Observational Health Data Sciences and Informatics (OHDSI)

Gold Standard Phenotype Library

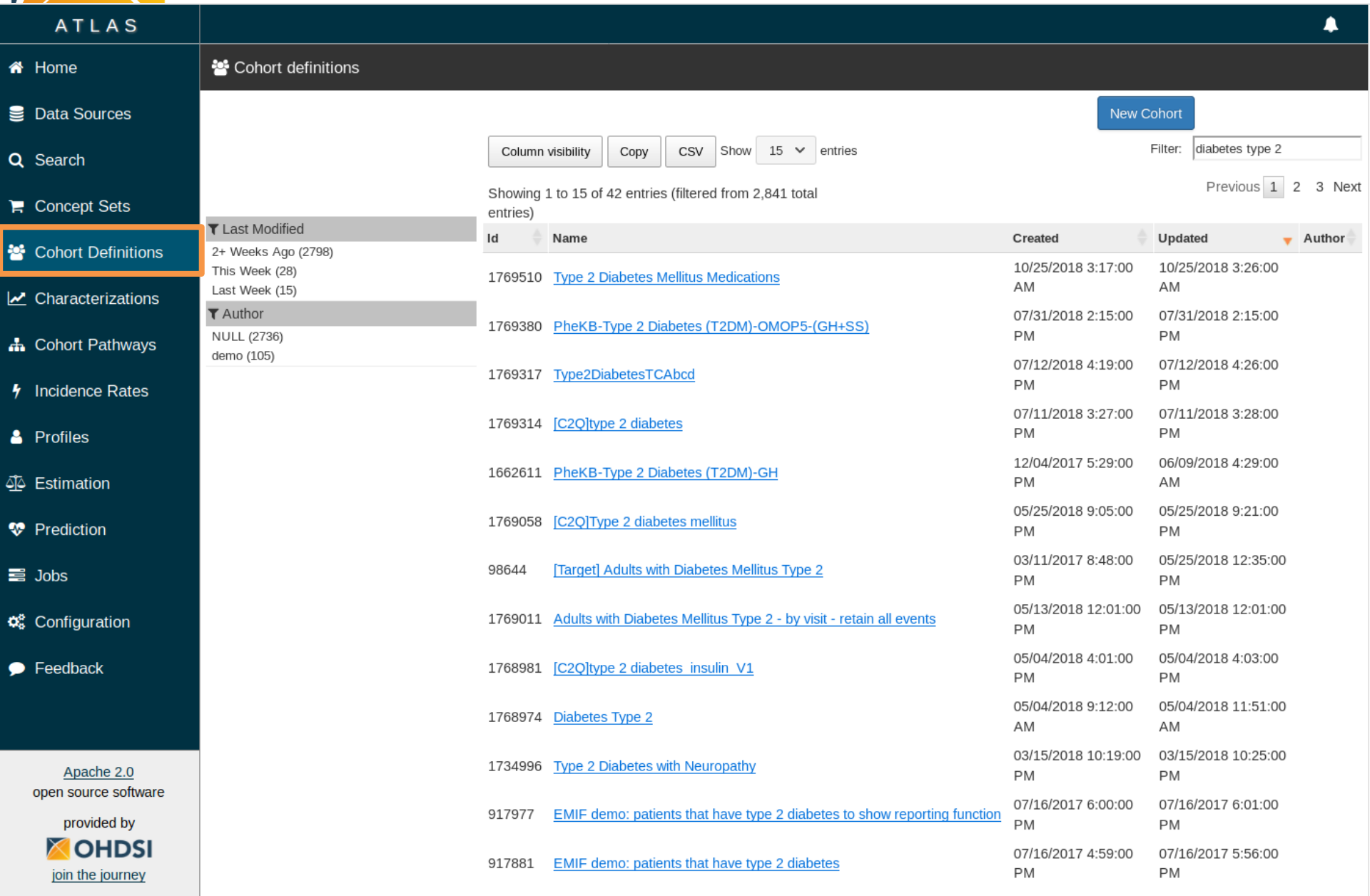
January 15, 2019

What is it, and why do we need it?

Suppose an investigator wishes to do a study using a cohort definition.

They may try going to Atlas to look for existing definitions.

