Collaborator Showcase: Rapid Fire Presentations

Katia Verhamme, MD, PhD

Associate Professor of Use and Analysis of Observational Data

Adoption of the OMOP Common Data Model in the UK

Alex Knight

Health Data Research UK, UK

Poster: 42

HDRUK
Health Data Research UK

Adoption of the OMOP Common Data Model in the UK

OHDSI Europe 2024

3/6/2024 | Alex Knight, Project Manager (Data Standards)

UK Landscape

HDRUK
Health Data Research UK

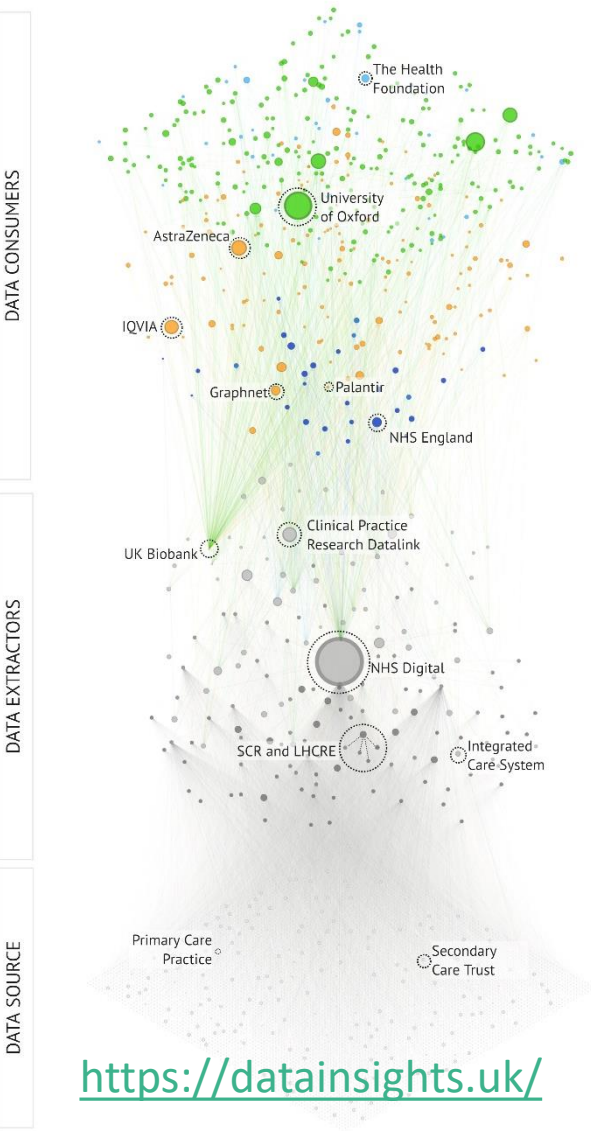
NHS

Research Secure Data Environment NETWORK

HDRUK
Health Data Research UK

UK Health Data Research Alliance

- The UK National Health Service holds detailed health data on 67 million people – potentially an invaluable resource for research
- However, data is distributed across many healthcare providers and formats
- NHS England establishing a network of secure data environments (SDEs/TREs) to support research
- Wales and Scotland already have their own SDEs
- The UK's national institute for health data science
- 103+ leading healthcare and research organisations
- Establishing best practice for the ethical use of UK health data for research at scale



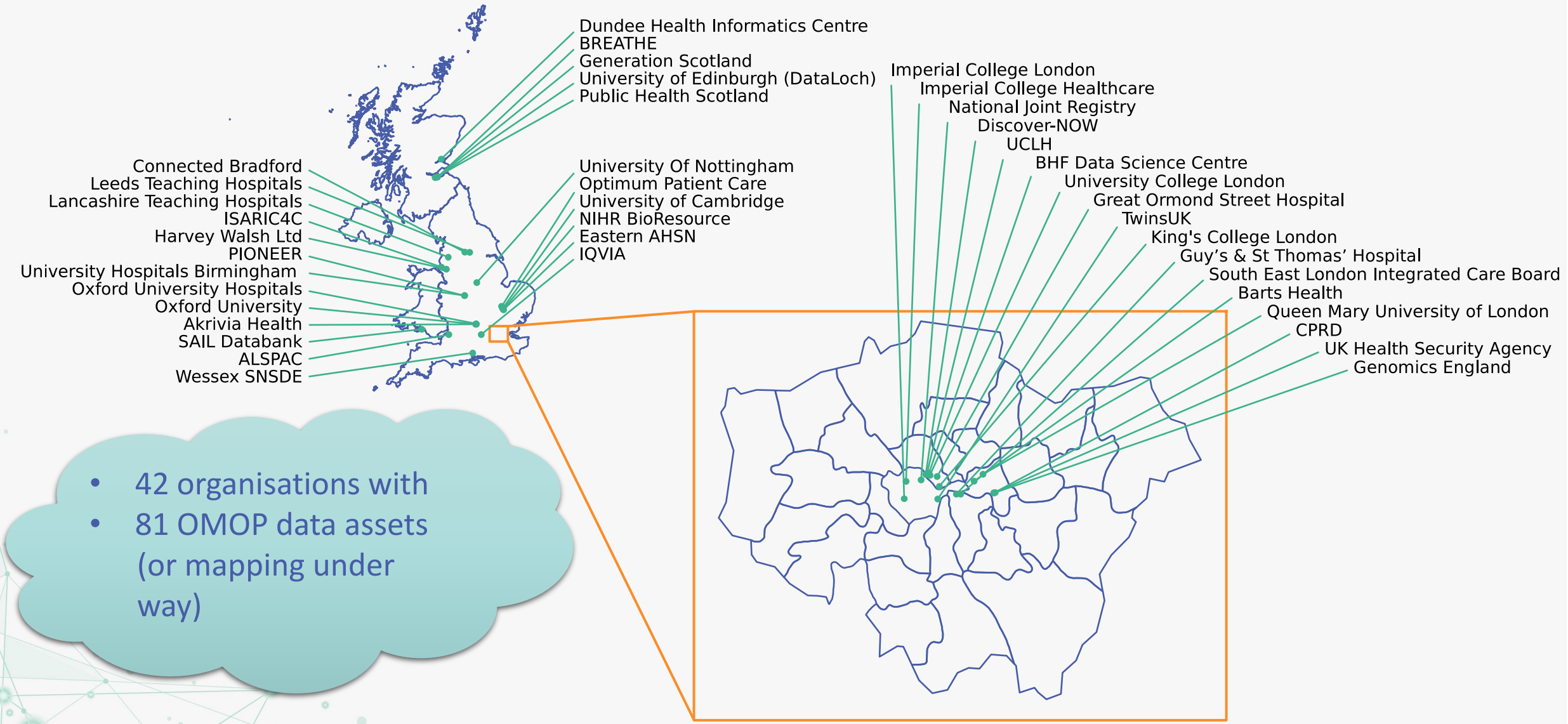
https://datainsights.uk/

Activities in support of OMOP adoption

HDRUK
Health Data Research UK

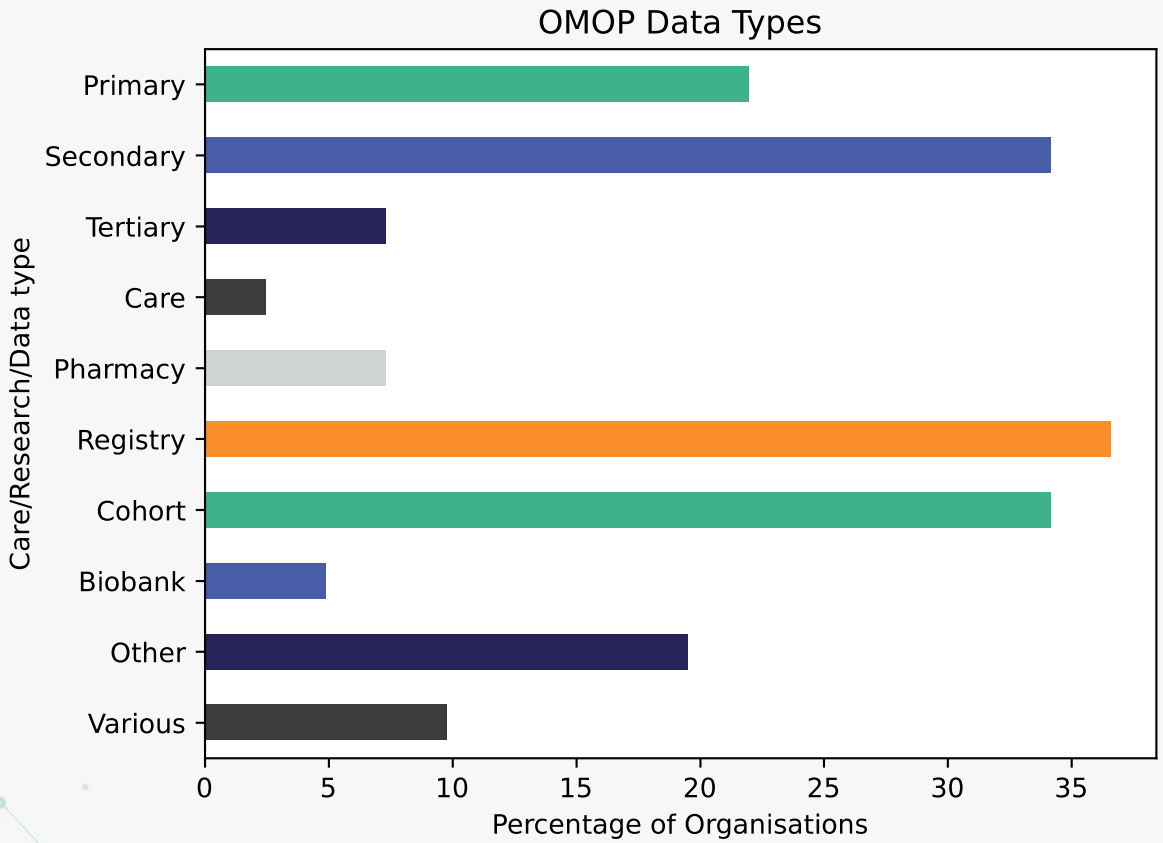
- Collaboration with EHDEN, funding 5 new data partners to transform to OMOP
- Partnership with OHDSI UK
- Collaboration with NHS England's network of Secure Data Environments to support their adoption of OMOP
- Survey of OMOP Landscape in the UK
- Community activities including OMOP events
- Developed the Health Data Research Innovation Gateway which acts as a single hub for metadata on UK data sets, (many in OMOP format).

UK Data Custodians with OMOP Data Sets



- 42 organisations with
- 81 OMOP data assets (or mapping under way)

Data Types of UK OMOP Data



Data types cover the full spectrum of routine care and research data types

Conclusions:

- Significant momentum behind OMOP adoption in the UK
- Significant barriers and limitations remain

Next steps include:

- Improve researcher/analyst skills and capabilities
- Identify use cases and priority research
- Agree a minimum viable OMOP dataset
- Work with national data collections and GP Data

Piloting the Transformation of Multiple Sclerosis Real-World Data to the OMOP CDM: Lessons Learned

Tina Parciak

Uhasselt, Belgium

Poster: 51

Piloting the Transformation of Multiple Sclerosis Real-World Data to the OMOP CDM: Lessons Learned

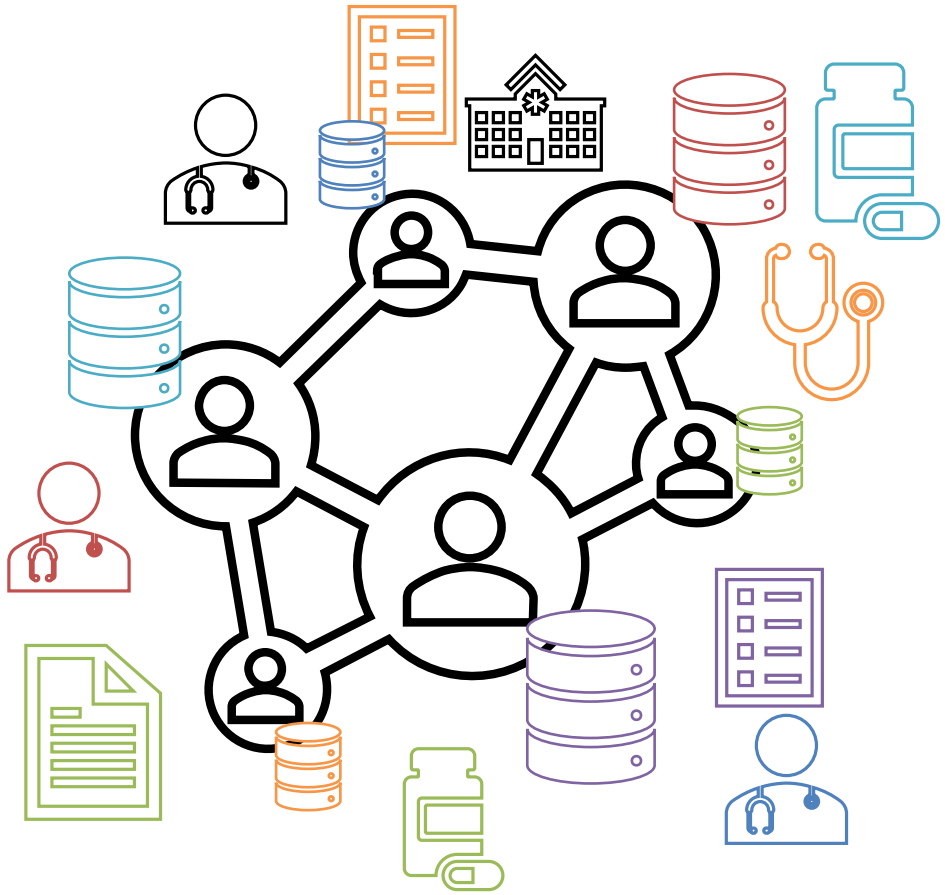
Tina Parciak[1,2,3], Kirstin Tümler[4], Alexander Stahmann[5], Emma Gesquiere[6], Freija Descamps[6], Liesbet M. Peeters[1,2,3]

[1] University MS Center (UMSC), Hasselt - Pelt, BELGIUM
[2] UHasselt, Biomedical Research Institute (BIOMED), Agoralaan, 3590 Diepenbeek, BELGIUM
[3] UHasselt, Data Science Institute (DSI), Agoralaan, 3590 Diepenbeek, BELGIUM
[4] German Center for Diabetes Research (DZD), Munich, Germany
[5] German MS Register by the German MS Society, MS Research and Project Development gGmbH (MSFP), Hanover, GERMANY
[6] edenceHealth NV, Veldkant 33 A, 2550 Kontich, BELGIUM

Background / Motivation

Multiple Sclerosis (MS)

Background / Motivation



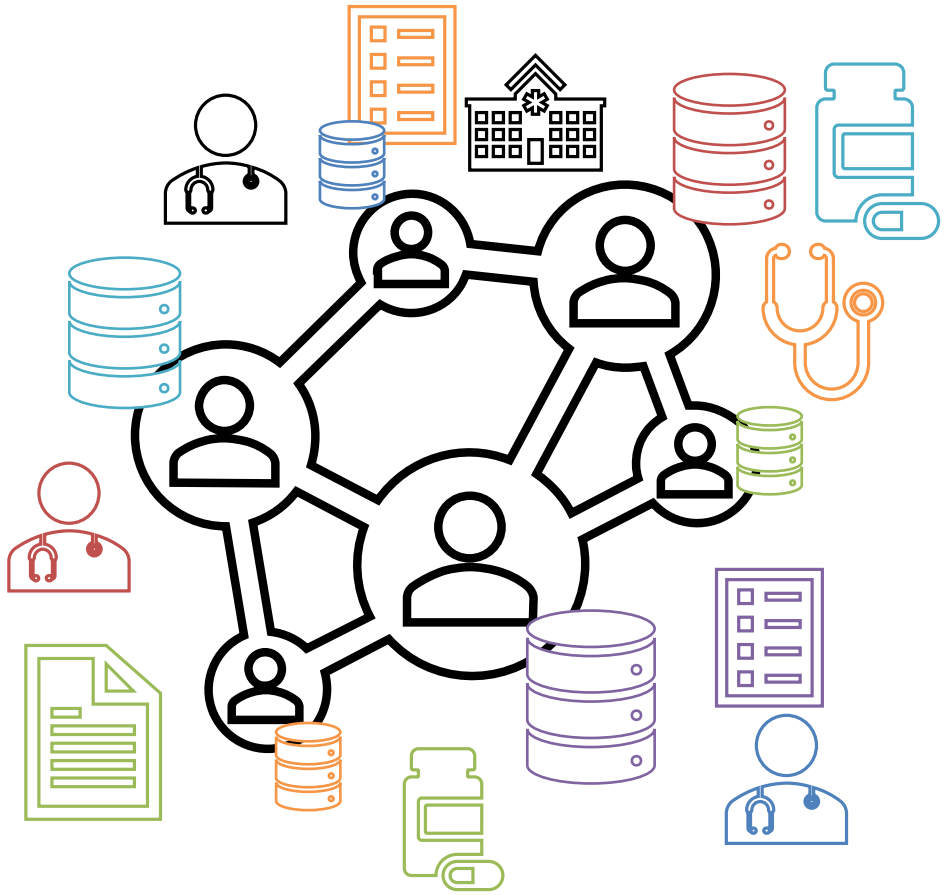
MS real-world data (RWD) coming from registries and cohorts

Background / Motivation

Real-world evidence (RWE) generation within MS community

Background / Motivation

RWE generation outside
MS community



How about the OMOP CDM?

