



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.

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-Besora² | Romain Griffier³ | Guillaume Verdy³ | Angela Leis⁴ | Juan Manuel Ramirez-Anguita⁵ | Miguel A. Mayer^{4,5} | James T. Brash⁶ | Sarah Seager⁶ | Rowan Parry⁷ | Annika Jodicke² | Talita Duarte-Salles^{7,8} | Peter R. Rijnbeek⁷ | Katia Verhamme⁷ | Alexandra Pacurariu⁹ | Daniel Morales^{9,10}  | Luis Pinheiro⁹  | Daniel Prieto-Alhambra^{2,7}  | Albert Prats-Urbe²

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

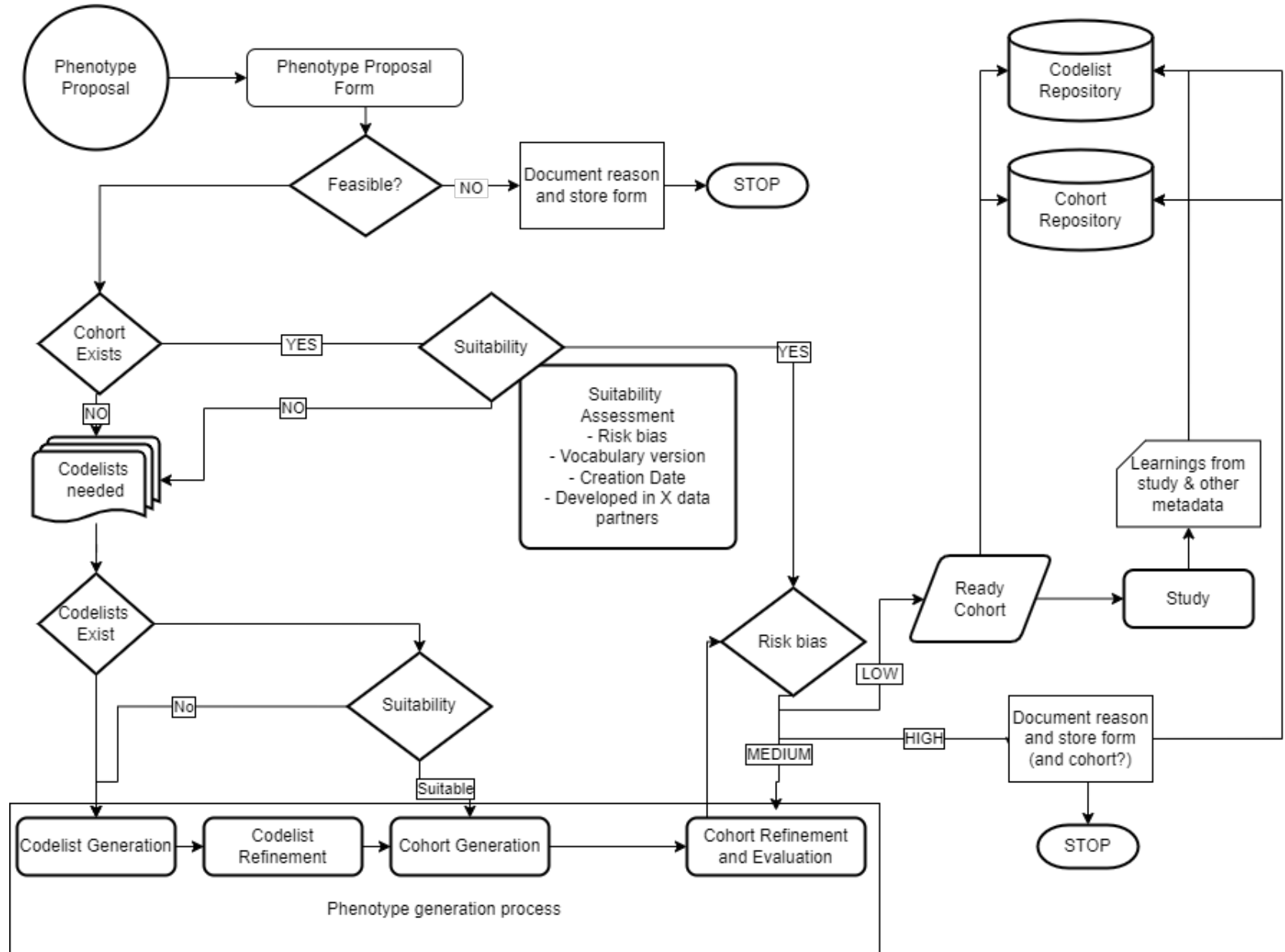
DARWIN EU data network strengths and DARWIN EU catalogue of standard analytics in mind

How?

