Heterogeneity in DARWIN EU: the truth vs the artifact

Dani Prieto-Alhambra
Deputy Director, DARWIN EU

Disclaimer

The views expressed in this presentation are my personal views and may not be understood or quoted as being made on behalf of or reflecting the position of the European Medicines Agency or one of its committees or working parties.

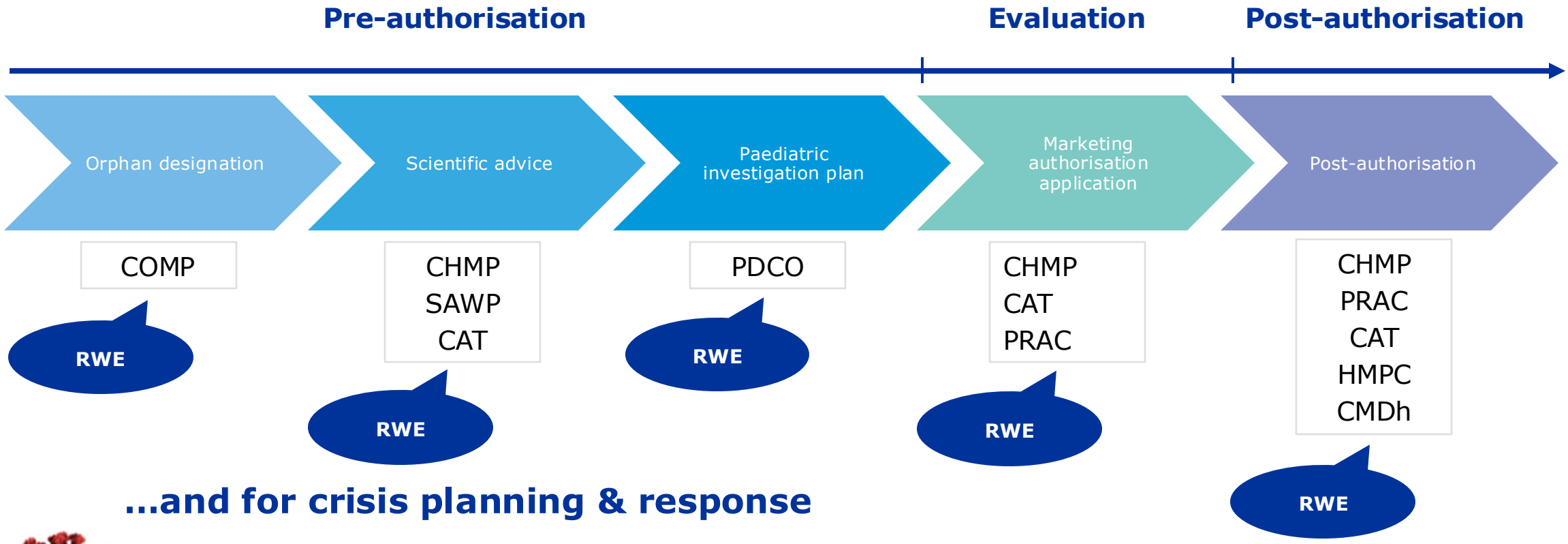
AGENDA

- What data? The DARWIN EU® network
- Standardising data – and analytics
- The phenotyping process in DARWIN EU®
- Is heterogeneity always a ‘curse’?

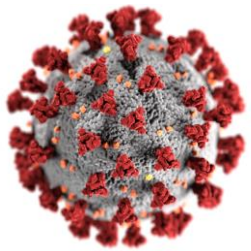
AGENDA

- What data? The DARWIN EU® network
- Standardising data – and analytics
- The phenotyping process in DARWIN EU®
- Is heterogeneity always a 'curse'?

Demand for data: across the medicinal product lifecycle

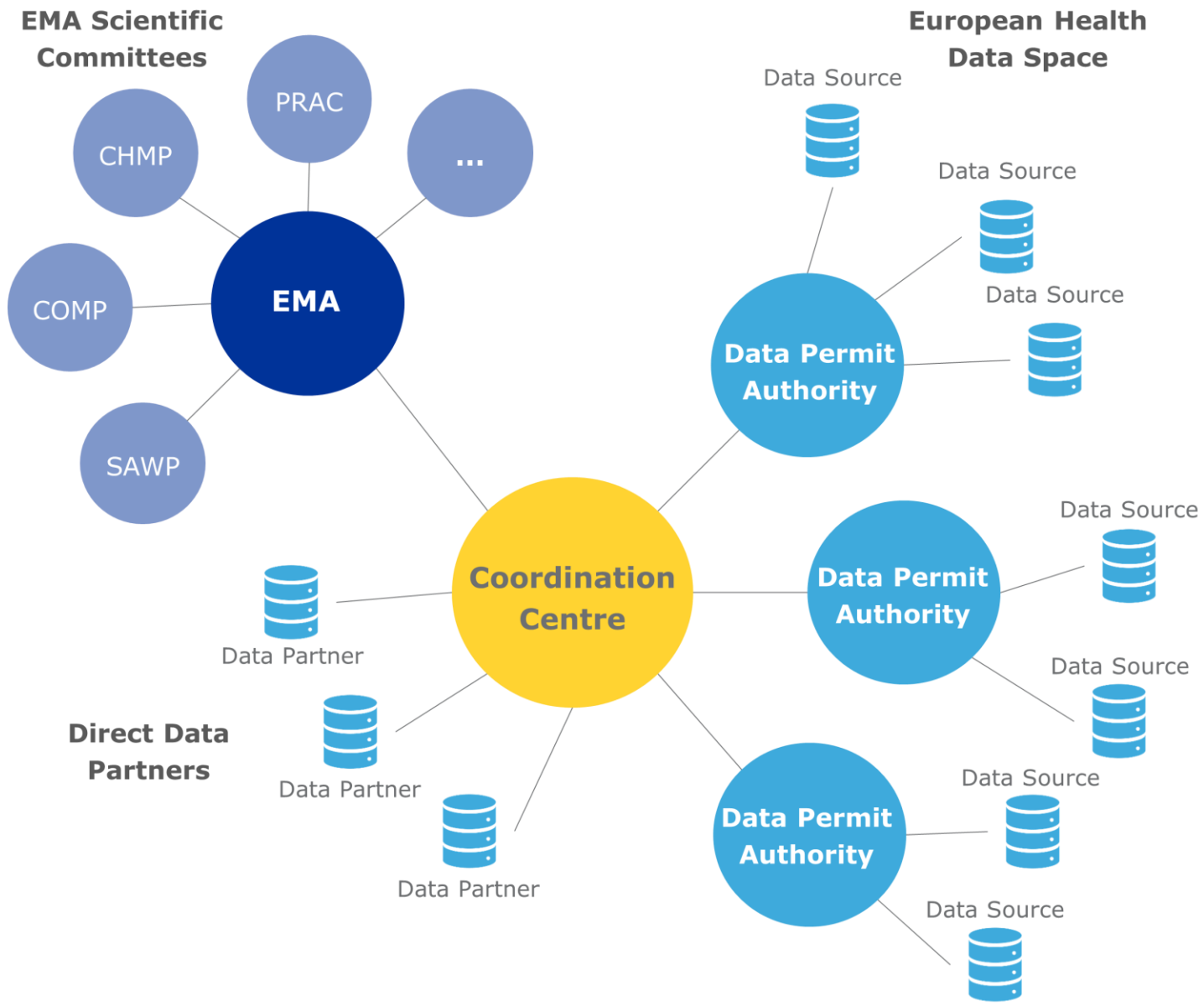


...and for crisis planning & response



- Monitoring the use of medicines to predict demand and shortages
- Understanding the disease natural history → development of vaccines and therapeutics
- Provide evidence for repurposing existing medicines
- Monitor the safety and effectiveness of vaccines and therapeutics post-authorisation