Materials & Method

MS DataConnect

German MS Registry

EHDEN

EUROPEAN HEALTH DATA & EVIDENCE NETWORK

Lessons learned

MS registry or cohort datasets
differ from EHR data.

Standardisation to OMOP CDM can still result
in heterogeneous outputs.

Lessons learned

Exchange of experiences and alignment for registry-type data transformations is necessary.

Transforming MS RWD demands substantial time investment and interdisciplinary knowledge.

Annotation-preserving machine translation of English corpora to validate Dutch clinical concept extraction tools

Tom Seinen

Erasmus MC, The Netherlands

Poster: 112

CLINICAL CONCEPT EXTRACTION

Temporary kidney enlargement in the newborn infant

Recognition	**Disorder**	**Person**	UMLS, SNOMED CT, OMOP vocab
+ Linking	**4104152**	**4046034**	

Many extraction **tools** exist for **English**

cTakes, MedTagger, MedCAT, QuickUMLS, …

Only a **few** tools for **Dutch**

UMCU: MedCAT, *EMC*: MedSpaCy

However: How good are these Dutch extraction tools?

Erasmus MC

ANNOTATED CLINICAL CORPORA

Ground truth: sample texts with manually annotated concepts

`Temporary kidney enlargement in the newborn infant`

C0542518: kidney enlargement **[10-28]**
C0021289: newborn infant **[36-50]**

Evaluation:

Does the tool extract the **correct concepts** at the **correct locations**?

English tools are evaluated using **English clinical corpora**
(MedMentions, ShARe/CLEF, i2b2, etc.)

Few problems:

- There are **no** large **Dutch** corpora…
- **Creation is difficult**: very **labor intensive**, **sensitive** patient data

Erasmus MC

TRANSLATION OF EXISTING CORPORA

Translation  **+ Alignment**  (Language dependent)

Temporary kidney enlargement in the newborn infant

Tijdelijke niervergroting bij de pasgeboren baby

C0542518: kidney enlargement **[10-28]**
C0021289: newborn infant **[36-50]**

C0542518: kidney enlargement **[???????]**
C0021289: newborn infant **[????????]**

Embedding  **+ Translation**  **+ Extraction**  (Language independent)

Temporary [[kidney enlargement][C0542518]] in the [[newborn infant][C0021289]]

Tijdelijke [[niervergroting][C0542518]] bij de [[pasgeboren baby][C0021289]]

Tijdelijke niervergroting bij de pasgeboren baby

C0542518: niervergroting **[11-25]**
C0021289: pasgeboren baby **[33-48]**

Erasmus MC

ANNOTATION-PRESERVING TRANSLATION

Google Cloud Translate API

- Dedicated translation model
- Out-of-the-box, Fast
- One model, no tweaking
- Makes formatting errors (1%)
- Keeps all annotations (1% missing)

OpenAI GPT-4 API

- SoTA generative model
- Needs prompting, Slower
- Endless optimalization possibilities
- Makes almost no formatting errors (<<1%)
- Ignores some annotations (1-6%)

`Temporary [[kidney enlargement][C0542518]] in the [[newborn infant][C0021289]]`

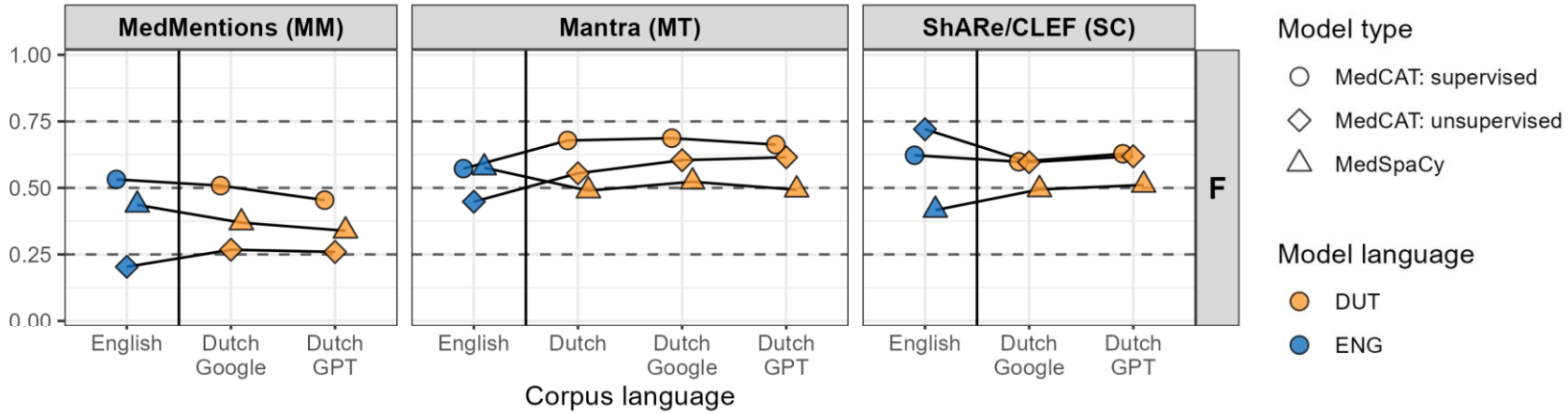
`…_[niervergroting]_C0542518]]…` `… bij de pasgeboren baby …`

Large agreement between the translations (BLEU~0.5, chrF~0.8)

Erasmus MC

CONCEPT EXTRACTION PERFORMANCE

Experiments: 2 extraction **tools (MedCAT, MedSpaCy)**, **3** different English **corpora**

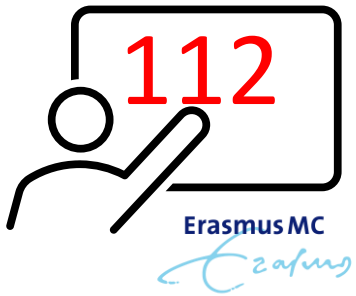


Good overall **performance** from both tools

No significant difference between **English** and the **Dutch translations**

Conclusion:

- **Generated**: Dutch annotated corpora, **Evaluated**: Dutch clinical extraction tools
- **Method: No manual annotations** needed & **language agnostic**.



Beyond Diagnosis Codes: A Weakly Supervised Learning Framework for Accurate Multimorbidity Identification in Electronic Health Records

Bernardo Neves

Luz Saúde, Portugal

Poster: 111

HOSPITAL DA LUZ
LEARNING HEALTH
TRAINING, RESEARCH & INNOVATION CENTER

Beyond Diagnosis Codes: A Weakly Supervised Learning Framework for Accurate Multimorbidity Identification in Electronic Health Records

Bernardo Neves[1,2,3], Jorge Cerejo[1], Simão Gonçalves[1], José Maria Moreira[1], Nuno A. da Silva[1], Francisca Leite[1], Mário J. Silva[3]

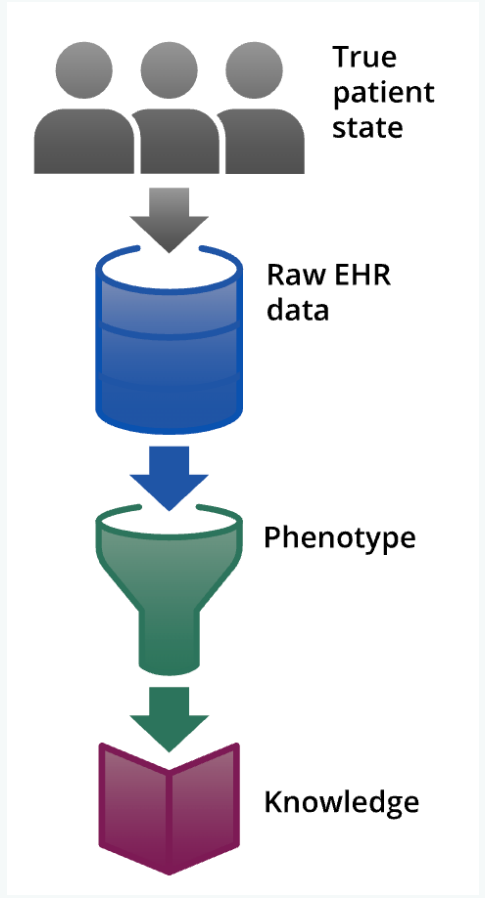
[1]Hospital da Luz Learning Health
[2]Departamento de Medicina Interna, Hospital da Luz de Lisboa
[3]INESC-ID, Instituto Superior Técnico, Universidade de Lisboa

Secondary use of EHRs

- Phenotyping chronic conditions typically relies on direct mentions (diagnosis codes), however there are many indirect surrogate markers (Labs, procedures, etc.)

- Expert validation is hard/unfeasable at scale: weakly-supervised learning is a possible approach[1]

- Goals:
  - To develop a phenotyping dictionary to identify common chronic conditions in EHRS using diverse data sources
  - To explore unsupervised approaches to validate phenotyping rules

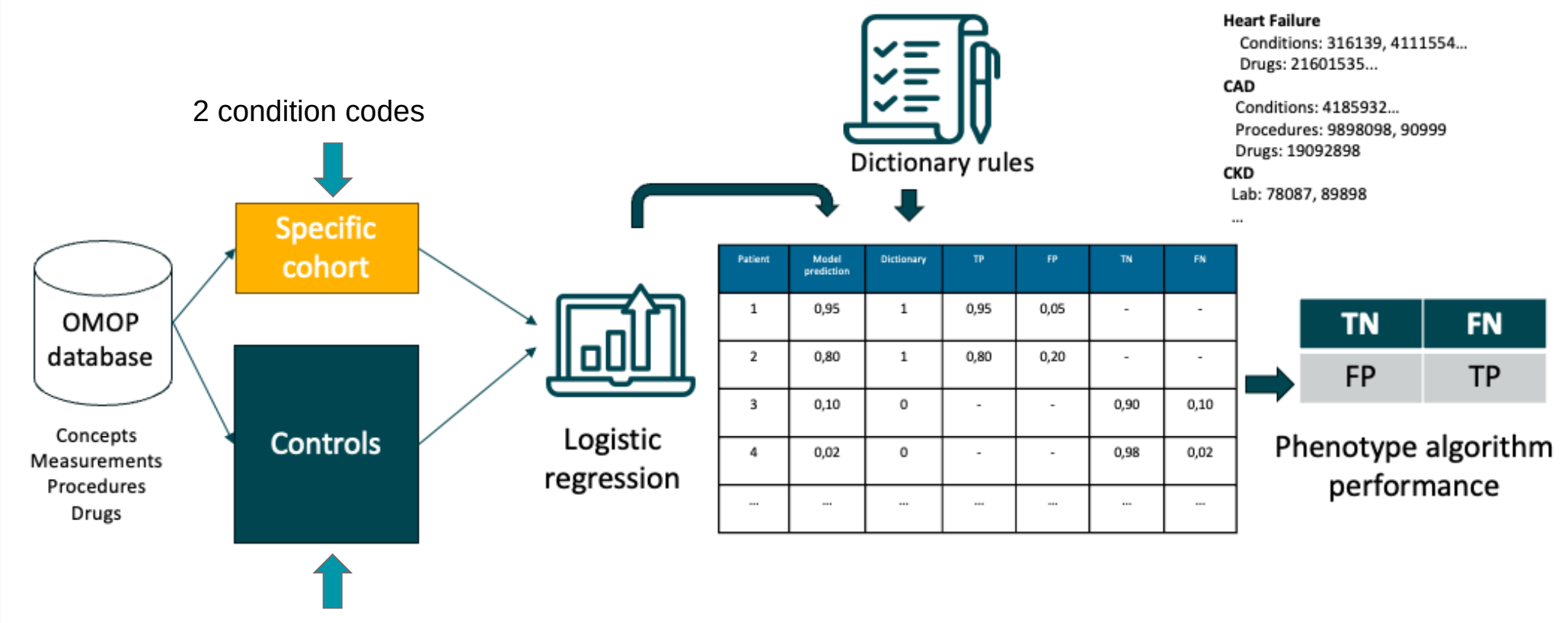


rethinkingclinicaltrials.org

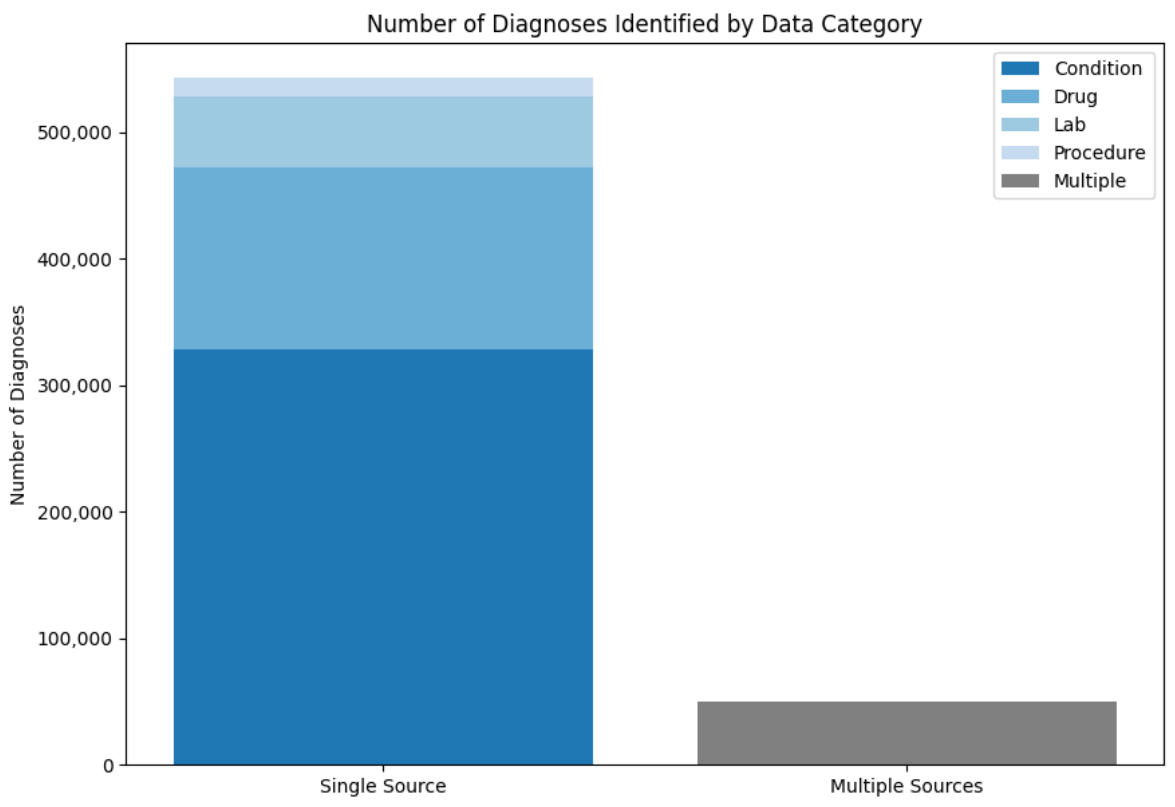
[1]Swerdel JN et. Al. J Biomed Inform. 2019 Sep;97:103258