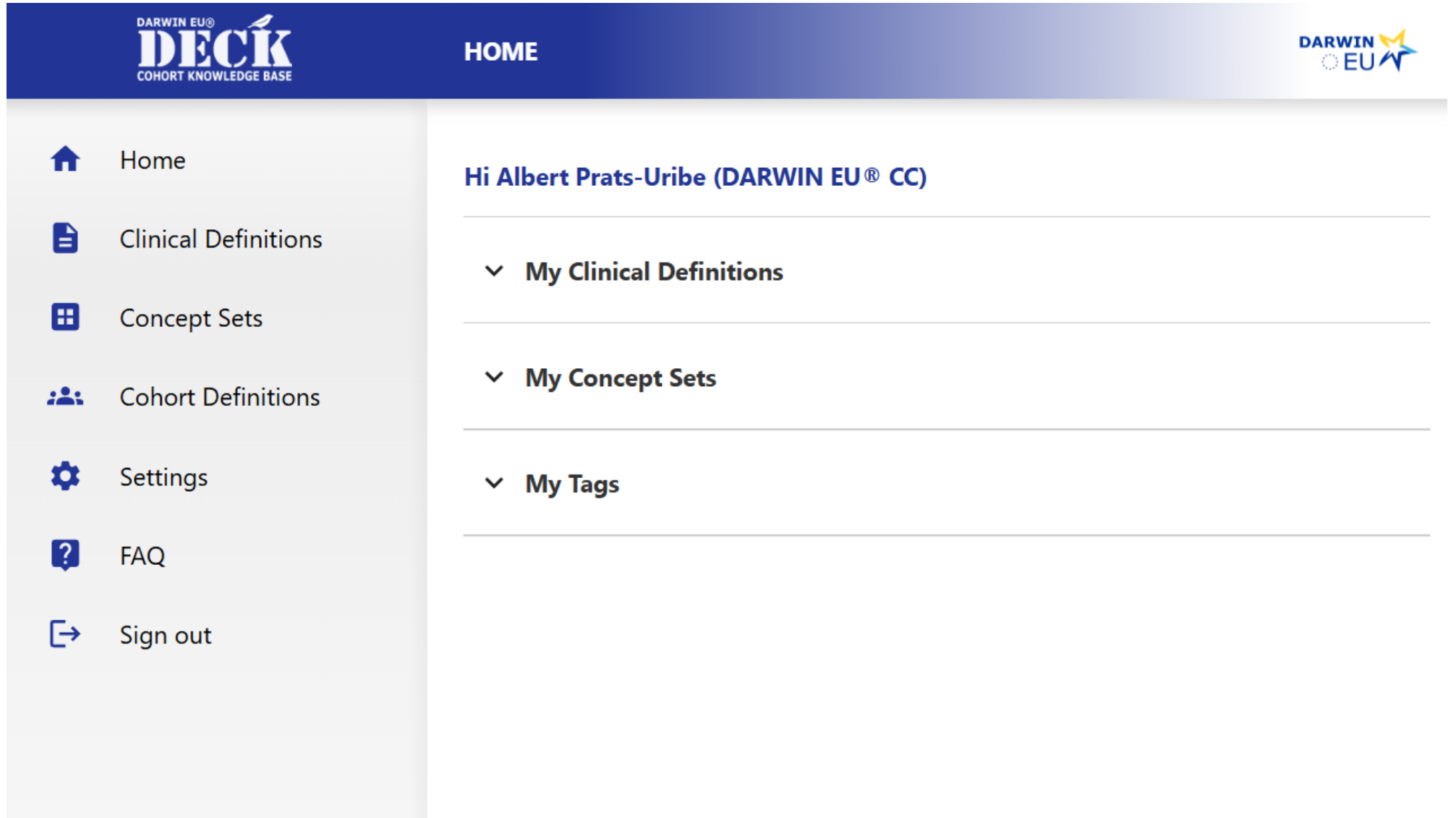
Co-creation DARWIN CC with **EMA** according to their needs and regulatory outcomes:

- Creation of a **DARWIN-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...

