28 Days, 28 Phenotypes

OHDSI

Phenotype Phebruary

forums.ohdsi.org

Join The Conversations!

# OHDSI Phenotype Phebruary: lessons learned

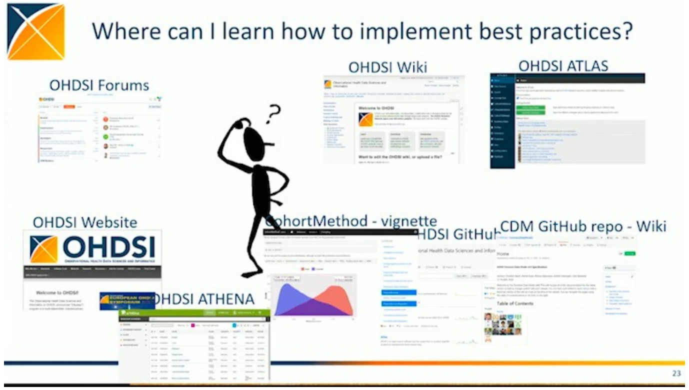
2022 OHDSI Symposium

14 October 2022 Bethesda MD

Azza Shoaibi, Joel Swerdel, Allan Wu, Gowtham Rao, Adam Black, Evan Minty, Asieh Golozar, Rupa Makadia, Jill Hardin, Erica Voss, Tiffany J. Callahan, Juan Banda, Anna Ostropolets, Claudia Pulgarin, Marcela Rivera, David Vizcaya, Patrick Ryan

# Phenotype development and evaluation is yet to become a completed chapter in the book of OHDSI

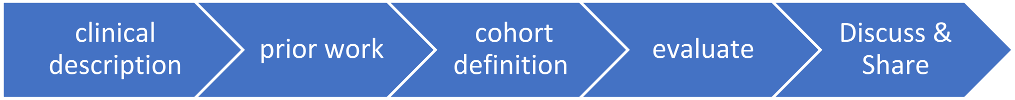
- Phenotypes are the foundational elements in almost every real-world analysis.
- The reliability of the generated evidence depends on the validity of the phenotypes.
- Yet, the science of phenotype development and evaluation is relatively immature.
- We have built best practices and end to end process, tools and packages for characterization, estimation and prediction.
- But for phenotyping- **“addressed in the limitation section.”**



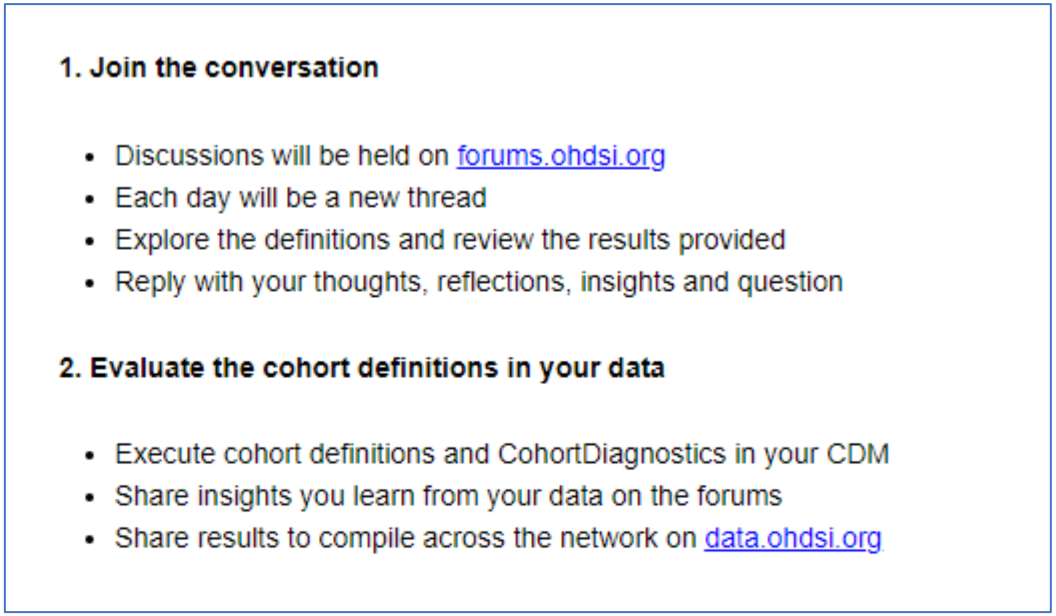
Where can I learn how to implement best practices?

