Key Achievements from the IMPRESIVE Project

RWE investigator meeting on July 25th

Steven Chin Yao Shen

Population Heath Data Center, National Cheng Kung University

成大群體健康數據中心
Population Health Data Center NCKU

IMPRESIVE

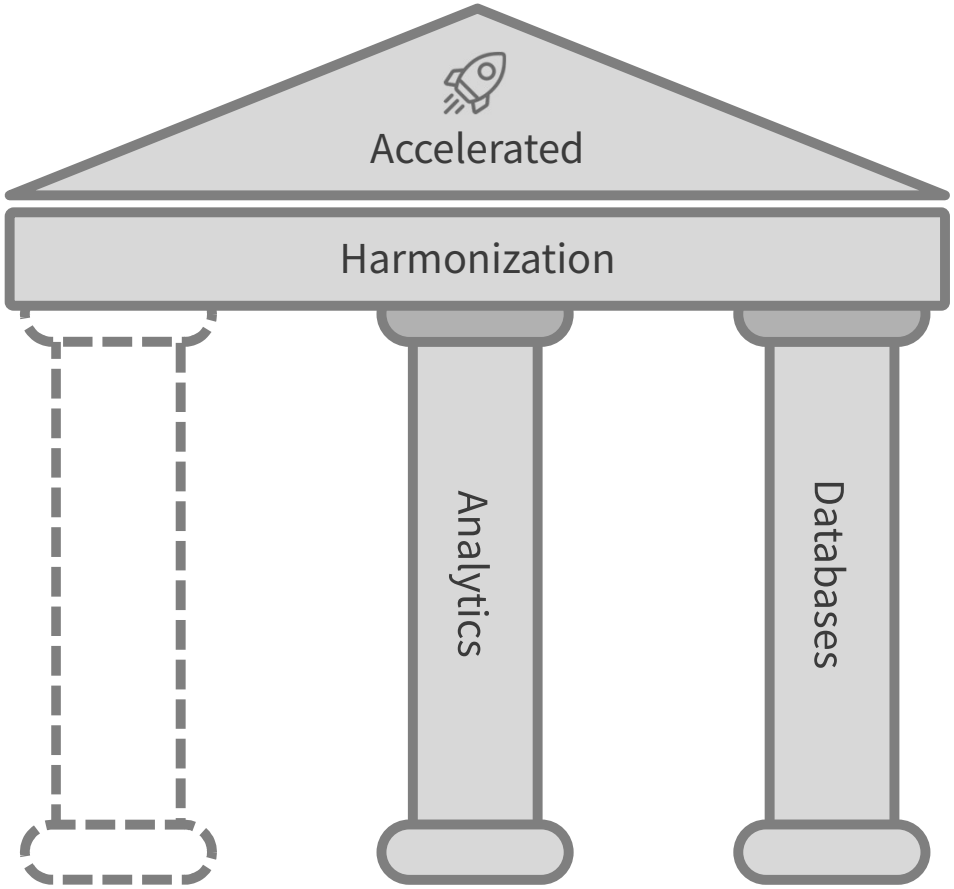
International **M**ulti-database study **PRE**paredness: Databases **S**tandardization, **I**ntegration and **V**isualization for timely **E**valuation on Disease and Treatment

Overview of BTA, IMPRESIVE 1-3 projects

	Common Protocol	Common Data Model (CDM)		
		OMOP CDM	Sentinel CDM	Sentinel CDM
Project name	BTA Project			
Target Diseases	Breast, prostate and lung cancer with bone metastasis			
Objectives	Utilization pattern and effectiveness of bone targeting agents			

Distributed Data Networks

Items	Common Protocol	Common Data Model
Data infrastructure investment	+	+++
Site-specific programming effort	+++	+
Perform tailored analysis to individual databases	+++	++
Develop reusable tools	+	+++
Consistent analysis across databases	++	+++
Study specific data management and cleaning	+++	++
Duration for study-specific analysis	++	+
Marginal cost per study	++	+

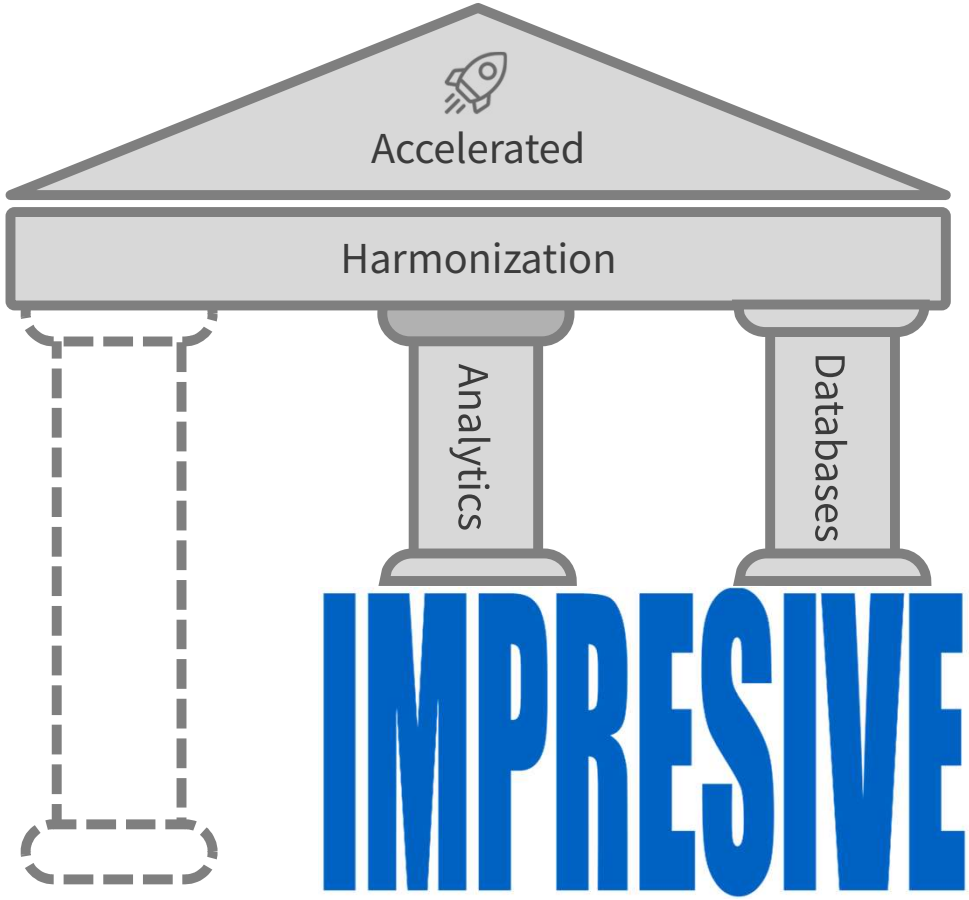


Goal: Accelerated evidence generation and application

Method: Harmonization is a key enabler of efficiency, consistency, and effective communication.

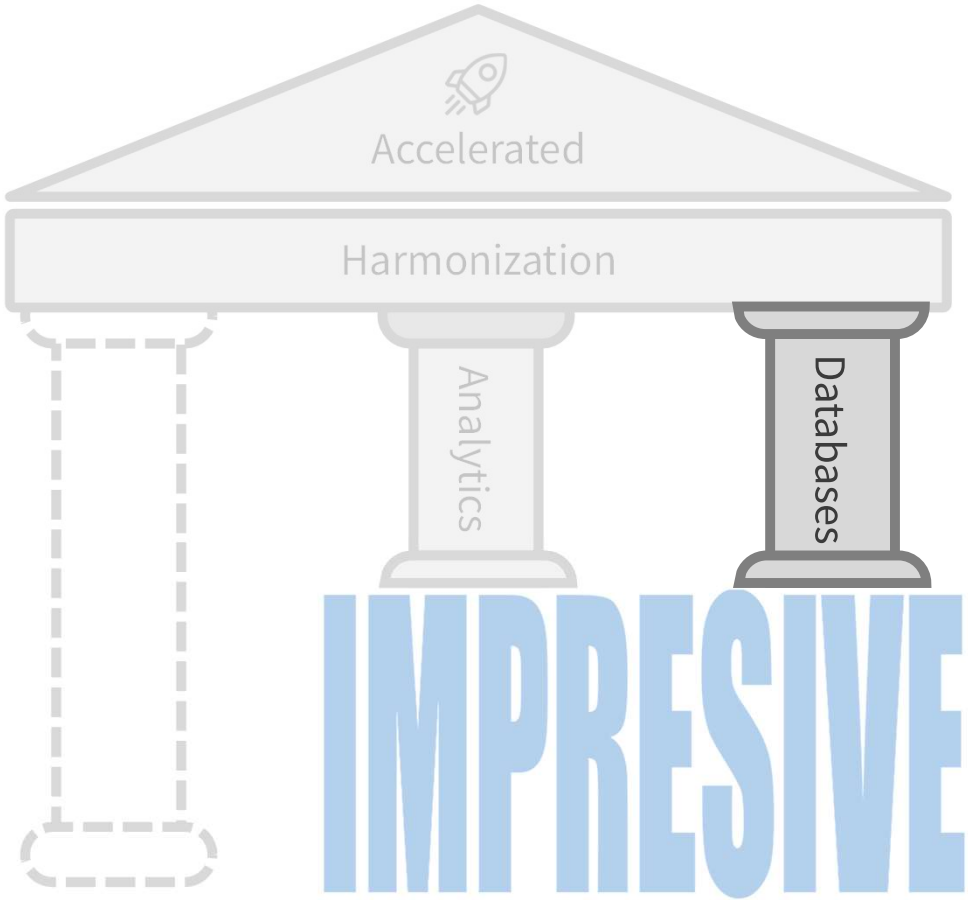
The 3-Pillar Model of Harmonization:

- **Database** – Standardization for consistent and interoperable data inputs
- **Analytics** – Reproducible and scalable execution of real-world studies
- [**To Be Revealed**] – will be unveiled at a later stage


Accelerated
Harmonization
Analytics
Databases
IMPRESIVE

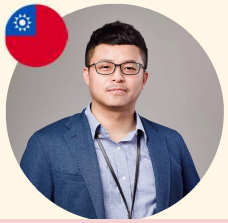
Overview of BTA, IMPRESIVE 1-3 projects

	Common Protocol	Common Data Model (CDM)		
		OMOP CDM	Sentinel CDM	Sentinel CDM
Project name	BTA Project	IMPRESIVE 1.0	IMPRESIVE 2.0 & 3.0	
Target Diseases	Breast, prostate and lung cancer with bone metastasis	Atherosclerotic cardiovascular disease (ASCVD)	Atherosclerotic cardiovascular disease (ASCVD)	Atopic dermatitis (AD) or prurigo nodularis (PN)
Objectives	Utilization pattern and effectiveness of bone targeting agents	Baseline characteristics and recurrent risk in different risk groups	Result consistency among different CDMs and different follow-up periods	Prevalence, baseline characteristics, and prescribing pattern


Accelerated
Harmonization
Analytics
Databases
IMPRESIVE

Collaboration

Academy: Claim Data



Prof. Edward Lai

National Health Insurance Research Database (NHIRD)

Prof. Ju-Young Shin

1. National Health Insurance Service Database (NHIS)
2. Health Insurance Review & Assessment Service Database (HIRA)

Prof. Hiraku Kumamaru

As consultant
Japan DeSC (provided by Amgen)

Technical Support

Prof. Wei-Sheng Wu

Department of Electrical Engineering, NCKU

Academy: Electronic Medical Records

Dr. Shih-Chieh Shao

Chang Gang Research Database (CGRD)

Prof. Celine Chui **Prof. Shirley Li**

Clinical Data Analysis and Reporting System (CDARS)

Clinical Support

Dr. Chaw-Ning Lee

Department of Dermatology, NCKUH

Dr. Cheng-Yang Hsieh

Department of Neurology, Sin-Lau Hospital

Established Disease Cohorts within the CDM Framework

Hospitalized ASCVD Cohort

NHIRD (2016 – 2022)	**CGRD** (2016 – 2023)	**NHIS** (2019 – 2023)	**CDARS** (2005 – 2019)	**DeSC*** (2014 – 2023)
N = **569,553**	N = **389,517**	N = **2,822,882**	N = **221,258**	N = **1,660,393**
Female: 194,246 (34%)	Female: 167,984 (43%)	Female: 1,225,655 (43%)	Female: 114,545 (52%)	Female: 805,939 (49%)

Atopic Dermatitis Cohort

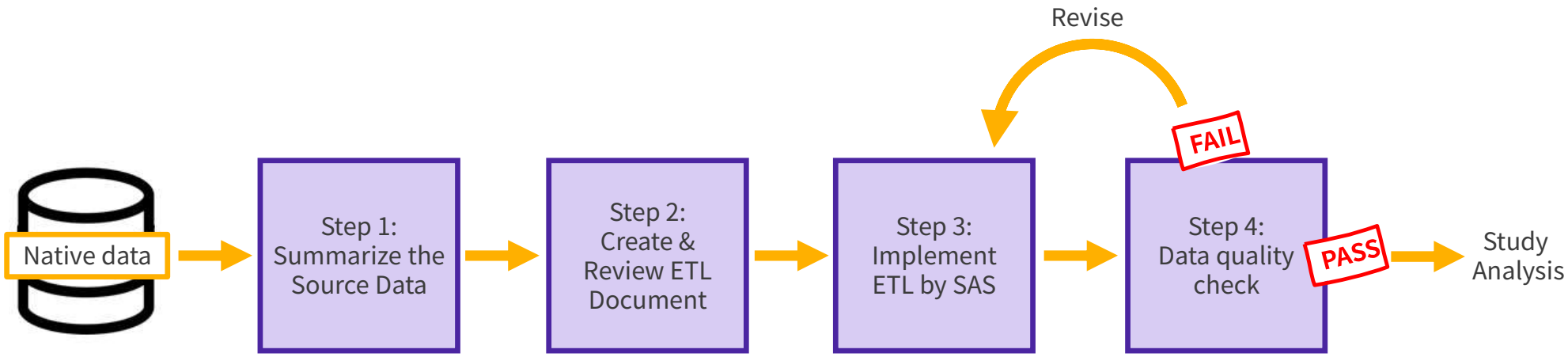
NHIRD (2017 – 2021)	**CGRD** (2016 – 2023)	**HIRA** (2019 – 2023)	**CDARS** (2017 – 2023)	**DeSC*** (2014 – 2023)
N = **491,767**	N = **65,320**	N = **3,229,218**	N = **4,585**	N = **16,26,398**
Female: 268,235 (54%)	Female: 31,020 (48%)	Female: 1,882,351 (58%)	Female: 2,342 (51%)	Female: 891,018 (55%)

Prurigo Nodularis Cohort

NHIRD (2017 – 2021)	**CGRD** (2016 – 2023)	**HIRA** (2019 – 2023)	**CDARS** (2017 – 2023)	**DeSC*** (2014 – 2023)
N = **75,081**	N = **4,103**	N = **420,379**	N = **6,350**	N = **28,554**
Female: 31,020 (41%)	Female: 1,605 (39%)	Female: 206,212 (49%)	Female: 3,910 (62%)	Female: 14,191 (50%)

* Under conversion

Extract, Transform & Loading (ETL) Process



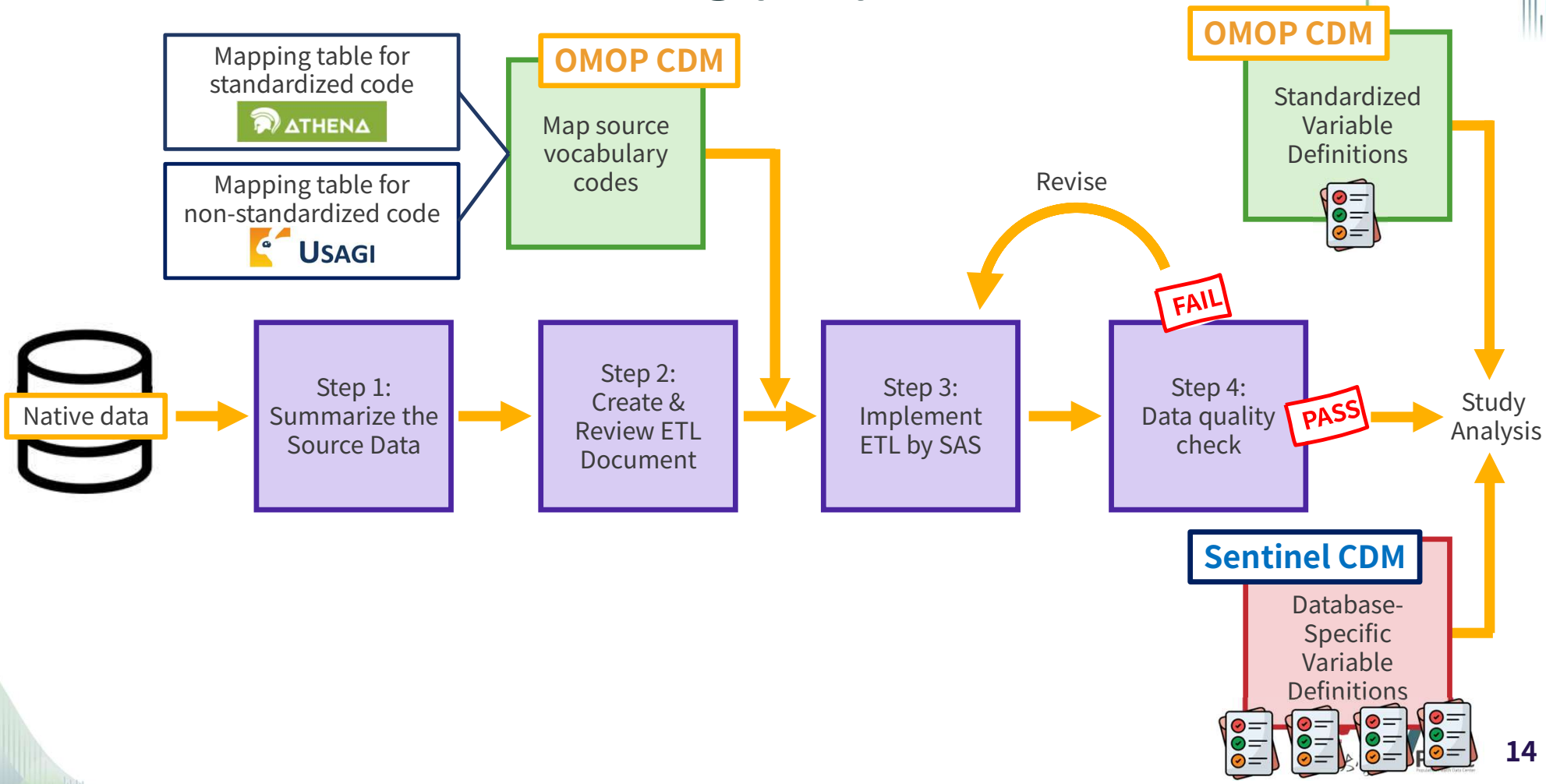
Assessment for Common Data Model Conversion

Red= Not available, Yellow= requires additional data or calculation, Green= Available for conversion