Which ones should they consider?



Atlas is a Sandbox

1) People sometimes use it to make “test” or “demo” definitions for exploration or testing.

Id	Name
922762	Hypertension_test
226235	Hyperglycemia -- Glucose Test > 200 MSHN Test
146	Hypertension Test
922763	Hypertension_test2
922764	Hypertension_test3
922765	Hypertension_test3
922783	Hypertension_test4
922784	Hypertension_test4
922785	Hypertension_test4

Who Created It?

2) Currently, authors are anonymous.

Showing 136 to 150 of 2,840 entries

Previous 1 … 9 10 11 … 190 Next

Id	Name	Created	Updated	Author
1769764	PLP - Ischemic stoke index date	12/18/2018 5:13:00 PM	12/18/2018 5:20:00 PM	
1769762	New Cohort Definition MM	12/18/2018 3:49:00 PM	12/18/2018 4:03:00 PM	
1769760	gender_with_code	12/18/2018 2:42:00 PM	12/18/2018 2:45:00 PM	
1769526	patient-level prediction of endometriosis women in ED with abdominal laparoscopy target definition	10/28/2018 5:09:00 PM	12/17/2018 5:55:00 PM	
1769756	PheKB-T2DM-ZZ	12/17/2018 3:52:00 AM	12/17/2018 7:19:00 AM	
1769754	Endometriosis Phenotype B	12/16/2018 7:46:00 PM	12/16/2018 7:46:00 PM	
142972	PheKB-ADHD-PJ	05/02/2017 6:43:00 PM	12/16/2018 7:32:00 AM	
1769753	PheKB-ADHD-ZZ	12/16/2018 7:23:00 AM	12/16/2018 7:27:00 AM	
1769750	PheKB-Crohn's-disease-ZZ	12/15/2018 5:27:00 AM	12/16/2018 5:25:00 AM	
143083	PheKB-Autism-[missing nlp]-PJ	05/02/2017 7:55:00 PM	12/15/2018 6:04:00 PM	
1769682	C: dialysis OR CKD OR transplant (excluding acute states)	12/03/2018 1:56:00 AM	12/14/2018 7:37:00 PM	

Versioning is Desired

3) The community wants different phenotype versions.

Id	Name
422	eMERGE PheKB Type 2 Diabetes phenotype algorithm (Northwestern University) Literal
341	eMERGE PheKB Type 2 Diabetes phenotype algorithm (Northwestern University) oldV2
336	eMERGE PheKB Type 2 Diabetes phenotype algorithm (Northwestern University) oldV1
343	eMERGE PheKB Type 2 Diabetes phenotype algorithm (Northwestern University) oldV3
1198483	beva_ch_mbc_v01
112062	Thia_Cohort_v2
112061	Sulf_Cohort_v2

Tagging is Desired

4) The community wants different ways to tag phenotypes.

Id	Name
143083	PheKB-Autism-[missing nlp]-PJ
341	eMERGE PheKB Type 2 Diabetes phenotype algorithm (Northwestern University) oldV2
336	eMERGE PheKB Type 2 Diabetes phenotype algorithm (Northwestern University) oldV1
343	eMERGE PheKB Type 2 Diabetes phenotype algorithm (Northwestern University) oldV3
99131	[eMerge] Peripheral Arterial Disease
1769380	PheKB-Type 2 Diabetes (T2DM)-OMOP5-(GH+SS)
1662611	PheKB-Type 2 Diabetes (T2DM)-GH
143430	PheKB-Height-PJ
98094	[PHEKB] Type 2 Diabetes

Need for Performance Metrics

5) How can we compare/contrast multiple definitions?

