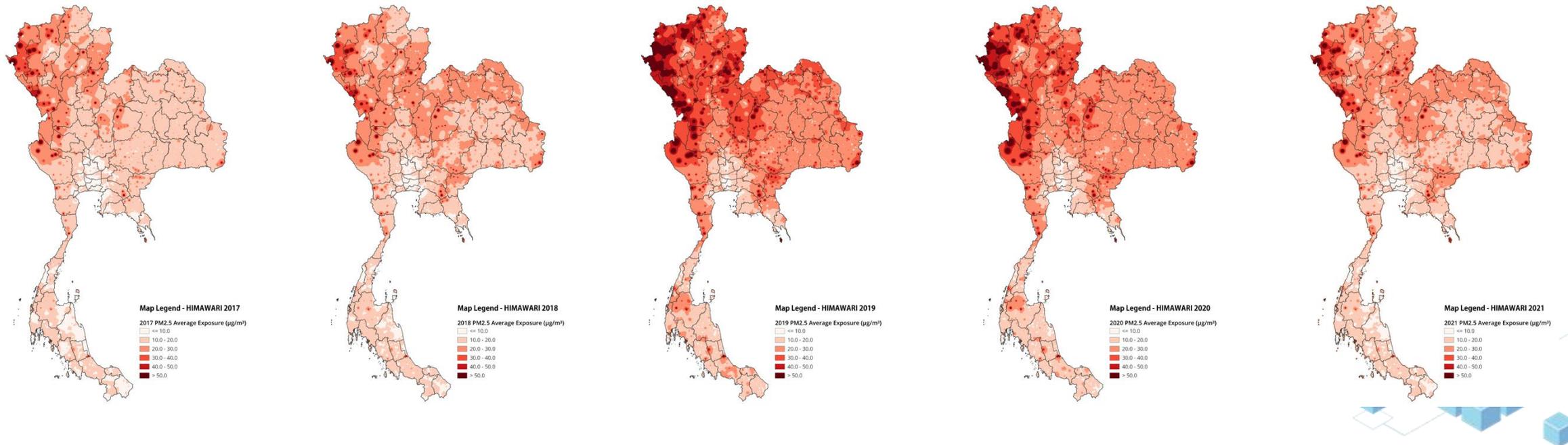




Exposure: PM2.5 Data

Himawari-8-AHI Aerosol Optical Depth (AOD) Level-3 data

- Preprocessed for hourly data for PM 2.5 by GISTDA at Subdistrict (tambon/khwaeng) level
- (Daily) Total PM 2.5 and (Daily) Average PM 2.5 for the area (Raw and Weighted for available measurements)
- Further processed for Annual Average and Annual Maximum at district level (amphoe/khet)





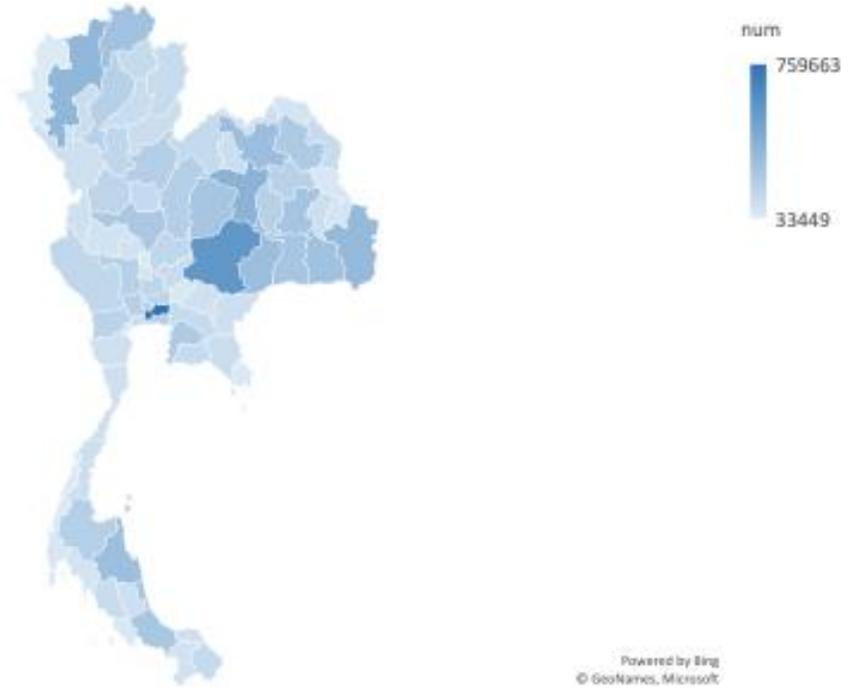
Location Data

National Health Security Office

- Patient residential information were not made available with investigators due to PDPA.