The screenshot shows the user interface of the Darwin EU DECK Cohort Knowledge Base. The top navigation bar is dark blue with the Darwin EU logo on the right and the text "HOME" in the center. Below the navigation bar is a sidebar on the left with a light gray background, containing a list of menu items with icons: Home (house icon), Clinical Definitions (document icon), Concept Sets (grid icon), Cohort Definitions (people icon), Settings (gear icon), FAQ (question mark icon), and Sign out (door icon). The main content area on the right has a white background and displays the user's name "Hi Albert Prats-Urbe (DARWIN EU® CC)" at the top. Below the name are three expandable sections, each with a downward-pointing chevron icon and the text "My Clinical Definitions", "My Concept Sets", and "My Tags".

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...

The screenshot shows the 'DECK COHORT KNOWLEDGE BASE' interface. The top navigation bar includes the DECK logo, the text 'CLINICAL DEFINITIONS', and the Darwin EU logo. A left-hand sidebar contains a menu with the following items: Home (house icon), Clinical Definitions (document icon), Concept Sets (grid icon), Cohort Definitions (people icon), Settings (gear icon), FAQ (question mark icon), and Sign out (exit icon). The main content area is titled 'New Clinical Definition' and contains several input fields: 'Name *' (with 'Test_Disease_1' entered and a user profile icon), 'Description', 'Study ID', 'Principal Investigator', 'PI Email' (with a user profile icon), and 'Deadline / Milestone' (with a calendar icon). A blue button labeled 'ADDITIONAL PI' is located at the bottom right of the form.

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

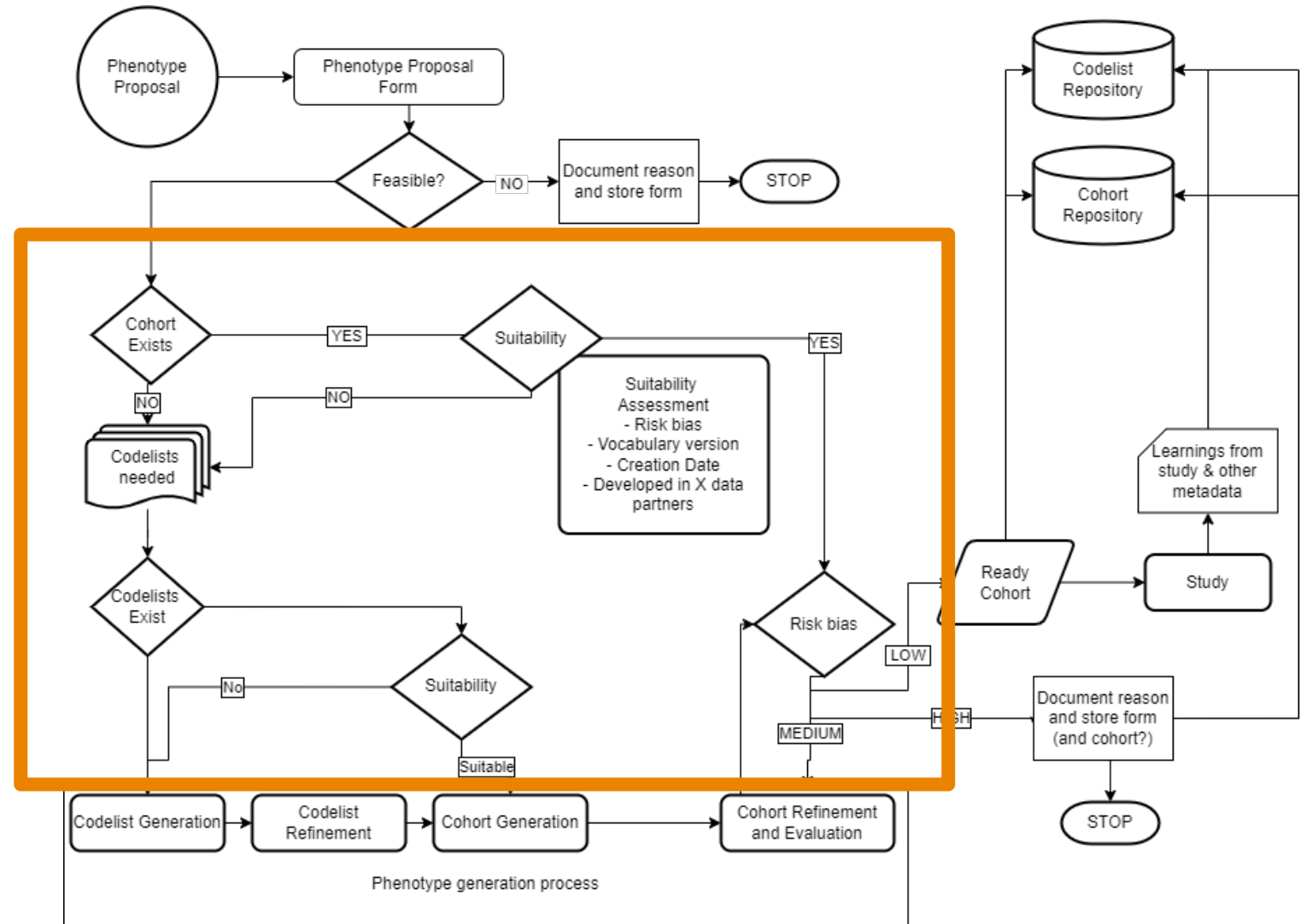
List all proposed concept sets and the proposed keywords that will be used for the search strategy. Also add domains to be searched in, and exclusions if known.