Showing 1 to 15 of 42 entries (filtered from 2,840 total entries)

Previous 1 2 3 Next

Id	Name	Created	Updated	Author
1769510	Type 2 Diabetes Mellitus Medications	10/25/2018 3:17:00 AM	10/25/2018 3:26:00 AM	
1769380	PheKB-Type 2 Diabetes (T2DM)-OMOP5-(GH+SS)	07/31/2018 2:15:00 PM	07/31/2018 2:15:00 PM	
1769317	Type2DiabetesTCAbcd	07/12/2018 4:19:00 PM	07/12/2018 4:26:00 PM	
1769314	[C2Q]type 2 diabetes	07/11/2018 3:27:00 PM	07/11/2018 3:28:00 PM	
1662611	PheKB-Type 2 Diabetes (T2DM)-GH	12/04/2017 5:29:00 PM	06/09/2018 4:29:00 AM	
1769058	[C2Q]Type 2 diabetes mellitus	05/25/2018 9:05:00 PM	05/25/2018 9:21:00 PM	
98644	[Target] Adults with Diabetes Mellitus Type 2	03/11/2017 8:48:00 PM	05/25/2018 12:35:00 PM	
1769011	Adults with Diabetes Mellitus Type 2 - by visit - retain all events	05/13/2018 12:01:00 PM	05/13/2018 12:01:00 PM	
1768981	[C2Q]type 2 diabetes_insulin_V1	05/04/2018 4:01:00 PM	05/04/2018 4:03:00 PM	
1768974	Diabetes Type 2	05/04/2018 9:12:00 AM	05/04/2018 11:51:00 AM	
1734996	Type 2 Diabetes with Neuropathy	03/15/2018 10:19:00 PM	03/15/2018 10:25:00 PM	
917977	EMIF demo: patients that have type 2 diabetes to show reporting function	07/16/2017 6:00:00 PM	07/16/2017 6:01:00 PM	
917881	EMIF demo: patients that have type 2 diabetes	07/16/2017 4:59:00 PM	07/16/2017 5:56:00 PM	
912968	mt - Type 2 Diabetes Mellitus (T2DM)	07/06/2017 6:18:00 PM	07/06/2017 6:22:00 PM	
98094	[PHEKB] Type 2 Diabetes	02/27/2017 4:11:00 PM	06/17/2017 7:44:00 PM	

Sandbox vs. Library

- Atlas is awesome: The sandbox is great for many community uses!
- The sandbox is not the same thing as a phenotype library.
- As OHDSI grows, there is a need to have a centralized home for rigorously evaluated phenotypes

Enter:

Gold Standard Phenotype Library!

(= GSPL)

For our investigator who wants to carry out a study on a cohort, which cohort definition(s) should he/she consider?

- This is *not* about making the decision outright for the investigator.
- It is about providing the setup and the tools required to **empower the investigator to make an informed decision.**

Wouldn't it be great if…

Available Phenotypes

- Addison's disease
- Asthma
- Autism
- Bipolar Mood Disorder
- Bronchiectasis
- Cardiac failure
- Cardiomyopathy
- Chronic obstructive pulmonary disorder
- Chronic renal disease
- Coronary artery disease
- Crohn's disease
- Diabetes insipidus
- Diabetes mellitus - Type 1 or Type 2
- Diabetes mellitus - Type 2 only
- Dysrhythmias
- Epilepsy
- Familial Hyperlipidemia
- Glaucoma
- Hemophilia
- Hyperlipidemia
- Hypertension
- Hypothyroidism
- Metastatic Cancer - Excludes basal cell
- Multiple sclerosis
- Parkinson's disease
- Rheumatoid arthritis
- Schizophrenia
- Systemic lupus erythematosus
- Ulcerative colitis

- Hypertension
  - **Primary Hypertension**
  - Secondary Hypertension
  - Severe Hypertension
  - …

- ✓ Author/Contact Info
- ✓ Submission/Approval Dates
- ✓ Objective
- ✓ Development Process (Gold Standard)
- ✓ Versions/Tags
- ✓ Site-specific Validation Metrics (Sensitivity, specificity, positive predictive value, etc.)
- ✓ Known caveats/cautions
- ✓ Peer review schedule

