Phenotype Phebruary kickoff

Looking back at Phenotype Phebruary 2022

https://www.ohdsi.org/phenotype-phebruary/

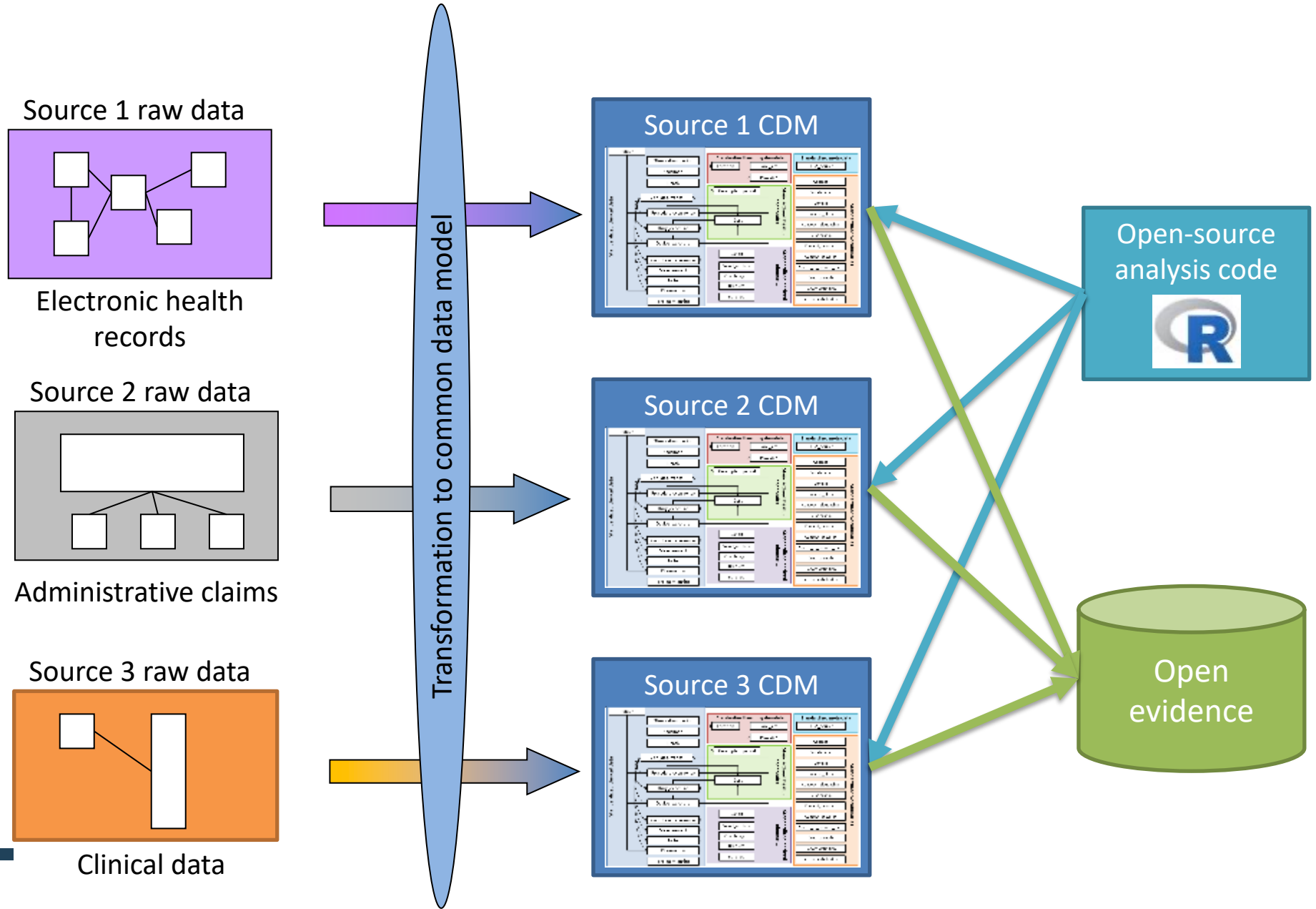
When poll is active, respond at **PollEv.com/patrickryan800**

What do you want to see accomplished during Phenotype Phebruary 2023?