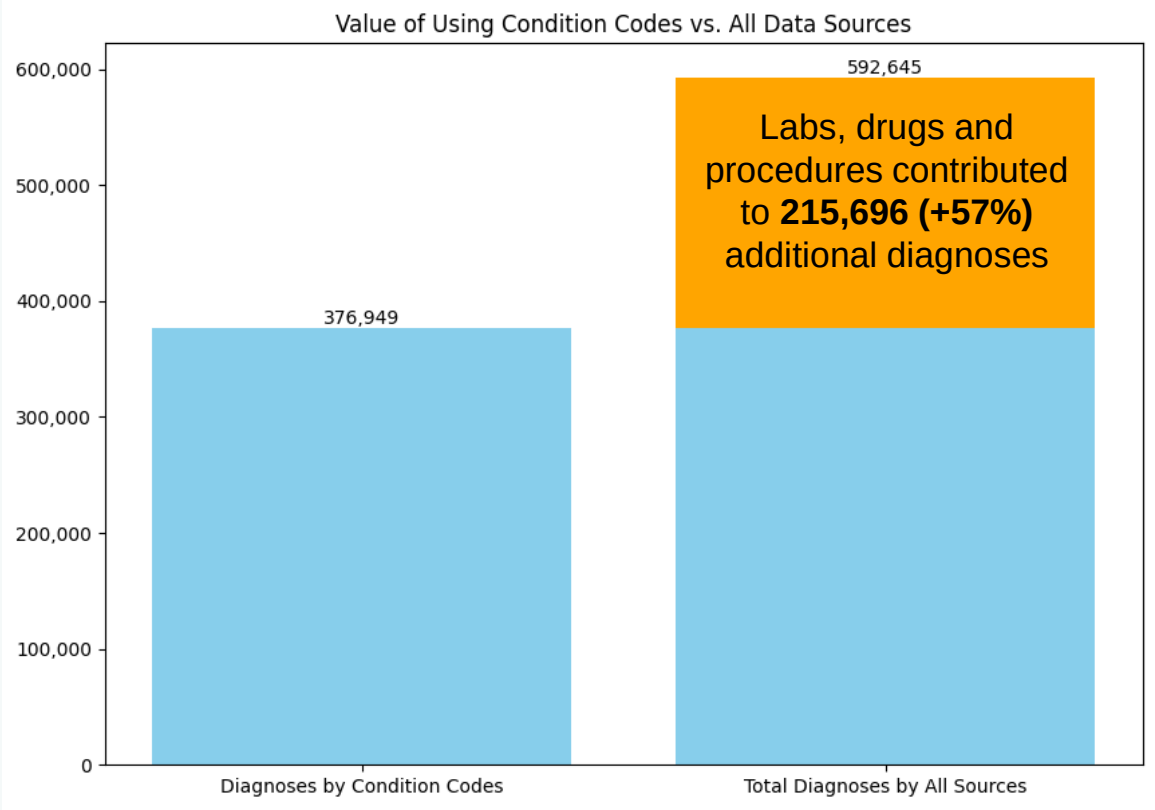
Methods

- Weakly-supervised learning phenotype validation



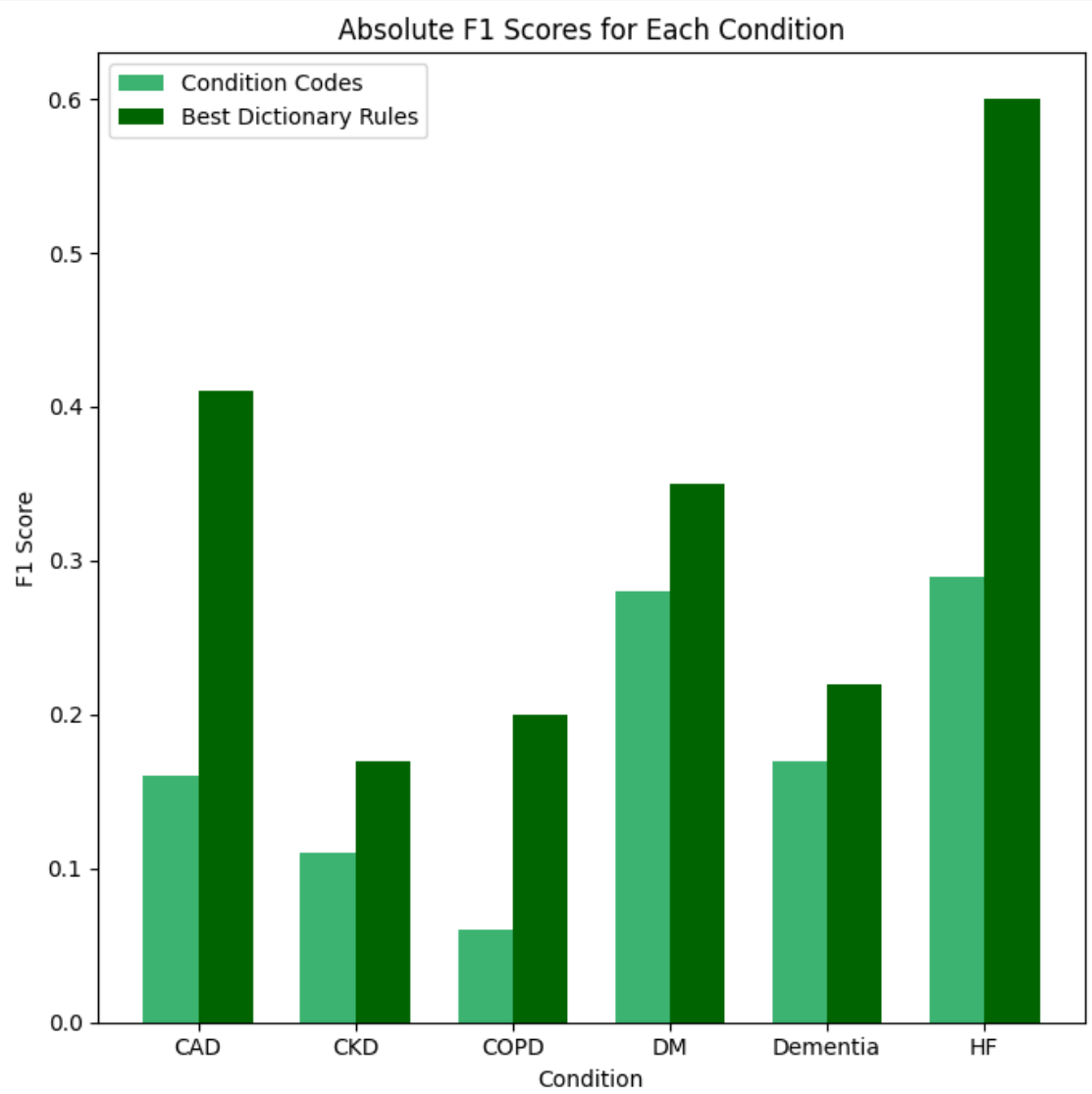
Results

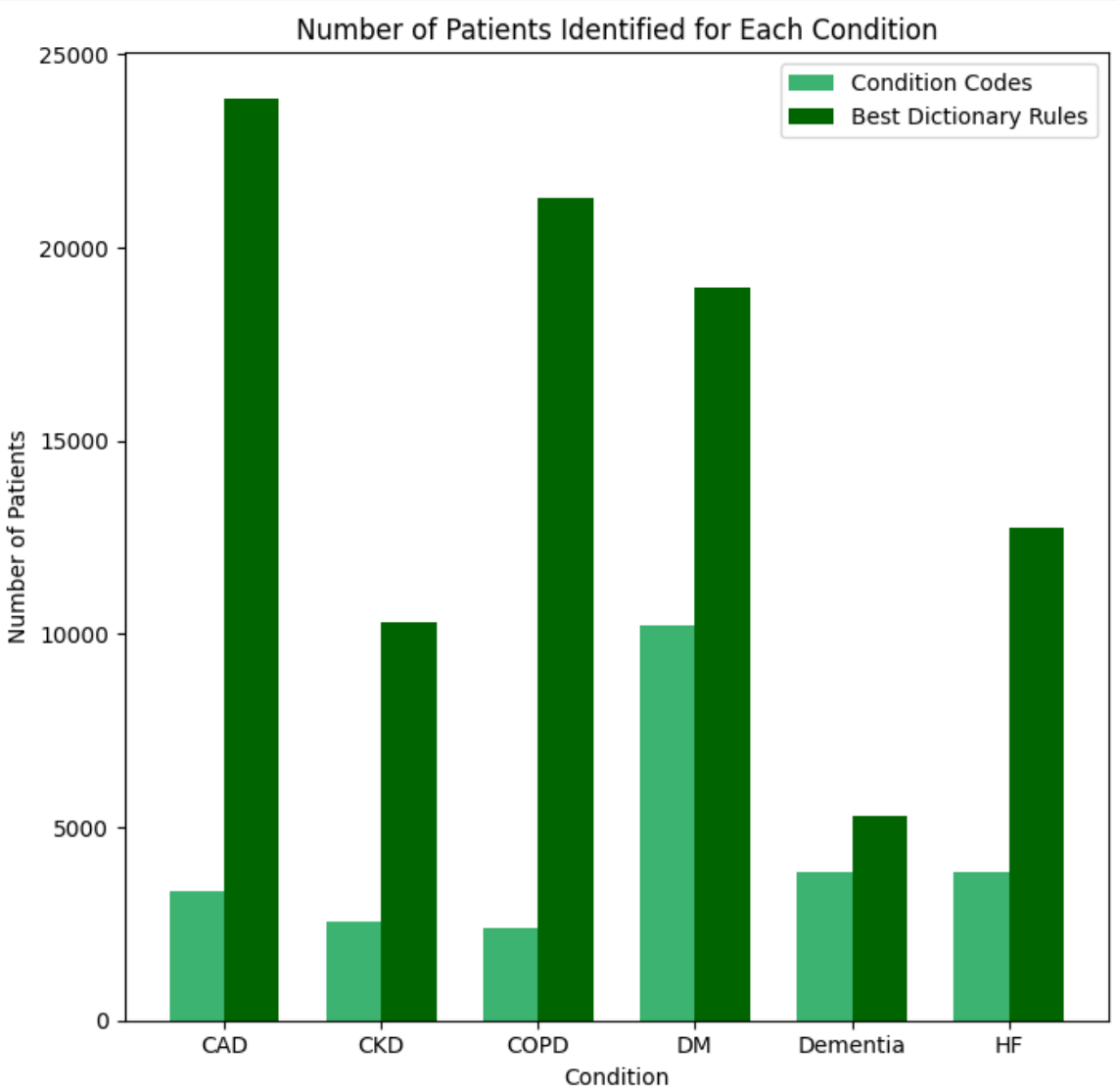




42,976,012 codes
7,452,487 hospital episodes
838,980 adult patients

Results


Absolute F1 Scores for Each Condition
Condition Codes
Best Dictionary Rules
F1 Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
CAD
CKD
COPD
DM
Dementia
HF
Condition


Number of Patients Identified for Each Condition
Condition Codes
Best Dictionary Rules
Number of Patients
0
5000
10000
15000
20000
25000
CAD
CKD
COPD
DM
Dementia
HF
Condition

Conclusions

Key points

- **Enhanced Detection Methodology**: Using expert-defined phenotyping rules with lab, procedure, and drug codes to identify patients with multimorbidity outperforms traditional condition codes.
- **Quantifiable Benefits**: Our automated method measures benefits from additional data sources, balancing true and false positives without manual labeling.
- **Future Improvements**: Integrate alternative data sources, like clinical text, to enhance recall and precision in phenotyping where condition coding is incomplete.

OHDSI meets Flowise to Streamline Biomedical Data Discovery and Analysis

João Almeida

University of Aveiro, Portugal

Software Demonstration

A Chatbot to Streamline Biomedical Data Discovery and Analysis

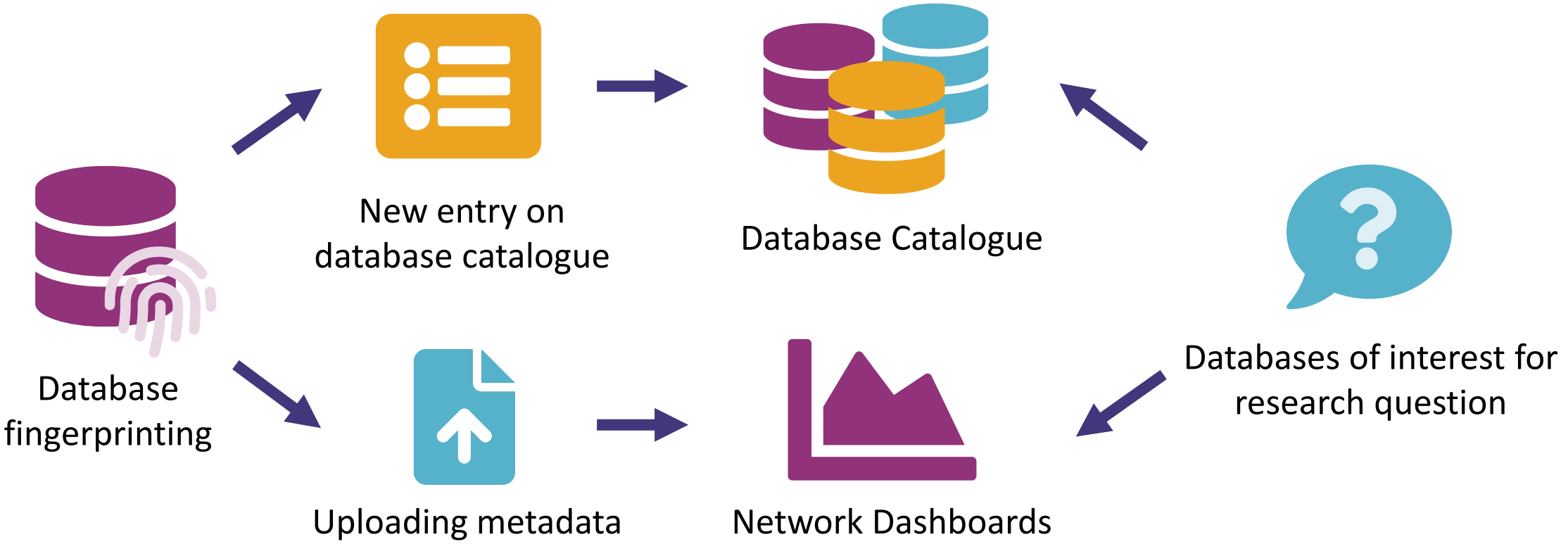
João Reis, **João Rafael Almeida**, Tiago Almeida,
José Luís Oliveira
University of Aveiro, Portugal

Motivation

- Population characterisation multicentre study
  - For instance, about Rheumatoid Arthritis patients
- Multicentre = multiple institutions
- Questions
  - Which datatypes are available?
  - Who are the contacts for the relevant databases?
  - What can we learn from the governance process?