DARWIN EU® is a federated **network** of **data**, **expertise** and **services** that supports better decision-making throughout the product lifecycle by generating reliable **evidence from real world healthcare data**



FEDERATED NETWORK PRINCIPLES

- Data stays **local**
- **Use of OMOP Common Data Model** (where applicable) to perform studies in a timely manner and increase consistency of results

Data Partners – Phase I

UK

1. Clinical Practice Research Datalink (CPRD GOLD)

Belgium

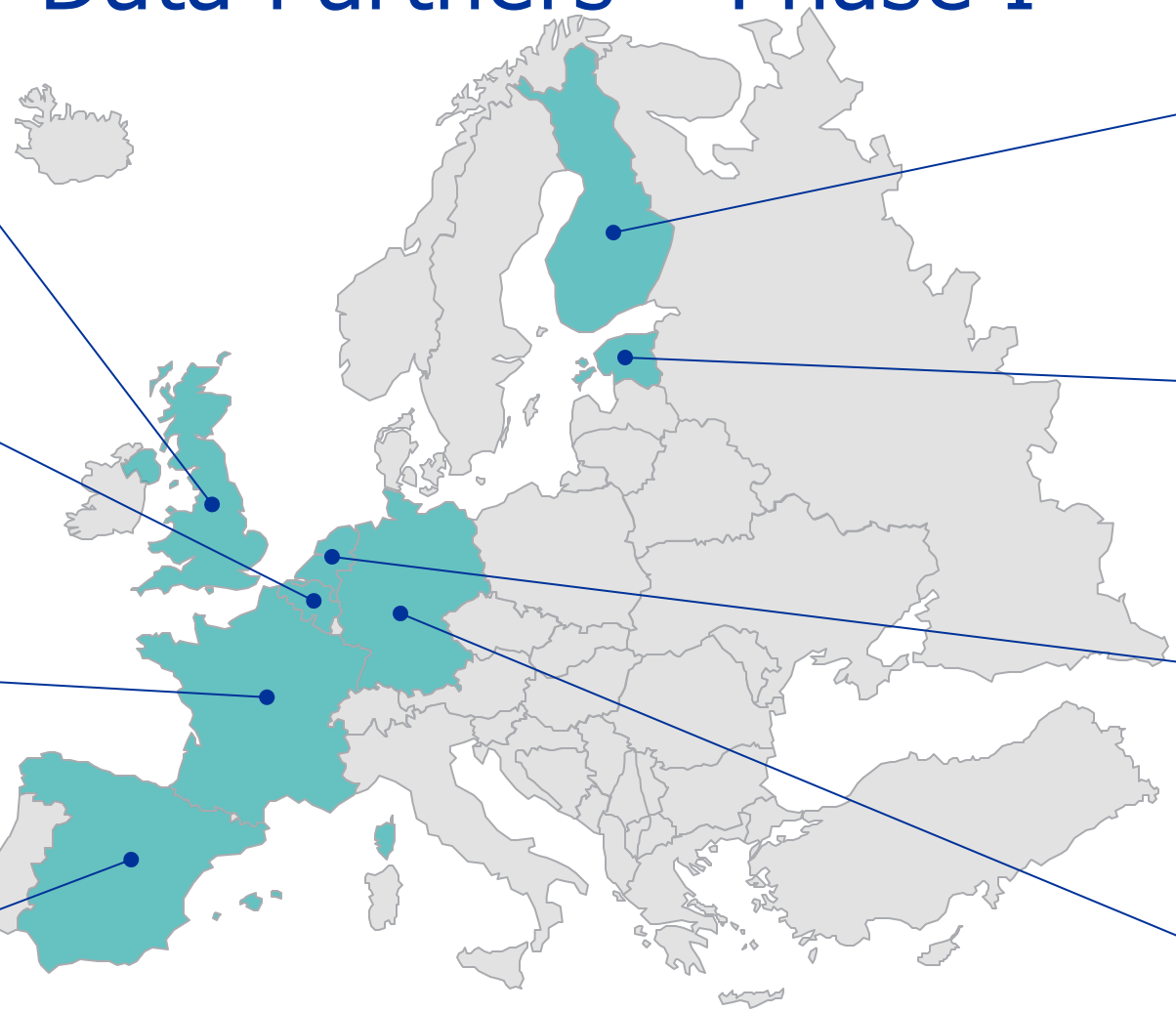
2. IQVIA Belgium Longitudinal Patient Data

France

3. Bordeaux University Hospital

Spain

4. IDIAPJGol
5. Parc Salut Mar Barcelona, Hospital del Mar (IMIM)



Finland

6. Auria Clinical Informatics at Hospital District of Southwest Finland (HDSF)

Estonia

7. University of Tartu (Biobank)

Netherlands

8. Integrated Primary Care Information
9. Netherlands Comprehensive Cancer Organisation

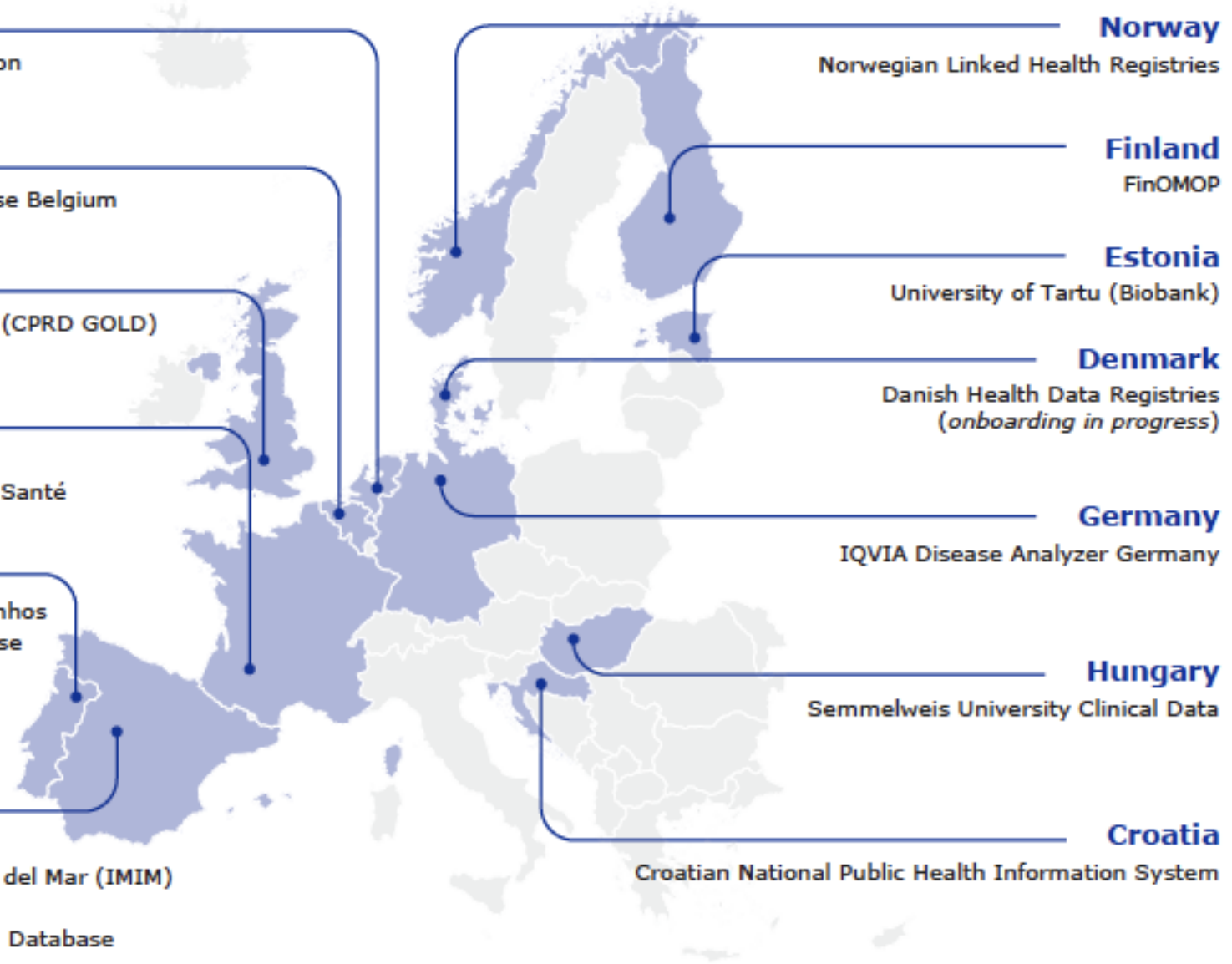
Germany

10. IQVIA Germany Disease Analyser

~26 million active patients

Currently **selecting Phase II DPs** via **open call for expression of interest**, then Phase III to follow