GSPL Objective

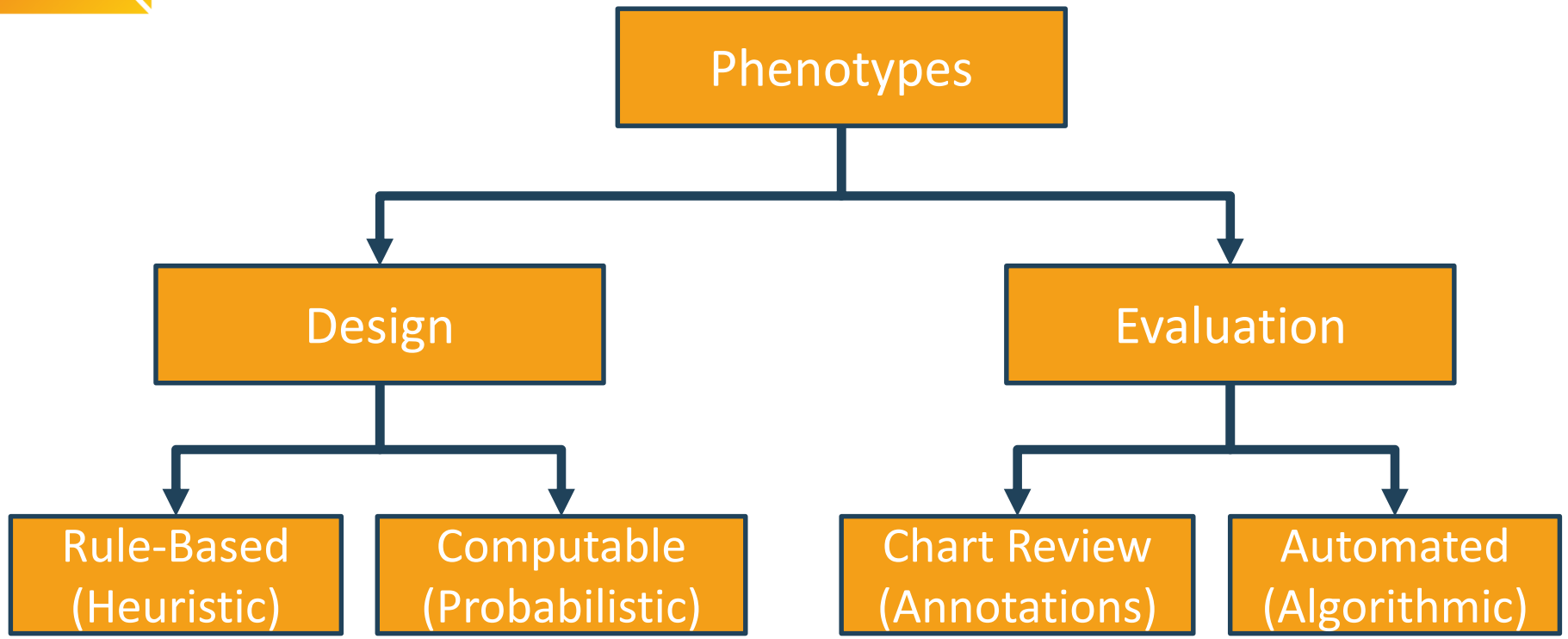
Objective: To enable members of the OHDSI community to find, evaluate, and utilize community-validated cohort definitions for research and other activities. These “gold standard” definitions will reside in a library, the entries of which are held to **specific standards of design and evaluation.**

- These standards are what the GSPL working group is ultimately working to establish, as well as the instantiation of the library itself.

What Makes a Phenotype a "Gold Standard"?