# " **Phenotype Phebruary"**: I realized that becoming a master of karate was not about learning 4,000 moves but about doing just a handful of moves 4,000 times." — Chet Holmes

- We collectively started a discussion on 28 phenotypes over 28 days
- Followed 5 step process:



clinical description → prior work → cohort definition → evaluate → Discuss & Share

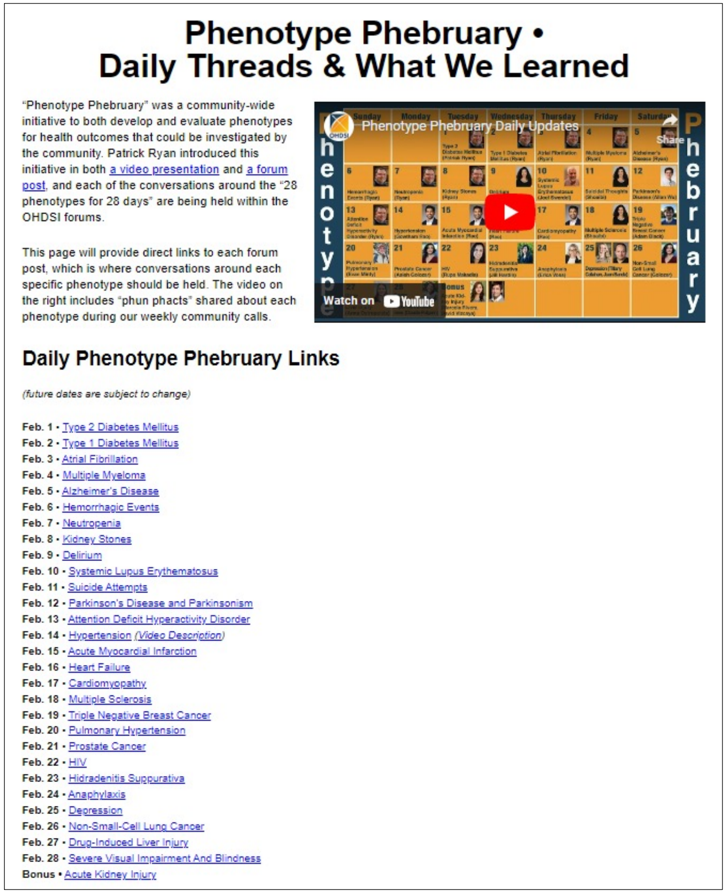


**1. Join the conversation**

- Discussions will be held on forums.ohdsi.org
- Each day will be a new thread
- Explore the definitions and review the results provided
- Reply with your thoughts, reflections, insights and question

**2. Evaluate the cohort definitions in your data**

- Execute cohort definitions and CohortDiagnostics in your CDM
- Share insights you learn from your data on the forums
- Share results to compile across the network on data.ohdsi.org



## Phenotype Phebruary • Daily Threads & What We Learned

"Phenotype Phebruary" was a community-wide initiative to both develop and evaluate phenotypes for health outcomes that could be investigated by the community. Patrick Ryan introduced this initiative in both a video presentation and a forum post, and each of the conversations around the "28 phenotypes for 28 days" are being held within the OHDSI forums.