Top

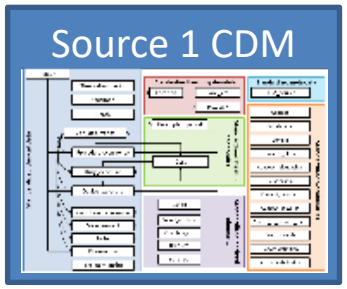
No responses received yet. They will appear here...

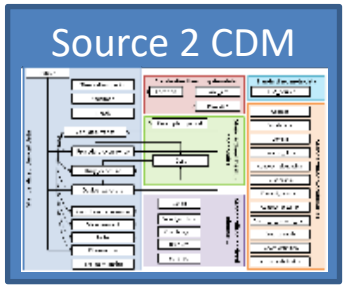
Common data model can enable standardized analytics across a distributed data network

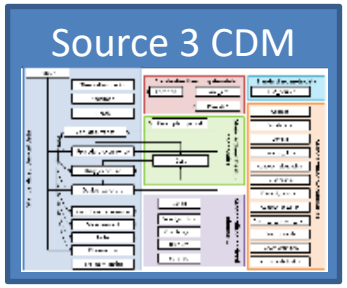


The journey to evidence

Standardized data







Cohort definition: a specification to identify the set of persons satisfying one or more criteria for a duration of time