Data Partners – Phase II



The Netherlands
Integrated Primary Care Information
Netherlands Cancer Registry

Belgium
IQVIA Longitudinal Patient Database Belgium

United Kingdom
Clinical Practice Research Datalink (CPRD GOLD)
UK BioBank

France
Bordeaux University Hospital
Système National des Données de Santé

Portugal
Unidade Local de Saúde de Matosinhos
Egas Moniz Health Alliance DataBase

Spain
SIDIAP
Parc Salut Mar Barcelona, Hospital del Mar (IMIM)
BIFAP
Valencia Health System Integrated Database

Norway
Norwegian Linked Health Registries

Finland
FinOMOP

Estonia
University of Tartu (Biobank)

Denmark
Danish Health Data Registries
(*onboarding in progress*)

Germany
IQVIA Disease Analyzer Germany

Hungary
Semmelweis University Clinical Data

Croatia
Croatian National Public Health Information System

Data Partners – Phase II - Provenance

- Primary care data linked to hospital admissions from ES (x3)
- Primary care data without hospital linkage from NL and UK

- Hospital EHR data from ES, FI, FR, HU, PT

- Outpatient data and / or claims from BE, DE, FR

- Biobank data from EE and the UK

- Nation-wide (sometimes linked) data from DK, FI, HR (Croatia), NO

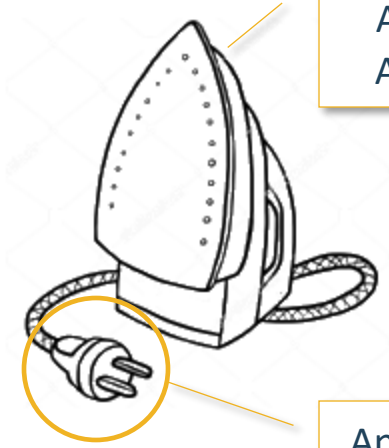
- Cancer registry x1

AGENDA

- What data? The DARWIN EU® network
- **Standardising data – and analytics**
- The phenotyping process in DARWIN EU®
- Is heterogeneity always a ‘curse’?

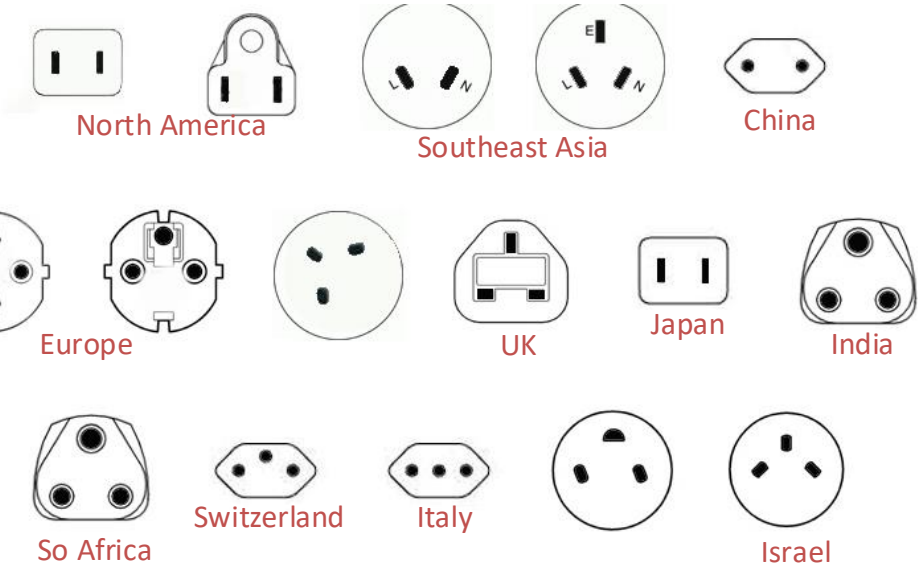
From standardised data to standardised analytics

"What's the adherence to my drug in the data assets I own?"

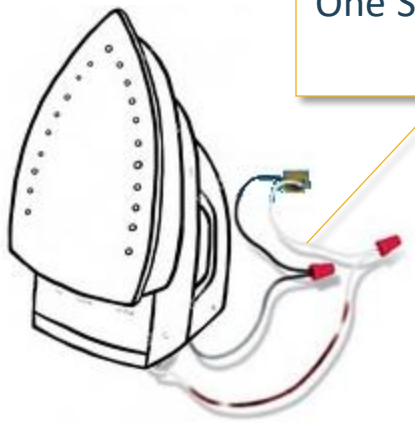


Analytical method:
Adherence to Drug

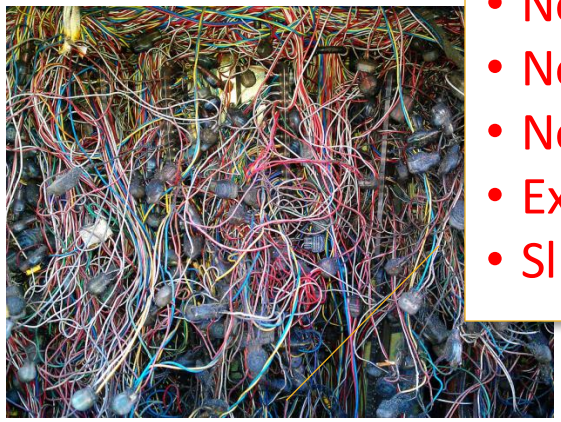
Application to data



Current solution:

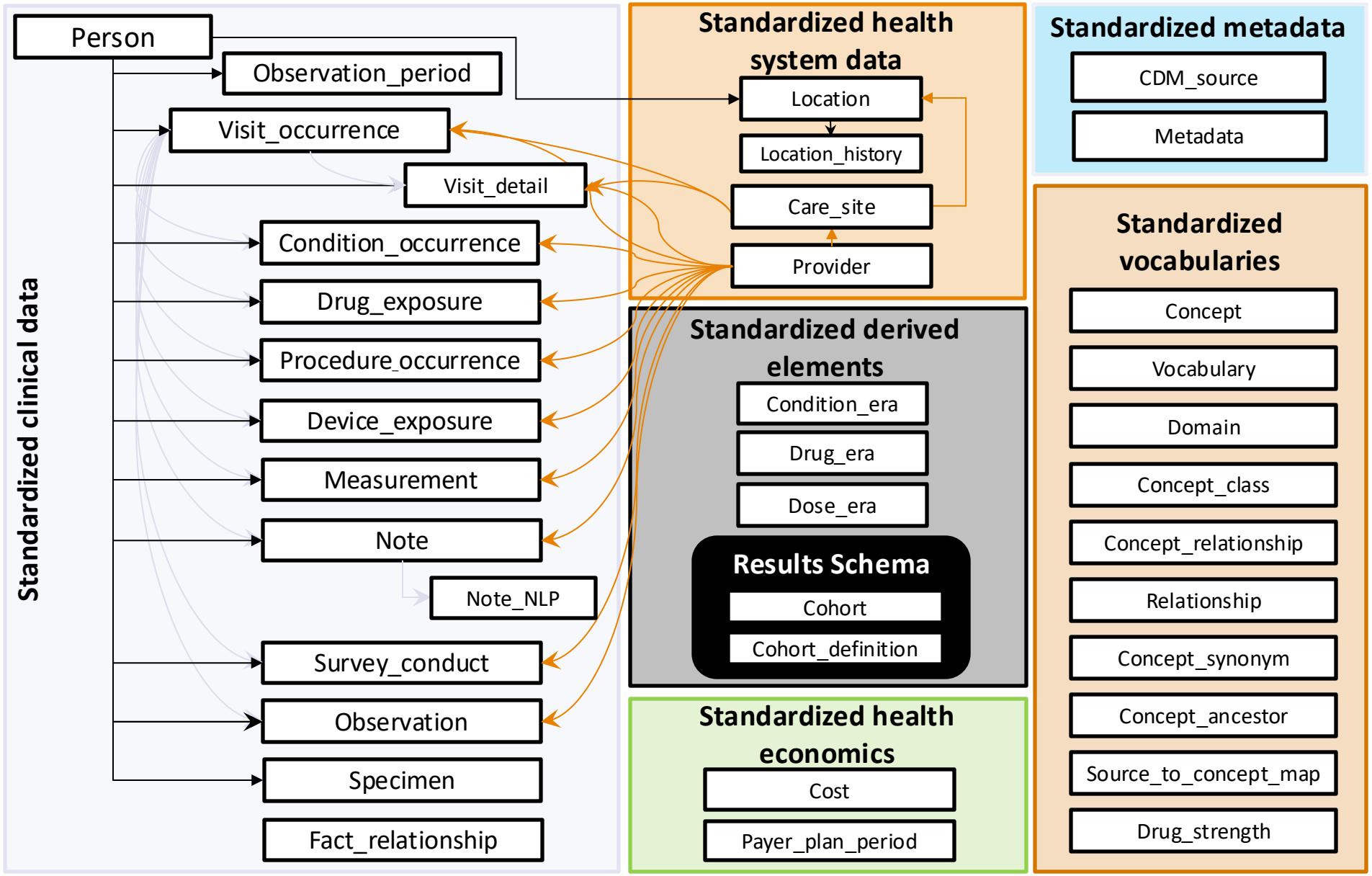


One SAS or R script for each study

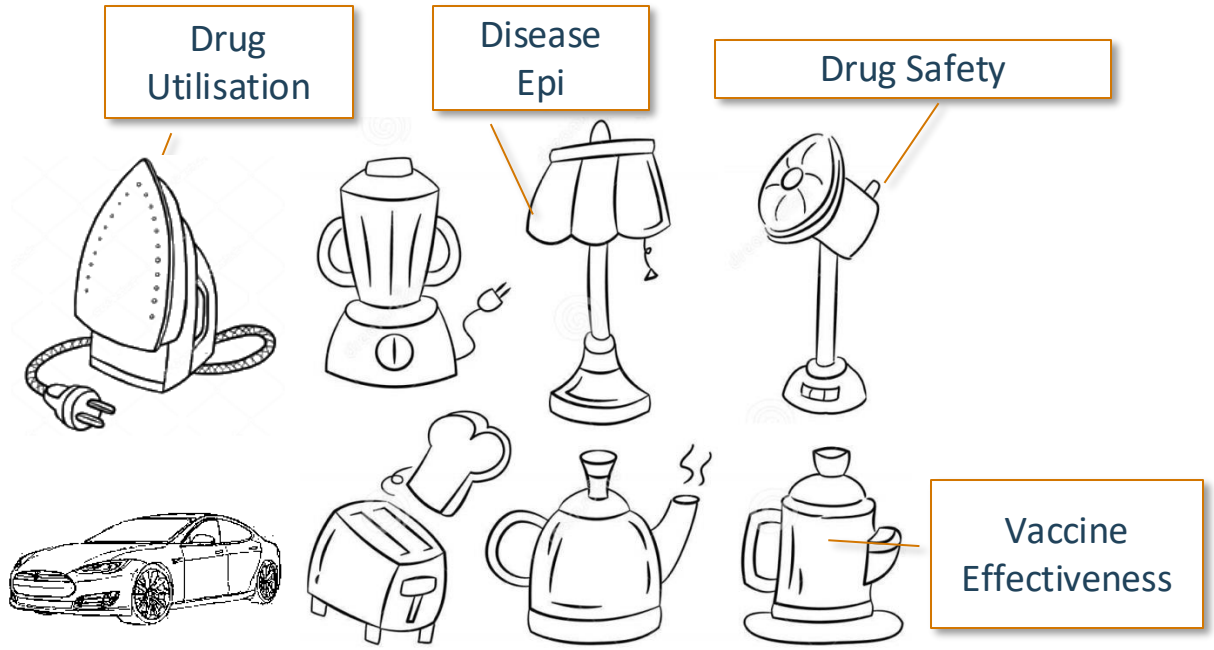


- Not scalable
- Not transparent
- Not reproducible
- Expensive
- Slow

One common data model: OMOP ver 5.4



From Data Standardization To Standardised Analytics



Standardized analytics → Standardized data

13

Catalogue of Standard Data Analyses

Off-the-shelf studies

These are mainly characterisation questions that can be executed with a generic protocol. This includes disease epidemiology, for example the estimation of the prevalence, incidence of health outcomes in defined time periods and population groups, or drug utilization studies at the population or patient level.

- Patient-level characterisation
- Patient-level DUS analyses
- Population-level DUS analyses
- Population-level descriptive epidemiology

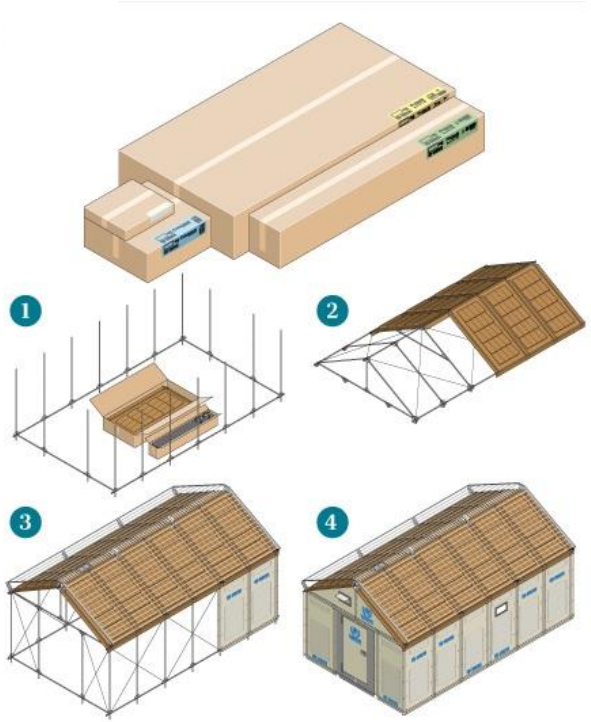
Complex

These are studies requiring development or customisation of specific study designs, protocols, analytics, phenotypes. This includes studies on the safety and effectiveness of medicines and vaccines.

