DARWIN EU

# Coordination Centre

## Standardised and Reproducible Phenotyping Using Distributed Analytics and Tools in DARWIN EU

Albert Prats Uribe
Senior Clinical Research Fellow in Public Health
Health Data Sciences
NDORMS, Oxford University (UK)
albert.prats-uribe@ndorms.ox.ac.uk

# Disclosure

This presentation represents the views of the DARWIN EU® Coordination Centre only and cannot be interpreted as reflecting those of the European Medicines Agency or the European Medicines Regulatory Network.

*Pharmacoepidemiology and Drug Safety*

WILEY

First published: 12 November 2024
https://doi.org/10.1002/pds.70042

ORIGINAL ARTICLE OPEN ACCESS

# Standardised and Reproducible Phenotyping Using Distributed Analytics and Tools in the Data Analysis and Real World Interrogation Network (DARWIN EU)

Francesco Dernie[1,2] | George Corby[1,2] | Abigail Robinson[1,2] | James Bezer[1,2] | Nuria Mercade-Besora[2] | Romain Griffier[3] | Guillaume Verdy[3] | Angela Leis[4] | Juan Manuel Ramirez-Anguita[5] | Miguel A. Mayer[4,5] | James T. Brash[6] | Sarah Seager[6] | Rowan Parry[7] | Annika Jodicke[2] | Talita Duarte-Salles[7,8] | Peter R. Rijnbeek[7] | Katia Verhamme[7] | Alexandra Pacurariu[9] | Daniel Morales[9,10] | Luis Pinheiro[9] | Daniel Prieto-Alhambra[2,7] | Albert Prats-Uribe[2]

[1]Medical Sciences Division, University of Oxford, Oxford, UK | [2]Pharmaco- and Device Epidemiology, Centre for Statistics in Medicines, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK | [3]Public Health Department, Medical Information Service, Medical Informatics and Archiving Unit (IAM), University Hospital of Bordeaux, Bordeaux, France | [4]Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Research Institute, Barcelona, Spain | [5]Management and Control Department, Consorci Mar Parc de Salut de Barcelona, Barcelona, Spain | [6]Real World Solutions, IQVIA, Brighton, UK | [7]Department of Medical Informatics, Erasmus University Medical Centre, Rotterdam, The Netherlands | [8]Fundació Institut Universitari per a la Recerca a l'Atencio Primaria de Salut Jordi Gol I Gurina (IDIAPJGol), Universitat Autonoma de Barcelona, Barcelona, Spain | [9]Real World Evidence Workstream, European Medicines Agency, Amsterdam, The Netherlands | [10]Division of Population Health and Genomics, University of Dundee, Dundee, UK

**Correspondence:** Daniel Prieto-Alhambra (d.prietoalhambra@darwin-eu.org)

**Received:** 10 January 2024 | **Revised:** 2 October 2024 | **Accepted:** 6 October 2024

# Why?

Improve **transparency, reproducibility** and **reliability**

Inform reuse of cohorts, **storing** all **metadata** needed to decide

Focus on **traceability**: Log of decisions, responsible person/s and their reasoning

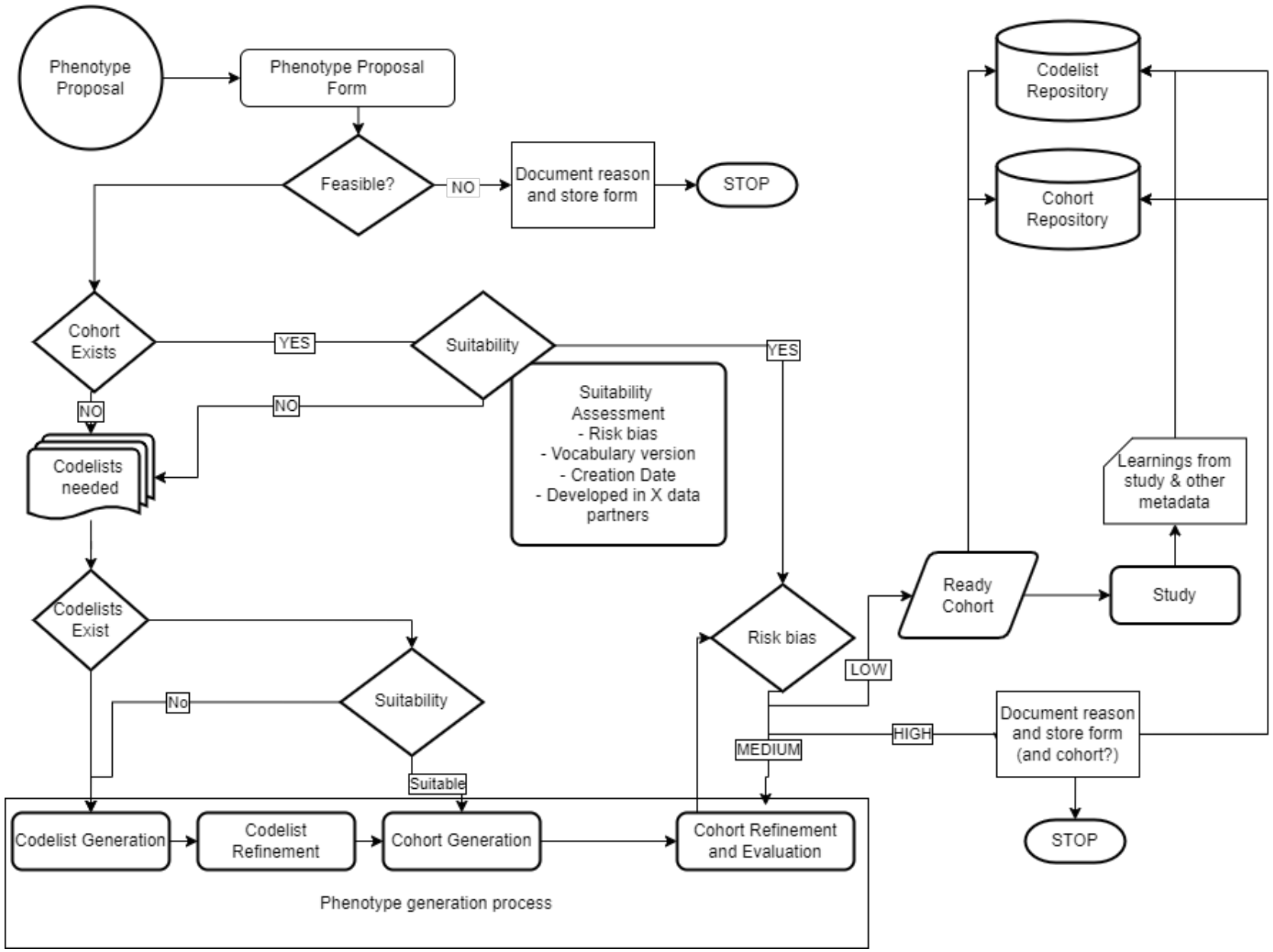
DARWIN EU data network strengths and DARWIN EU catalogue of standard analytics in mind