Where do we find data?

EHDEN Network Dashboards

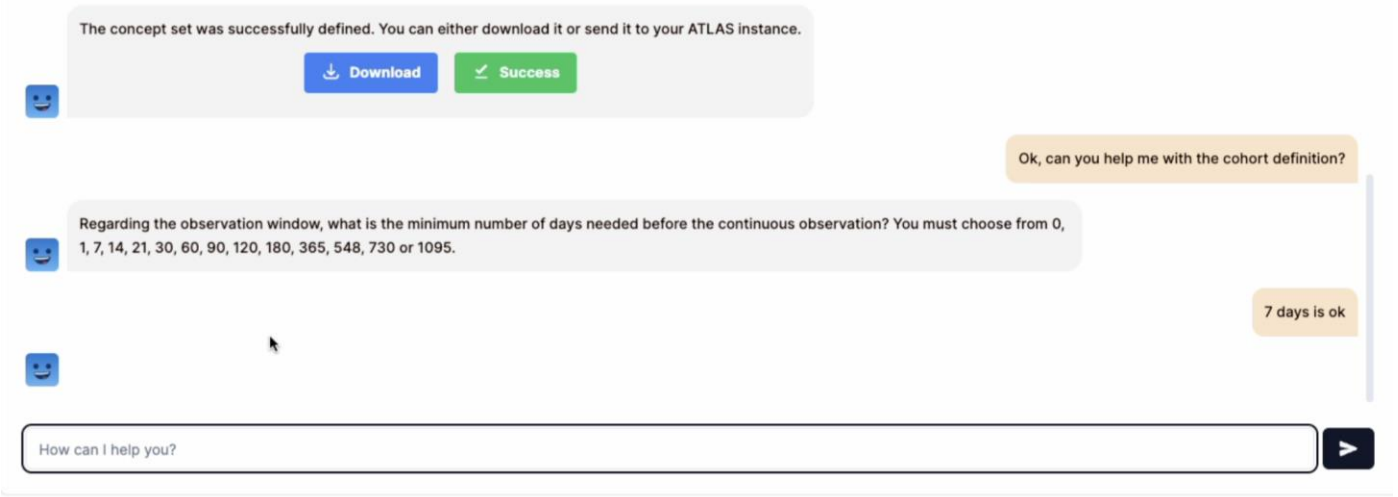


Data Custodian

Researcher

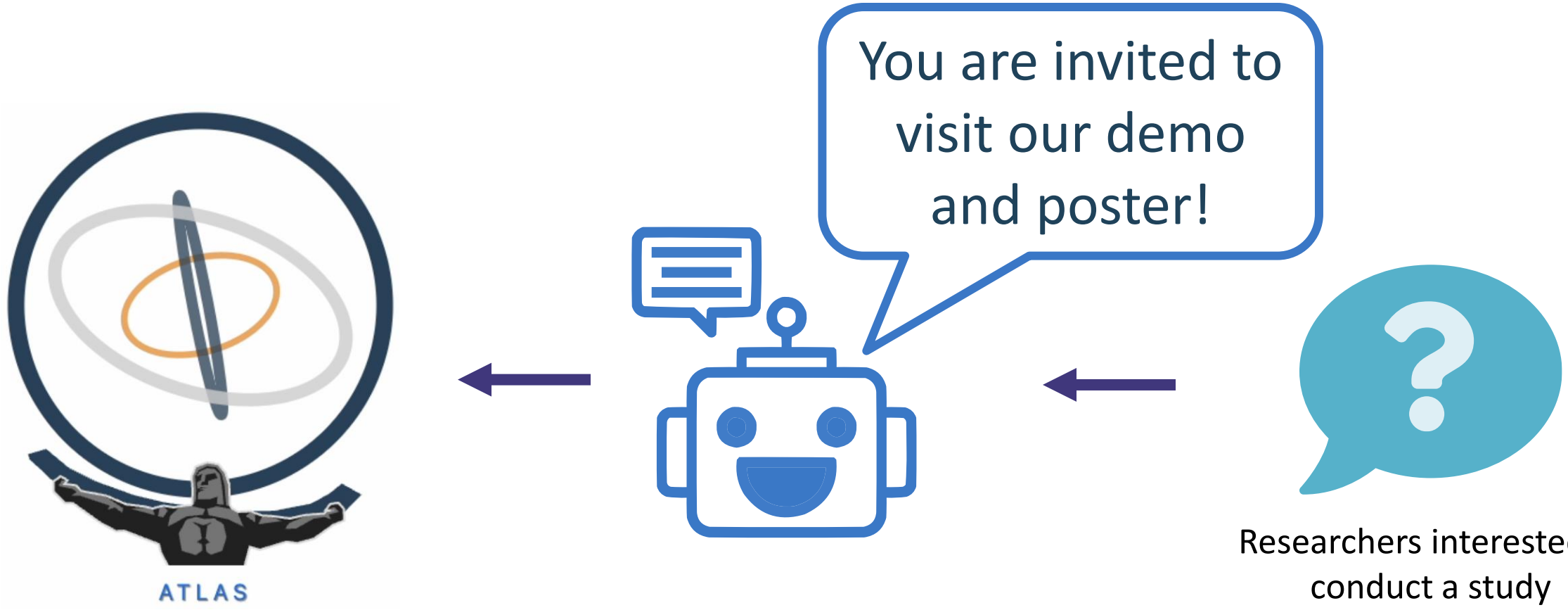
Chatbot overview

- Enable **discovery** and basic **feasibility enquiries**
  - Using EHDEN Network Dashboards data
- Compare basic characteristics
  - With other Data Partners in the network
- Provide suggestions of **most suitable databases**
  - Information Retrieval
  - Large Language Models
  - RAG techniques



Future directions

- Chatbot to help defining studies



ReportGenerator: Automating study reports and visualization apps for DARWIN EU® research

Cesar Barboza Gutierrez

Erasmus MC, The Netherlands

Software Demonstration

ReportGenerator

How to automate reports for off-the-shelf studies

Cesar Barboza, Ger Inberg and Ross Williams

OHDSI Europe Symposium 2024

DARWIN EU provides real-world evidence from across Europe

Developing tools to scale up to:

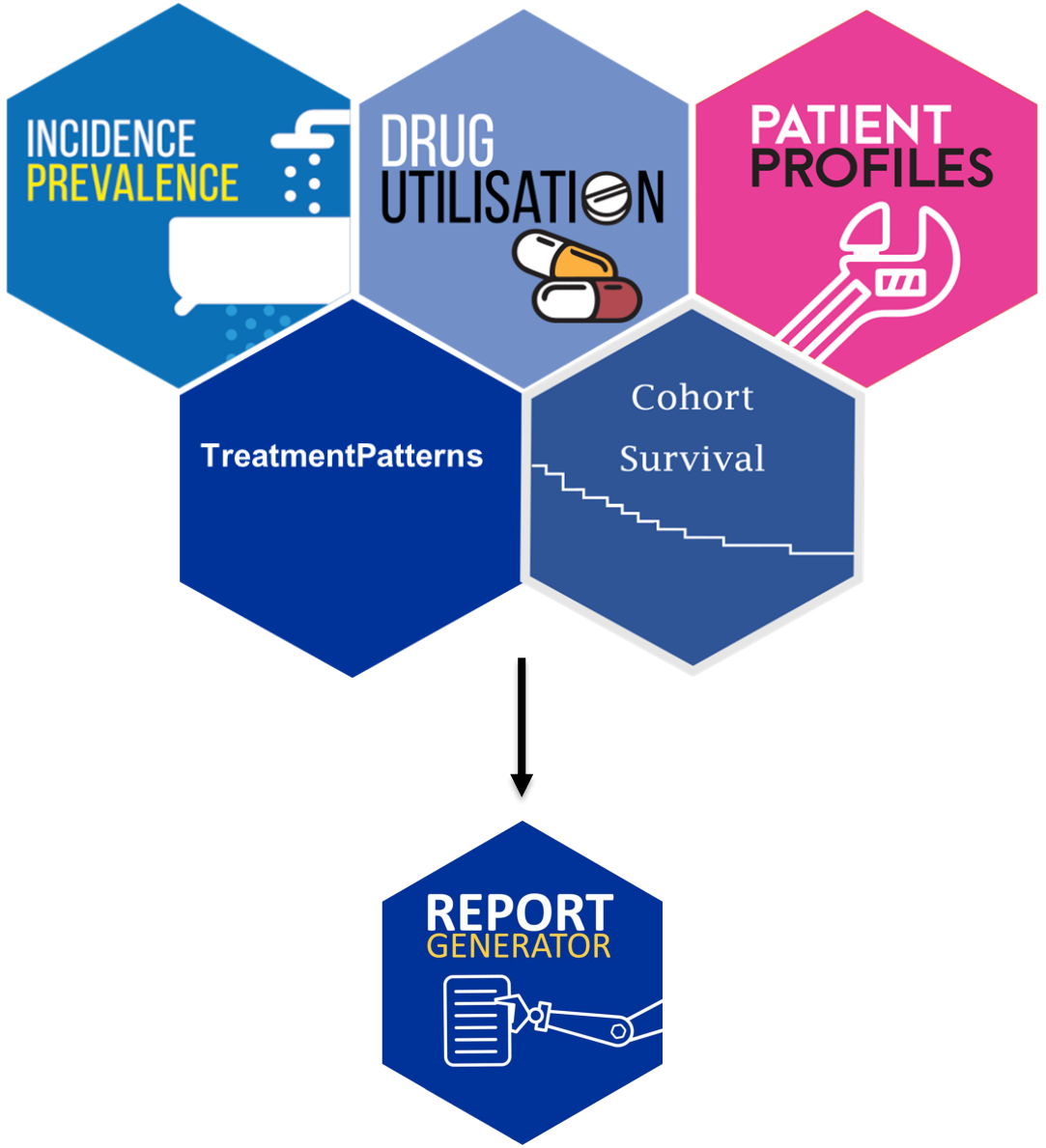
Analyzing Data
from
40 data partners
and
130 million patients

140
Studies delivered annually
by 2025

The Problem is complexity

It increases if we want to integrate results from multiple analytical packages

- Deliver standard reports to the European Medicines Agency (EMA).
- Reports should have a standardized format and be able to show results from multiple analyses.



ReportGenerator

The tool aims to:

1. Supports the Principal Investigators with creating tables and figures for the study report
2. Facilitate the automatic generation of a Shiny app for each study

Summarised Result

A standard output format for study results which makes it easier to bind results from multiple data partners

```
summary(cdm) |>
  dplyr::glimpse()
#> Rows: 12
#> Columns: 16
#> $ result_id        <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1
#> $ cdm_name         <chr> "example_cdm", "example_cdm", "example_cdm", "example…
#> $ result_type      <chr> "cdm_snapshot", "cdm_snapshot", "cdm_snapshot", "cdm_…
#> $ package_name     <chr> "omopgenerics", "omopgenerics", "omopgenerics", "omop…
#> $ package_version  <chr> "0.1.1", "0.1.1", "0.1.1", "0.1.1", "0.1.1", "0.1.1",…
#> $ group_name       <chr> "overall", "overall", "overall", "overall", "overall"…
#> $ group_level      <chr> "overall", "overall", "overall", "overall", "overall"…
#> $ strata_name      <chr> "overall", "overall", "overall", "overall", "overall"…
#> $ strata_level     <chr> "overall", "overall", "overall", "overall", "overall"…
#> $ variable_name    <chr> "snapshot_date", "person_count", "observation_period_…
#> $ variable_level   <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA
#> $ estimate_name    <chr> "value", "count", "count", "source_name", "version", …
#> $ estimate_type    <chr> "date", "integer", "integer", "character", "character…
#> $ estimate_value   <chr> "2024-03-09", "1", "1", "", NA, "5.3", "", "", "", ""…
#> $ additional_name  <chr> "overall", "overall", "overall", "overall", "overall"…
#> $ additional_level <chr> "overall", "overall", "overall", "overall", "overall"…
```

Data Visualization

Each analytical package provides its own tables and figures to display results according to DARWIN's Catalog of Standard Analytics

CDM name	Strata name	Strata level	Variable name	Variable level	Estimate name	Drug		
						Cisatracurium	Dexamethasone	Diazepam
CDWBordeaux	Icu status	Icu	Number records	Na	N	306	3,812	2,796
			Number subjects	Na	N	276	3,505	2,688
			Duration	Na	Median [Q25 - Q75]	1.00 [1.00 - 2.00]	2.00 [1.00 - 3.00]	2.00 [1.00 - 4.00]
					Range	1.00 to 25.00	1.00 to 49.00	1.00 to 72.00
			Initial daily dose	Na	Median [Q25 - Q75]	Na [na - na]	Na [na - na]	Na [na - na]
					Range	Inf to -inf	Inf to -inf	Inf to -inf

Tables and Figures

Incidence/Prevalence Plots

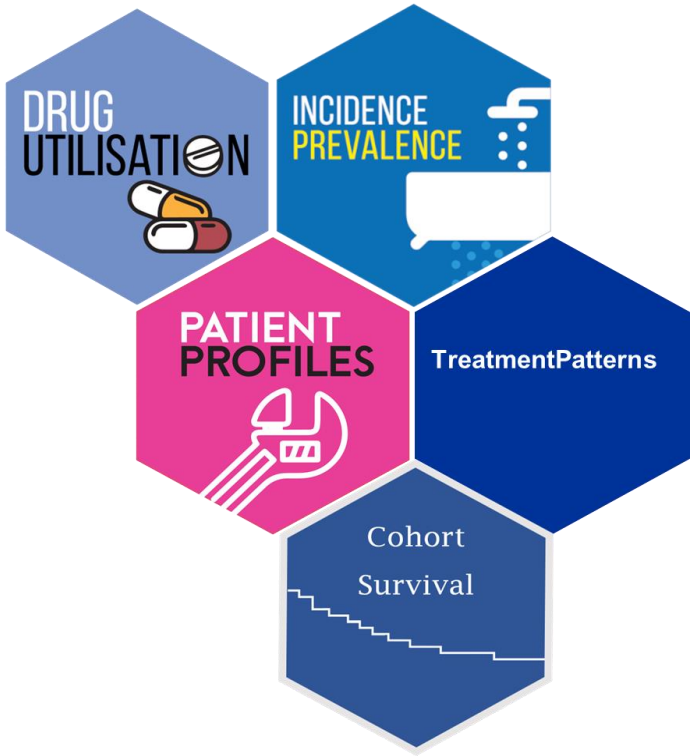
Survival Plots from CohortSurvival

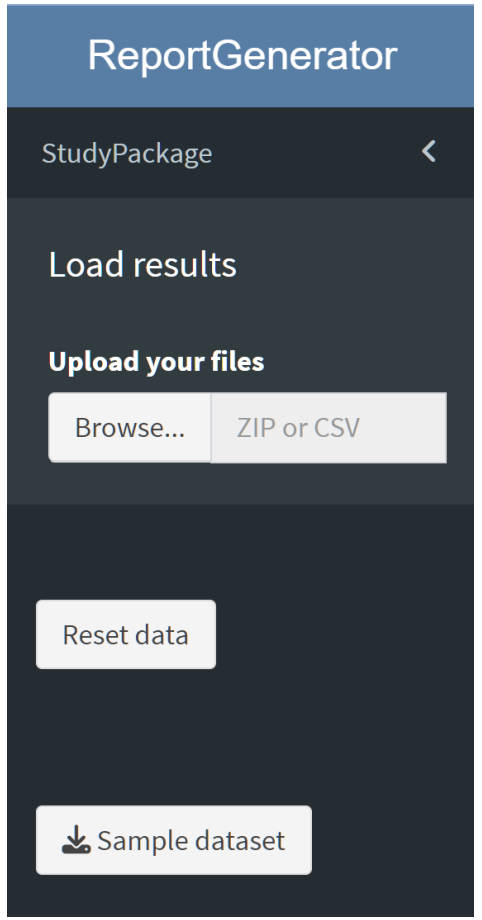
Characterisation tables for PatientProfiles