- Prevalent user active comparator cohort studies
- New user active comparator cohort
- Self-controlled case risk interval
- Self-controlled case series
- Time series analyses and Difference-in-difference studies
- RMM effectiveness

Hacking a hard problem: Scaling up

Off the shelf
Fully STD

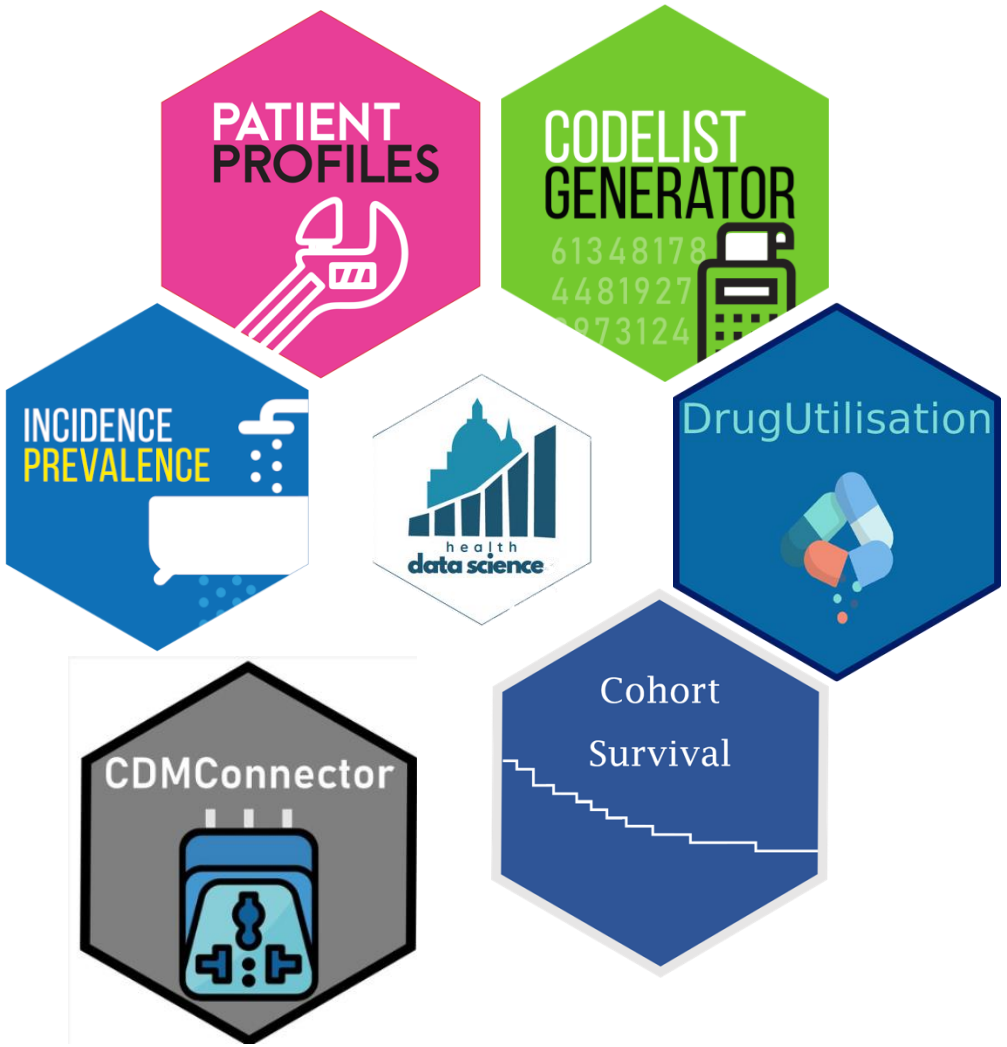


Complex
Pre-specified
(some bespoke)

Very complex
Bespoke code

Building the tools

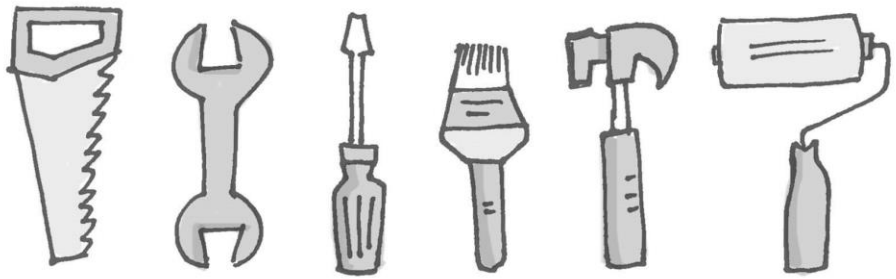
Developing OMOP-Std Analytical Pipelines