# How?

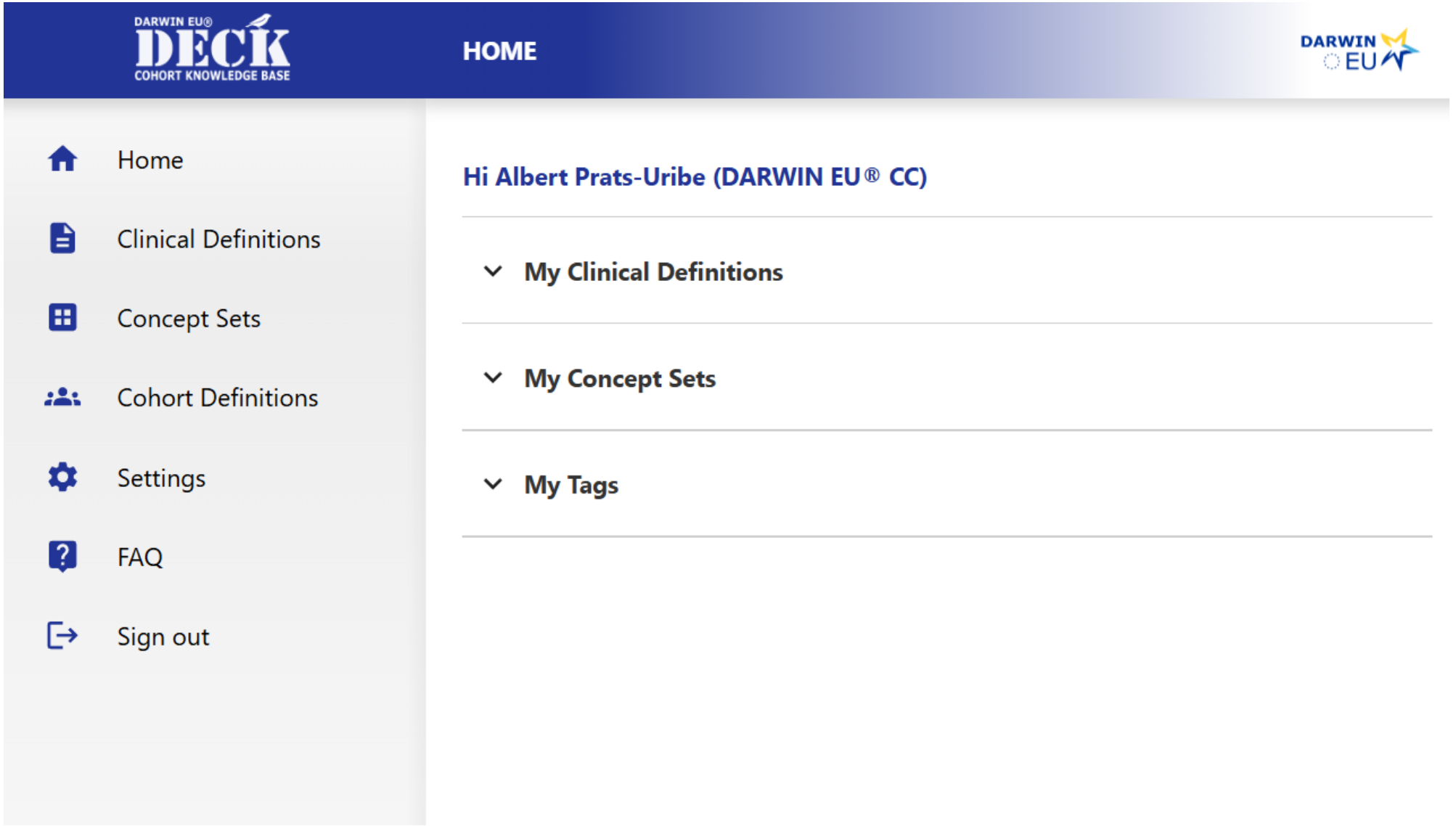
**Co-creation DARWIN CC** with **EMA** according to their needs and regulatory outcomes:

- Creation of a D**ARWIN-EMA phenotyping workgroup**
- Regular meetings with **EMA/NCA pharmacovigilance specialists** to review and improve the process
- In person training sessions and discussions with EMA RWD team

# What?
# The process:



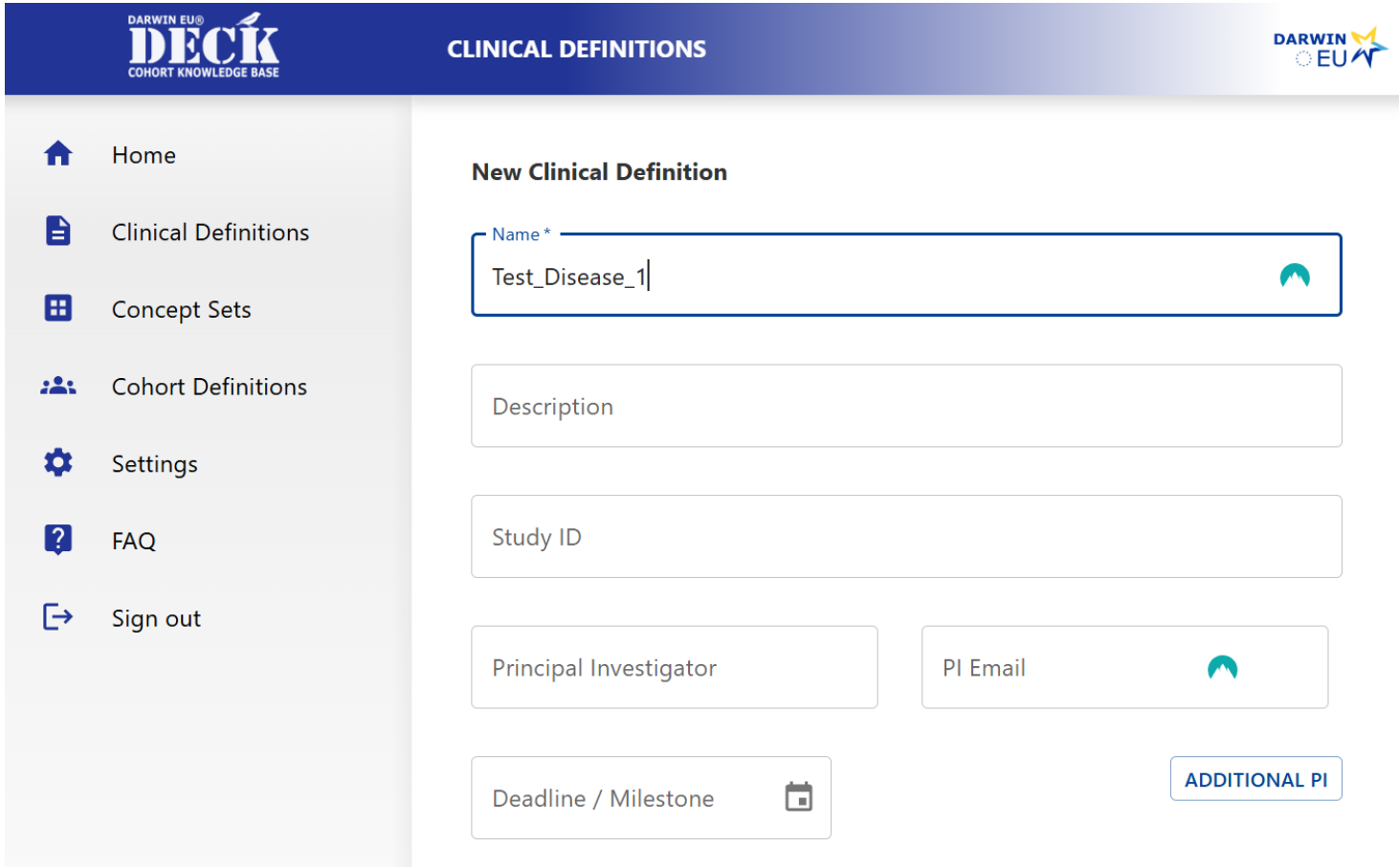
# And a tool to follow it…



# Step 1,2,3: Phenotype Proposal & Feasibility

**Phenotype Proposal Form**

Includes information on the requestor, a summary of the phenotype, the intended use (e.g., for incidence/prevalence studies, or drug utilisation, and the databases it will be used on) and timelines, and…



# Step 1,2,3: Clinical Description and Optimisation Requirements

A **Clinical description** summarising the disease's epidemiology, presenting symptoms, treatments, and potential strengthening or disqualifying factors

OVERVIEW | CLINICAL DESCRIPTION | CONCEPT SETS | COHORT DEFINITIONS | DISCUSSION

## Clinical Description

**Overview:**

A comprehensive introduction to the clinical phenotype, providing a general understanding of the condition or disease.