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

OVERVIEW CONCEPTS LOG VERSIONS REVIEWS

name ↑	domain	broad	narrow	prevalent	isExcluded	includeDescendants	comments	delete
Acute systemic lupus erythem...	Condition	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Autosomal systemic lupus eryth...	Condition	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Bullous systemic lupus erythema...	Condition	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Cheilitis due to lupus erythemat...	Condition	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Chilblain lupus erythematosus	Condition	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Chorea co-occurrent and due to ...	Condition	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Chorea in systemic lupus erythe...	Condition	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Cutaneous lupus erythematosus	Condition	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		
Demyelination of central nervou...	Condition	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>		

Rows per page: 100 1-92 of 92

[+ ADD CONCEPT](#) [SAVE AS SPECIFIC VERSION](#)

DARWIN EU DECK CONCEPT KNOWLEDGE BASE

CONCEPT SETS → SYSTEMIC LUPUS ERYTHEMATOSUS

OVERVIEW CONCEPTS LOG VERSIONS REVIEWS

[EXPORT TO CSV](#)

version	description	created by	date		
before_review	Original Codelist Generator List	Albert Prats-Urbe (DARWIN EU® CC)	7/17/2023		
Review by d.prietoalhambra@darwin-eu.org		Albert Prats-Urbe (DARWIN EU® CC)	7/17/2023		
v1	Version after review	Albert Prats-Urbe (DARWIN EU® CC)	7/17/2023		
Review by james.bezer@spc.ox.ac.uk		Albert Prats-Urbe (DARWIN EU® CC)	8/16/2023		
Review by abigail.robinson@spc.ox.ac.uk		Albert Prats-Urbe (DARWIN EU® CC)	7/25/2023		
Review by abigail.robinson@spc.ox.ac.uk		Albert Prats-Urbe (DARWIN EU® CC)	8/16/2023		
Copy of Review by abigail.robinson@spc.ox.ac.uk	Copy of undefined	James Bezer	10/8/2023		

Home
Clinical Definitions
Concept Sets
Cohort Definitions
Settings
FAQ
Sign out

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.

The screenshot shows the DECK Cohort Knowledge Base interface. The top navigation bar includes the DECK logo and the breadcrumb path: COHORT DEFINITIONS → CKD → V0. A sidebar on the left contains navigation links: Home, Clinical Definitions, Concept Sets, Cohort Definitions, Settings, FAQ, and Sign out. The main content area is titled "Cohort Definition Review Results" and contains several input fields for review parameters: "Reviewer *", "Database" (a dropdown menu), "Run date" (with a calendar icon), and a list of diagnostic categories: "flavour counts and overlaps", "orphan concepts", "incidence over time", "characteristics", "characteristics comparison", and "notes". The bottom left corner of the interface displays "DECK Portal v1.0.0-rc2".