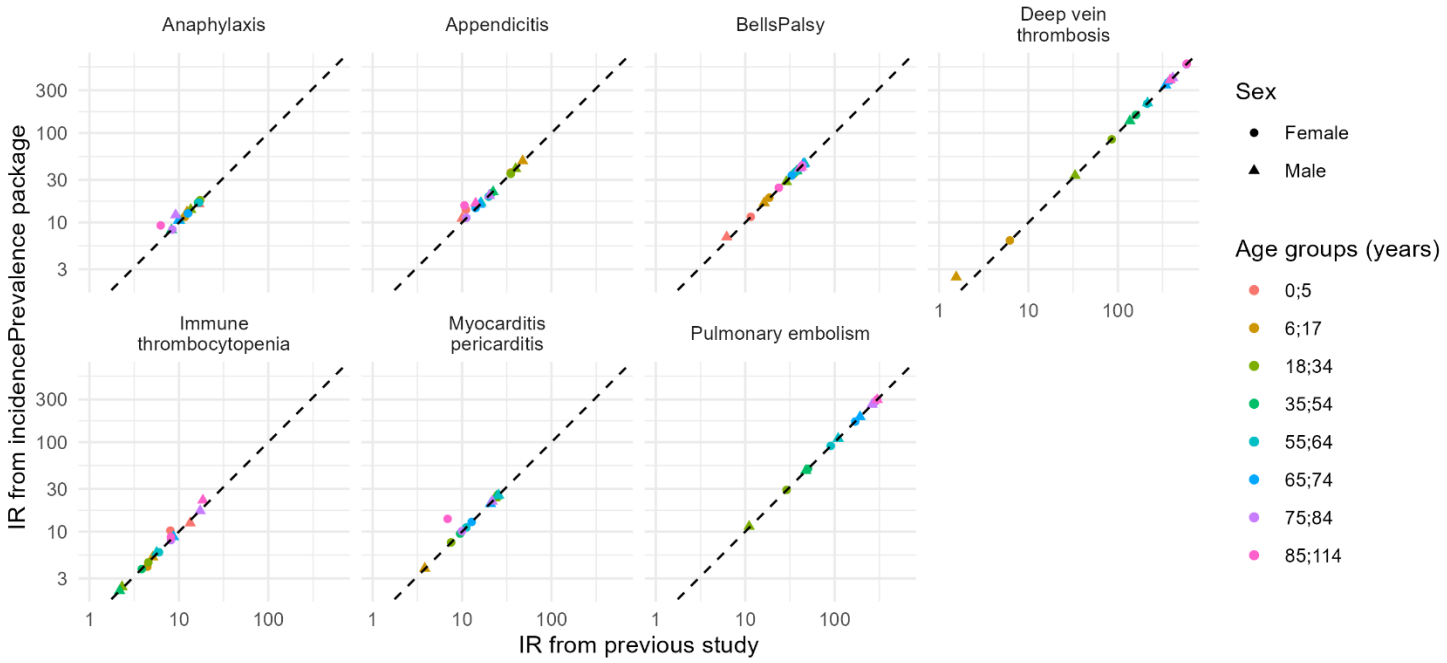
Off-the-shelf studies

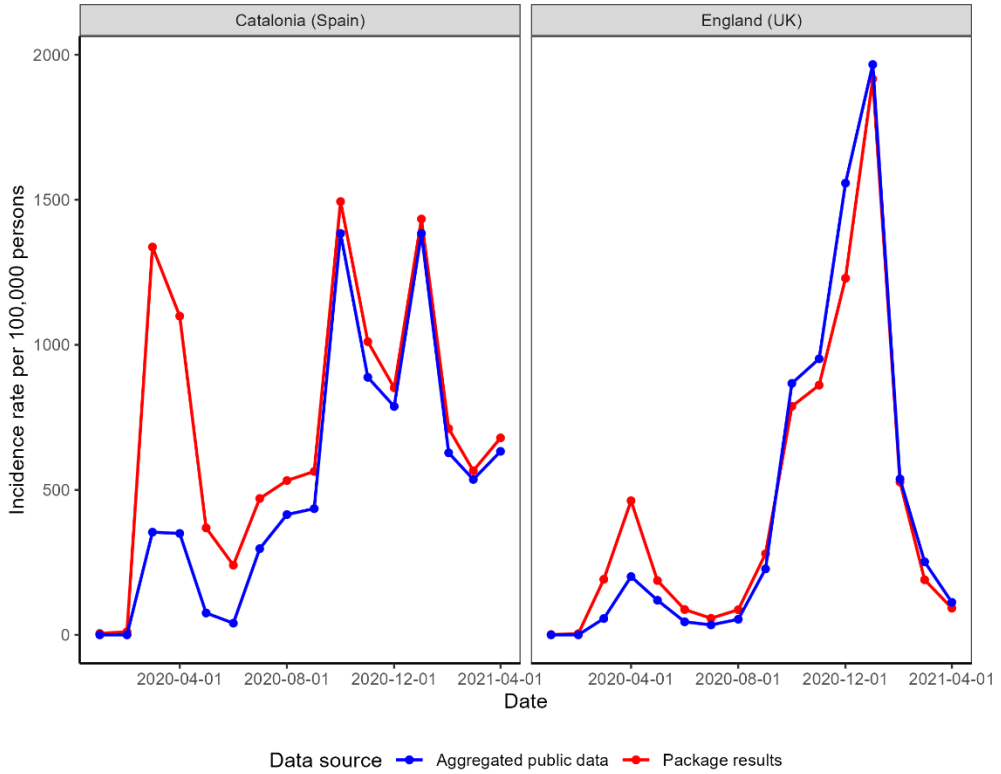
These are mainly characterisation questions that can be executed with a generic protocol. This includes disease epidemiology, for example the estimation of the prevalence, incidence of health outcomes in defined time periods and population groups, or drug utilization studies at the population or patient level.

- Patient-level characterisation
- Patient-level DUS analyses
- Population-level DUS analyses
- Population-level descriptive epidemiology



Software validation



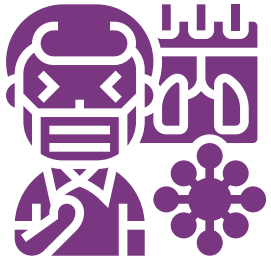


AGENDA

- What data? The DARWIN EU® network
- Standardising data – and analytics
- **The phenotyping process in DARWIN EU®**
- Is heterogeneity always a ‘curse’?

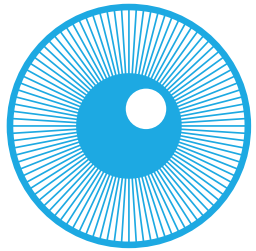
Phenotyping in RWD: The Problem

Finding patients that:



- Have or had certain disease

- Use/d certain treatment (drug, vaccine, surgeries…)



- With certain characteristic/s

DEFINITION OF A PHENOTYPE = **COHORT LOGIC**

- Have or had certain disease

- Notice verbal tense: phenotypes are cohorts (time-bound features), not just lists of codes

The Process and Tools

1 **Title**

2 **Standardised and reproducible phenotyping using distributed analytics and tools in**
3 **the Data Analysis and Real World Interrogation Network (DARWIN EU®)**

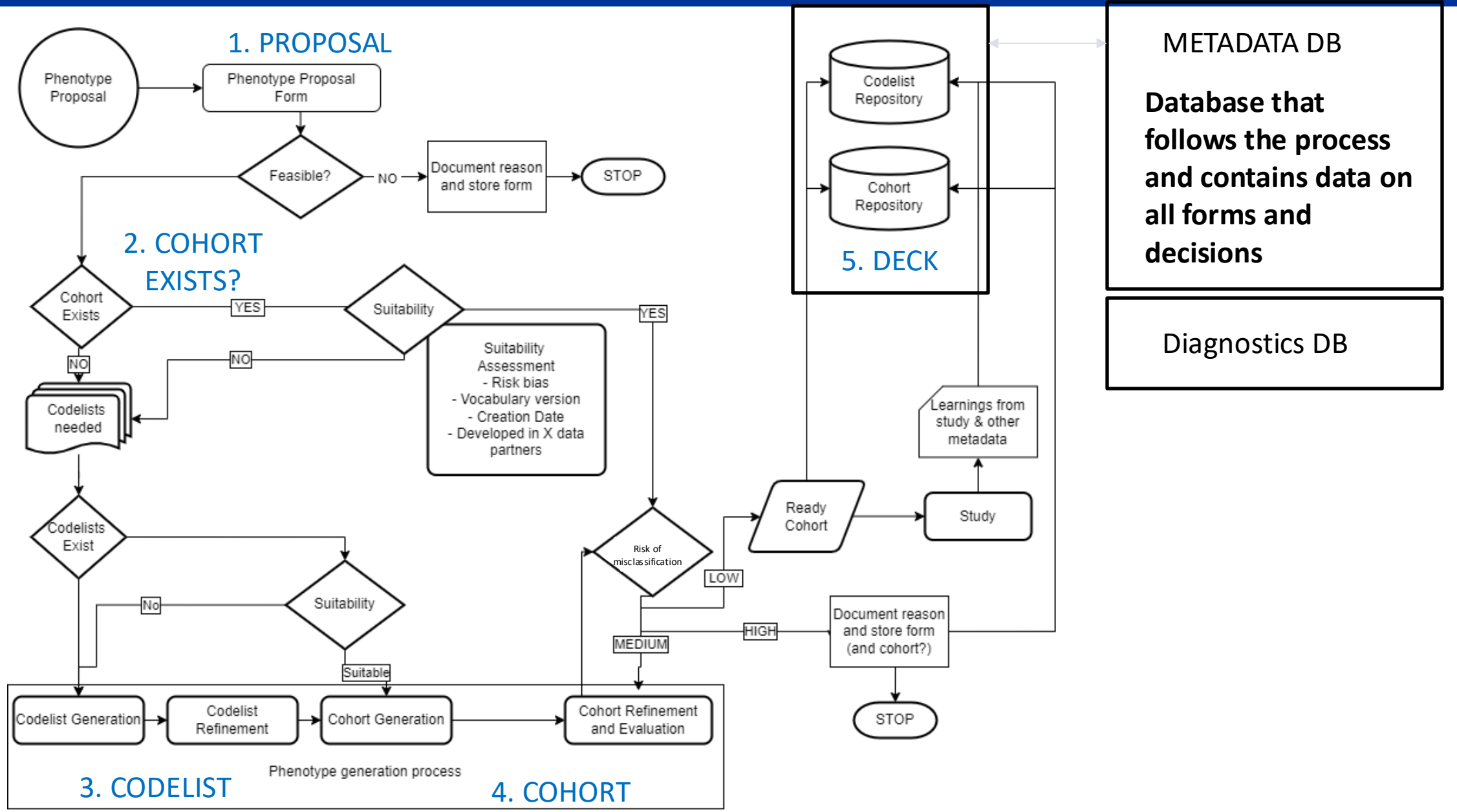
4

5 **Authors**

6 Francesco Dernie[1,3], George Corby[1,3], Abigail Robinson[1,3], James Bezer[1,3], Rowan Parry[2],
7 Annika Jodicke[3], Talita Duarte-Salles[2,4], Peter R Rijnbeek[2], Katia Verhamme[2], Alexandra
8 Pacurariu[5], Daniel Morales[5,6], Luis Pinheiro[5], Daniel Prieto-Alhambra[2,3]*, and Albert Prats-
9 Uribe[3]