**Presentation:**

Description of the typical signs and symptoms that patients with the clinical phenotype may experience.

**Epidemiology:**

Add known population measures of the disease: Age distribution, sex distribution, and incidence and prevalence from previous studies

**Assessment:**

Explanation of the diagnostic tests and procedures used to assess and confirm the presence of the clinical phenotype. This may include laboratory tests, imaging studies, and other diagnostic techniques. Criteria and factors considered in establishing a diagnosis of the clinical phenotype, including specific markers, characteristics, or clinical findings.

**Therapeutic Plan:**

An outline of the treatment and management strategies commonly employed for patients with the clinical phenotype. This may include medications, therapies, surgical interventions, and supportive care measures.

**Prognosis:**

Factors that influence the prognosis or expected outcome of individuals with the clinical phenotype. This section may identify both positive and negative prognostic indicators. Potential complications or disease progression patterns associated with the clinical phenotype. This section may describe the development of related conditions or transformation into other diseases.

**Disqualifiers:**

Any information useful to rule out the correct coding of the proposed phenotype. Typical examples include differential diagnoses (e.g. psoriatic arthritis instead of rheumatoid arthritis) or treatments with a differential indication (e.g. treatment with adalimumab for a person identified as suffering from osteoarthritis)

**Strengtheners:**

Any information useful to strengthen the likelihood that the identified person suffers the condition/phenotype of interest. Typically, these will include compatible tests (e.g. rheumatoid factor for rheumatoid arthritis), procedures (e.g. breast biopsy for breast cancer), or treatments (e.g. aromatase inhibitor therapy initiation for breast cancer)

**Brighton Collaboration Definition or/and existing codings**

Add here the BC definition if it exists or to the medDRA codes.

## Phenotyping plan

**Resultant Proposed Logic**

Do we need 1 or more concept sets? Any temporal logic needed?

**Proposed Flavours**

How many variants or flavours of this phenotype do we need?

**Proposed Search strategy**

List all proposed concept sets and the proposed keywords that will be used for the search strategy. Also add domains to be searched in, and exclusions if known.

EDIT

# Step 4,5,6: Reusability of previous cohorts/concept sets

**Search for a Suitable Cohort &**

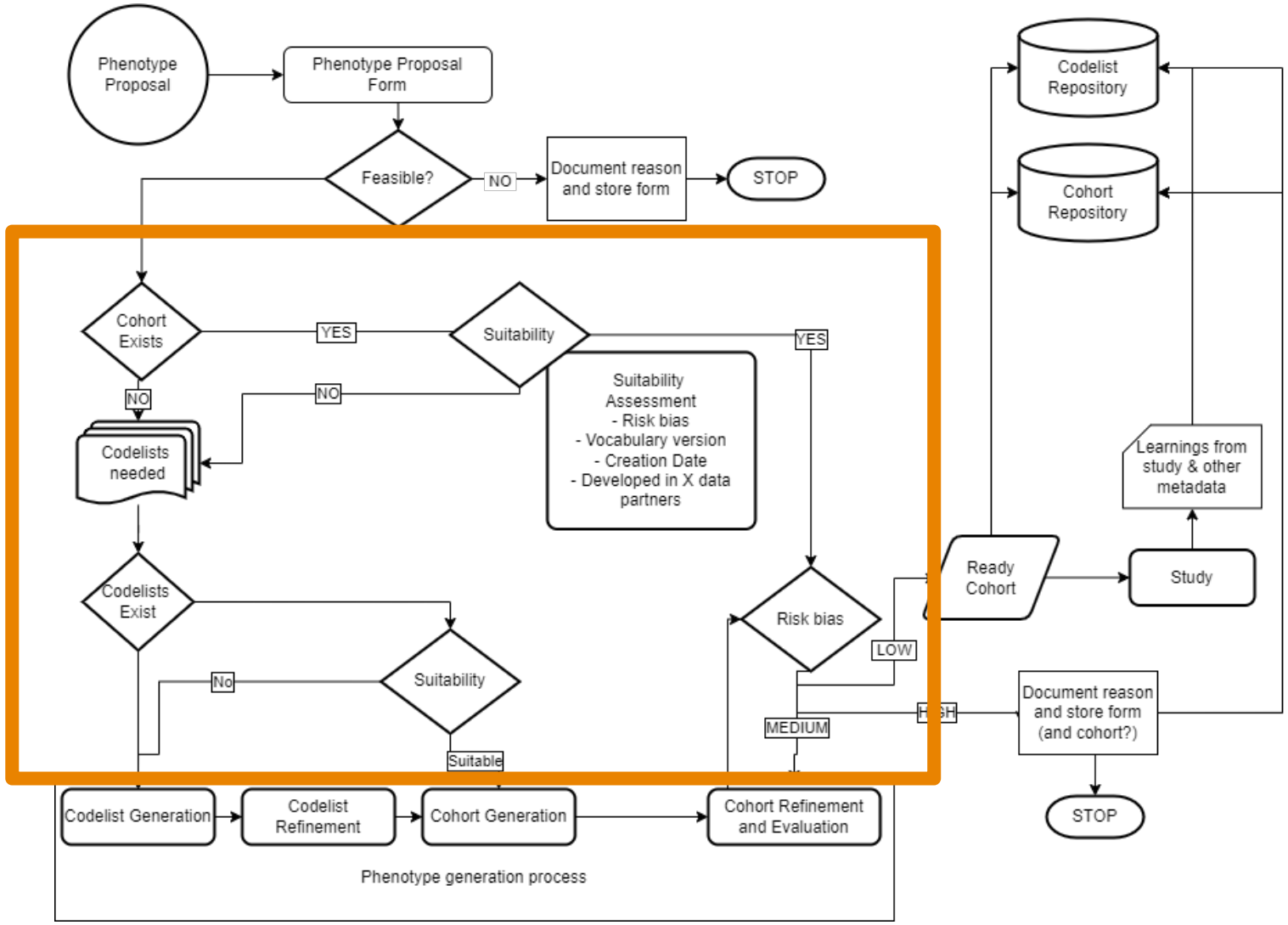
**Suitability and Relevance to Study of**