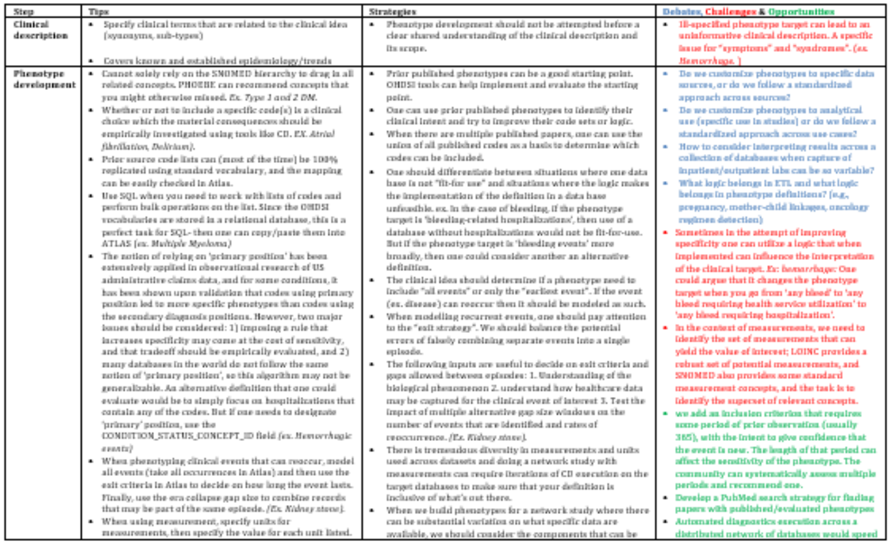
This page will provide direct links to each forum post, which is where conversations around each specific phenotype should be held. The video on the right includes "phun phacts" shared about each phenotype during our weekly community calls.