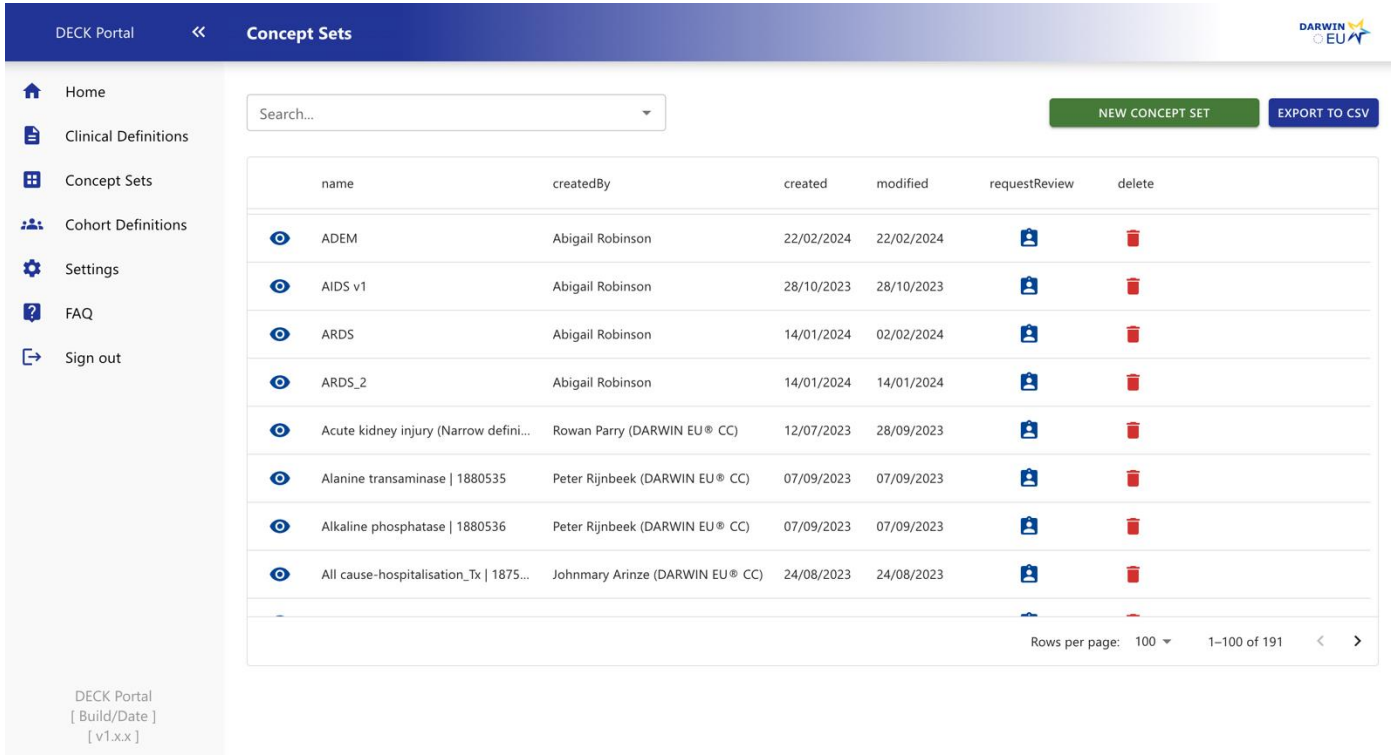
10 *Corresponding Author: d.prietoalhambra@darwin-eu.org

F Dernie et al. In press PDS 2024



STORAGE & TRACEABILITY

- Code lists review
- Storage of cohort definitions in all iterations
- All metadata stored and reusable for future studies



More detail tomorrow 13:45h: "Fit-for-Purpose Real-World Data: Principles and Developments"

AGENDA

- What data? The DARWIN EU® network
- Standardising data – and analytics
- The phenotyping process in DARWIN EU®
- **Is heterogeneity always a 'curse'?**

Different data provenance should lead to different rates

Received: 9 September 2022 | Revised: 11 April 2023 | Accepted: 13 April 2023
DOI: 10.1002/pds.5631

ORIGINAL ARTICLE

WILEY

Assessing heterogeneity of electronic health-care databases: A case study of background incidence rates of venous thromboembolism

Martin Russek[1] | Chantal Quinten[1] | Valentijn M. T. de Jong[1,2] | Catherine Cohet[1] | Xavier Kurz[1]

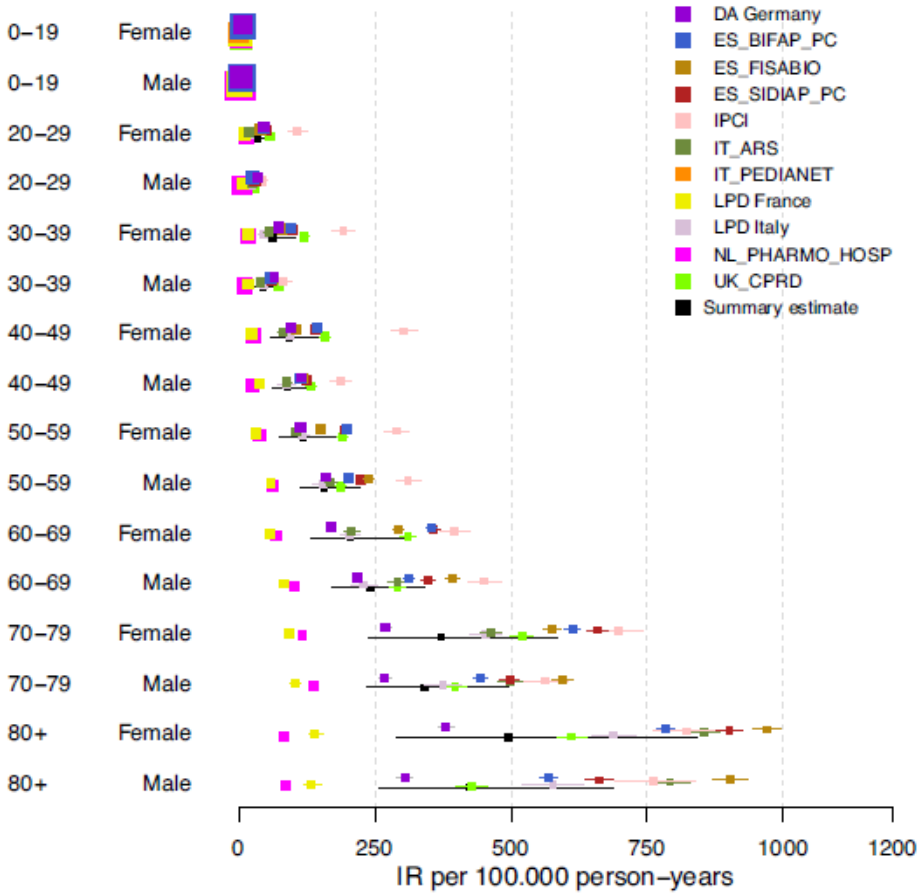
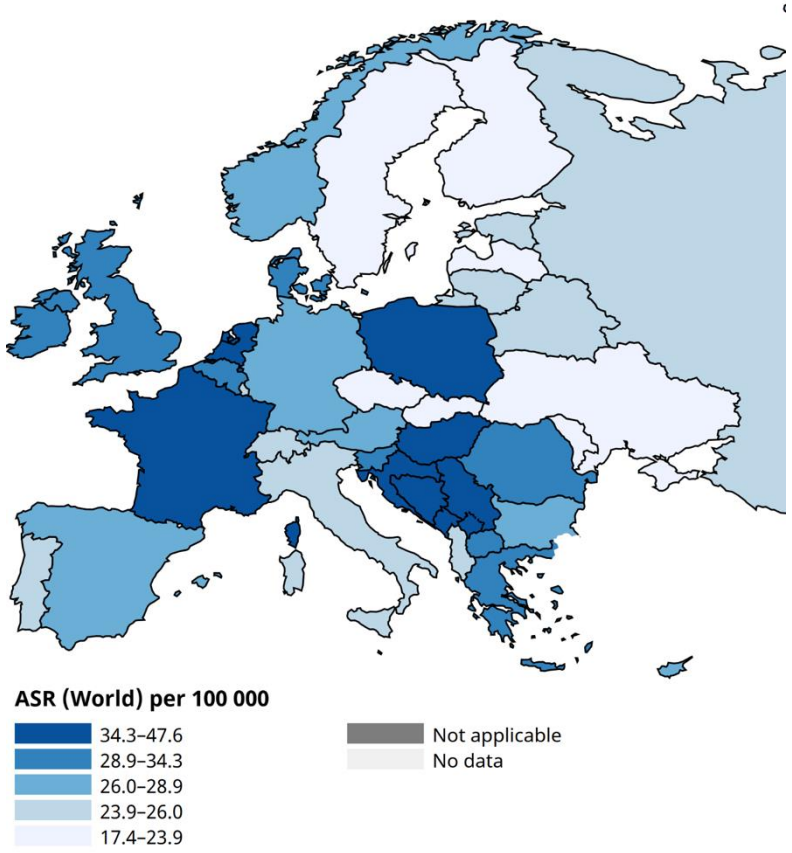
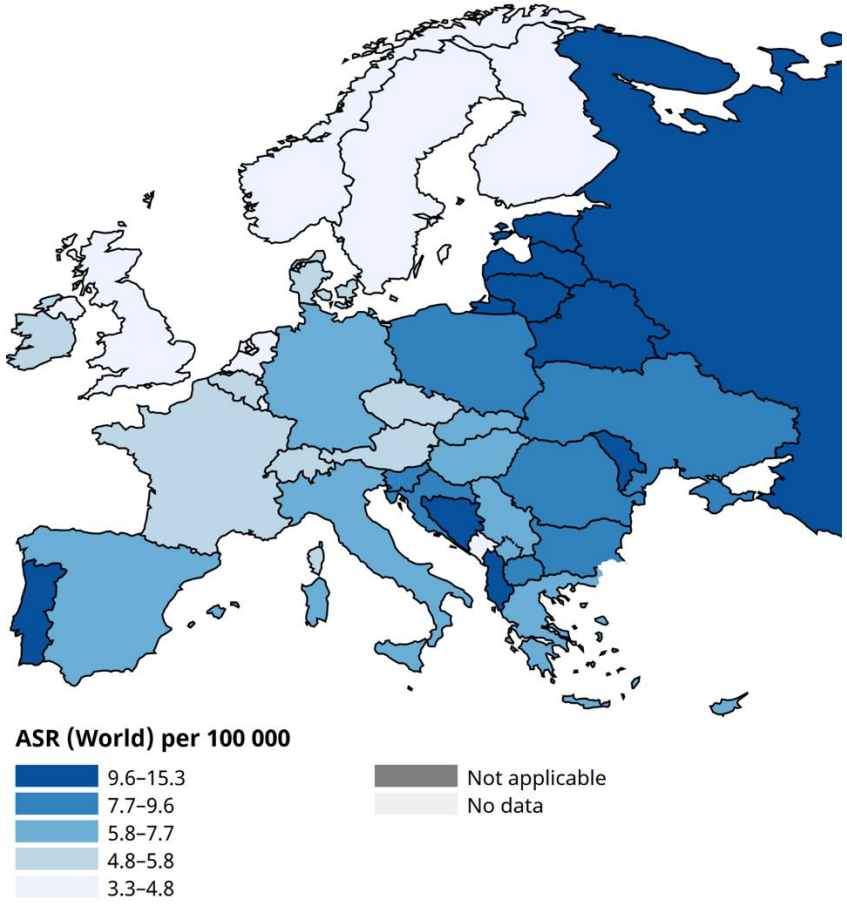