**Interest**

**Search for a Suitable Concept/s set &**

**Suitability and Relevance to Study of**

**Interest**



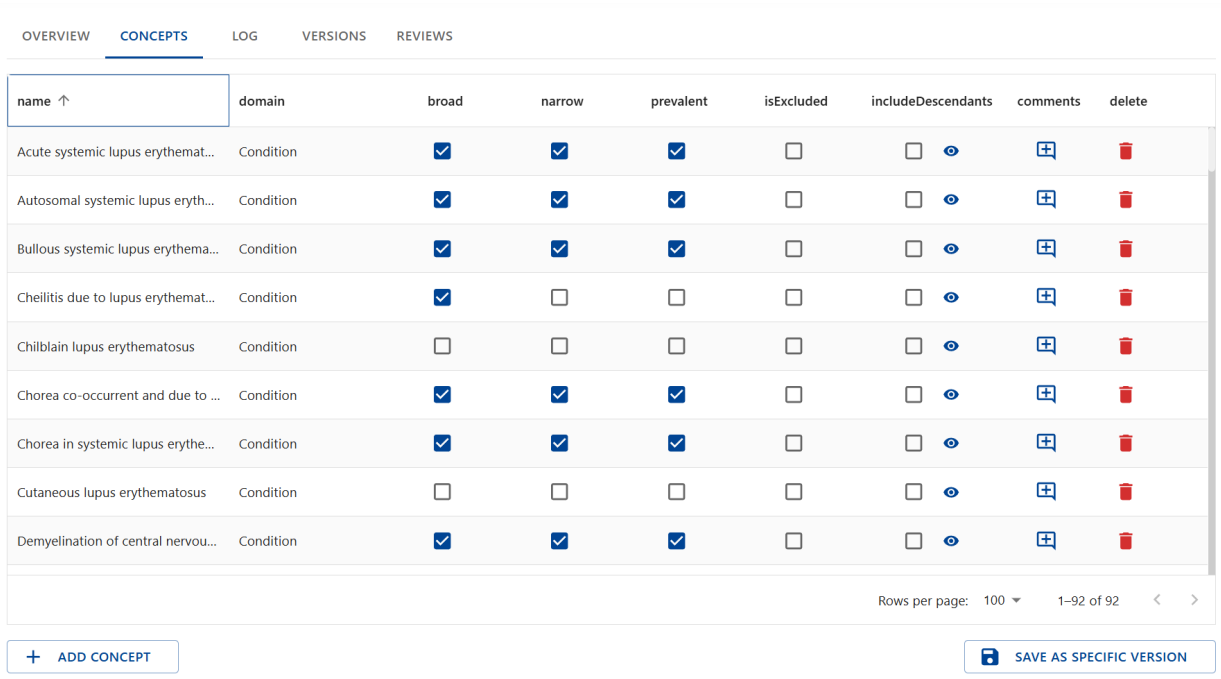
# Step 6 - 11: Phenotype generation and evaluation