Sunburst plots and Sankey Diagrams for TreatmentPatterns

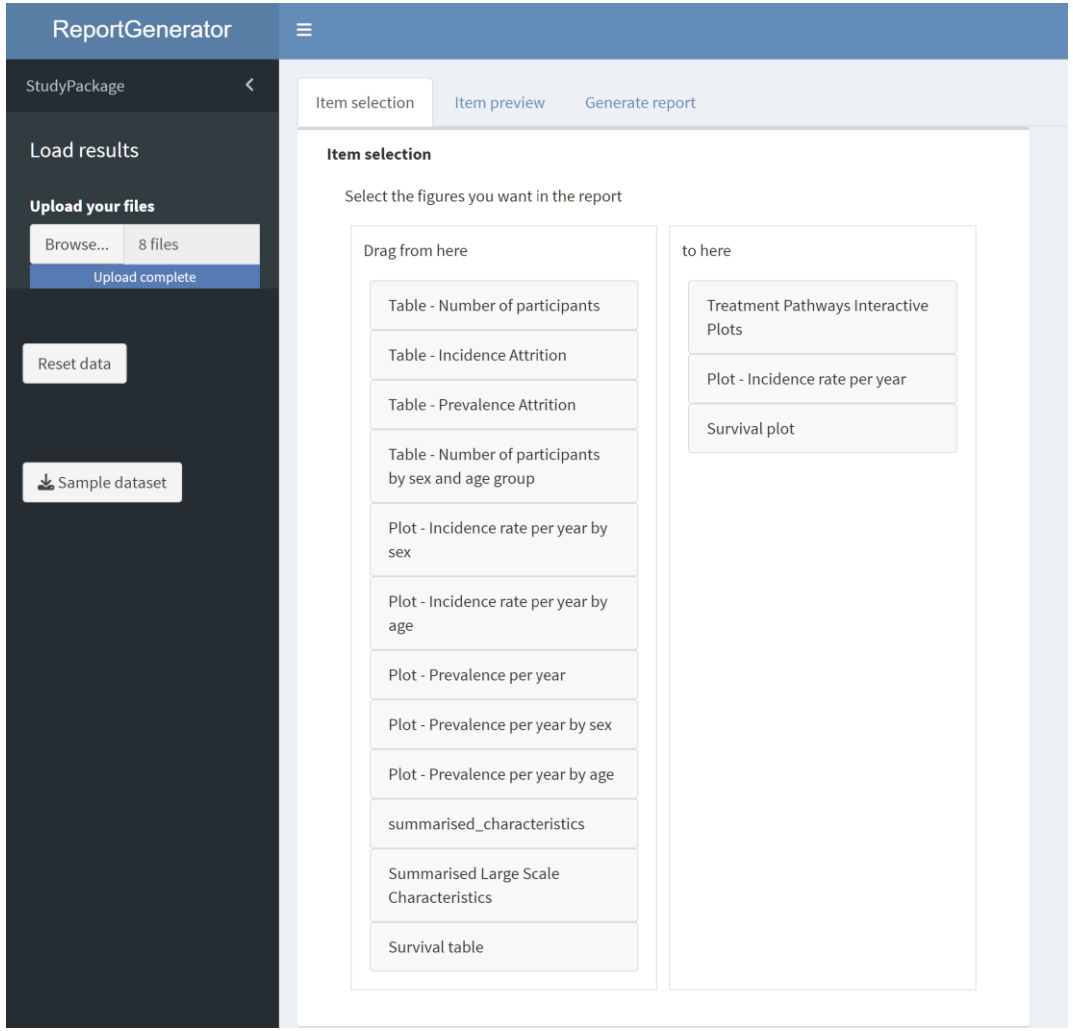
Functionality

1. Load data

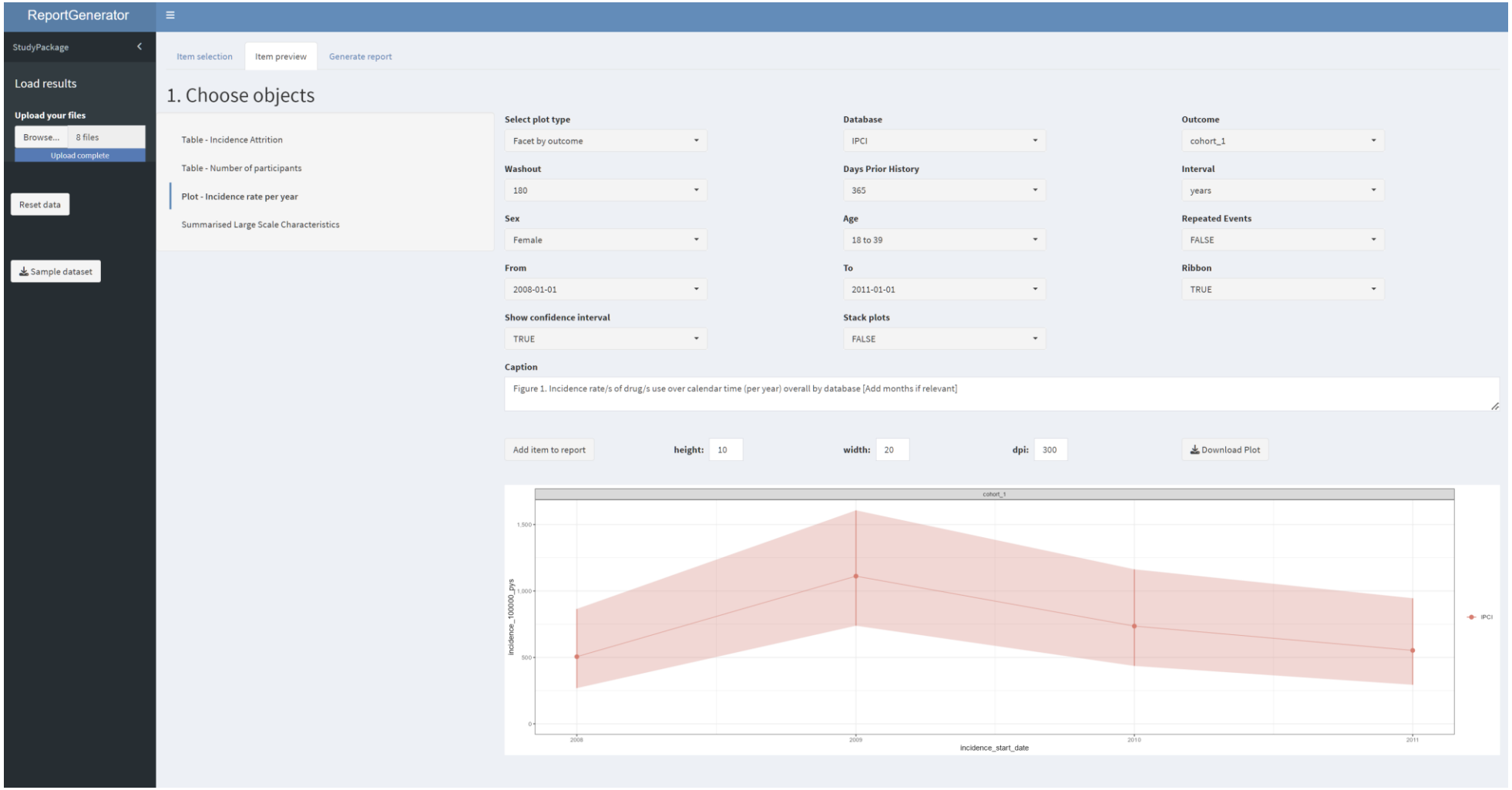




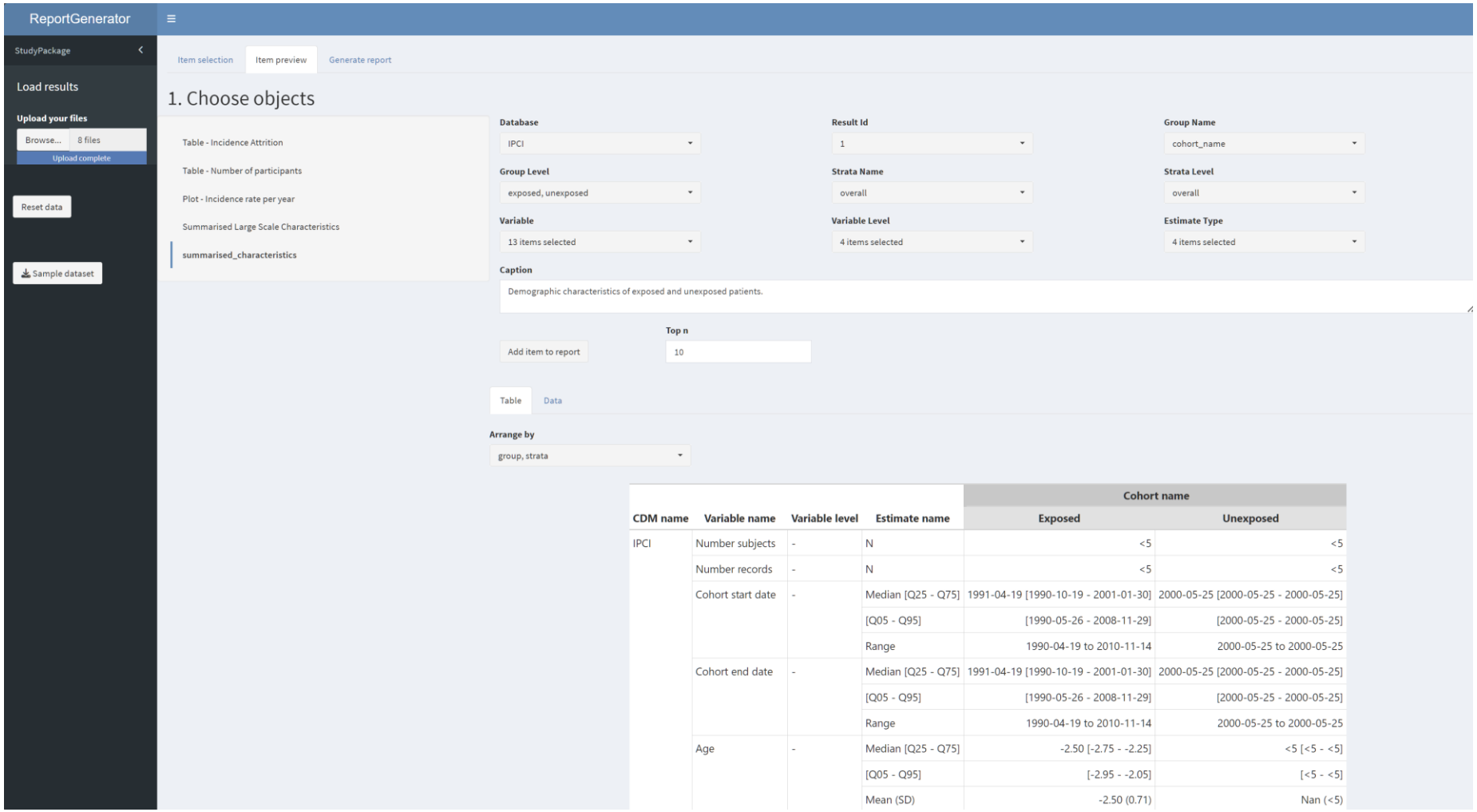
2. Interactive item selection



3. Visualization dashboard and item preview

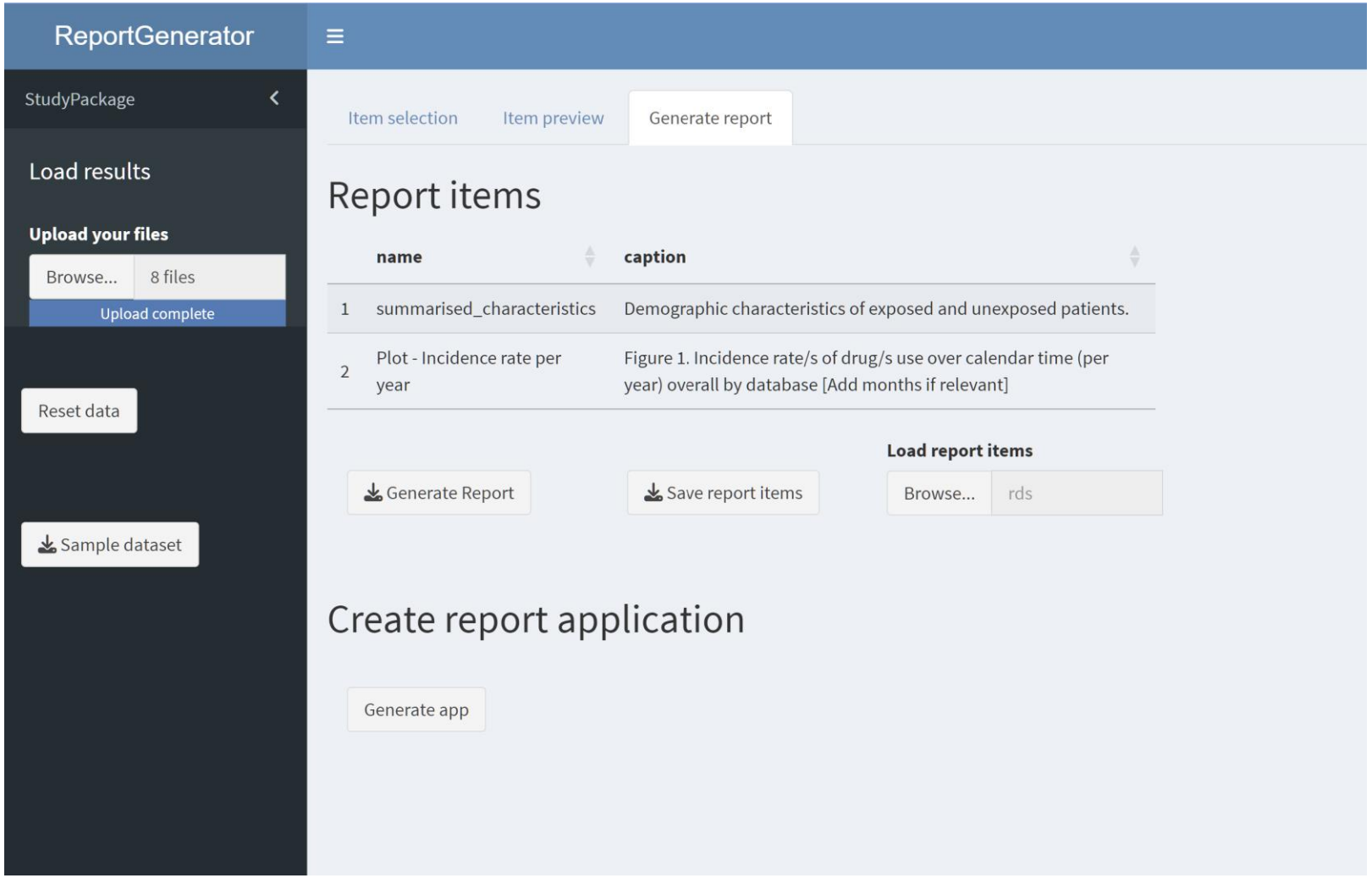


3. Visualization dashboard and item preview

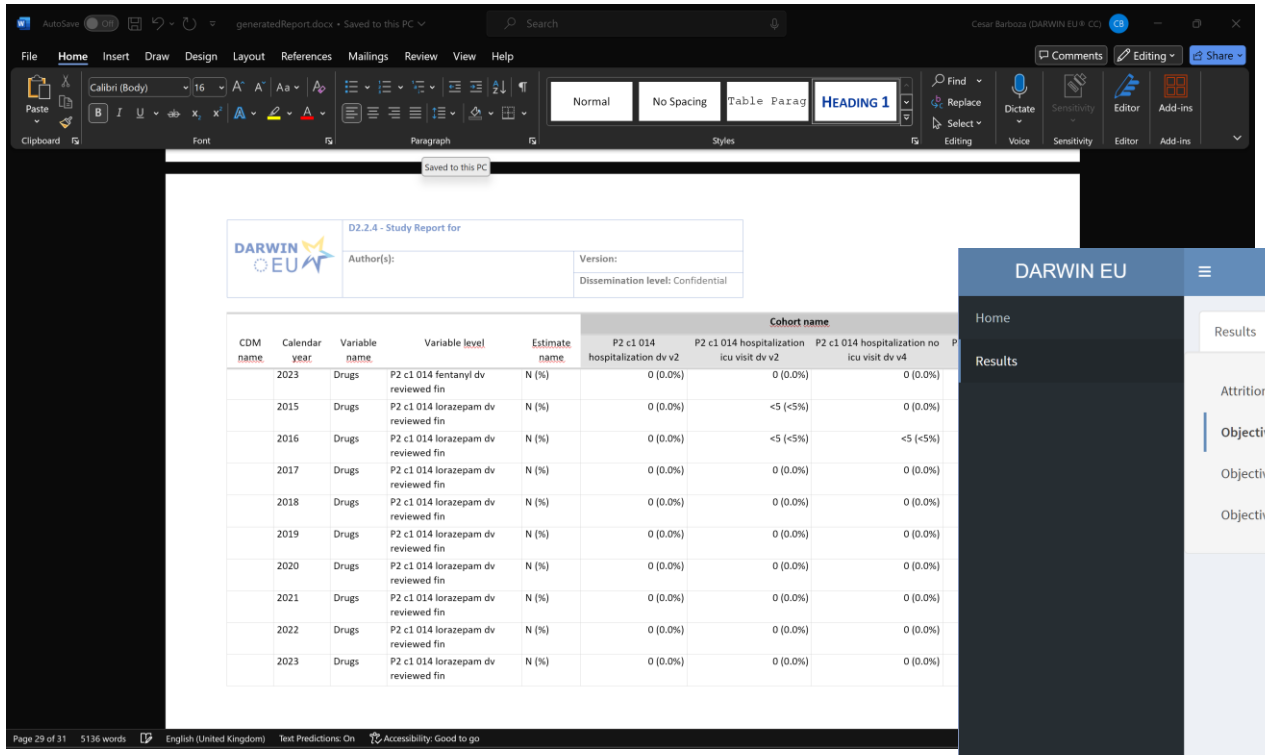


CDM name	Variable name	Variable level	Estimate name	Cohort name: Exposed	Cohort name: Unexposed
IPCI	Number subjects	-	N	<5	<5
	Number records	-	N	<5	<5
	Cohort start date	-	Median [Q25 - Q75]	1991-04-19 [1990-10-19 - 2001-01-30]	2000-05-25 [2000-05-25 - 2000-05-25]
			[Q05 - Q95]	[1990-05-26 - 2008-11-29]	[2000-05-25 - 2000-05-25]
			Range	1990-04-19 to 2010-11-14	2000-05-25 to 2000-05-25
	Cohort end date	-	Median [Q25 - Q75]	1991-04-19 [1990-10-19 - 2001-01-30]	2000-05-25 [2000-05-25 - 2000-05-25]
			[Q05 - Q95]	[1990-05-26 - 2008-11-29]	[2000-05-25 - 2000-05-25]
			Range	1990-04-19 to 2010-11-14	2000-05-25 to 2000-05-25
	Age	-	Median [Q25 - Q75]	-2.50 [-2.75 - -2.25]	<5 [<5 - <5]
			[Q05 - Q95]	[-2.95 - -2.05]	[<5 - <5]
			Mean (SD)	-2.50 (0.71)	Nan (<5)

4. Menu to generate a Shiny app or a Word report



Word report with tables and figures



D2.2.4 - Study Report for

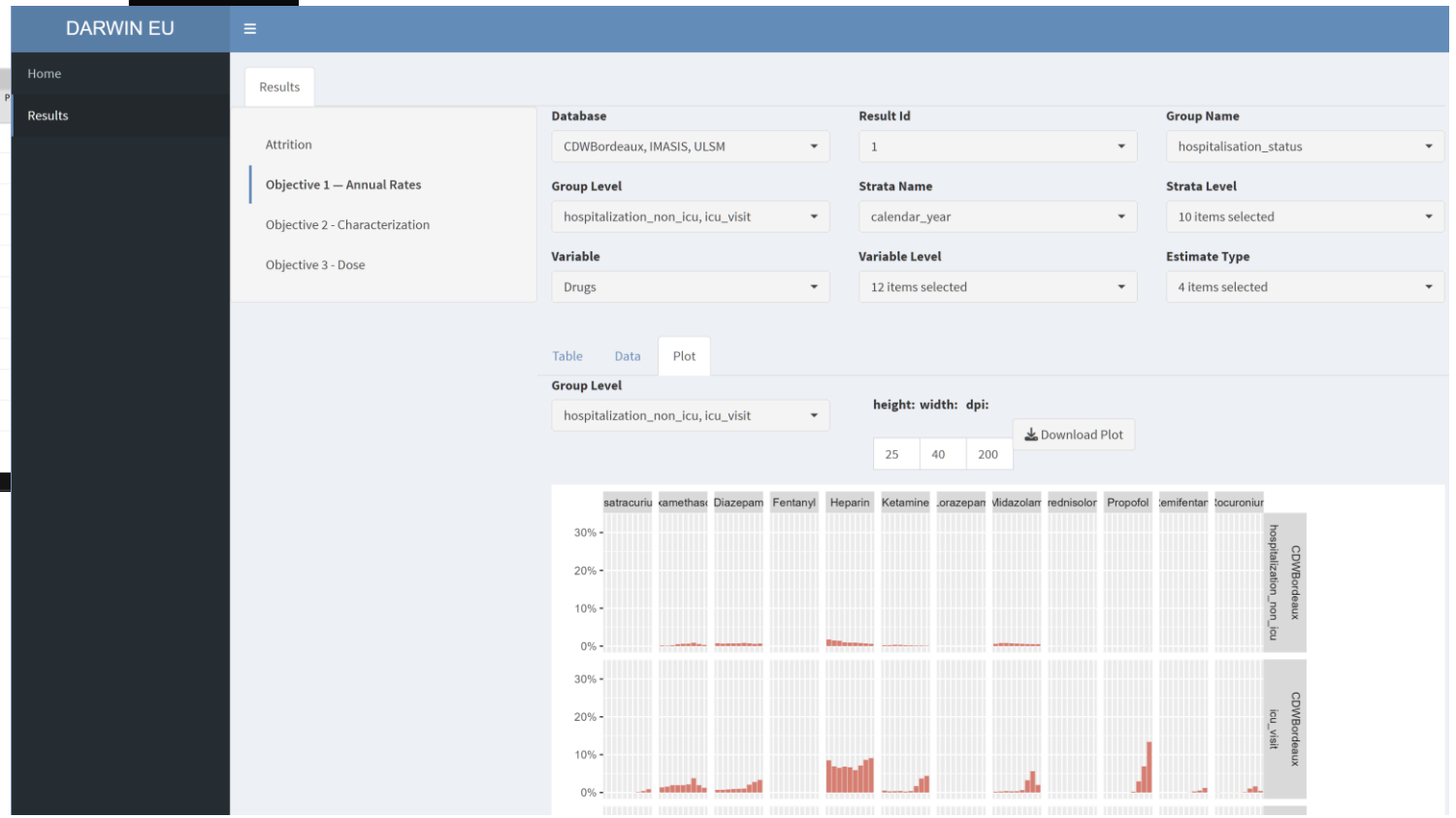
Author(s):

Version:

Dissemination level: Confidential

CDM name	Calendar year	Variable name	Variable level	Estimate name	Cohort name: P2 c1 014 hospitalization dv v2	P2 c1 014 hospitalization icu visit dv v2	P2 c1 014 hospitalization no icu visit dv v4
	2023	Drugs	P2 c1 014 fentanyl dv reviewed fin	N (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2015	Drugs	P2 c1 014 lorazepam dv reviewed fin	N (%)	0 (0.0%)	<5 (<5%)	0 (0.0%)
	2016	Drugs	P2 c1 014 lorazepam dv reviewed fin	N (%)	0 (0.0%)	<5 (<5%)	<5 (<5%)