### Daily Phenotype Phebruary Links

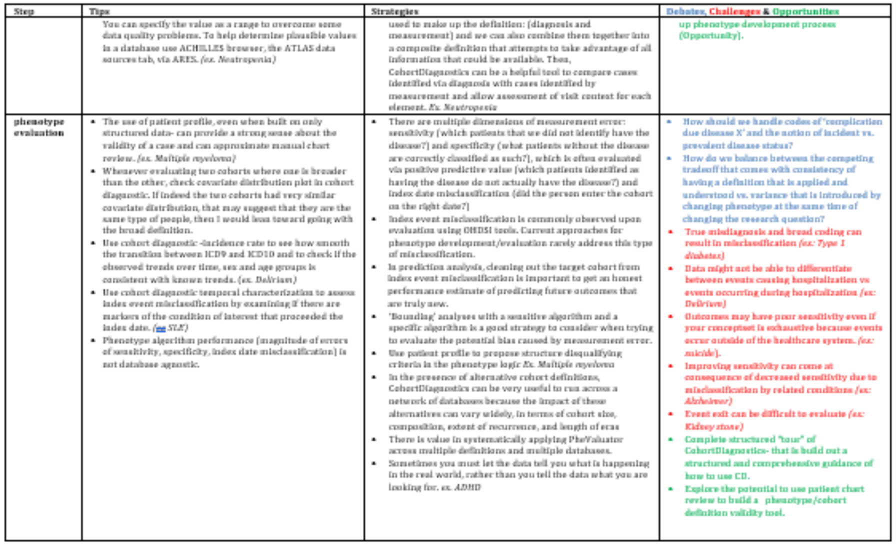
*(future dates are subject to change)*

**Feb. 1** • Type 2 Diabetes Mellitus
**Feb. 2** • Type 1 Diabetes Mellitus
**Feb. 3** • Atrial Fibrillation
**Feb. 4** • Multiple Myeloma
**Feb. 5** • Alzheimer's Disease
**Feb. 6** • Hemorrhagic Events
**Feb. 7** • Neutropenia
**Feb. 8** • Kidney Stones
**Feb. 9** • Delirium
**Feb. 10** • Systemic Lupus Erythematosus
**Feb. 11** • Suicide Attempts
**Feb. 12** • Parkinson's Disease and Parkinsonism
**Feb. 13** • Attention Deficit Hyperactivity Disorder
**Feb. 14** • Hypertension *(Video Description)*
**Feb. 15** • Acute Myocardial Infarction
**Feb. 16** • Heart Failure
**Feb. 17** • Cardiomyopathy
**Feb. 18** • Multiple Sclerosis
**Feb. 19** • Triple Negative Breast Cancer
**Feb. 20** • Pulmonary Hypertension
**Feb. 21** • Prostate Cancer
**Feb. 22** • HIV
**Feb. 23** • Hidradenitis Suppurativa
**Feb. 24** • Anaphylaxis
**Feb. 25** • Depression
**Feb. 26** • Non-Small-Cell Lung Cancer
**Feb. 27** • Drug-Induced Liver Injury
**Feb. 28** • Severe Visual Impairment And Blindness
**Bonus** • Acute Kidney Injury

# 15 phenotypes were developed, evaluated and discussed and we learned few things



| Step | Tips | Strategies | Debates, Challenges & Opportunities |
|---|---|---|---|
| **Clinical description** | • Specify clinical terms that are related to the clinical idea (synonyms, sub-types)<br>• Covers known and established epidemiology/trends | • Phenotype development should not be attempted before a clear shared understanding of the clinical description and its scope. | • Ill-specified phenotype target can lead to an uninformative clinical description. A specific issue for "symptoms" and "syndromes". *(ex. Hemorrhage. )* |
| **Phenotype development** | • Cannot solely rely on the SNOMED hierarchy to drag in all related concepts. PHOEBE can recommend concepts that you might otherwise missed. *Ex. Type 1 and 2 DM.*<br>• Whether or not to include a specific code(s) is a clinical choice which the material consequences should be empirically investigated using tools like CD. *EX. Atrial fibrillation, Delirium).*<br>• Prior source code lists can (most of the time) be 100% replicated using standard vocabulary, and the mapping can be easily checked in Atlas.<br>• Use SQL when you need to work with lists of codes and perform bulk operations on the list. Since the OHDSI vocabularies are stored in a relational database, this is a perfect task for SQL- then one can copy/paste them into ATLAS *(ex. Multiple Myeloma)*<br>• The notion of relying on 'primary position' has been extensively applied in observational research of US administrative claims data, and for some conditions, it has been shown upon validation that codes using primary position led to more specific phenotypes than codes using the secondary diagnosis positions. However, two major issues should be considered: 1) imposing a rule that increases specificity may come at the cost of sensitivity, and that tradeoff should be empirically evaluated, and 2) many databases in the world do not follow the same notion of 'primary position', so this algorithm may not be generalizable. An alternative definition that one could evaluate would be to simply focus on hospitalizations that contain any of the codes. But if one needs to designate 'primary' position, use the CONDITION_STATUS_CONCEPT_ID field *(ex. Hemorrhagic events)*<br>• When phenotyping clinical events that can reoccur, model all events (take all occurrences in Atlas) and then use the exit criteria in Atlas to decide on how long the event lasts. Finally, use the era collapse gap size to combine records that may be part of the same episode. *(Ex. Kidney stone).*<br>• When using measurement, specify units for measurements, then specify the value for each unit listed. | • Prior published phenotypes can be a good starting point. OHDSI tools can help implement and evaluate the starting point.<br>• One can use prior published phenotypes to identify their clinical intent and try to improve their code sets or logic.<br>• When there are multiple published papers, one can use the union of all published codes as a basis to determine which codes can be included.<br>• One should differentiate between situations where one data base is not "fit-for use" and situations where the logic makes the implementation of the definition in a data base unfeasible. ex. In the case of bleeding, if the phenotype target is 'bleeding-related hospitalizations', then use of a database without hospitalizations would not be fit-for-use. But if the phenotype target is 'bleeding events' more broadly, then one could consider another an alternative definition.<br>• The clinical idea should determine if a phenotype need to include "all events" or only the "earliest event". If the event (ex. disease) can reoccur then it should be modeled as such.<br>• When modelling recurrent events, one should pay attention to the "exit strategy". We should balance the potential errors of falsely combining separate events into a single episode.<br>• The following inputs are useful to decide on exit criteria and gaps allowed between episodes: 1. Understanding of the biological phenomenon 2. understand how healthcare data may be captured for the clinical event of interest 3. Test the impact of multiple alternative gap size windows on the number of events that are identified and rates of reoccurrence. *(Ex. Kidney stone).*<br>• There is tremendous diversity in measurements and units used across datasets and doing a network study with measurements can require iterations of CD execution on the target databases to make sure that your definition is inclusive of what's out there.<br>• When we build phenotypes for a network study where there can be substantial variation on what specific data are available, we should consider the components that can be | • Do we customize phenotypes to specific data sources, or do we follow a standardized approach across sources?<br>• Do we customize phenotypes to analytical use (specific use in studies) or do we follow a standardized approach across use cases?<br>• How to consider interpreting results across a collection of databases when capture of inpatient/outpatient labs can be so variable?<br>• What logic belongs in ETL and what logic belongs in phenotype definitions? (e.g., pregnancy, mother-child linkages, oncology regimen detection)<br>• Sometimes in the attempt of improving specificity one can utilize a logic that when implemented can influence the interpretation of the clinical target. *Ex: hemorrhage:* One could argue that it changes the phenotype target when you go from 'any bleed' to 'any bleed requiring health service utilization' to 'any bleed requiring hospitalization'.<br>• In the context of measurements, we need to identify the set of measurements that can yield the value of interest; LOINC provides a robust set of potential measurements, and SNOMED also provides some standard measurement concepts, and the task is to identify the superset of relevant concepts.<br>• we add an inclusion criterion that requires some period of prior observation (usually 365), with the intent to give confidence that the event is new. The length of that period can affect the sensitivity of the phenotype. The community can systematically assess multiple periods and recommend one.<br>• Develop a PubMed search strategy for finding papers with published/evaluated phenotypes<br>• Automated diagnostics execution across a distributed network of databases would speed |