FIGURE 1 Age–gender-stratified IR estimates and 95% CIs* for VTE by database and pooled. *Due to the CIs being too small compared with the size of the square, some of the CIs are not noticeable in the figure.

And Europe is heterogeneous…





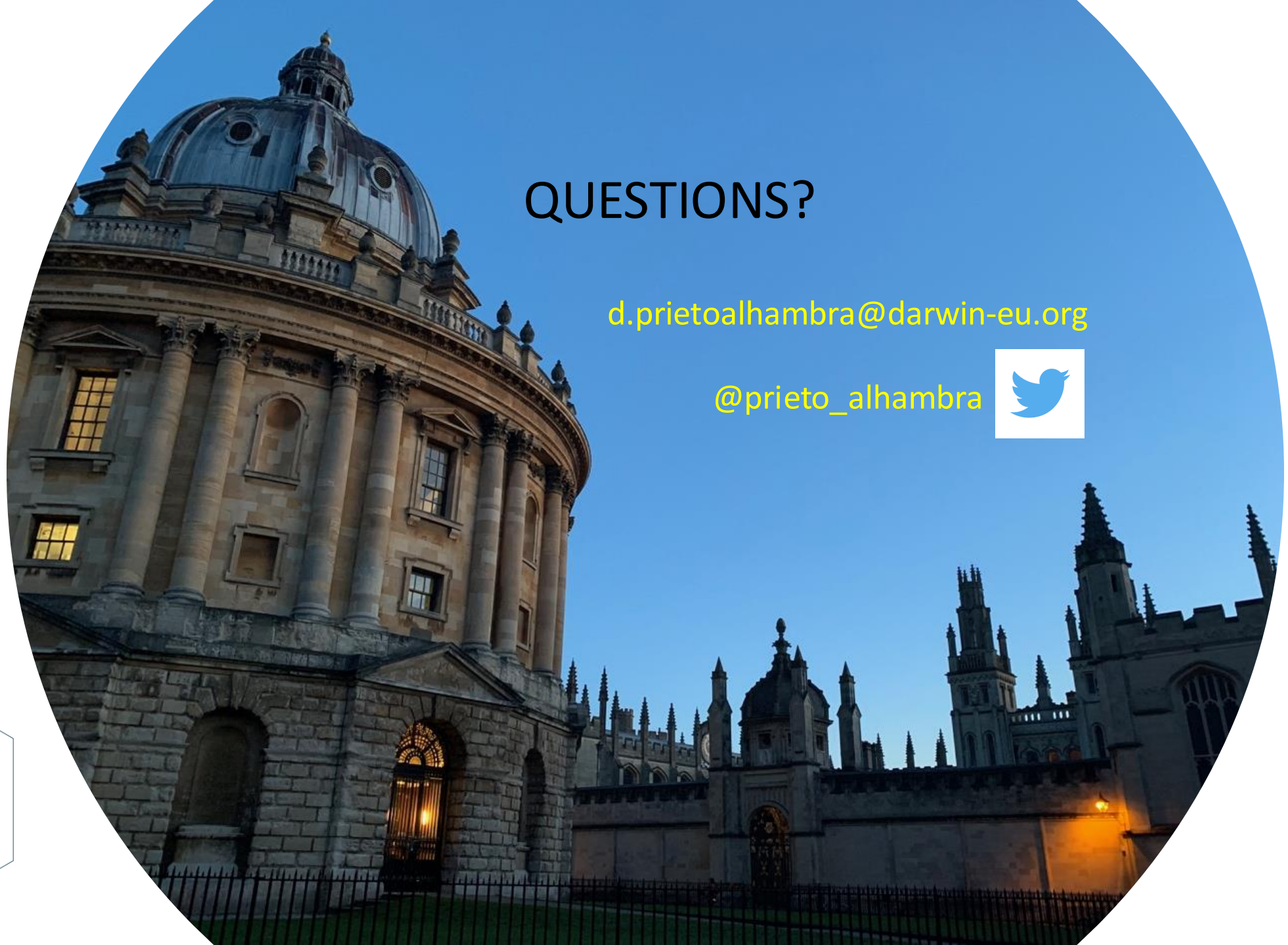
Source:

WHO IARC GLOBOCAN

https://gco.iarc.fr/

Lung Cancer Age-std rates

Stomach cancer ASR



QUESTIONS?

d.prietoalhambra@darwin-eu.org

@prieto_alhambra

health data sciences