	2017	Drugs	P2 c1 014 lorazepam dv reviewed fin	N (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2018	Drugs	P2 c1 014 lorazepam dv reviewed fin	N (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2019	Drugs	P2 c1 014 lorazepam dv reviewed fin	N (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2020	Drugs	P2 c1 014 lorazepam dv reviewed fin	N (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2021	Drugs	P2 c1 014 lorazepam dv reviewed fin	N (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2022	Drugs	P2 c1 014 lorazepam dv reviewed fin	N (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
	2023	Drugs	P2 c1 014 lorazepam dv reviewed fin	N (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

Shiny project with full functionality



Next steps:

- Complete integration with summarisedResult format
- Interactive selection of features when generating the Shiny project
- Provide enough functionality to support results for all studies

Thank you!

Analysis of Lung Cancer Patient Treatment with Immune Checkpoint Inhibitors Using Natural Language Processing for Data Extraction from Electronic Health Records

Clara L. Oeste and Annelies Verbiest

Lynxcare, Belgium

Poster: 67

EUROPEAN OHDSI SYMPOSIUM

EUROPE

Analysis of Lung Cancer Patient Treatment with Immune Checkpoint Inhibitors using NLP for Data Extraction from EHRs

Dr. Annelies Verbiest, MD, PhD

Department of Oncology, Antwerp University Hospital, Antwerp, Belgium

Annelies.Verbiest@uza.be

Clara L. Oeste, PhD

LynxCare Inc., Leuven, Belgium

clara.oeste@lynx.care

June 1 - 3 2024

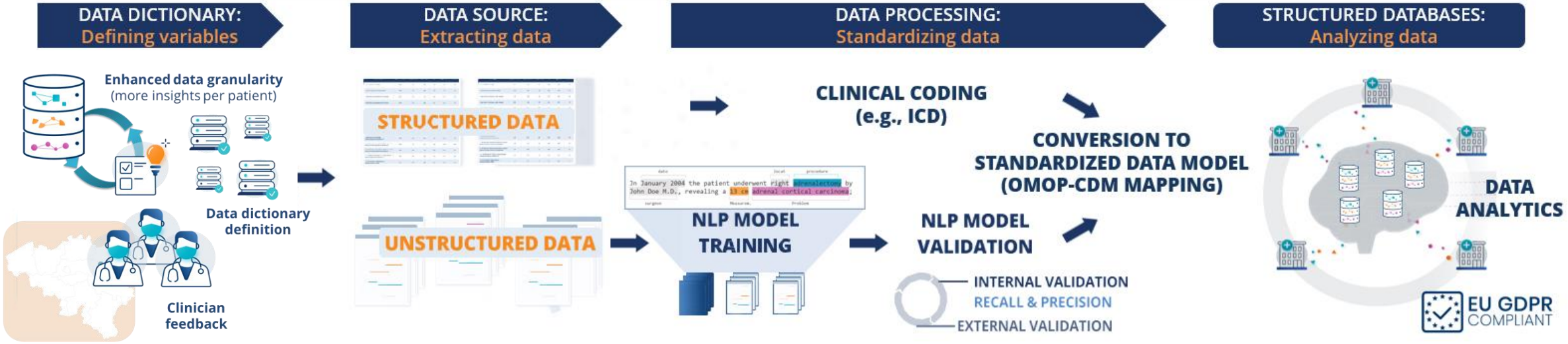
Rotterdam

Oncology Research Group Setup

A Belgian Network for Oncology Research

Objectives of the research group:

- Set up a pipeline to develop large-scale **OMOP-CDM granular warehouses** (GDPR compliant)
- Bring together a **consortium of hospitals** (EHDEN project)
- Perform project-specific **data dictionary and validation**



Study Design

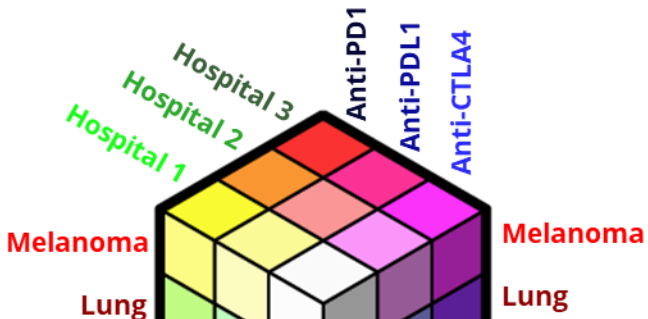
A Multidimensional Data Warehouse

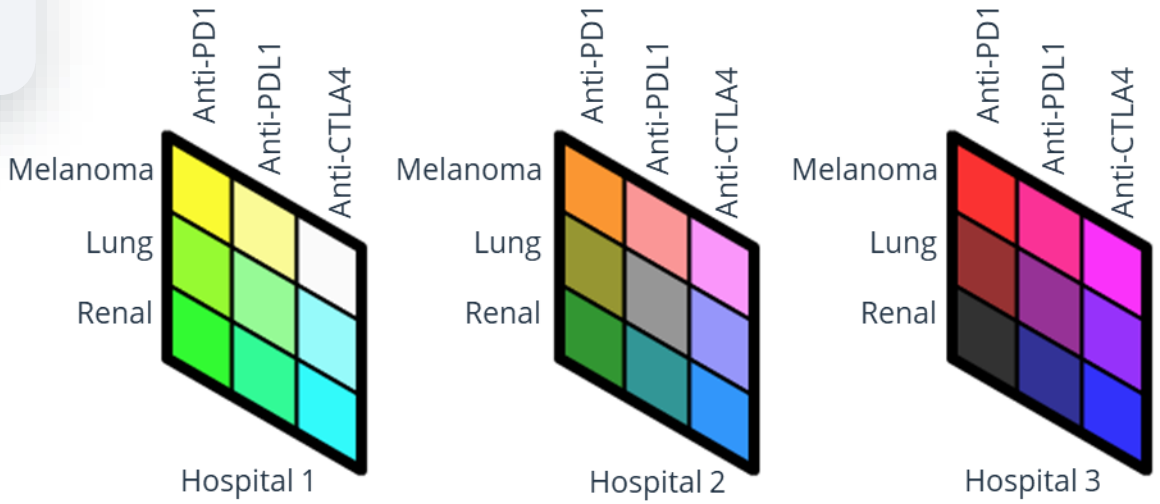
Assessing the use of immune checkpoint inhibitors (ICI) across different:

- Cancer types
- ICIs
- Treatment settings
- Hospitals
- Performance status
- ...