| Step | Tips | Strategies | Debates, Challenges & Opportunities |
|---|---|---|---|
|  | You can specify the value as a range to overcome some data quality problems. To help determine plausible values in a database use ACHILLES browser, the ATLAS data sources tab, via ARES. *(ex. Neutropenia)* | used to make up the definition: (diagnosis and measurement) and we can also combine them together into a composite definition that attempts to take advantage of all information that could be available. Then, CohortDiagnostics can be a helpful tool to compare cases identified via diagnosis with cases identified by measurement and allow assessment of visit context for each element. *Ex. Neutropenia* | up phenotype development process (Opportunity). |
| **phenotype evaluation** | • The use of patient profile, even when built on only structured data- can provide a strong sense about the validity of a case and can approximate manual chart review. *(ex. Multiple myeloma)*<br>• Whenever evaluating two cohorts where one is broader than the other, check covariate distribution plot in cohort diagnostic. If indeed the two cohorts had very similar covariate distribution, that may suggest that they are the same type of people, then I would lean toward going with the broad definition.<br>• Use cohort diagnostic -incidence rate to see how smooth the transition between ICD9 and ICD10 and to check if the observed trends over time, sex and age groups is consistent with known trends. *(ex. Delirium)*<br>• Use cohort diagnostic temporal characterization to assess index event misclassification by examining if there are markers of the condition of interest that proceeded the index date. *(ex SLE)*<br>• Phenotype algorithm performance (magnitude of errors of sensitivity, specificity, index date misclassification) is not database agnostic. | • There are multiple dimensions of measurement error: sensitivity (which patients that we did not identify have the disease?) and specificity (what patients without the disease are correctly classified as such?), which is often evaluated via positive predictive value (which patients identified as having the disease do not actually have the disease?) and index date misclassification (did the person enter the cohort on the right date?)<br>• Index event misclassification is commonly observed upon evaluation using OHDSI tools. Current approaches for phenotype development/evaluation rarely address this type of misclassification.<br>• In prediction analysis, cleaning out the target cohort from index event misclassification is important to get an honest performance estimate of predicting future outcomes that are truly new.<br>• 'Bounding' analyses with a sensitive algorithm and a specific algorithm is a good strategy to consider when trying to evaluate the potential bias caused by measurement error.<br>• Use patient profile to propose structure disqualifying criteria in the phenotype logic *Ex. Multiple myeloma*<br>• In the presence of alternative cohort definitions, CohortDiagnostics can be very useful to run across a network of databases because the impact of these alternatives can vary widely, in terms of cohort size, composition, extent of recurrence, and length of eras<br>• There is value in systematically applying PheValuator across multiple definitions and multiple databases.<br>• Sometimes you must let the data tell you what is happening in the real world, rather than you tell the data what you are looking for. ex. *ADHD* | • How should we handle codes of "complication due disease X" and the notion of incident vs. prevalent disease status?<br>• How do we balance between the competing tradeoff that comes with consistency of having a definition that is applied and understood vs. variance that is introduced by changing phenotype at the same time of changing the research question?<br>• True misdiagnosis and broad coding can result in misclassification *(ex: Type 1 diabetes)*<br>• Data might not be able to differentiate between events causing hospitalization vs events occurring during hospitalization *(ex: Delirium)*<br>• Outcomes may have poor sensitivity even if your conceptset is exhaustive because events occur outside of the healthcare system. *(ex: suicide)*.<br>• Improving sensitivity can come at consequence of decreased sensitivity due to misclassification by related conditions *(ex: Alzheimer)*<br>• Event exit can be difficult to evaluate *(ex: Kidney stone)*<br>• Complete structured "tour" of CohortDiagnostics- that is build out a structured and comprehensive guidance of how to use CD.<br>• Explore the potential to use patient chart review to build a phenotype/cohort definition validity tool. |