Target Tables	Required Column	Taiwan- NHIRD	Korea- NHIS/HIRA	Japan- DeSC	Hong Kong- CDARs	Taiwan- CGRD
DEMOGRAPHIC	Person ID					
	Sex					
	Year of Birth					
	Race	Assume all Taiwanese	Assume all Korean	Assume all Japanese	Assume all Chinese	Assume all Taiwanese
ENROLLMENT	Person ID					
	Date of enrollment					
	Date of disenrollment	†	†		†	†
ENCOUNTER	Person ID					
	Type of healthcare visits					
	Start date of the visit					
	End date of the visit		$			
DIAGNOSIS	Person ID					
	Diagnosis code	ICD9/10CM	ICD10CM/KCD-7	ICD10CM		ICD9/10CM
	Diagnosis date					
PRESCRIBING	Person ID					
	Drug code	Domestic code	Domestic code	Domestic code	Domestic code	Domestic code
	Start date of the drug					
	End date of the drug	‖	‖	‖		‖
PROCEDURE	Person ID					
	Procedure Code	ICD & Domestic code	Domestic code	Domestic code		ICD & Domestic code
	Date of Procedure					
LABORATORY RESULT	Person ID		*			
	Laboratory data		Domestic code*	Domestic code	Domestic code	Domestic code
	Date of examination		*			
DEATH	Person ID					
	Death date		*			
	Cause of death		*			

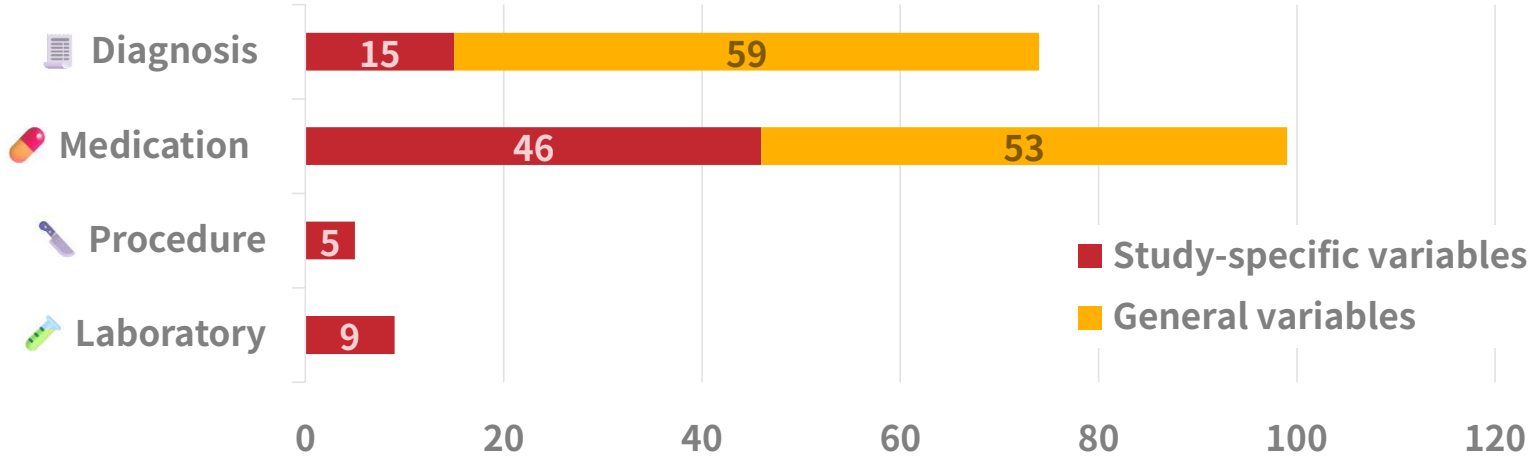
†Assume the last date of the database or the death date. $By adding number hospitalized days, if hospitalized. ‖By adding drug days. *HIRA without the information

Extract, Transform & Loading (ETL) Process



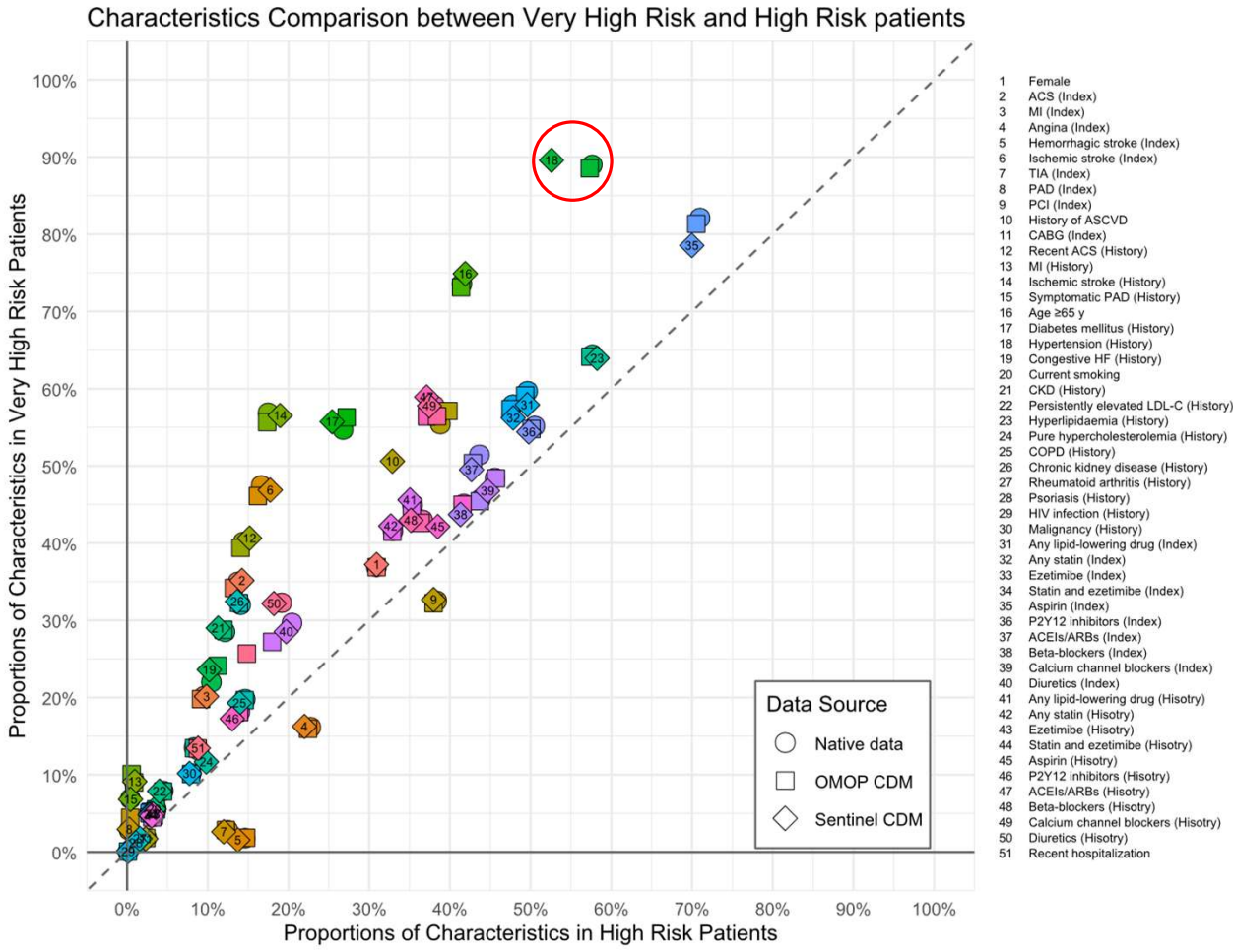
Coding Libraries

A site-specific repository of variable definitions, where each variable is associated with a clinician-reviewed set of codes



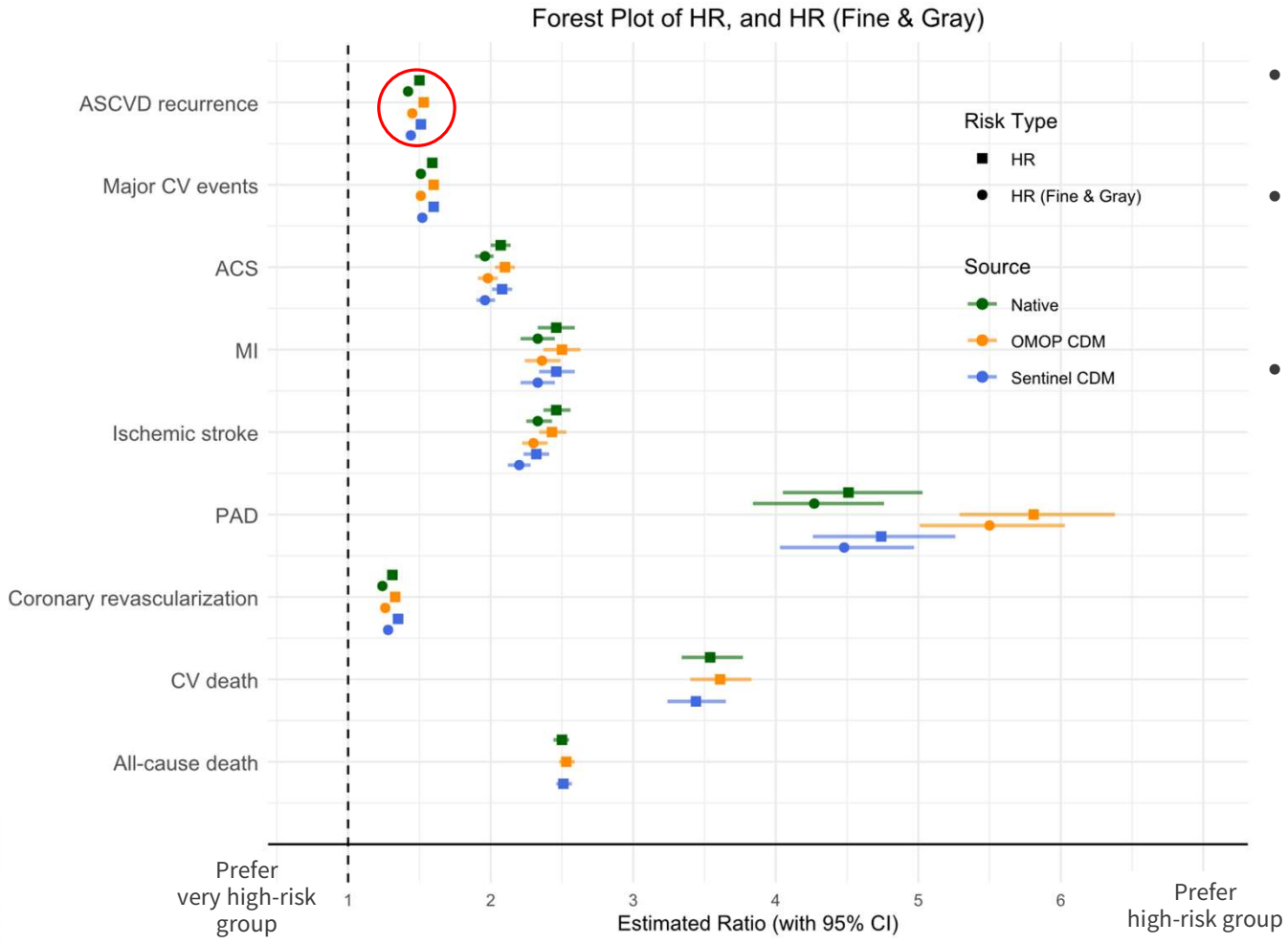
	NHIRD	CGRD	NHIS/HIRA	CDARS	DeSC
Ischemic stroke	I63	I63	I63, I64	433, 434	I63
Statin	A039307100, A039403100, A039582100, A039601100, A040060100, A040367100, A040541100, A040719100, A040723100, A041319100...	P6A080M, P6A085M, P6A086M, P7A208M, PCF067M, PCF069M, PFA090M, PFA137M, PFC001M, PFC002M...	111501, 111502, 111503, 111504, 472300, 472400, 472500, 502201, 502202, 502203, 502204, 518900, 524000, 524100, 527000...	ATOR02, ATOR01, SIMV01, SIMV04, SIMV02, ATOR03, ROSU01, PRAV02, FLUV02, S00033, ROSU02, PRAV01, LOVA01, ATOR04, FLUV03...	622204901, 622098501, 622161901, 622475100, 622165701, 622076501, 622170101, 622170201, 622076401, 62218670...