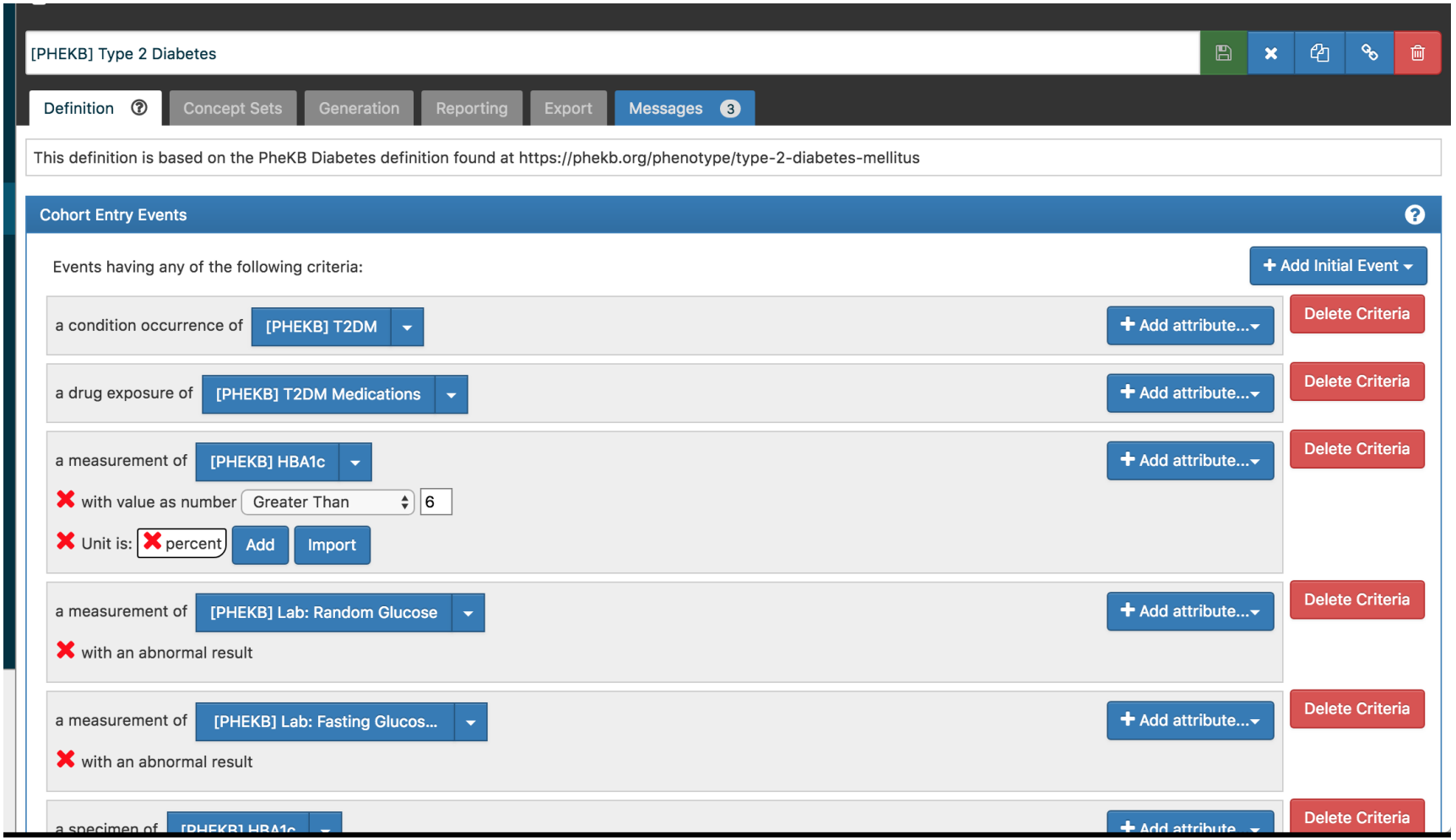
- A phenotype that has been:
  - **Designed** with best practices
  - **Evaluated** with best practices
  - **Documented** with best practices

Heuristic and Computational



\+ Library Architecture and Implementation

Heuristic Phenotype Design



Computable Phenotype Design

- Phenotype Discovery using Machine Learning (Aphrodite – Juan Banda)