... and in DIFFERENT COHORTS

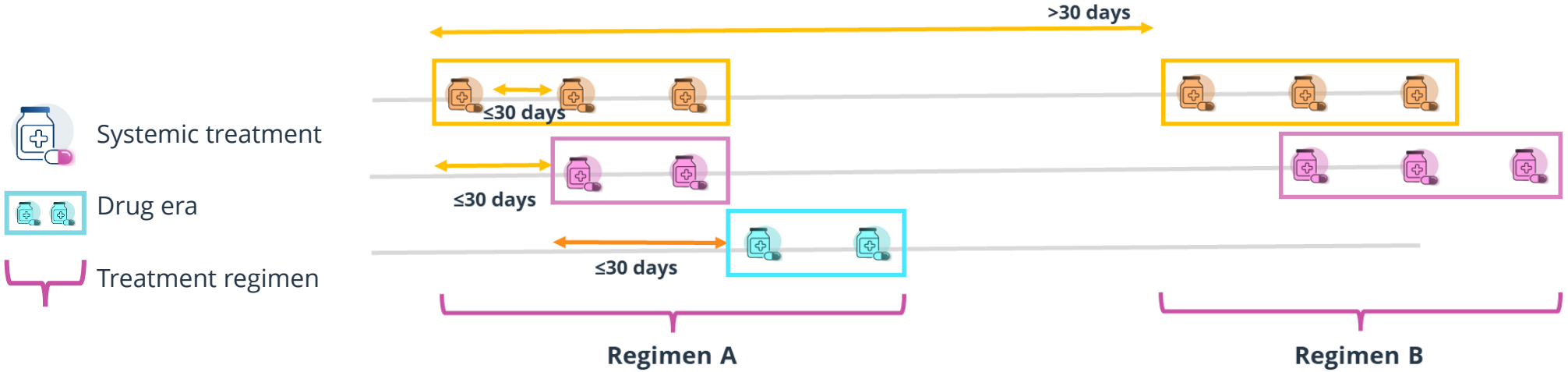
ALL | **Patients with adverse events** | **Frail patients**

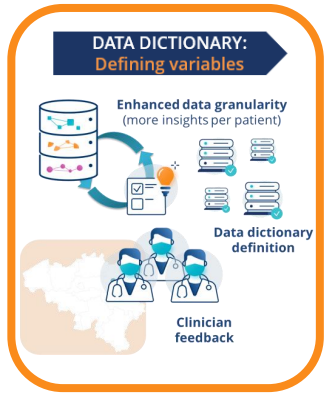




Data Dictionary Example

Determining patient treatments (OncoRegimenFinder)





SYSTEMIC TREATMENT

- **Calculation logic:**
  - All antineoplastic agents
  - Follows ATHENA hierarchy
- **Data source:** structured
- **OMOP table:** Drug exposure

DRUG ERA

- **Calculation logic:**
  - Span of time when Person is exposed to active ingredient.
  - Successive periods of exposure: ERA
  - Max time window between exposures: 30 days
- **Data source:** structured
- **OMOP table:** Drug exposure

TREATMENT REGIMEN

- **Calculation logic:**
  - Follows **OncoRegimenFinder**
  - Occurrence of ≥ 2 drug eras within 30 days: combined into REGIMEN
- **Data source:** structured
- **OMOP table:** Drug era

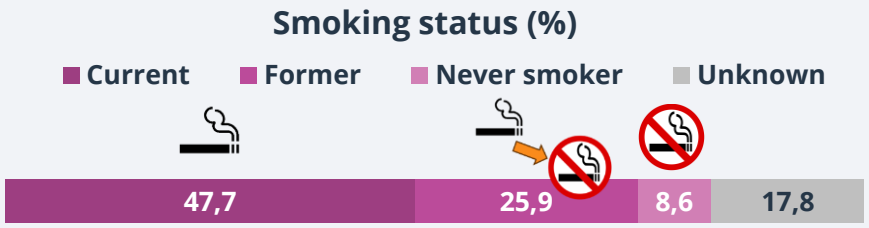
TREATMENT LINE

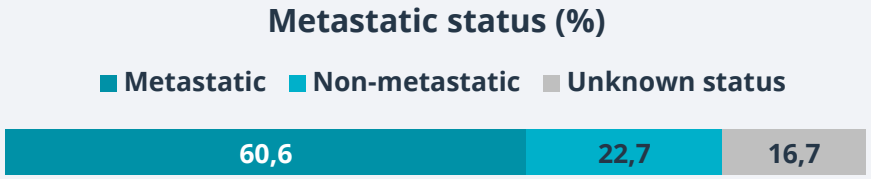
- **Calculation logic:**
  - First regimen after initial cancer diagnosis
  - Change in line if change in regimen: when new drug introduced
- **Data source:** structured
- **OMOP table:** Episode

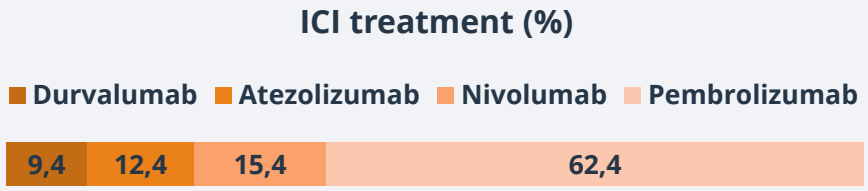
Clinical validation at a glance

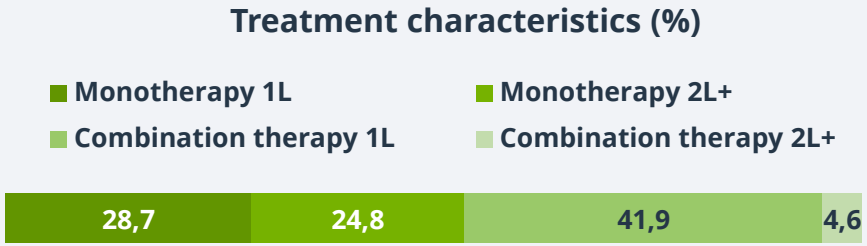
Results

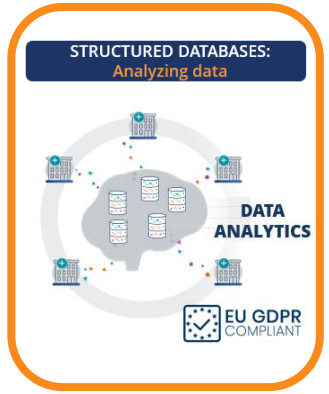
Hospitals	3
Data sources	10
Lung cancer patients	730
Median age	67
Sex (male)	67%
ICI administrations	8145
Median OS	22 months

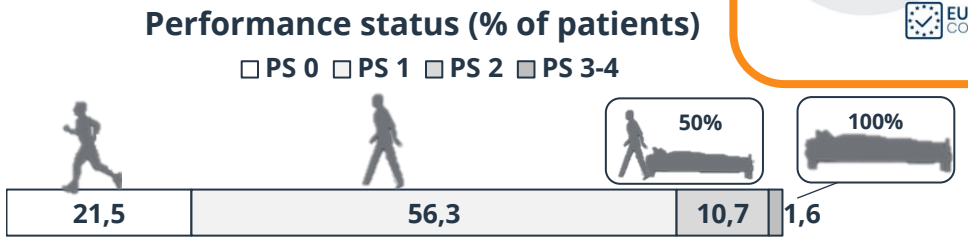


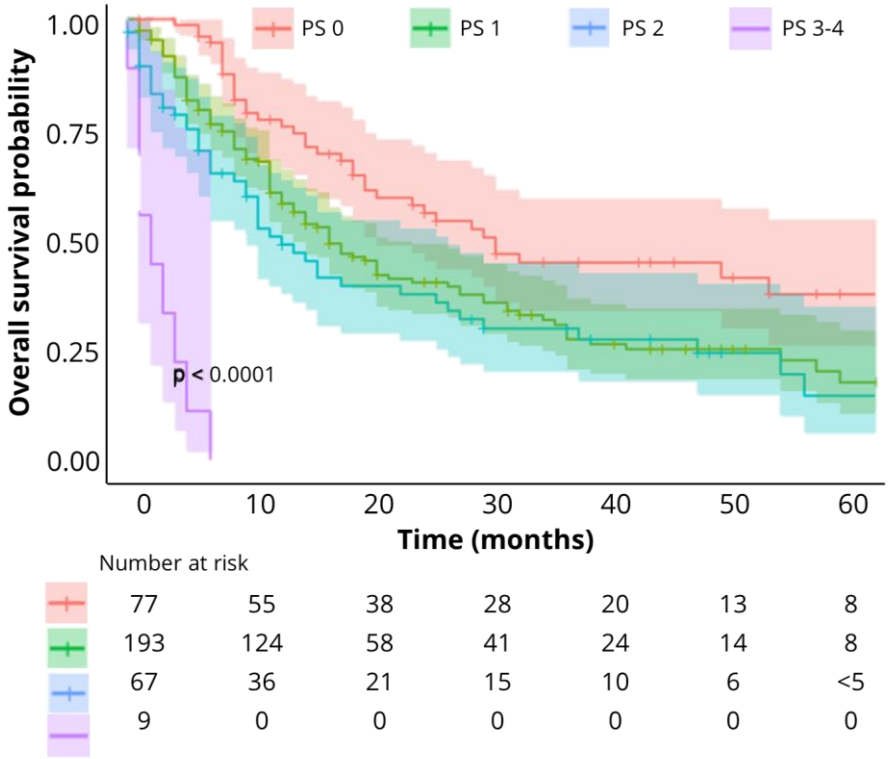












An Exploration of Ovarian Cancer Therapy Sequence Utilization in Treatment-naïve Women from 2008-2020

Whitney Burton

Taipei Medical University, Taiwan

Poster: 24

An Exploration of Ovarian Cancer Therapy Sequence Utilization in Treatment-naive Women from 2008-2020

Whitney Burton, Quynh Nguyen, Mohammad Solihuddin Muhtar,
Christianus Heru Setiawan, Septi Melisa, & Jason Hsu
Student, Taipei Medical University
College of Management
College of Pharmacy

Background

- In women, ovarian cancer is the 8th leading cancer and the 8th cancer-related mortality cause[1]
- Regionally, women located in Europe and Southeast Asia at at-increased vulnerability for the disease[1,2,3]
- Low and medium-income counties experience disproportional mortality rates in juxtaposition with incident rates[2]
- Global treatment guidelines call for appropriate surgical staging and debulking surgery followed by biomedical interventions (i.e., oral or intravenous)[1]

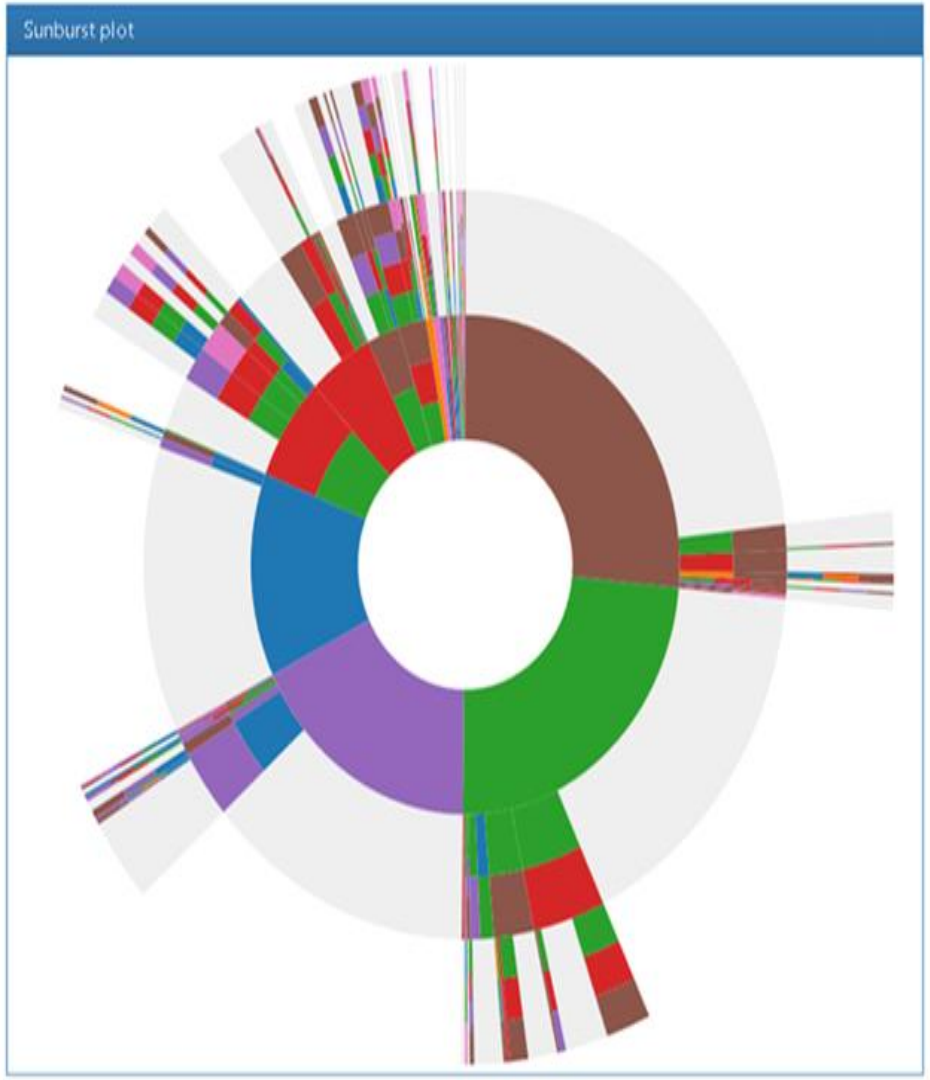
References:

1. Armstrong, D. K., Alvarez, R. D., Bakkum-Gamez, J. N., Barroilhet, L., Behbakht, K., Berchuck, A., ... & Engh, A. M. (2021). Ovarian cancer, version 2.2020, NCCN clinical practice guidelines in oncology. *Journal of the National Comprehensive Cancer Network*, *19*(2), 191-226.
2. Cabasag, C. J., Fagan, P. J., Ferlay, J., Vignat, J., Laversanne, M., Liu, L., ... & Soerjomataram, I. (2022). Ovarian cancer today and tomorrow: A global assessment by world region and Human Development Index using GLOBOCAN 2020. *International Journal of Cancer*, *151*(9), 1535-1541.
3. Lowe, K. A., Chia, V. M., Taylor, A., O'Malley, C., Kelsh, M., Mohamed, M., ... & Goff, B. (2013). An international assessment of ovarian cancer incidence and mortality. *Gynecologic oncology*, *130*(1), 107-114.