**Thematic learnings:** clinical description, phenotype development and phenotype evaluation. The themes identified belonged to **5 different types of lessons**: tips, strategies, debatable topics, challenges, and opportunities.

# Lessons learned: Phenotype development

## Tips/strategies

Evaluate ***all*** types of measurement error

Use patient profile to get a sense of validity. Identify disqualifying patterns.

Explore in CD:
temporal stability, expected trends, patients composition, index event misclassification

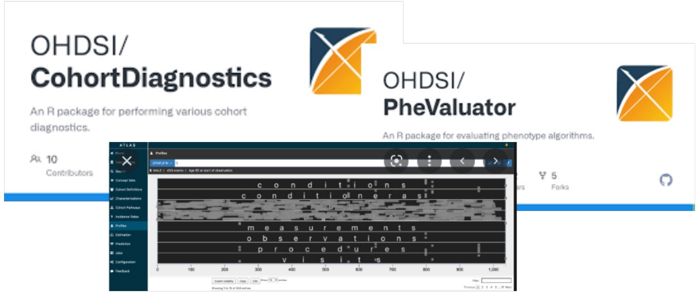
Estimate measurement error & quantify trade-offs by PheValuator or APHRODITE

## Challenges

Subjective

Time-consuming and complex

Lack an approach to evaluate exit criteria & washout periods for reoccurring events



OHDSI/
CohortDiagnostics

An R package for performing various cohort diagnostics.

10 Contributors

OHDSI/
PheValuator

An R package for evaluating phenotype algorithms.

5 Forks

What objective diagnostic criteria can we apply to determine fitness-for-use'?