Application to two phenotypes (with a simple version)

Pancreatic Cancer

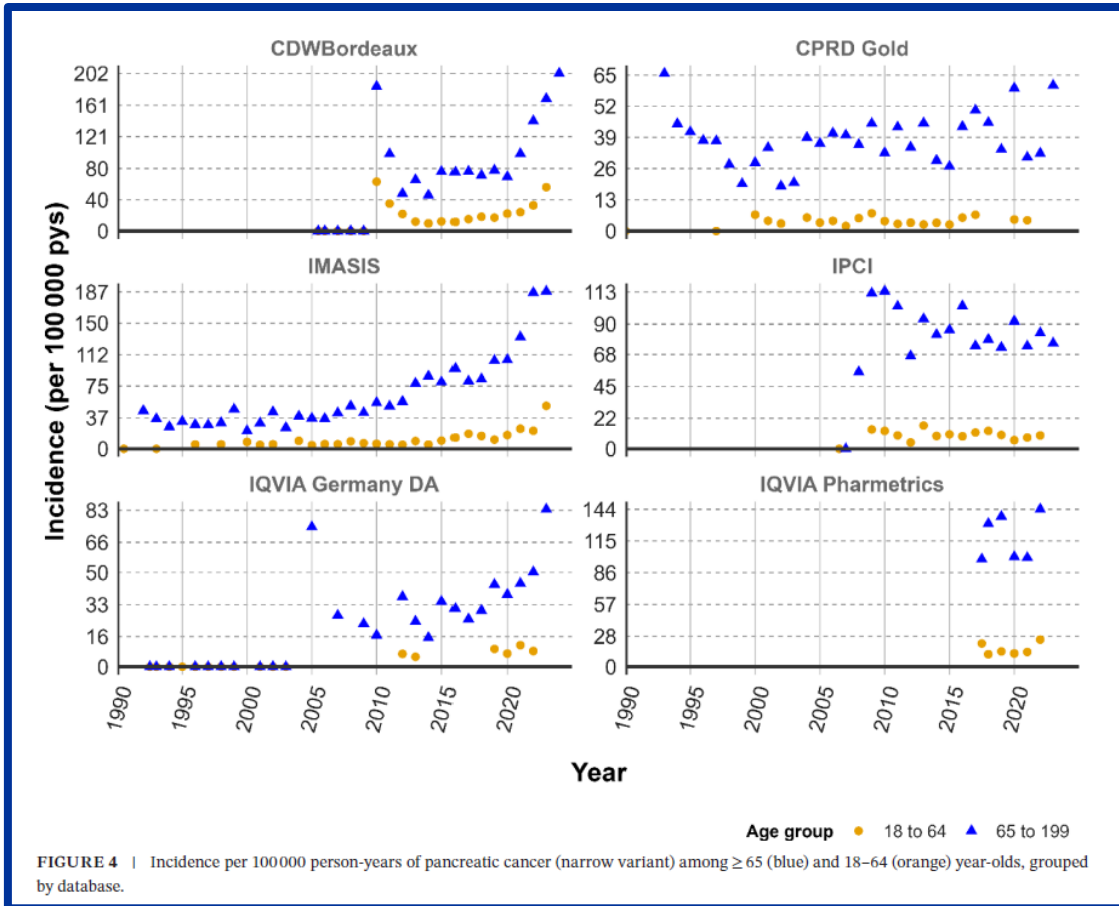


FIGURE 4 | Incidence per 100000 person-years of pancreatic cancer (narrow variant) among ≥65 (blue) and 18-64 (orange) year-olds, grouped by database.