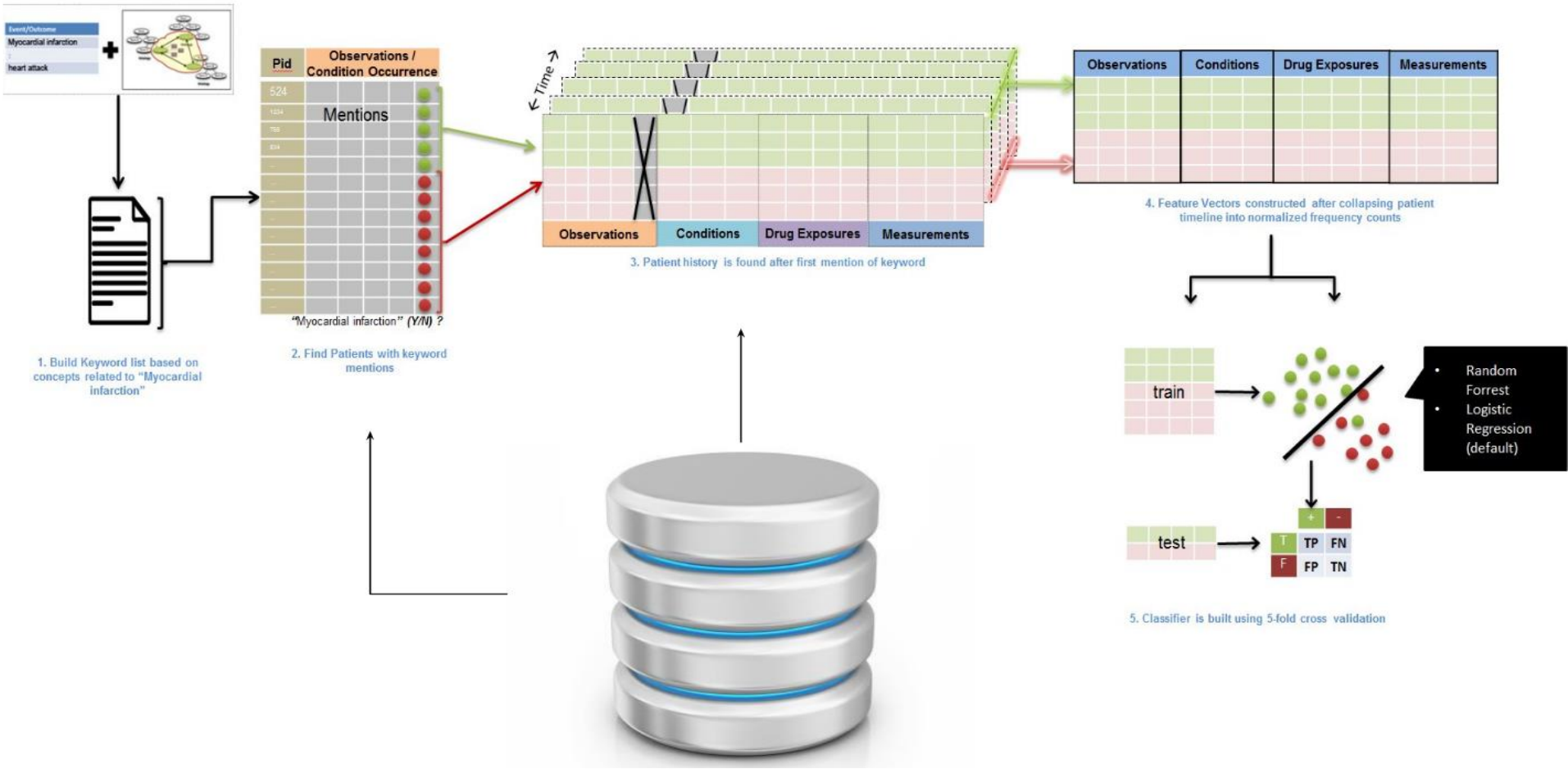


Chart Review / Evaluation

- Annotation Tool – Trey Schneider