Similar Distribution of Characteristics before and after CDM conversion

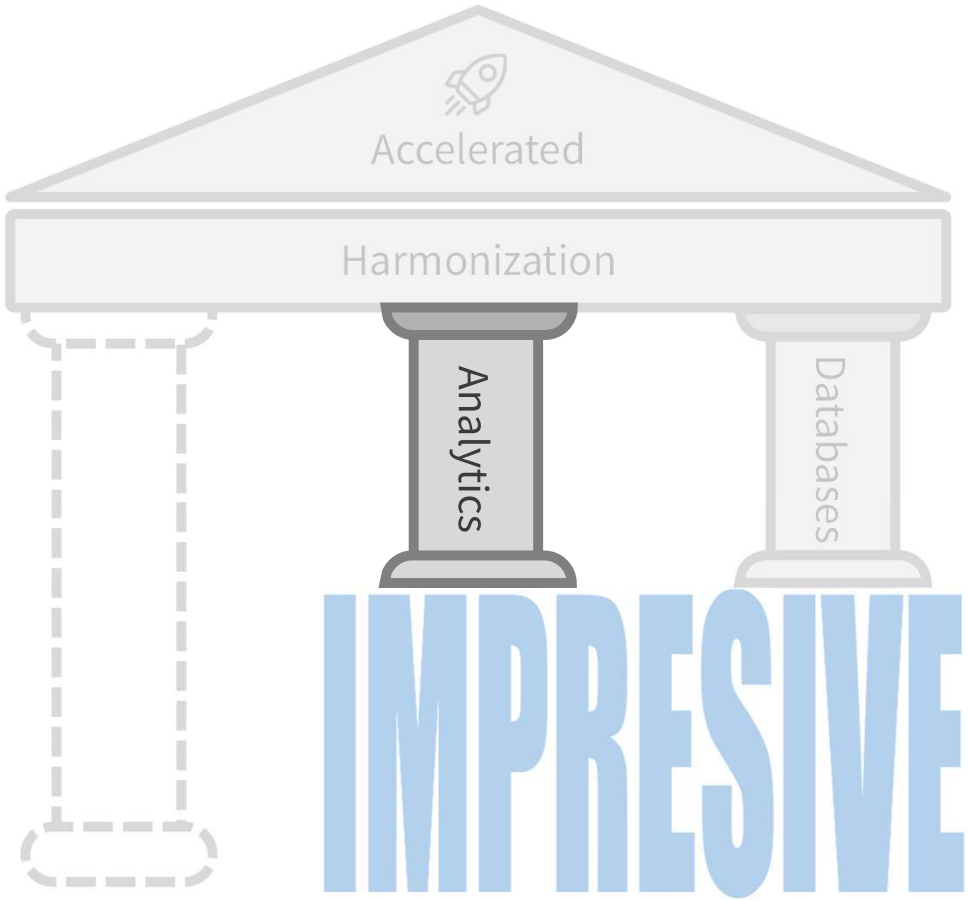


- Compares characteristic prevalence between very high-risk and high-risk groups
- Uses three data sources, each shown by a different dot shape
- Each variable is shown in a different color
- Same-color shapes are grouped closely
- Distribution is similar before and after CDM conversion

Similar Distribution of Hazard Ratio after CDM conversion



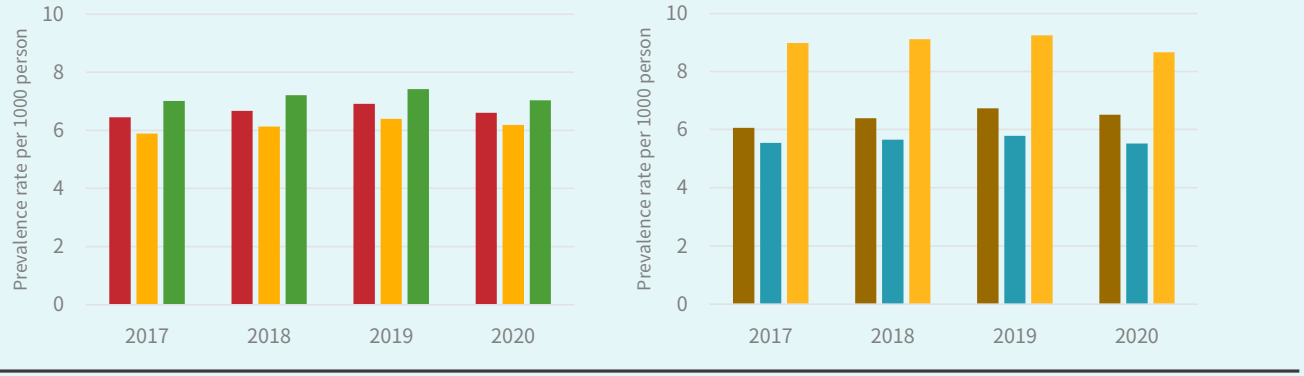
- Different colors in different data sources
- Most outcomes show consistent results across data sources
- Exception: Peripheral Arterial Disease (PAD)
  - Usually recorded as “Unspecified peripheral vascular disease” in clinical practice
  - Mapping the unspecified record resulted in inflated record numbers.
  - Affect the classification between very high-risk and high-risk groups


Accelerated
Harmonization
Analytics
Databases
IMPRESIVE

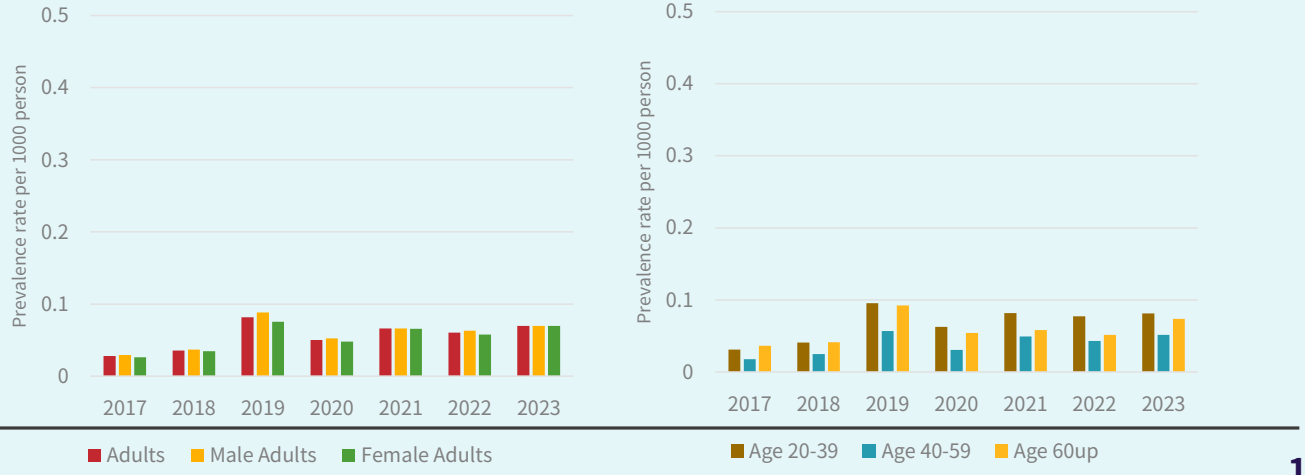
Analytic Modules

- Incidence & Prevalence
- Prescribing Pattern
- Clinical Characteristics
- Inferential Analysis