Standardized analytics

T, {E} → Treatment pathways

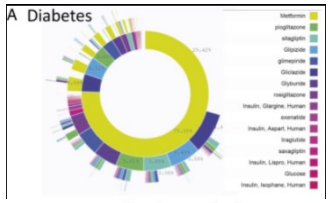
T, O → Incidence rate analysis

T, C, O → Comparative cohort design

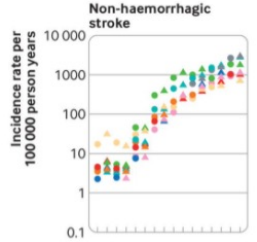
T, O → Self-controlled case series

T, O → Patient-level prediction

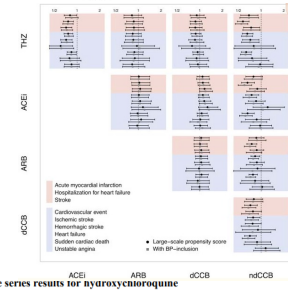
Impactful results



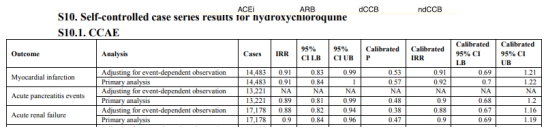
Hripcsak et al
PNAS 2016



Li et al
BMJ 2021



Suchard et al
Lancet 2019

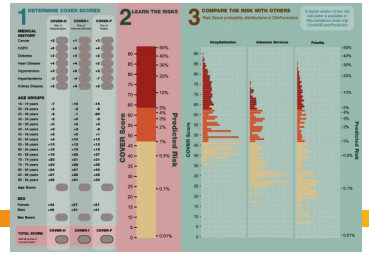


S10. Self-controlled case series results for hydroxychloroquine

S10.1. CCAE

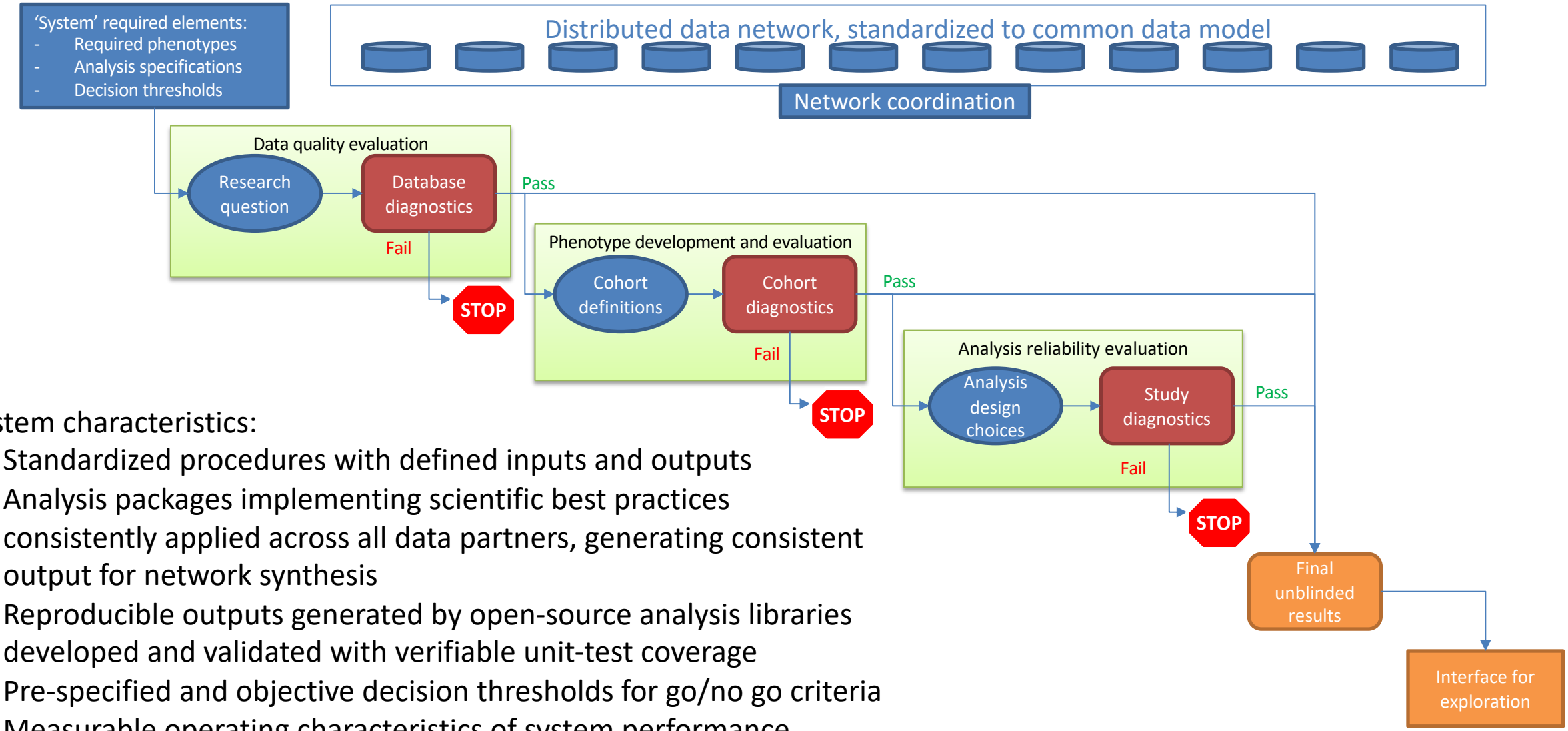
Outcome	Analysis	Cases	IRR	95% CI LB	95% CI UB	Calibrated P	Calibrated IRR	Calibrated 95% CI LB	Calibrated 95% CI UB
Myocardial infarction	Adjusting for event-dependent observation	14,483	0.91	0.83	0.99	0.53	0.91	0.69	1.21
	Primary analysis	14,483	0.91	0.84	1	0.57	0.92	0.7	1.22
Acute pancreatitis events	Adjusting for event-dependent observation	13,221	NA	NA	NA	NA	NA	NA	NA
	Primary analysis	13,221	0.89	0.81	0.99	0.48	0.9	0.68	1.2
Acute renal failure	Adjusting for event-dependent observation	17,178	0.88	0.82	0.94	0.38	0.88	0.67	1.16
	Primary analysis	17,178	0.9	0.84	0.96	0.47	0.9	0.69	1.19

Lane et al Lancet Rheumatology 2020



Williams et al
BMC MRM 2022

Engineering open science systems that build trust into the real-world evidence generation and dissemination process



System characteristics:

- Standardized procedures with defined inputs and outputs
- Analysis packages implementing scientific best practices consistently applied across all data partners, generating consistent output for network synthesis
- Reproducible outputs generated by open-source analysis libraries developed and validated with verifiable unit-test coverage
- Pre-specified and objective decision thresholds for go/no go criteria
- Measurable operating characteristics of system performance