Study Design

- Aim: To characterize real-world ovarian cancer therapy sequence utilization patterns
- Dataset: Taipei Medical University Clinical Research Database (TMUCRD)
- Study Population: 1,190 women treatment-naïve women from 1/1/2008-31/12/2020
- Tools:
- Outcon
  - 1st three medication pathways
  - 3 pathways: Chemotherapy, Targeted therapy, and Hormone therapy classes
    - 13 sub-class for the analysis: Alkaloids, Alkylating agents, Anthracyclines, Antimetabolites, Taxanes, Anti-estrogens, Aromatase inhibitors, PD-L1 inhibitors, CKD inhibitors, HER2, monoclonal antibodies, mammalian target of rapamycin (mTORs), and multi-target inhibitors)

Preliminary Findings



Implications:

- Leveraging of OHDSI tools enabled a standardized analysis of treatment pathways
- Improves our understanding of therapy utilization patterns
- This study can guide resource allocation within healthcare systems by highlighting gaps in treatment, accessibility, and clinical alignment/misalignment with guidelines

Acknowledgments

- Research team at Taipei Medical University's College of Management
- National Science and Technology Council of Taiwan
- Early insight from international partners
  - Dr. Seng Chan You, Yonsei University
  - Dr. Nicole Pratt, University of South Australia
  - Dr. Celine Sze Ling Chui, The University of Hong Kong

Call to Action

*"We're told to **sit with discomfort and that it's normal** until it's too late. Then, we're questioned about not being proactive…I won't be siloed or silenced any longer. I am one person. A solo datapoint. **But if we come together, we can make a dataset, and that set makes all the difference**. It has the power to change fortunes and futures from prediction models to improve the quality of care. Let's make a **representative data tapestry that reflects us and changes the narrative.** Let's change the course of ovarian and gynecological cancers."*

-39 yo Female

Baseline Characterization and Treatment Pathways of Patients With Alport Syndrome Across Geographies: Exploring a Rare Disease in a Multi-Database Retrospective Cohort Study

Katrin Manlik

Bayer AG, Germany

Poster: 79

Rapid Fire Presentation
OHDSI Europe Symposium 2024

Baseline Characterization and Treatment Pathways of Patients With Alport Syndrome Across Geographies: Exploring a Rare Disease in a Multi-Database Retrospective Cohort Study

Katrin Manlik, Glen James, Andrea Scalise, Charlie Scott, Daloha Rodriguez Molina, David Vizcaya (Bayer)

Rotterdam, June 03, 2024

Rare genetic kidney disease

Usually manifests in early childhood

Mutations in the COL4A3, COL4A4 and COL4A5 genes lead to defective collagen production

Alport syndrome

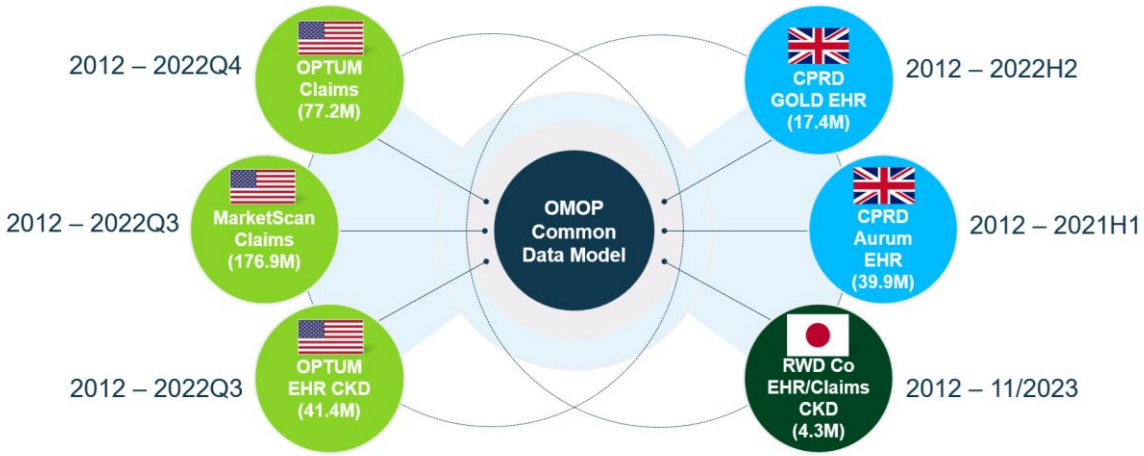
Patients treated with Angiotensin-converting enzyme inhibitors (ACEi), Angiotensin receptor blockers (ARB) and Sodium-glucose cotransporter-2 inhibitors (SGLT2i) to delay onset of kidney failure

Hematuria, proteinuria, ocular abnormalities, hearing loss, progressive loss of kidney function leading to kidney failure

Research question:
What are the characteristics of patients diagnosed with Alport syndrome and how are these patients treated in a real-world setting across different countries?

Methods

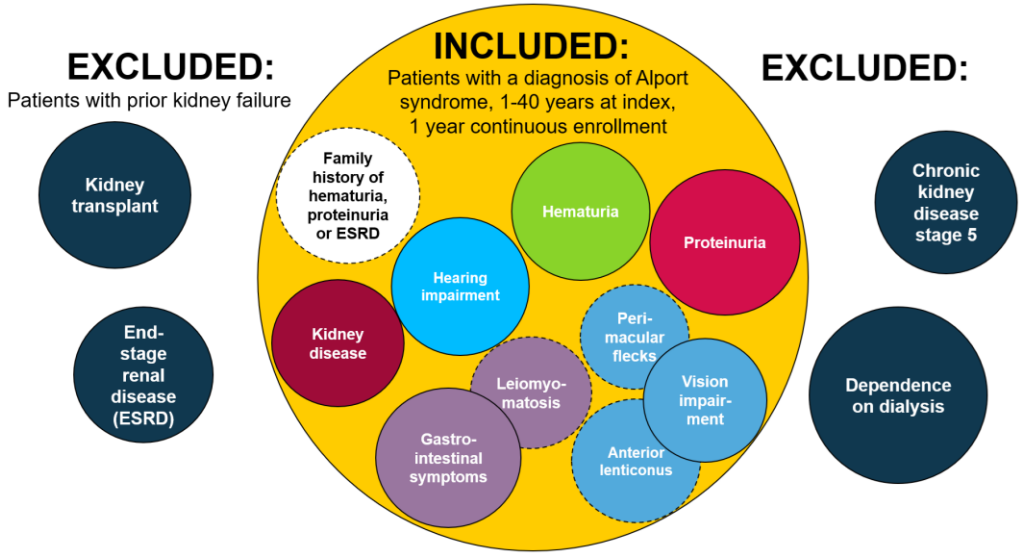
Data sources



Attrition table

	CPRD Aurum EHR	CPRD GOLD EHR	MarketScan Claims	OPTUM Claims	OPTUM EHR	RWD Co EHR Claims
Total # of patients in database (in million)	39.9M	17.4M	176.9M	77.2M	41.4M	4.3M
Step 1: Inclusion criteria: At least 1 diagnosis code of Alport syndrome	542	283	2116	1613	2424	100
Step 2: Inclusion criteria: Age between 1 and 40 years	398	212	1303	784	1302	65
Step 3: Inclusion criteria: 1 year of continuous enrolment before index date ([-365,0] days)	310	153	696	370	948	35
Step 4: Study start date 01-Jan-2012	162	59	696	370	904	19
Step 5: Exclusion criteria: Exclude patients with kidney failure prior to or at index date **Final Cohort**	**158**	**58**	**585**	**314**	**688**	**16**

Inclusion & Exclusion Criteria and Comorbidites of interest



A longitudinal retrospective cohort study

- ✓ 6 OMOP databases from 3 countries, each analyzed separately
- ✓ Study start date 01-JAN-2012
- ✓ Inclusion criteria:
  - 1 diagnosis code for AS
  - age between 1 and 40 years at index
  - at least 12 months of continuous enrolment
- ✓ Exclusion criteria:
  - prior kidney failure before or on index

Results – Baseline characteristics

Overall **1819** AS patients were identified from 6 databases across 3 countries

Variable	CPRD Aurum EHR	CPRD GOLD EHR	MarketScan Claims	OPTUM Claims	OPTUM EHR	RWD Co EHR Claims
Country	UK	UK	USA	USA	USA	Japan
Patient Count	158	58	585	314	688	16
Female %	52.5	41.4	54.7	51.3	57.6	56.2
Age at diagnosis (in years)						
Overall Median (IQR) Age	13 (8-28)	14 (7-25)	19 (10-29)	19 (10-32)	23 (13-32)	24 (17-26)
Female Median (IQR) Age	16 (9-30)	14 (9-30)	23 (12-32)	24 (11-33)	27 (18-33)	24 (24-25)
Male Median (IQR) Age	11 (6-24)	14 (5-18)	16 (9-26)	16 (9-25)	17 (9-28)	23 (4-26)

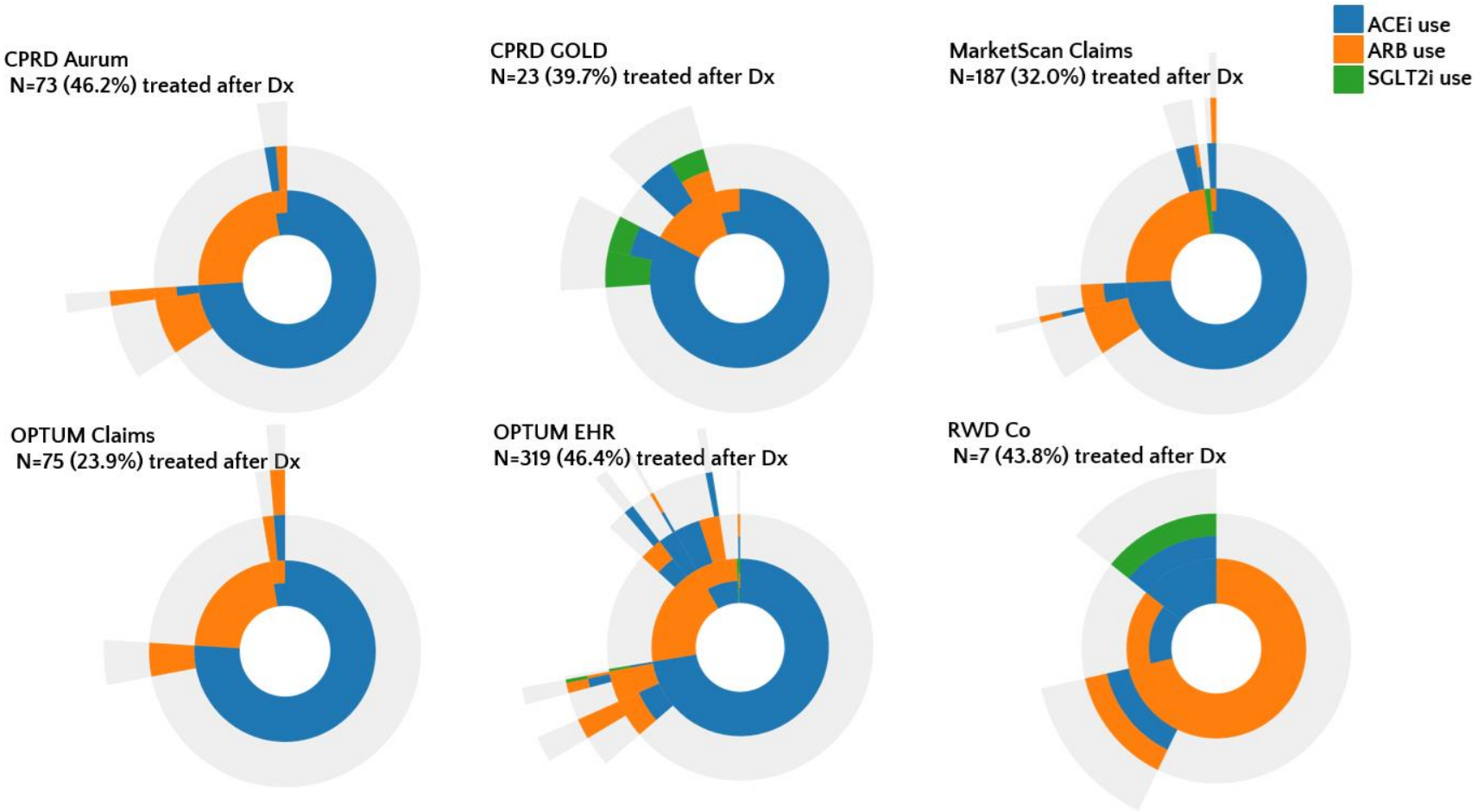
Demographics

- A **higher proportion of females** with AS in all DBs except UK CPRD GOLD.
- In the **US**, patients were diagnosed with AS around the age of 20. **Male** patients were **7-10 years younger** than females at index.
- In the **UK**, patients were diagnosed with AS in their early teens, in **Japan** around the age of 24.

Comorbidities

- Across all data sources hematuria (12-56%), proteinuria (6-44%) and kidney disease (22-69%) were common.
- **Arterial hypertension** ranged from 5 to 44%.
- **Hearing impairment** was more prevalent in **males** compared to females in all databases.
- **Hematuria** and **kidney disease** were more prevalent in **females** compared to males in the US.
- **Vision impairment** was prevalent in up to a quarter of patients (3-25%).

Results - Treatment pathways after diagnosis



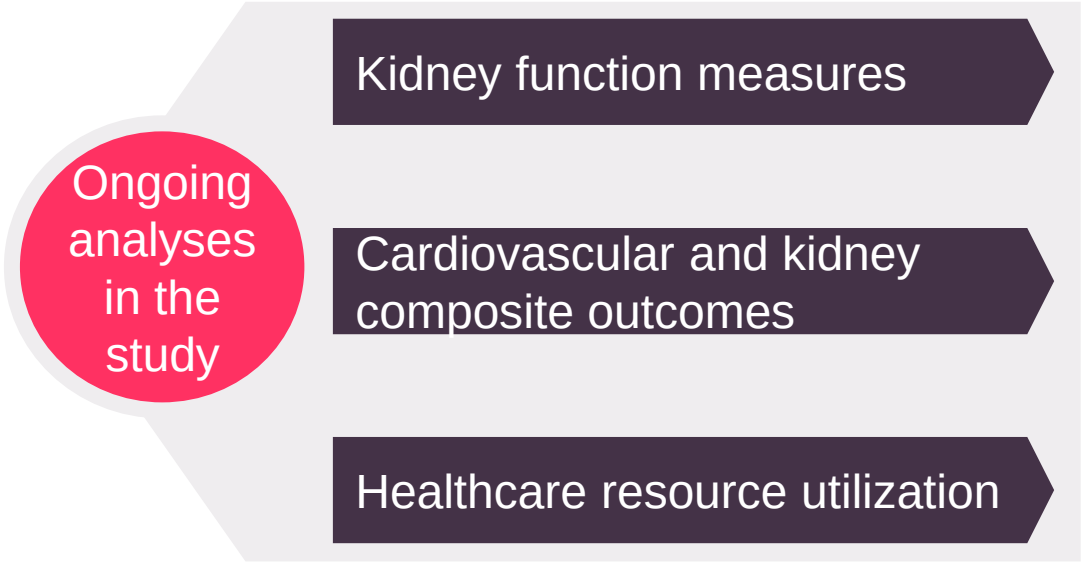
- **ACEi most frequently used** 1st line therapy in US and UK – around 3/4 of patients.
- **ARBs second most frequently used** 1st line therapy in US and UK – nearly 1/4 of patients, but most frequently used in Japan.
- **SGLT2i** were **rarely used** in the AS population.
- **Less than half of patients** were **treated** with cardiorenal protective therapies after diagnosis.

Conclusions & Next steps

1. **Alport syndrome**, a rare genetic kidney disease, shows **notable gender and regional differences** in patient characteristics.

2. This study provides **new insights into demographics, clinical characteristics, and treatment utilization** of patients with AS. These data are useful to gain knowledge about the disease, provide better support to clinicians and healthcare providers and most importantly, improve patient's quality of life.

3. Use of **OMOP data sources and OHDSI tools** provides an excellent opportunity to gain **insights into rare diseases** across multiple geographies and healthcare settings in a standardized approach.



We are eager to expand the study to further databases, Data Partners are welcome to participate!

Thank you!

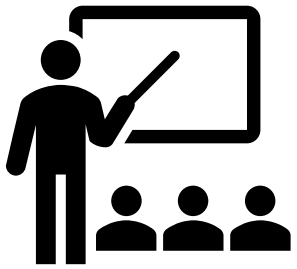
katrin.manlik@bayer.com

Collaborator Showcase: Rapid Fire Presentations

THANK YOU!

Lunch, Collaborator Showcase, and Early Investigator meetings

La Fontaine & Odyssee Room



The Collaborator Showcase
13:00

Queen's Lounge

Early Investigators Mentor Meetings
14:00

We will be back here at 16:00!