# Lessons learned: Phenotype evaluation

## Tips/strategies

Model the clinical idea not the analytical use case

Code selection is a clinical choice. The material consequences should be empirically investigated

Notions like "primary position" need to be standardized and evaluated

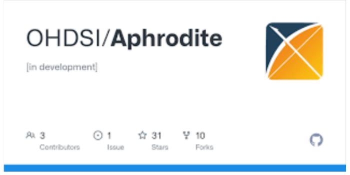
Differentiate between situations where a data is not “fit-for use” and situations where logic is not "fit" for the data

## Opportunities

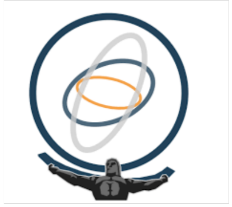
Systematically assess multiple look back periods and recommend one.

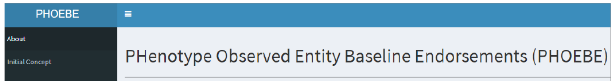
Combine a heuristic-based approach (APHRODITE) with a rule-based approach (Atlas)

Develop a PubMed search strategy to find published/evaluated phenotypes


OHDSI/**Aphrodite**

[in development]



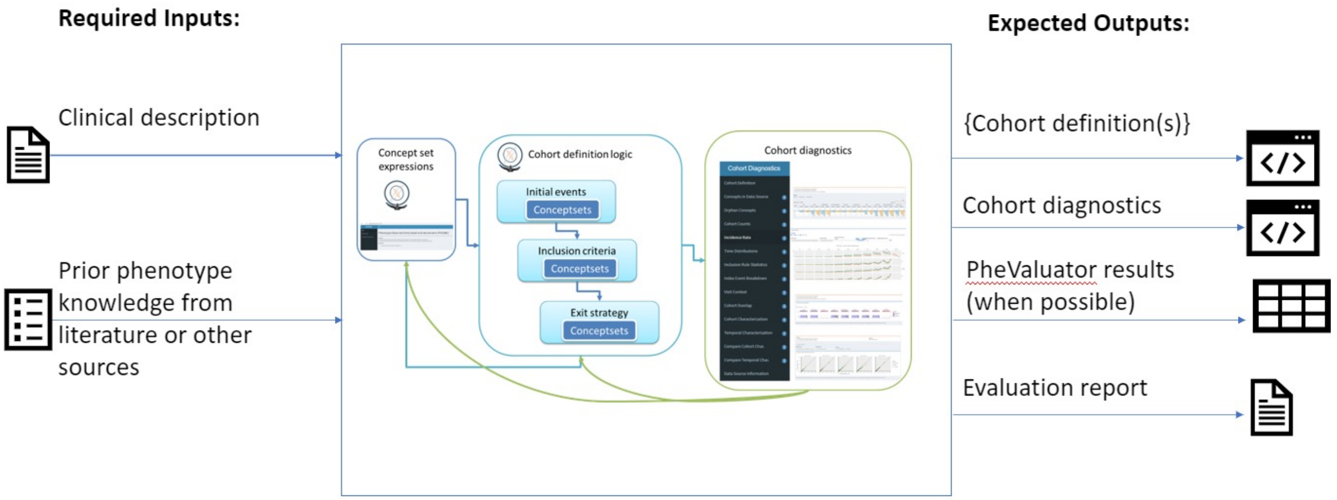


PHenotype Observed Entity Baseline Endorsements (PHOEBE)

**Do we customize phenotypes to specific data sources/analytical use ?**

# The choices are NOT:

## 1: Code list
## 2: Code list with chart reviews



**Required Inputs:**

Clinical description

Prior phenotype knowledge from literature or other sources

**Expected Outputs:**

{Cohort definition(s)}

Cohort diagnostics

PheValuator results (when possible)

Evaluation report

- Phenotyping is complex, multidimensional and requires exchange of knowledge, learnings and insights across collaborators from different background and expertise
- Large scale characterization (e.g.CD), Diagnostic predictive models (e.g., PheValuator) and structured review of patient’s profile are potentially effective and novel strategies for phenotype evaluation.
- We are getting closer to a standardized process. **But** further collaboration is needed to formalize a **scalable and reproducible** processes and establish **empirically-driven objective diagnostics**