OHDSI's definition of 'cohort'

Cohort = a set of persons who satisfy one or more inclusion criteria for a duration of time

Cohort era = a continuous period during which a person has satisfied a cohort's inclusion criteria

Cohort definition = the specification for how to identify a cohort

Questions to answer when defining a cohort

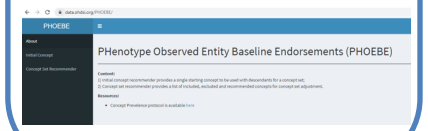
- What event(s) let you enter the cohort?
- What inclusion criteria are applied to those events?
- For each event, how long do you satisfy the inclusion criteria?
- How should events be combined into cohort eras?

Concept Set Expressions

- Concept Set = logical expression to represent a list of concepts in the OHDSI vocabularies
  - List of 1 or more concepts
  - optional operators for each concept in list:
    - Descendants = uses CONCEPT_ANCESTOR to identify standard concepts which have descendant ancestral relationship with selected concepts
    - Exclude = remove concept (and optionally descendants) from list
    - Mapped = use CONCEPT_RELATIONSHIP to materialize non-standard concepts for all included concepts
- A conceptset expression can be materialized into a list of concepts using any instance of the OHDSI vocabularies
  - JSON expression executed via webAPI into standard SQL query

A phenotype development and evaluation workflow

Concept set expressions



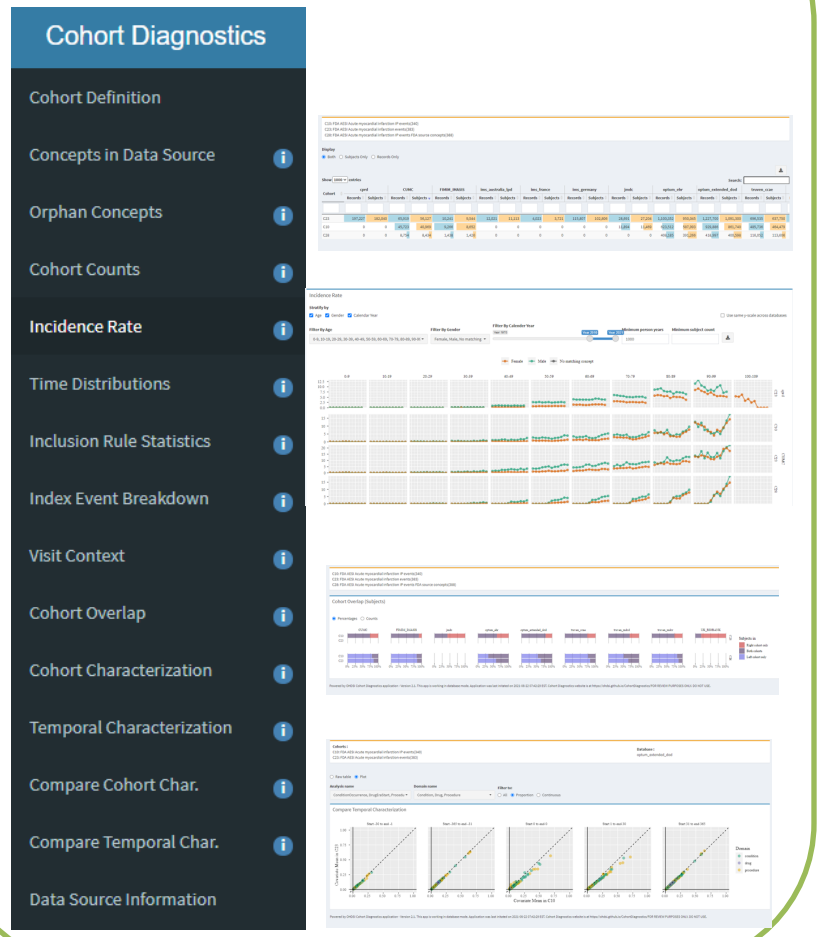
Cohort definition logic

Initial events
- Conceptsets

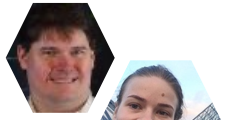
Inclusion criteria
- Conceptsets

Exit strategy
- Conceptsets

Cohort diagnostics



OHDSI open-source community tools to support phenotype development and evaluation process