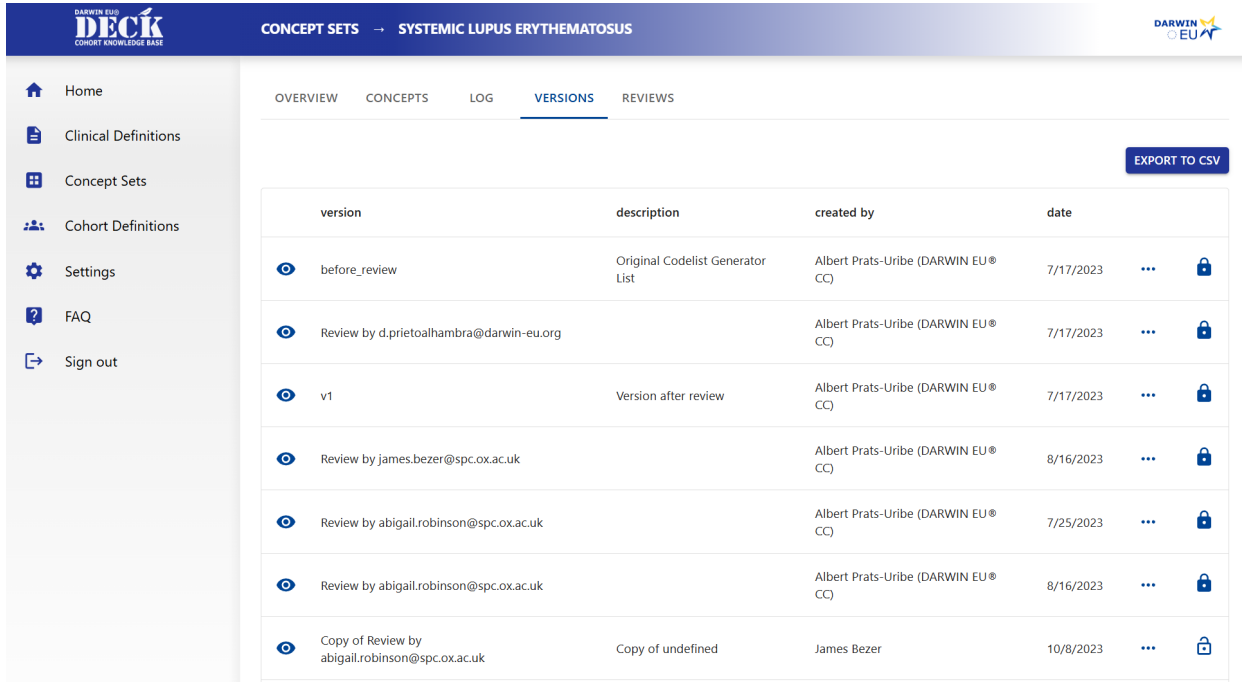
## Step 7: Generate the Concept Set

Systematic generation, saving the keywords used, vocabulary version, and steps followed

## Step 8: Refine the Concept Set

Two reviewers with a clinical/pharmacy background (PAs) and/or topic expertise and present them with the initial concept set from Step 7 together with the provided PPF and clinical description