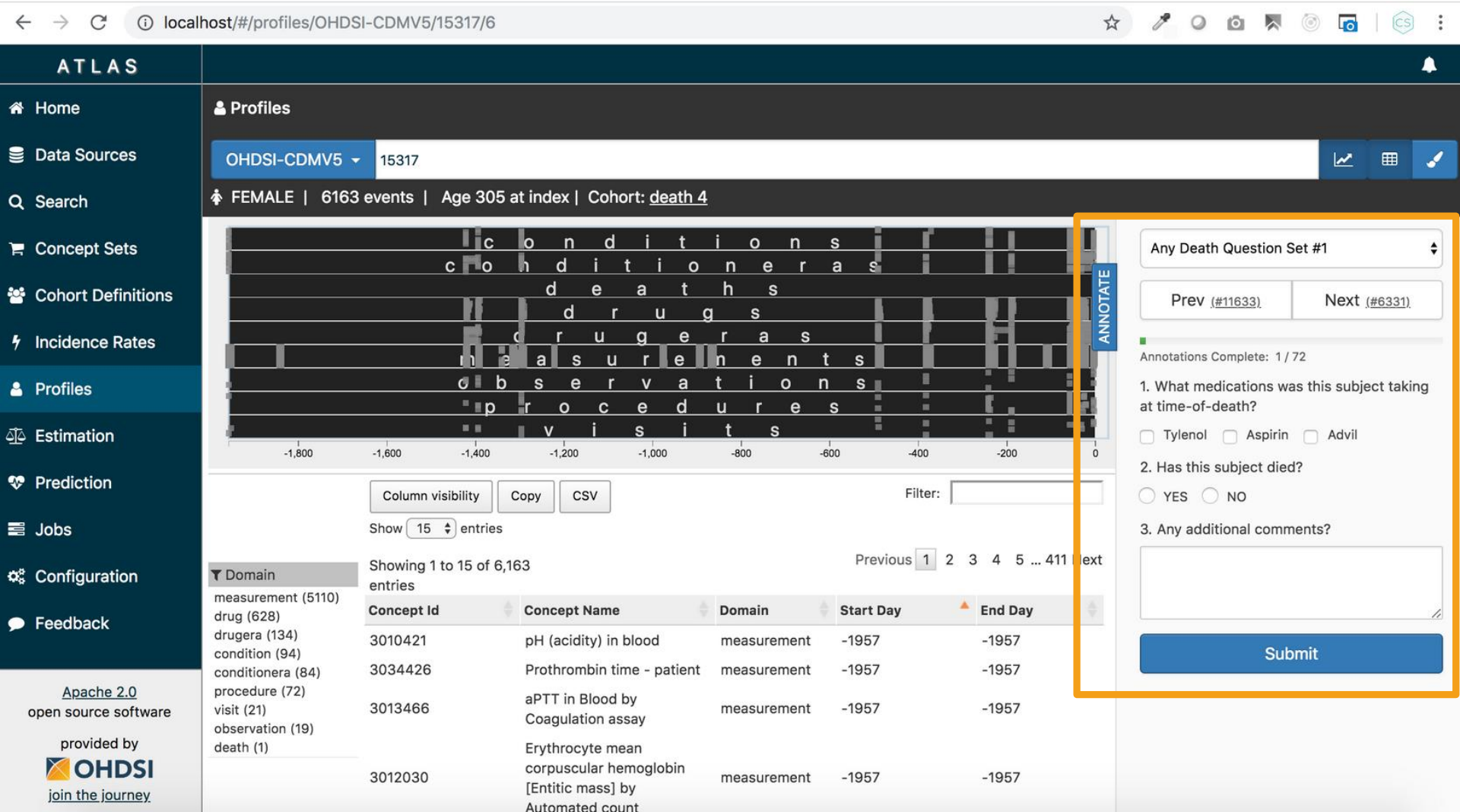
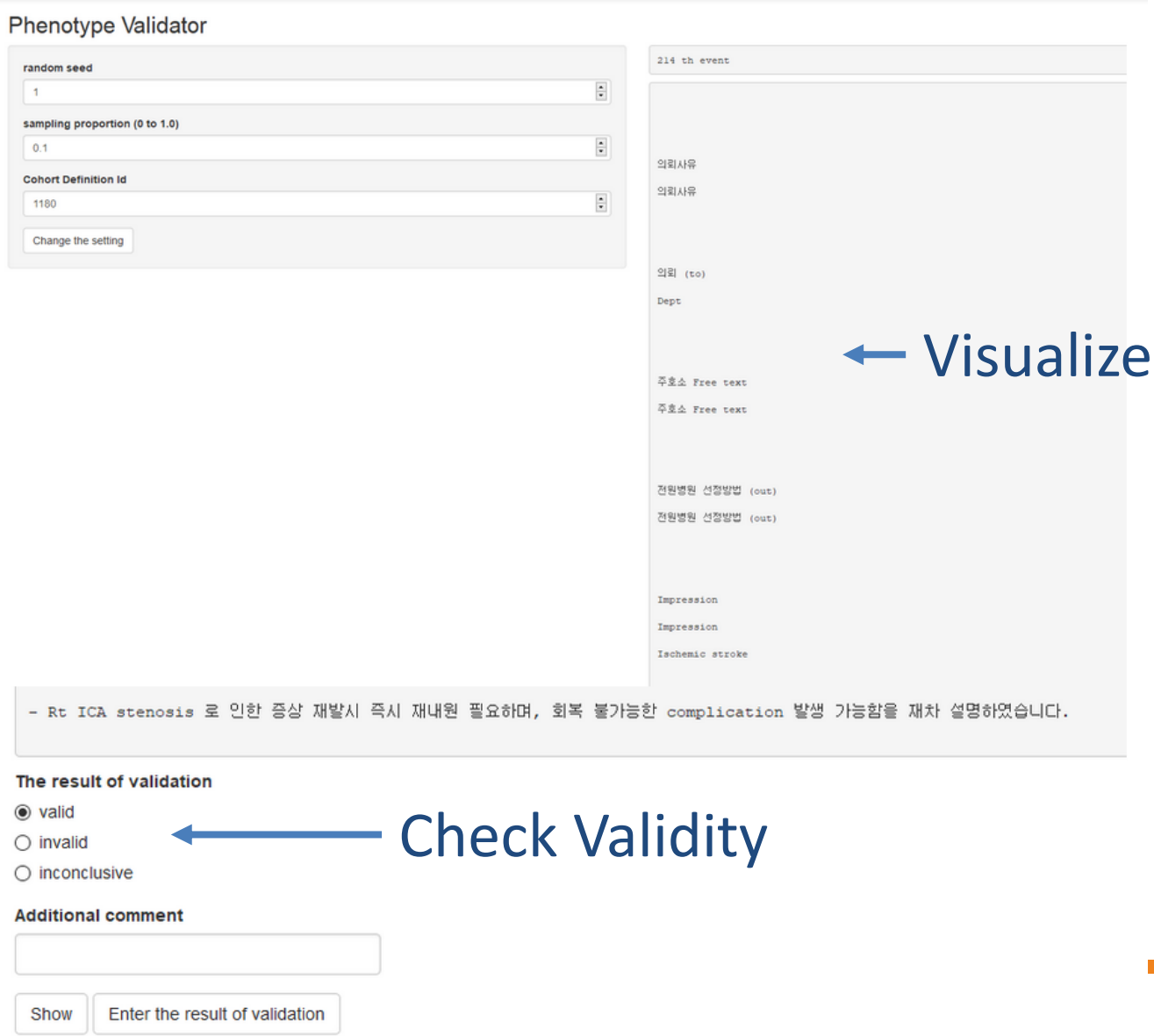