Phenotype definition tools:

- ATLAS
  - Concept set expressions – with recommendations from PHOEBE2.0
  - Cohort Definitions – to design a rule-based cohort definition
  - Profiles – to review individual cases
- CapR - cohort definition application programming in R, to design rule-based cohort definitions consistent with CIRCE JSON specifications
- APHRODITE - to develop a probabilistic phenotype by training a prediction model using noisy labels



Phenotype evaluation tools:

- CohortExplorer – to review individual cases
- CohortDiagnostics – to evaluate phenotype algorithms using population-level characterization to identify sensitivity/specificity errors and index date misspecification
- PheValuator - to evaluate a phenotype algorithm (estimate sensitivity/specificity/PPV) by training a prediction model and creating a probabilistic reference standard

Phenotype Library

2. Classification and Diagnosis of Diabetes: *Standards of Medical Care in Diabetes—2021*

American Diabetes Association

Diabetes Care 2021;44(Suppl. 1):S15–S33 | https://doi.org/10.2337/dc21-S002

CLASSIFICATION

Diabetes can be classified into the following general categories:

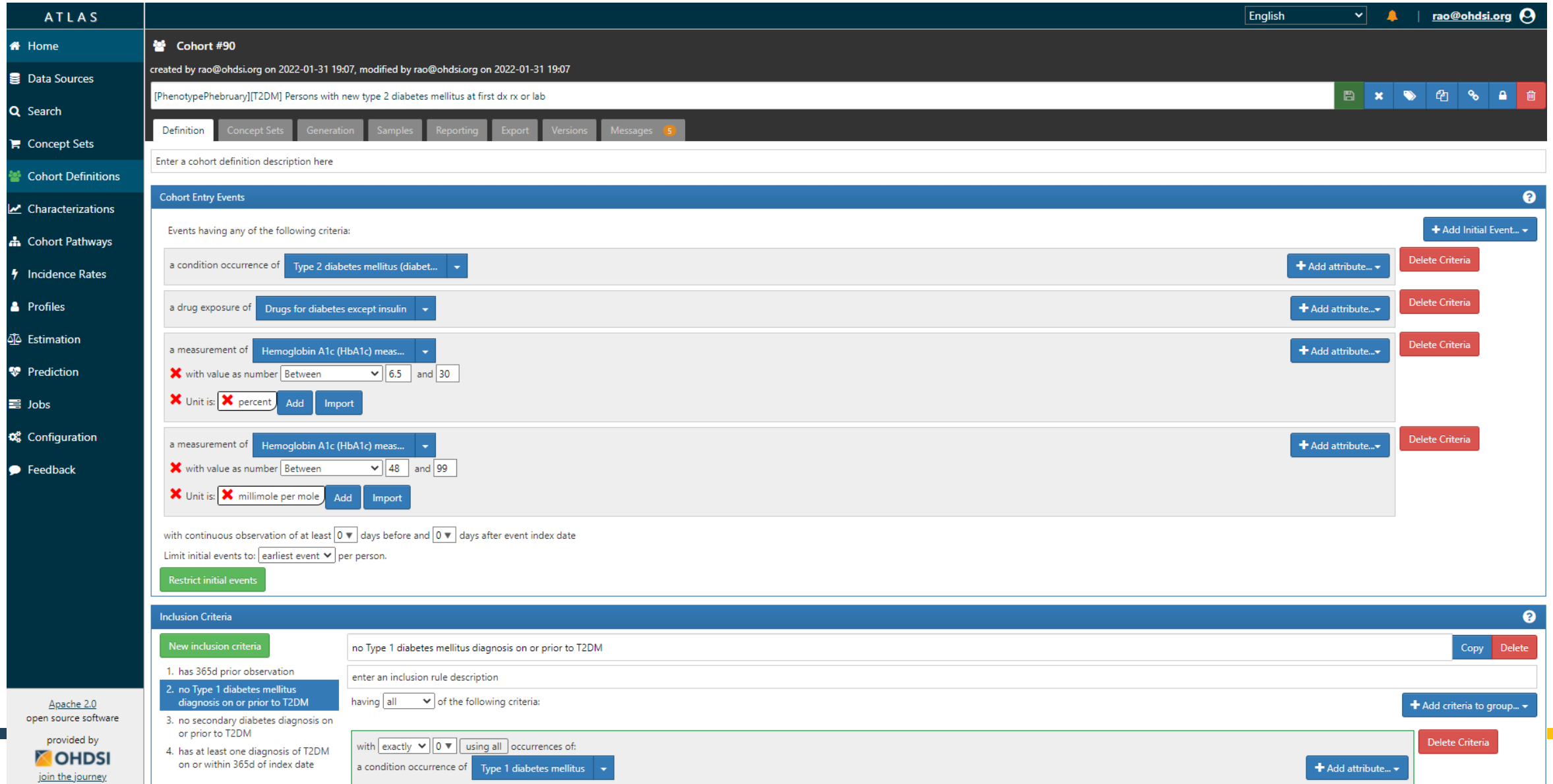
1. Type 1 diabetes (due to autoimmune β-cell destruction, usually leading to absolute insulin deficiency, including latent autoimmune diabetes of adulthood)
2. Type 2 diabetes (due to a progressive loss of adequate β-cell insulin secretion frequently on the background of insulin resistance)
3. Specific types of diabetes due to other causes, e.g., monogenic diabetes syndromes (such as neonatal diabetes and maturity-onset diabetes of the young), diseases of the exocrine pancreas (such as cystic fibrosis and pancreatitis), and drug- or chemical-induced diabetes (such as with glucocorticoid use, in the treatment of HIV/AIDS, or after organ transplantation)
4. Gestational diabetes mellitus (diabetes diagnosed in the second or third trimester of pregnancy that was not clearly overt diabetes prior to gestation)