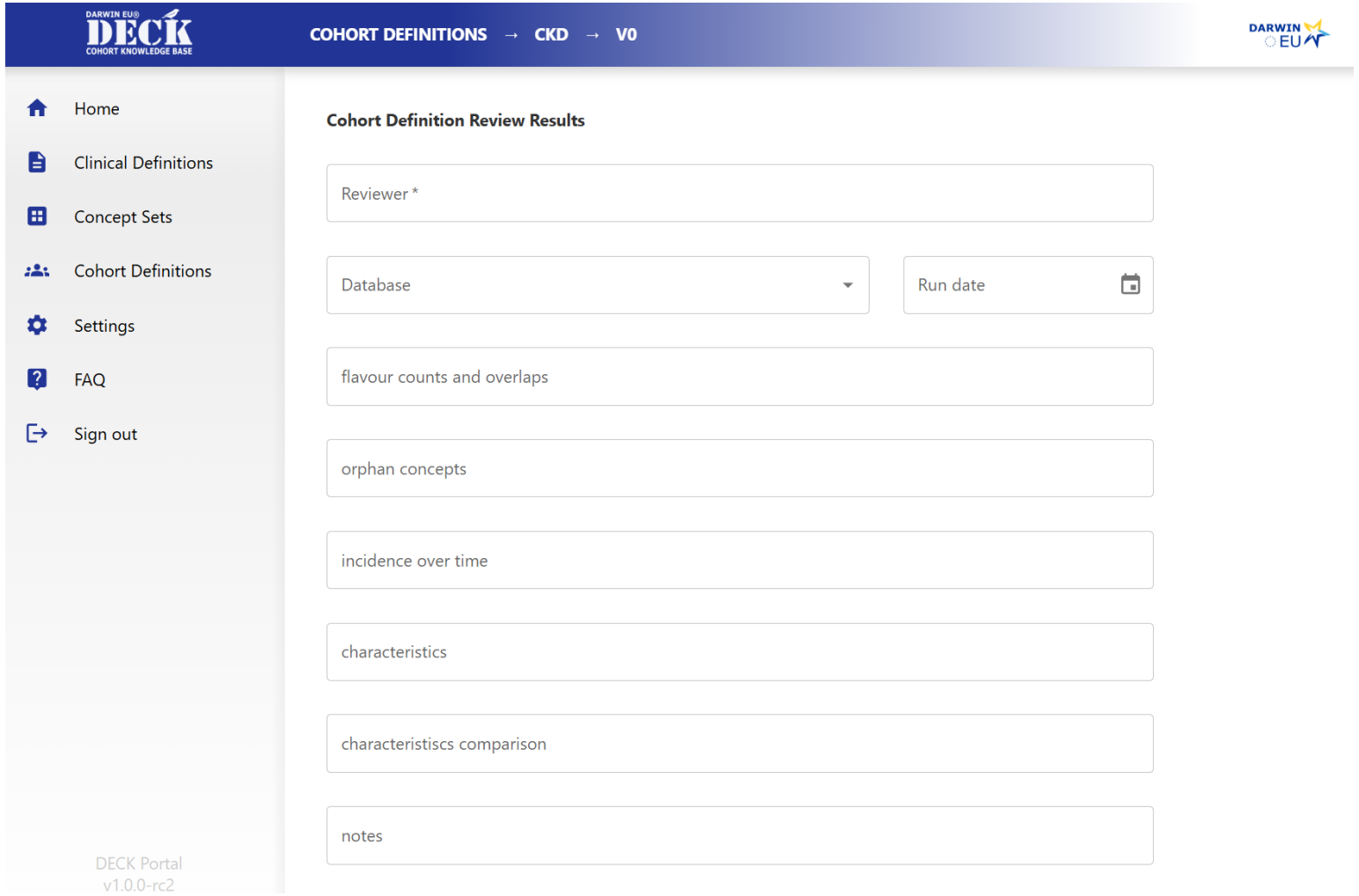



# Step 9 - 14: Cohort generation, evaluation, study and storage

**Step 9: Generating the Cohort**
Add all the logic to generate cohorts using ATLAS or CAPR and fed back into the DECK

**Steps 10 and 11: New Phenotype Evaluation**
Running cohort diagnostics and storing results for each data partner in the DECK, with recommendations and decision to approve and move the phenotype forward.

**Steps 12-14: Storage, Sign off for study, and review and storage of study learnings.**



# Application to two phenotypes (with a simple version)

## Pancreatic Cancer

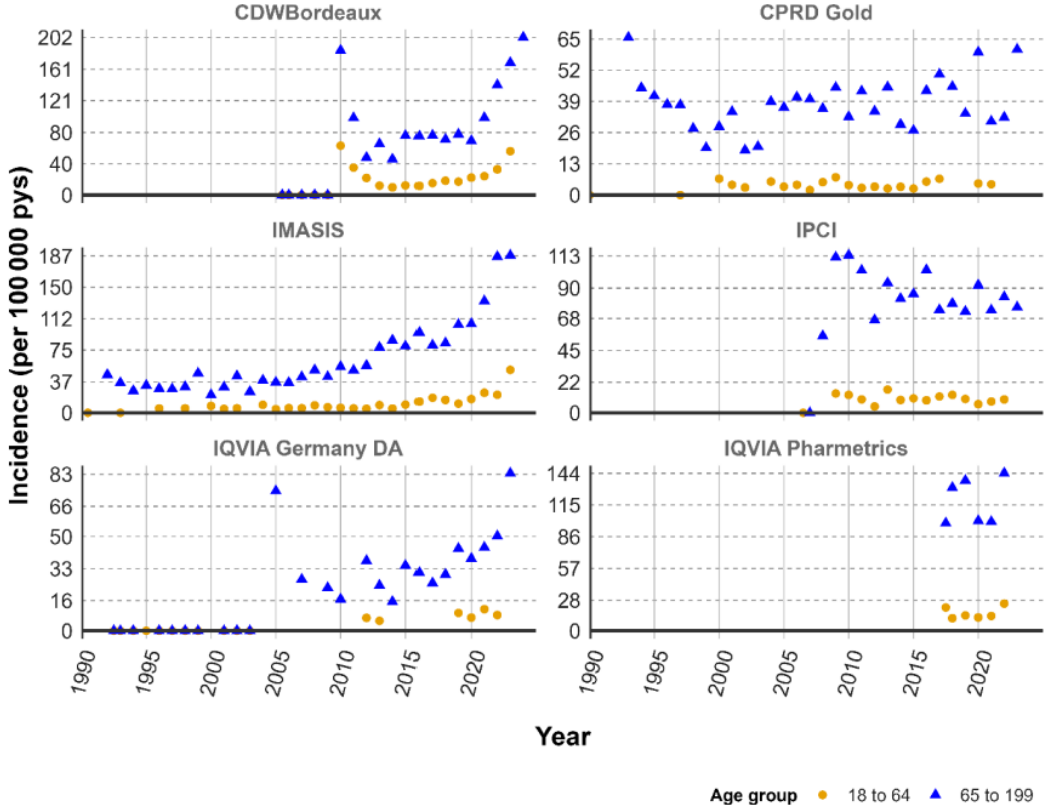



FIGURE 4 | Incidence per 100000 person-years of pancreatic cancer (narrow variant) among ≥ 65 (blue) and 18–64 (orange) year-olds, grouped by database.