NHIRD



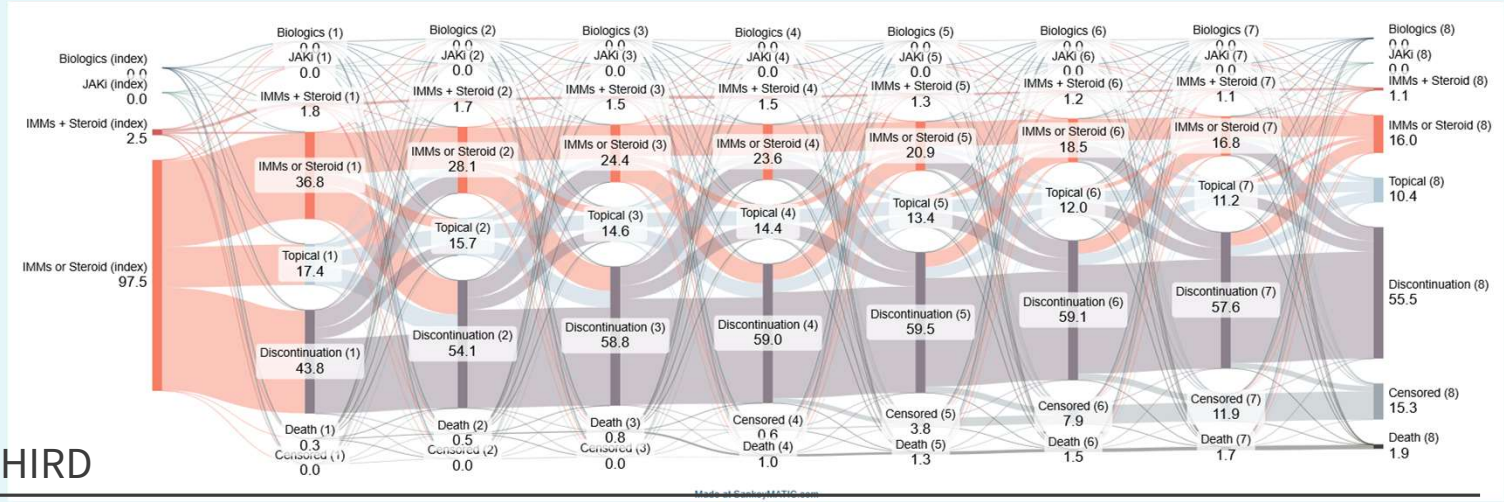
CDARS



Analytic Modules

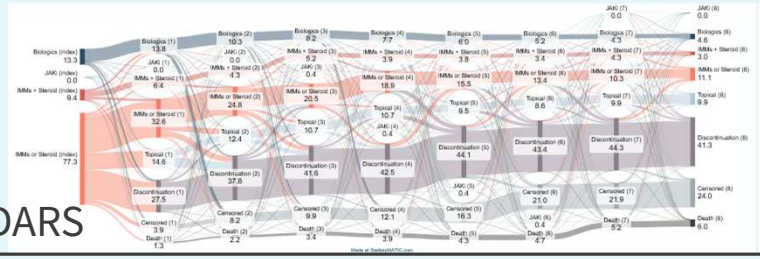
- Incidence & Prevalence
- Prescribing Pattern
- Clinical Characteristics
- Inferential Analysis

Atopic Dermatitis



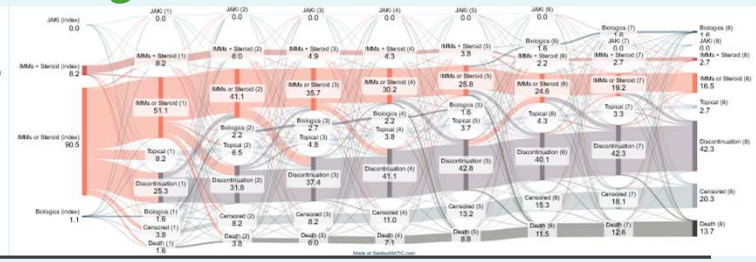
NHIRD

Atopic Dermatitis



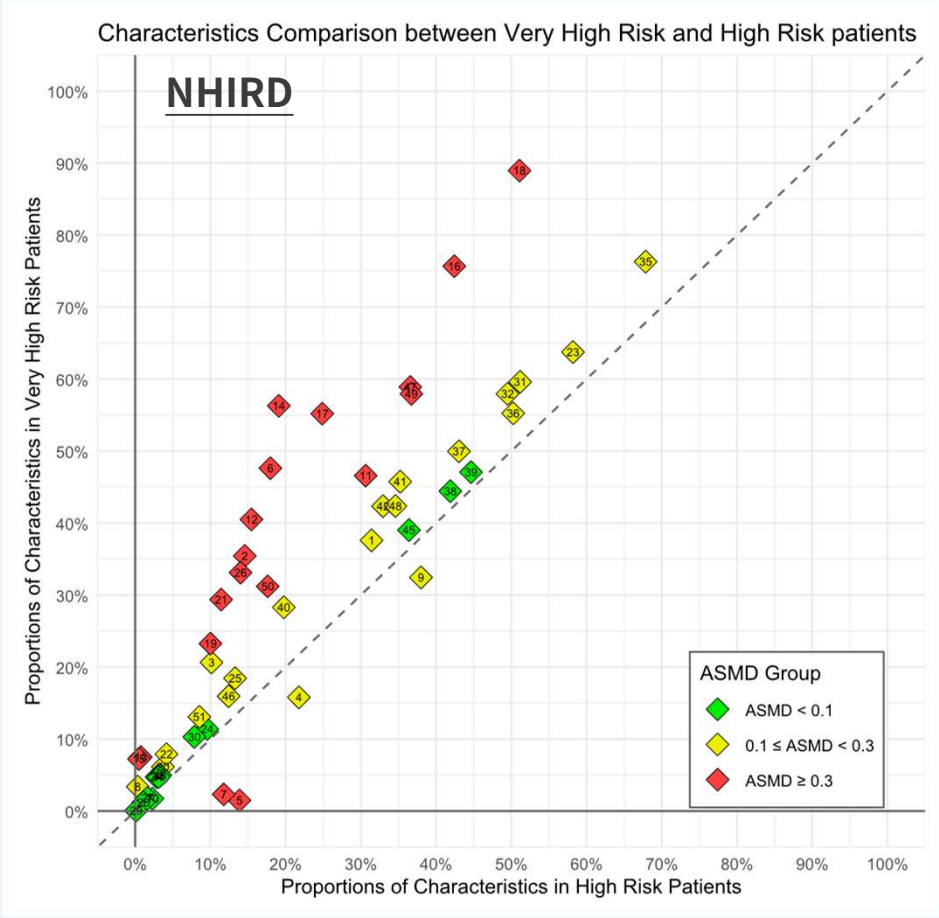
CDARS

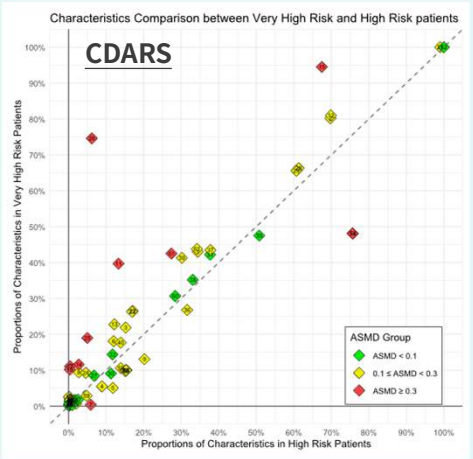
Prurigo Nodularis

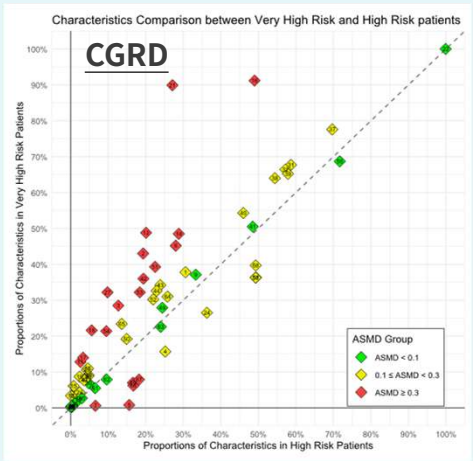


Analytic Modules

- Incidence & Prevalence
- Prescribing Pattern
- Clinical Characteristics
- Inferential Analysis