Table 2.2—Criteria for the diagnosis of diabetes

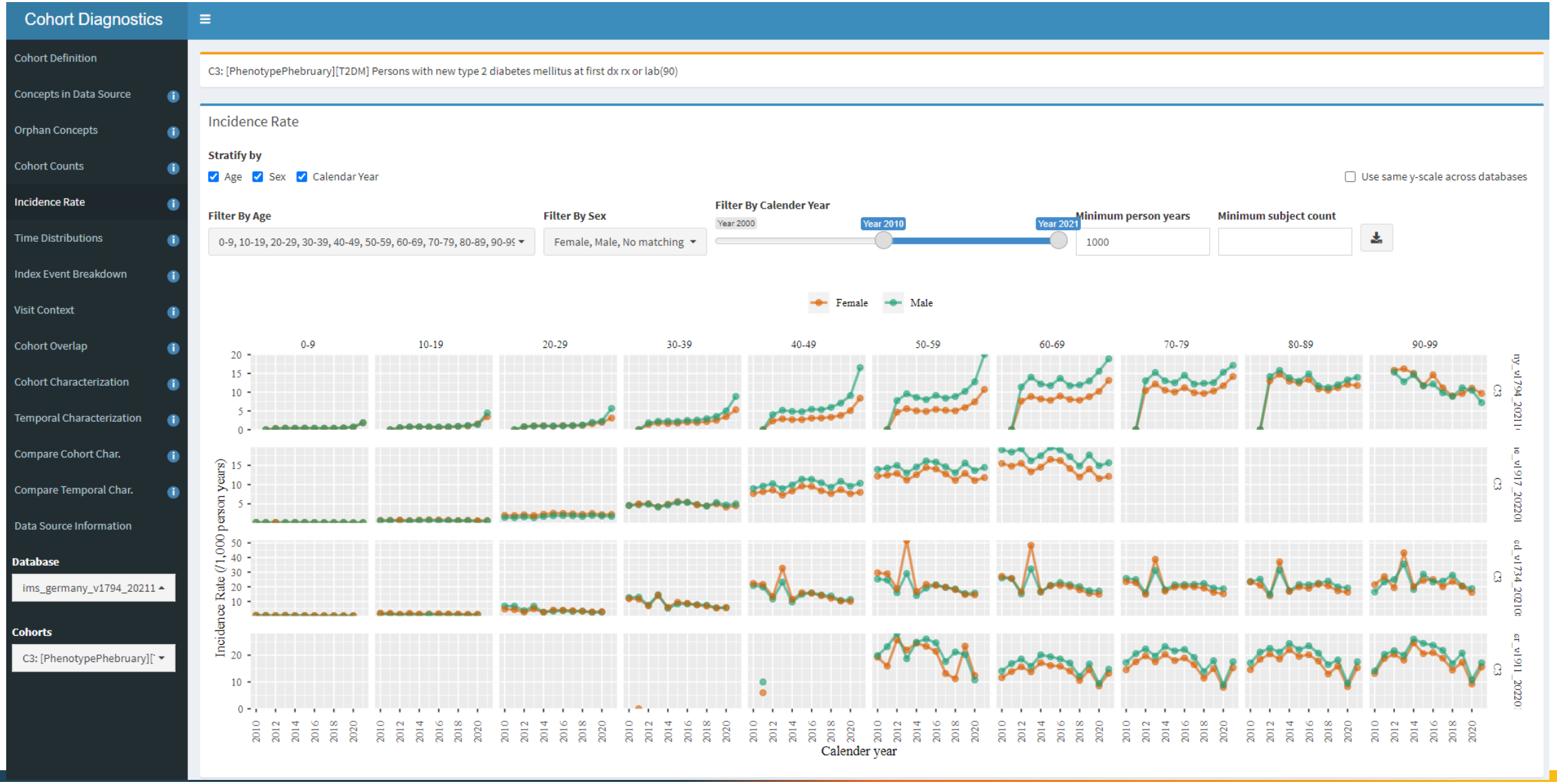
FPG ≥126 mg/dL (7.0 mmol/L). Fasting is defined as no caloric intake for at least 8 h.*
OR
2-h PG ≥200 mg/dL (11.1 mmol/L) during OGTT. The test should be performed as described by WHO, using a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water.*
OR
A1C ≥6.5% (48 mmol/mol). The test should be performed in a laboratory using a method that is NGSP certified and standardized to the DCCT assay.*
OR
In a patient with classic symptoms of hyperglycemia or hyperglycemic crisis, a random plasma glucose ≥200 mg/dL (11.1 mmol/L).