From each visit or prescription data,

- The location of hospital where the patient's health coverage is assigned to, is inferred as patient's residential location.
- It can be mapped to province (Changwat), district (amphoe/khet) and subdistrict (tambon/khwaeng) levels.



Code	No. Subjects	Name	Province	District	Subdistrict	NHSO ZONE
10685	51,784	รพ.สมุทรปราการ	สมุทรปราการ	เมืองสมุทรปราการ	ปากน้ำ	6





Analysis: Cox Proportional Hazards Model

- Individual level – Demographics and Cardiovascular profile
- District level – PM2.5 measurements
- Time level – Time varying covariates at annual level
- PM 2.5 at current time, or cumulative exposure (since 50 years of age) until current process.

Hashed ID	Visit Date	Gender	Age	Atrial Fibrillation	PM 2.5 (Annual Max)	Cumulative PM 2.5 (Annual Max)	Row Type
A	2015 – 10 – 08	F	50	0	102.08	102.08	Index
A	2016 – 09 – 27	F	51	1	234.14	366.22	Follow up
A	2017 – 09 – 27	F	52	1	182.99	519.21	Follow up
A	2018 – 09 – 25	F	53	1	188.09	707.30	Follow up
A	2019 – 08 – 13	F	54	1	275.69	982.99	Follow up
A	2020 – 09 – 17	F	55	1	277.11	12060.10	Dementia

B





[Preliminary] Univariate Analysis

PM 2.5 Data

- [Preliminary Analysis]
- Random effects of geographical difference is estimated by random intercept at Amphoe level.
- Random intercepts have median variance of 0.2435 (IQR 0.2418 – 0.2498; range 0.2398 – 0.2542).

PM 2.5	Annual Aggregation	At current process			Cumulative until current process		
		Univariate HR	P Value	Time (Hours)	Univariate HR	P Value	Time (Hours)
Standard PM2.5 Measure	Annual Average	0.9936 (0.9929,0.9944)	< 0.0001	0.74	0.9974 (0.9973,0.9975)	< 0.0001	0.76
	Annual Maximum	0.9997 (0.9997,0.9997)	< 0.0001	0.86	0.9998 (0.9998,0.9998)	< 0.0001	0.9
Average PM2.5	Annual Average	0.9665 (0.9636,0.9694)	< 0.0001	0.75	0.988 (0.9875,0.9886)	< 0.0001	0.82
	Annual Maximum	0.9992 (0.9991,0.9992)	< 0.0001	0.9	0.9995 (0.9995,0.9995)	< 0.0001	1.02
Weighted Standard PM2.5 per hour	Annual Average	0.9971 (0.9956,0.9985)	< 0.0001	0.91	0.9975 (0.997,0.998)	< 0.0001	0.85
	Annual Maximum	0.9999 (0.9999,0.9999)	< 0.0001	0.88	0.9999 (0.9999,0.9999)	< 0.0001	0.8
Weighted Average PM2.5	Annual Average	0.9996 (0.9993,0.9999)	0.01446	0.8	0.9995 (0.9994,0.9996)	< 0.0001	0.9
	Annual Maximum	0.9999 (0.9999,0.9999)	0.00050	0.87	0.9999 (0.9999,0.9999)	< 0.0001	0.88