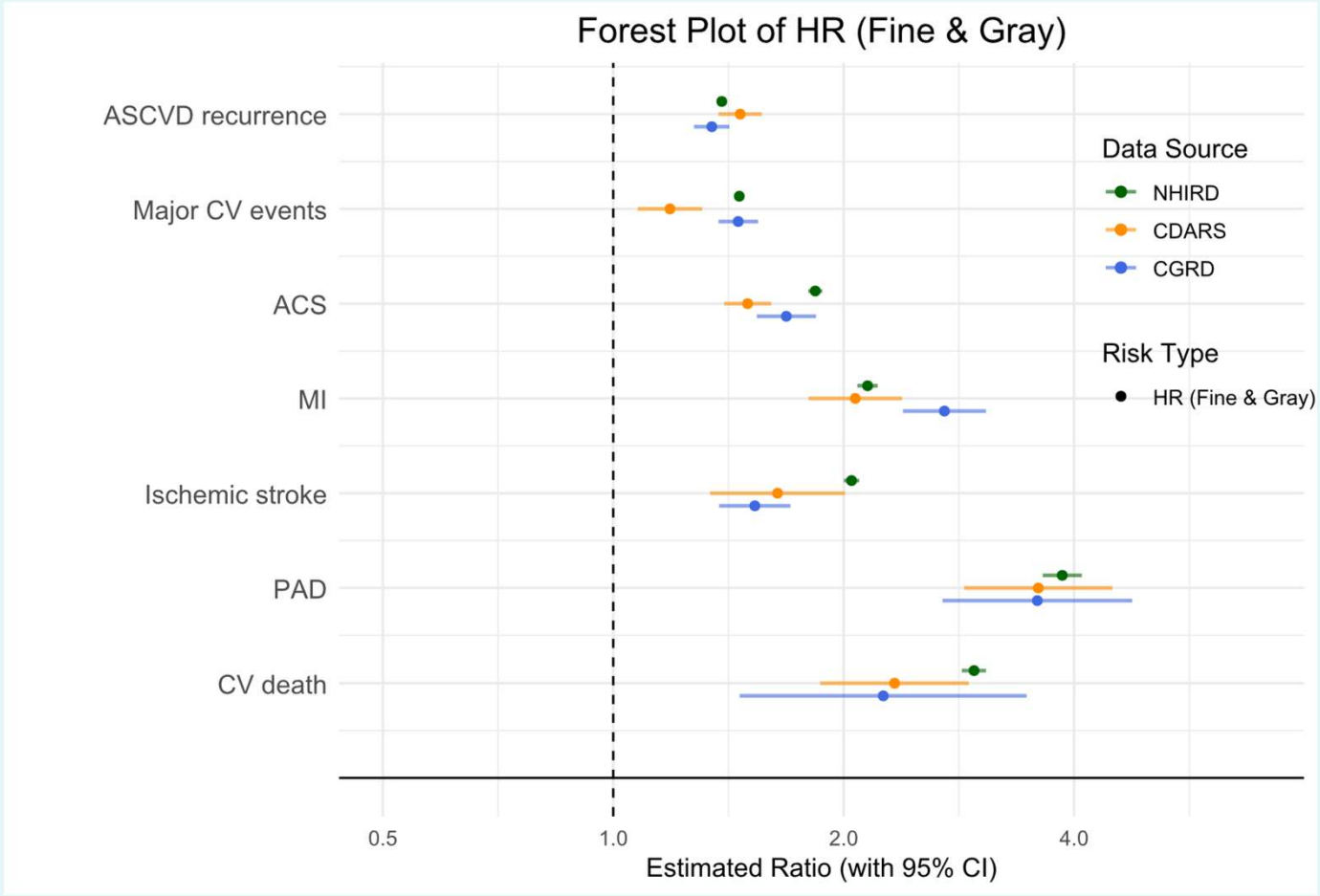




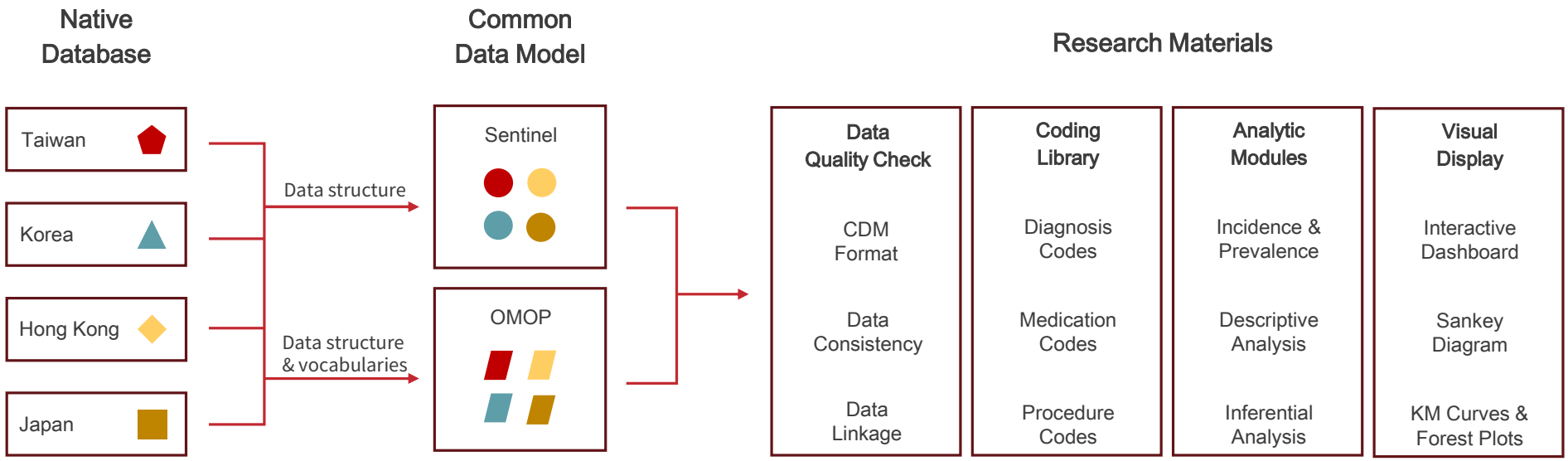


Analytic Modules

- Incidence & Prevalence
- Prescribing Pattern
- Clinical Characteristics
- Inferential Analysis



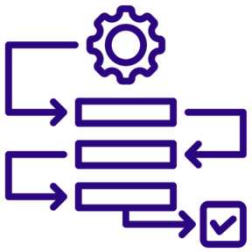
Key Achievements in IMPRESIVE projects



Insights

Adaptable & Reusable

- ETL program fit for different cohorts
- CDM can be converted to another CDM (e.g. VESALIUS-REAL)
- Coding library and analytic modules are adaptable to different studies



Standardized Execution

- Reduces variability in results due to implementation differences
- Makes it easier to reproduce, validate, or extend the findings

Lower Cost

- Speeds up study deployment and analysis
- Reduces duplicated effort across sites or studies

Manpower Requirement

- Single-center study
- Single-center reproducibility assessment

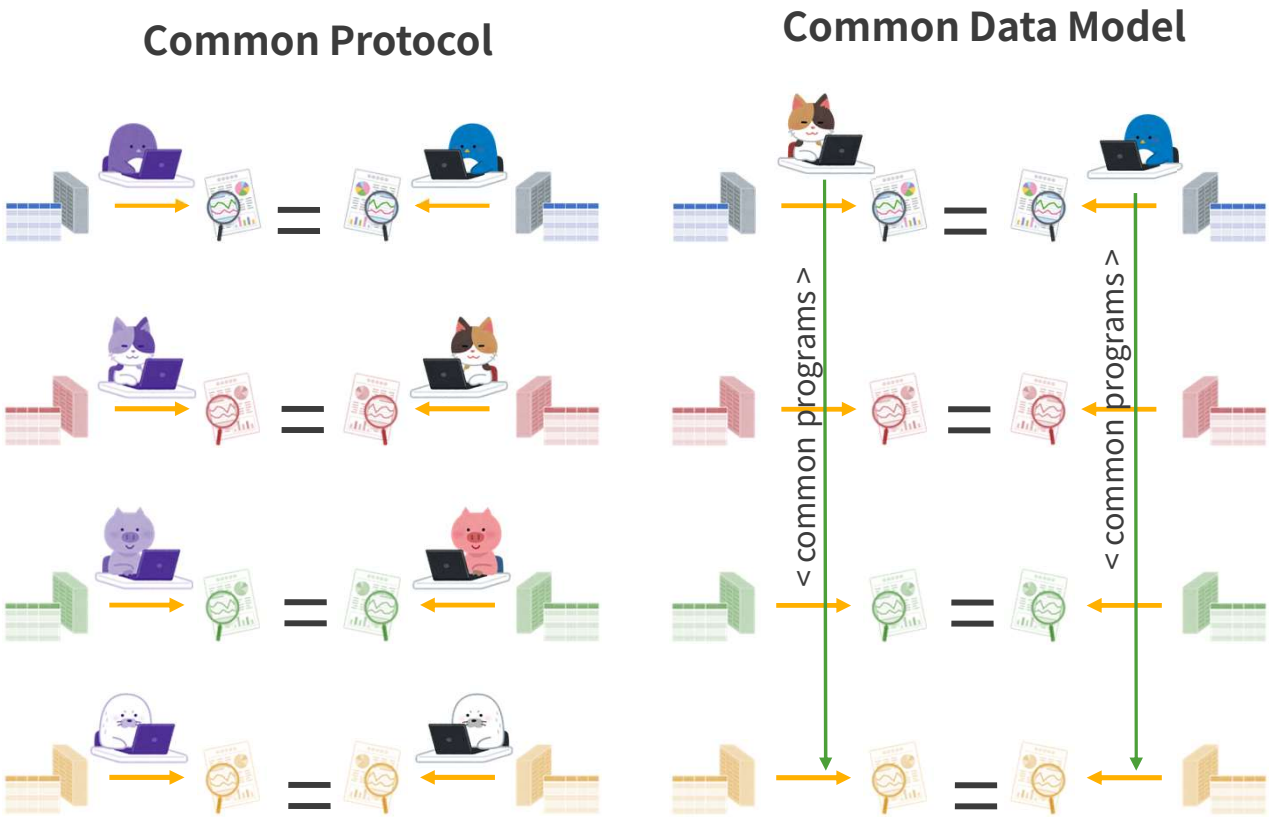
Common Protocol

Common Data Model

	Common Protocol	Common Data Model
Number of Programmers:	2	2

Manpower Requirement

- Multiple-center study with reproducibility assessment



Number of Programmers: 8 2

Challenges

Data Heterogeneity

- Inconsistent availability of key data domains (e.g., labs, prescriptions, death data)
- Complex variable definitions require customized algorithms at each site