SLE

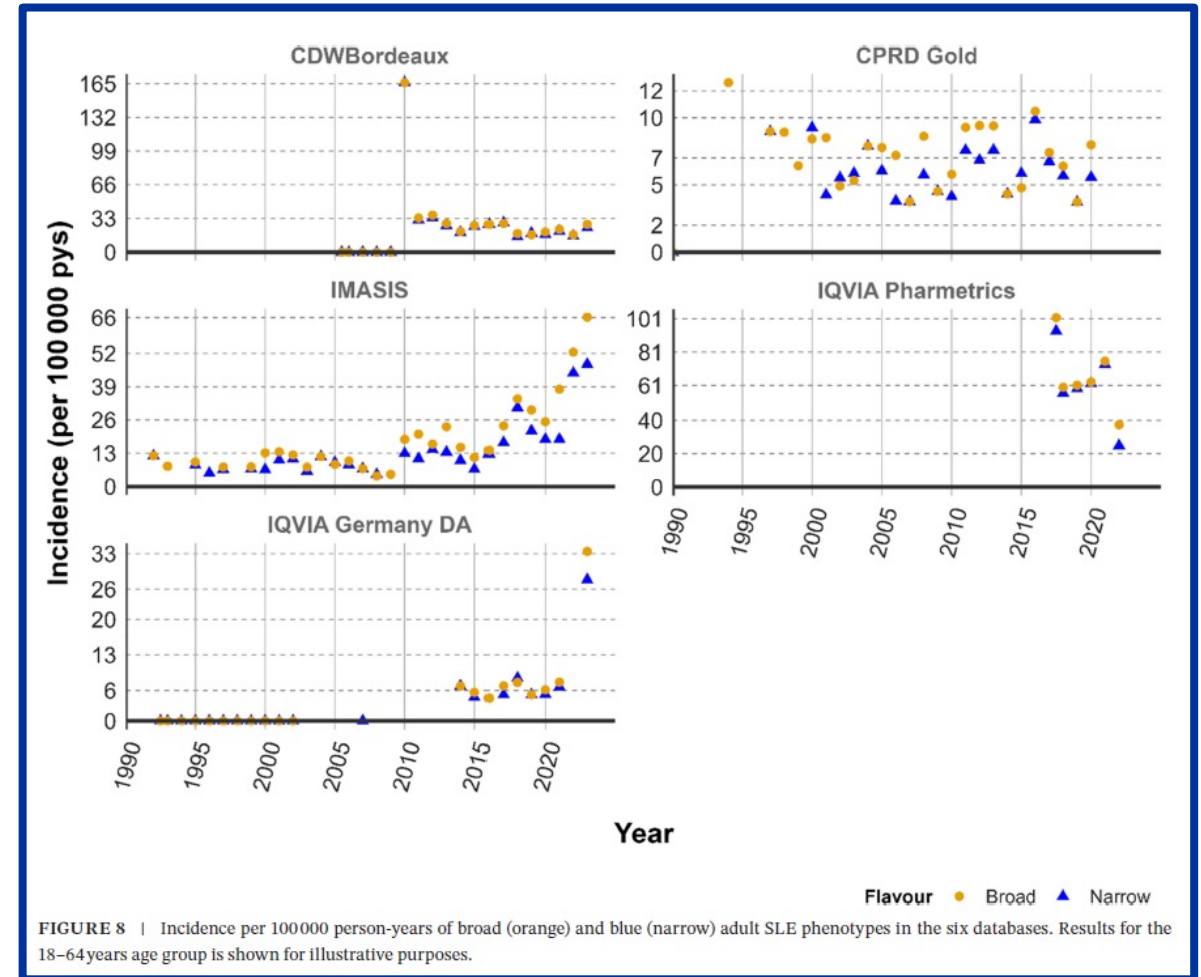





FIGURE 8 | Incidence per 100000 person-years of broad (orange) and blue (narrow) adult SLE phenotypes in the six databases. Results for the 18-64 years age group is shown for illustrative purposes.

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-Besora² | Romain Griffier³ | Guillaume Verdy³ | Angela Leis⁴ | Juan Manuel Ramirez-Anguila⁵ | Miguel A. Mayer^{4,5} | James T. Brash⁶ | Sarah Seager⁶ | Rowan Parry⁷ | Annika Jodicke² | Talita Duarte-Salles^{7,8} | Peter R. Rijnbeek⁷ | Katia Verhamme⁷ | Alexandra Pacurariu⁹ | Daniel Morales^{9,10}  | Luis Pinheiro⁹  | Daniel Prieto-Alhambra^{2,7}  | Albert Prats-Uribe²

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