[Preliminary] Univariate Analysis

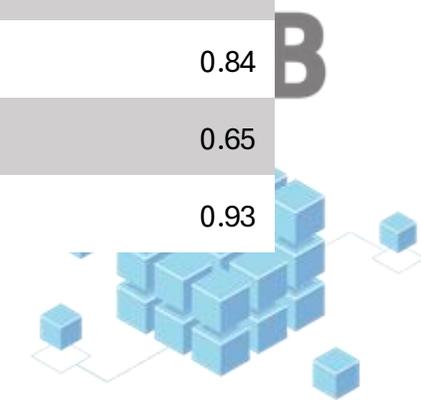
Demographics and Comorbidity Data

Variable		Univariate HR	P Value	Time (Hours)
Gender	Male	1		
	Female	1.0678 (1.0554,1.0804)	< 0.0001	0.65

Variable		Univariate HR	P Value	Time (Hours)
Stroke	None	1		
	Other	3.0222 (2.9634,3.0822)	< 0.0001	0.96
	Ischemic	3.0018 (2.9396,3.0653)	< 0.0001	0.96
Traumatic Brain Injury		2.4595 (2.3786,2.5431)	< 0.0001	1.11

- Dyslipidemia – Pure hypercholesterolemia, Pure hyperglyceridemia, Mixed
- Other Stroke – Hemorrhagic Stroke, Stroke sequelae of cerebrovascular disease, Other cerebrovascular conditions, Stroke non-specified

Variable	Univariate HR	P Value	Time (Hours)
Age, years	1.0787 (1.0782,1.0793)	< 0.0001	0.9
Hypertension	0.9784 (0.9664,0.9905)	0.00050	0.66
Type 2 Diabetes	0.9227 (0.9104,0.9351)	< 0.0001	0.77
Chronic Kidney Disease	1.2444 (1.2247,1.2645)	< 0.0001	0.75
Coronary Artery Disease	1.424 (1.3921,1.4566)	< 0.0001	1.15
Peripheral Vascular Disease	1.0304 (1.0072,1.0541)	0.00975	0.65
Heart Failure	1.4854 (1.444,1.5281)	< 0.0001	1.33
Atrial Fibrillation	1.6373 (1.5899,1.686)	< 0.0001	0.84
Dyslipidemia	0.9619 (0.9496,0.9744)	< 0.0001	0.65
Obesity	0.5294 (0.4919,0.5697)	< 0.0001	0.93





[Preliminary] Univariate Analysis

Demographics and Comorbidity Data

Variable	Univariate HR	P Value	Time (Hours)
COPD	1.2501 (1.2134,1.2879)	< 0.0001	0.64
Rhinitis	0.9771 (0.9485,1.0067)	0.12890	0.79
Psychiatric Disorders			
Anxiety	2.1263 (2.0745,2.1793)	< 0.0001	0.9
Bipolar affective disorder	4.7018 (4.3799,5.0474)	< 0.0001	0.8
Depression	3.4452 (3.3699,3.5221)	< 0.0001	0.85
Schizophrenia	3.5608 (3.4537,3.6713)	< 0.0001	0.7

Variable	Univariate HR	P Value	Time (Hours)
Audio-Visual Impairment	1.3955 (1.3687,1.4228)	< 0.0001	0.71
Substance Abuse (Overall)	0.8234 (0.8035,0.8438)	< 0.0001	0.97
Sleep Disorders (overall)	1.7501 (1.7006,1.8009)	< 0.0001	0.73
STD (HIV + Syphilis)	0.4258 (0.3895,0.4656)	< 0.0001	0.82

- Substance Abuse – Alcohol, Opioid, Cannabis, Sedatives or hypnotics, Cocaine + Other stimulants, Hallucinogen, Tobacco, Volatile solvent, Other substances
- Sleep Disorders – Insomnias, Sleep-wake cycles, Sleep apnea

- Rhinitis – Vasomotor and allergic rhinitis, Chronic rhinitis, nasopharyngitis and pharyngitis
- Audio-Visual Impairment - Hearing impairment, Visual impairment (Blindness, binocular; Severe, binocular; Moderate, binocular; Mild, binocular)





Continuing Works

- Paradoxical observations are being addressed
- Careful modelling is required for PM 2.5 index.
- Additional feature specifications
 - Hypertension, Diabetes : (yes/ no) \rightarrow (yes with medication/ yes without medication/ no)
- Confounding features should be taken into account, such as cardiovascular conditions, dyslipidemia and obesity.
- Several features might be under-coded, such as substance abuse and obesity.