## SLE

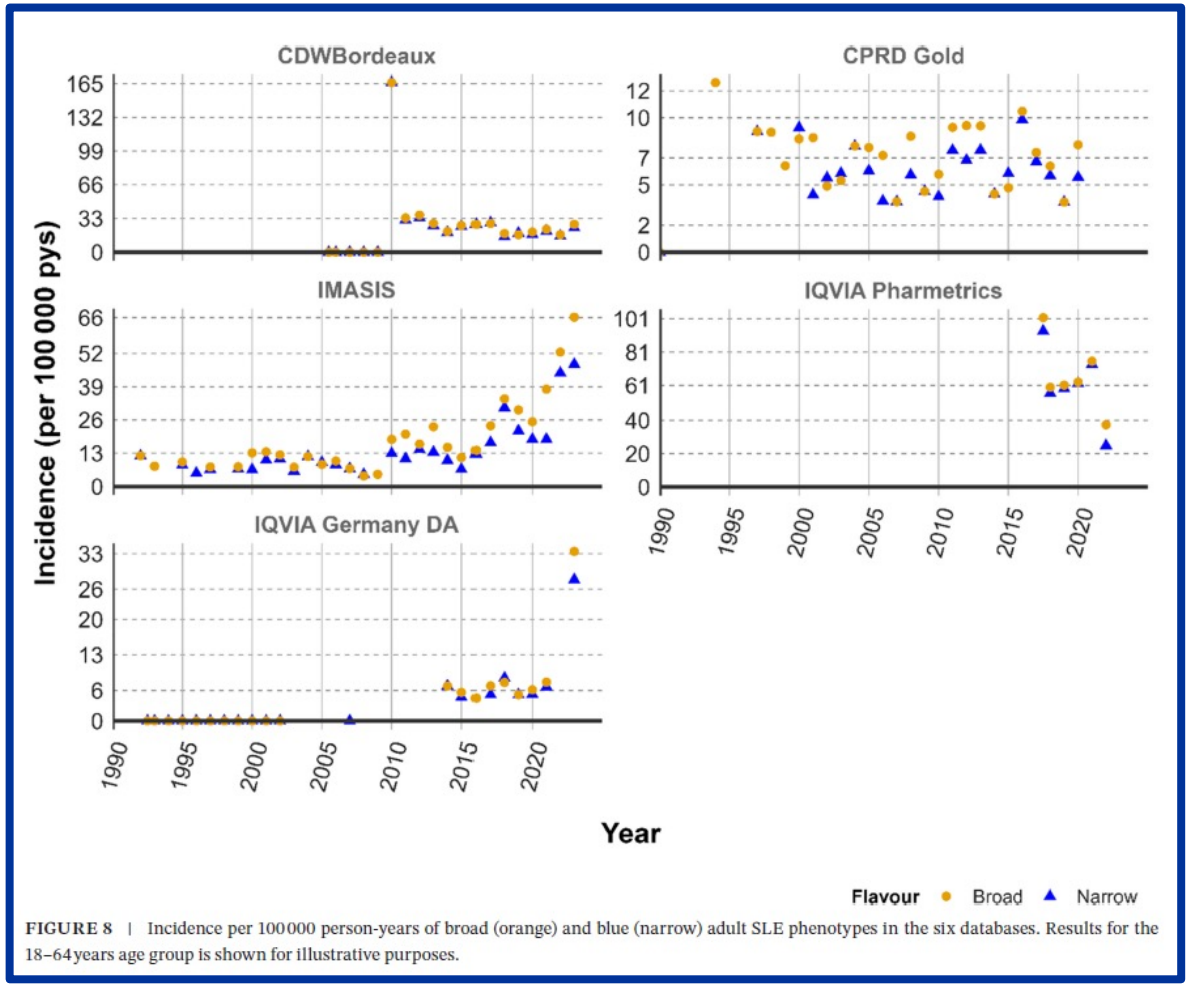



FIGURE 8 | Incidence per 100000 person-years of broad (orange) and blue (narrow) adult SLE phenotypes in the six databases. Results for the 18–64 years age group is shown for illustrative purposes.

*Pharmacoepidemiology and Drug Safety*

WILEY

# The paper:

ORIGINAL ARTICLE OPEN ACCESS

## Standardised and Reproducible Phenotyping Using Distributed Analytics and Tools in the Data Analysis and Real World Interrogation Network (DARWIN EU)

Francesco Dernie[1,2] | George Corby[1,2] | Abigail Robinson[1,2] | James Bezer[1,2] | Nuria Mercade-Besora[2] | Romain Griffier[3] | Guillaume Verdy[3] | Angela Leis[4] | Juan Manuel Ramirez-Anguita[5] | Miguel A. Mayer[4,5] | James T. Brash[6] | Sarah Seager[6] | Rowan Parry[7] | Annika Jodicke[2] | Talita Duarte-Salles[7,8] | Peter R. Rijnbeek[7] | Katia Verhamme[7] | Alexandra Pacurariu[9] | Daniel Morales[9,10] | Luis Pinheiro[9] | Daniel Prieto-Alhambra[2,7] | Albert Prats-Uribe[2]

[1]Medical Sciences Division, University of Oxford, Oxford, UK | [2]Pharmaco- and Device Epidemiology, Centre for Statistics in Medicines, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK | [3]Public Health Department, Medical Information Service, Medical Informatics and Archiving Unit (IAM), University Hospital of Bordeaux, Bordeaux, France | [4]Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Research Institute, Barcelona, Spain | [5]Management and Control Department, Consorci Mar Parc de Salut de Barcelona, Barcelona, Spain | [6]Real World Solutions, IQVIA, Brighton, UK | [7]Department of Medical Informatics, Erasmus University Medical Centre, Rotterdam, The Netherlands | [8]Fundació Institut Universitari per a la Recerca a l'Atencio Primaria de Salut Jordi Gol I Gurina (IDIAPJGol), Universitat Autonoma de Barcelona, Barcelona, Spain | [9]Real World Evidence Workstream, European Medicines Agency, Amsterdam, The Netherlands | [10]Division of Population Health and Genomics, University of Dundee, Dundee, UK

**Correspondence:** Daniel Prieto-Alhambra (d.prietoalhambra@darwin-eu.org)

**Received:** 10 January 2024 | **Revised:** 2 October 2024 | **Accepted:** 6 October 2024

https://doi.org/10.1002/pds.70042

# Questions?

**a.prats-uribe@darwin-eu.org**