DCCT, Diabetes Control and Complications Trial; FPG, fasting plasma glucose; OGTT, oral glucose tolerance test; WHO, World Health Organization; 2-h PG, 2-h plasma glucose. *In the absence of unequivocal hyperglycemia, diagnosis requires two abnormal test results from the same sample or in two separate test samples.

Creating T2DM definition(s) in ATLAS



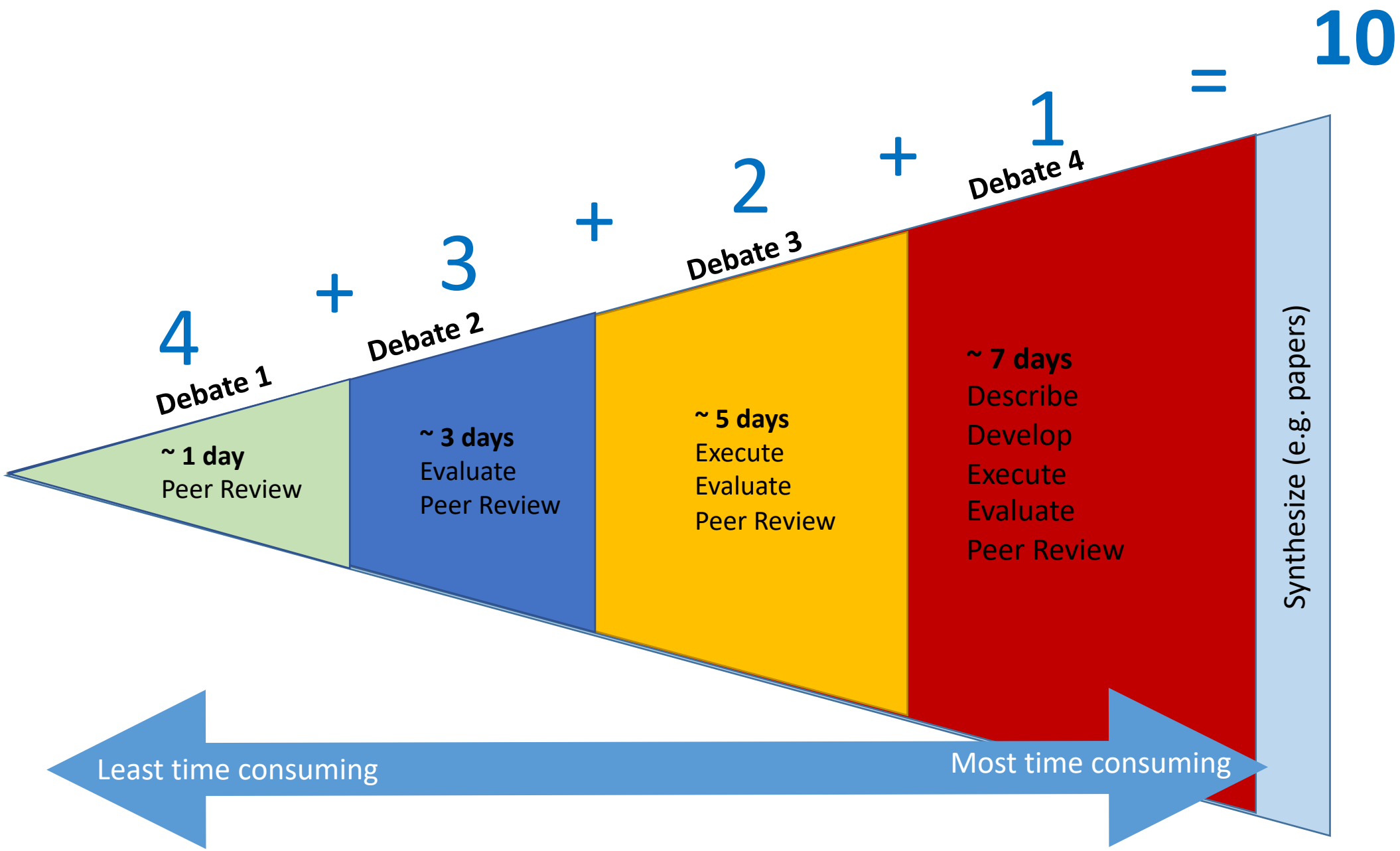
Evaluating T2DM definitions using CohortDiagnostics



Four Weeks. Four Debates. 10 Completed phenotypes

Stagger
Focused
Harden
Complete
Make useful

Debates



2023 FEBRUARY

CALENDAR YEAR | CALENDAR MONTH | MONDAY: FIRST DAY OF WEEK

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunda
30	31	01	02	03	04	
		P 1	P 2	P 3	P 4	
06	07	08	09	10	11	
P 5	OHDSI Community Call	P 6		P 7		
13	14	15	16	17	18	
P 8	OHDSI Community Call		P 9			
20	21	22	23	24	25	
P 10	OHDSI Community Call					
27	28	01	02	03	04	
	OHDSI Community					

Start 4 phenotypes

Debate 1: What is peer review? What is the value of peer review? What should be in peer review? Who should do the peer review?

Start 3 phenotypes

Debate 2: What is evaluation and why should we empirically evaluate cohort definitions? What to look for and perform an evaluation? What are the operating characteristics of a cohort definition? How helpful are the tools (Cohort Diagnostics, PheValuator, Cohort Explorer)? What is a gold standard and do we truly need it?

Start 2 phenotypes

Debate 3: What makes cohort definitions reusable? Do we customize phenotype development to data and/or analytical use case (study) or, is phenotyping an independent activity? What terminology should we use when we describe operating characteristics of a cohort definitions (e.g., sensitive cohort, specific cohort). What is the value of OHDSI PL and what should be in it?

Start 1 phenotypes

Debate 4: The role of probabilistic modelling in phenotype development and evaluation. Can we sustain this level of effort for rule-based phenotyping, does it scale?

Synthesize

Debate 5: Is phenotyping a scientific work that is worthy of scientific dissemination via publication? How was comparing to Gold standard useful, feasible, limitations

OHDSI Collaboration Opportunity Spotlight

https://www.pharmacoepi.org/meetings/annual-conference/