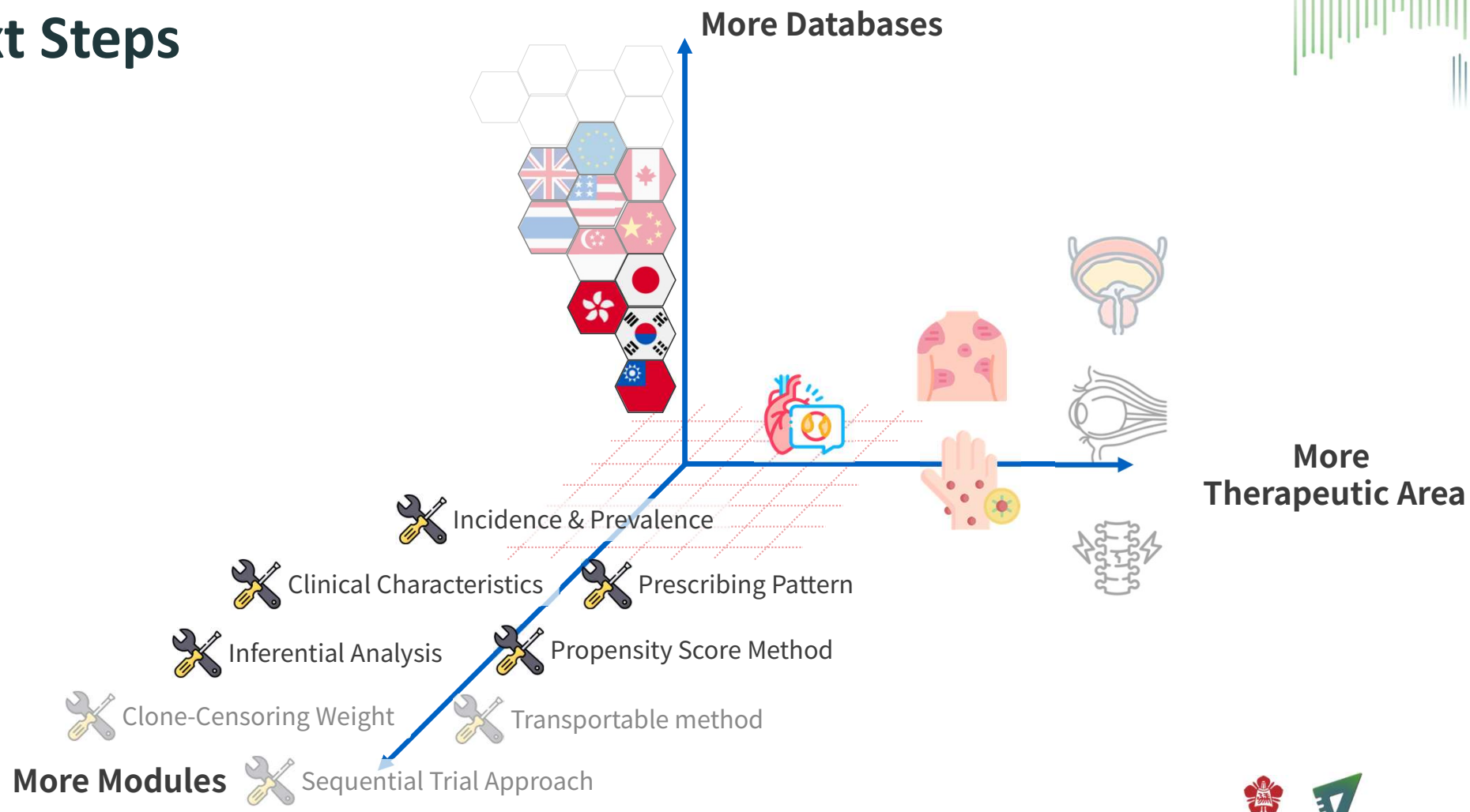
Technical and Infrastructure Limitations

- Software/tool familiarity varies
- Varying levels of CDM experience and ETL automation
- Limited data storage at some sites

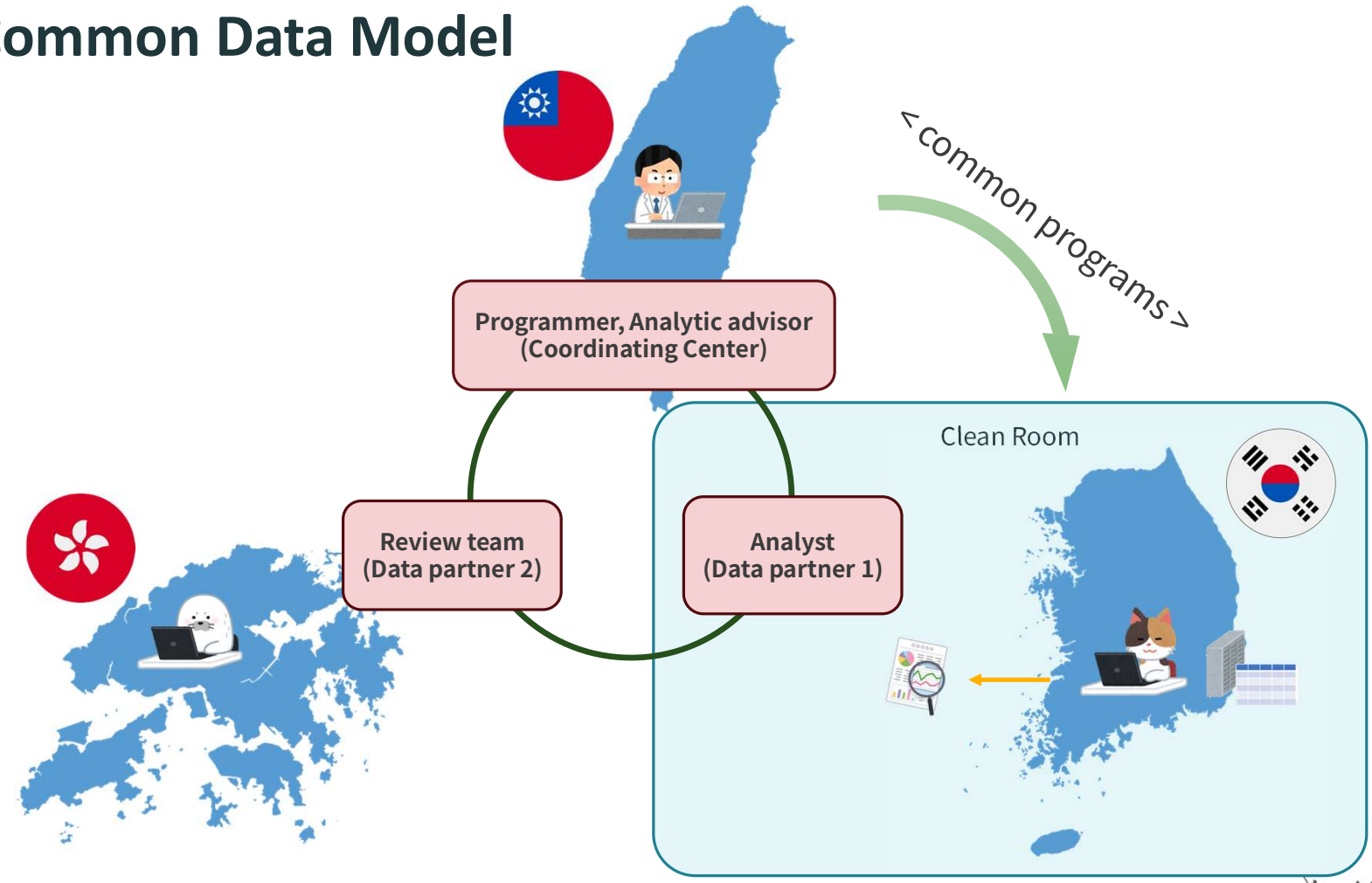
CDM Sustainability

- Short database access periods
- Limited resources for long-term platform maintenance

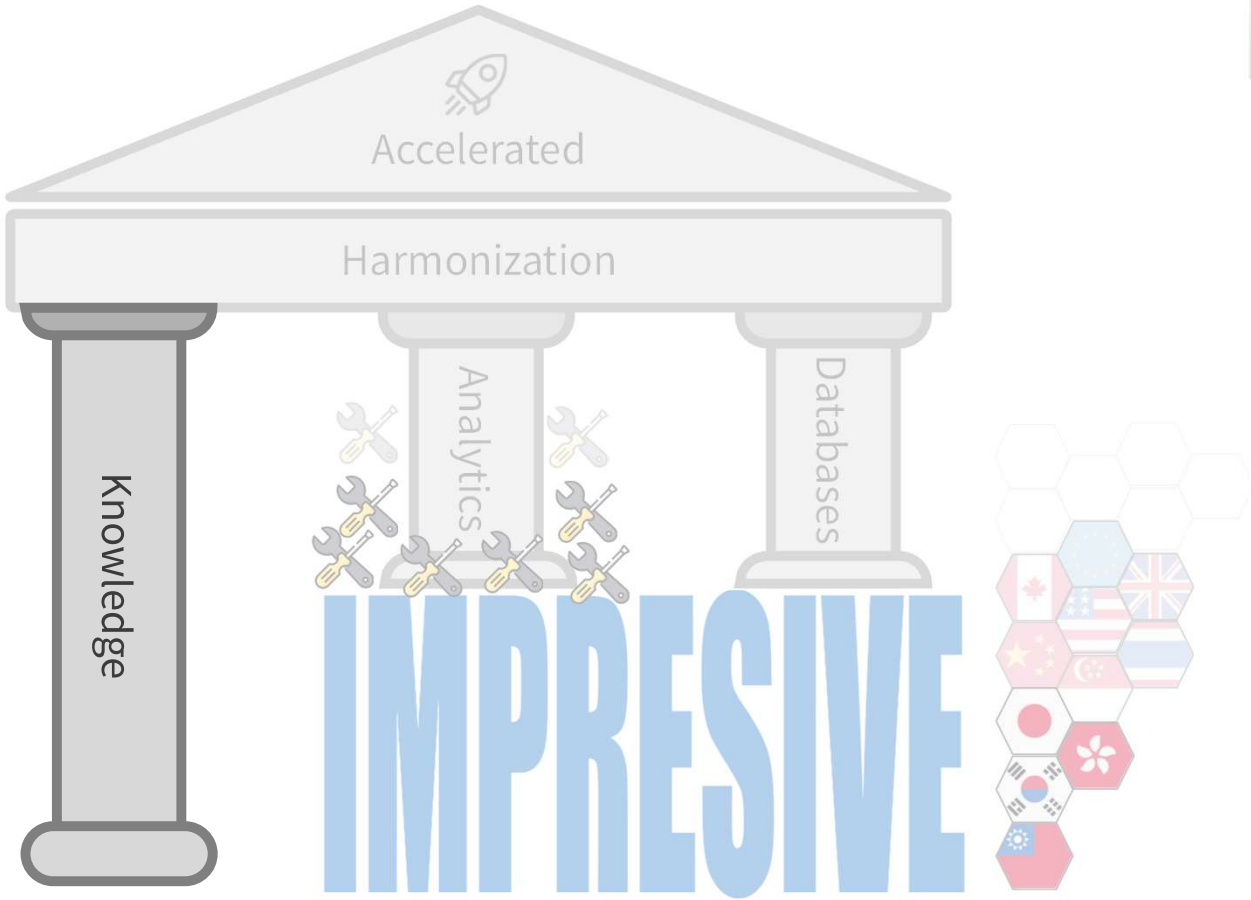
Next Steps


More Databases
More Therapeutic Area
Incidence & Prevalence
Clinical Characteristics
Prescribing Pattern
Inferential Analysis
Propensity Score Method
Clone-Censoring Weight
Transportable method
More Modules
Sequential Trial Approach

Gating Framework with Common Data Model



Knowledge – bridges the gap between evidence generation and decision-making



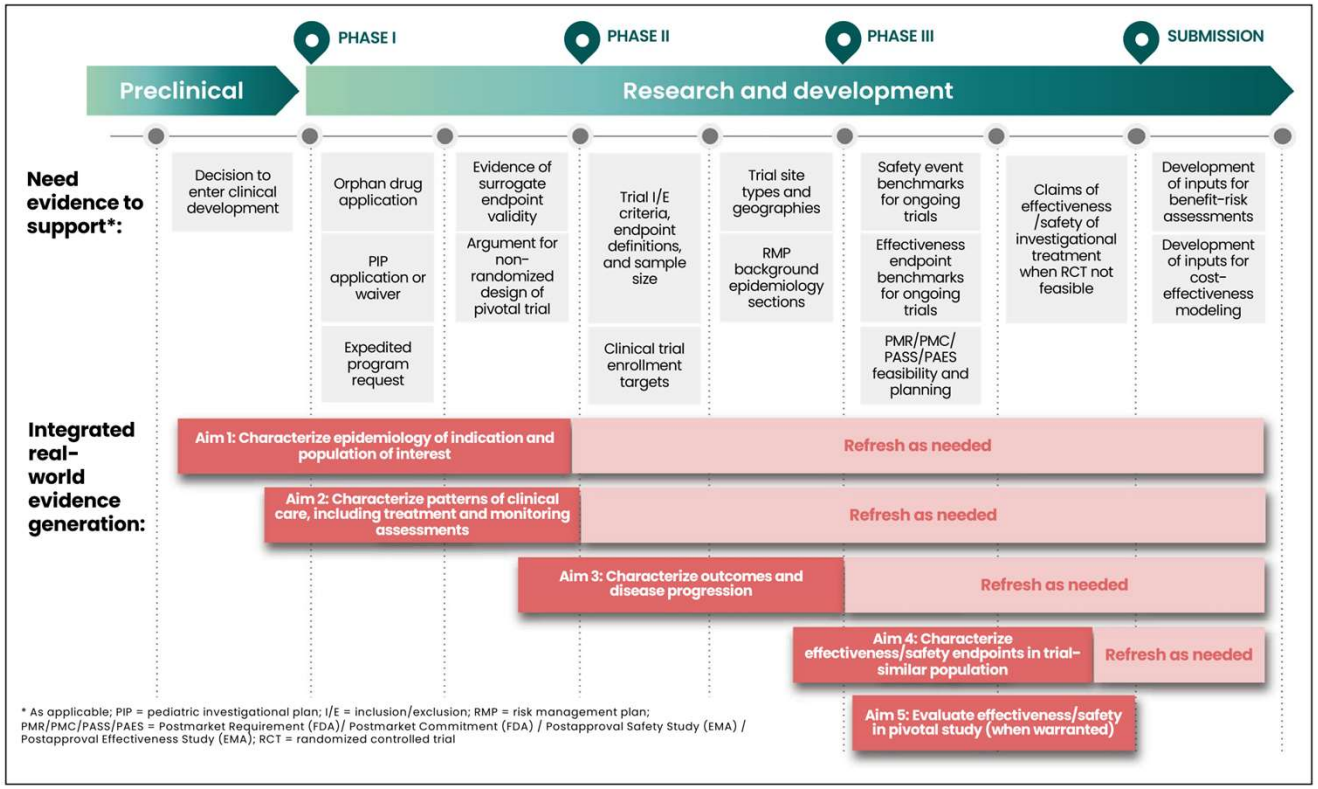


Figure 1 Integrating RWE into clinical development: needs and RWE evidence generation strategy.

Knowledge Harmonization is the Foundation

Harmonization is Important

- **Aligns understanding** of RWE across clinicians, regulators, and researchers
- Standardizes terminology and evidence thresholds for **broader acceptance**
- Enables **effective dissemination**
- **Minimizes misinterpretation** or misuse

Key Harmonization Strategies

- **Conceptual alignment** on RWE principles, study designs, and fit-for-purpose use
- **Standardized frameworks**: HARPER template, common cohort definitions
- **Cross-sector engagement** to co-create understanding (e.g., clinician-regulator-researcher dialogues)

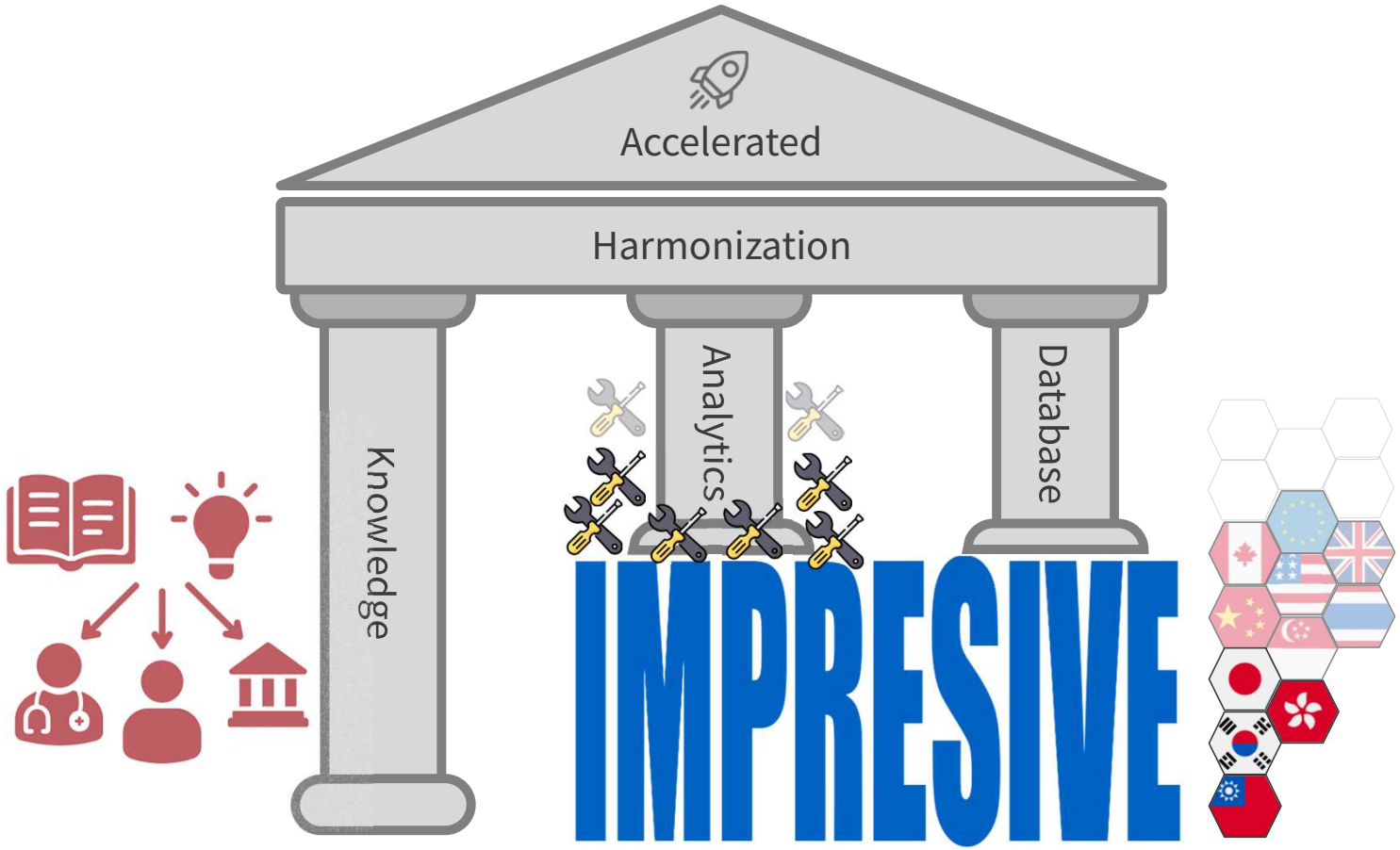
Dissemination with Purpose: Driving the Use of RWE

Strategic Approaches

1. **Education & Communication**
   - Workshops, webinars, and CME targeting healthcare professionals
   - Policy briefs and simplified summaries for regulators
2. **Timely Integration into Decision-Making**
   - Embedding RWE into clinical decision support tools
   - Using fit-for-purpose RWE in regulatory or reimbursement discussions

From Understanding to Action

- Harmonized knowledge → Trusted message → Real-world impact


Accelerated
Harmonization
Knowledge
Analytics
Database
IMPRESIVE

Thank You!