Chart Review / Evaluation

Shiny App – Seng Chan You

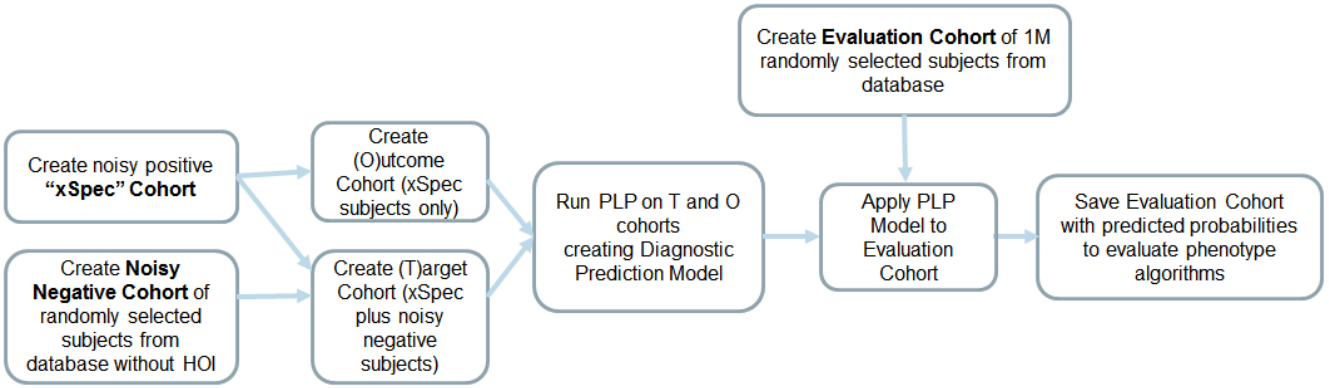


← Visualize Discharge Notes

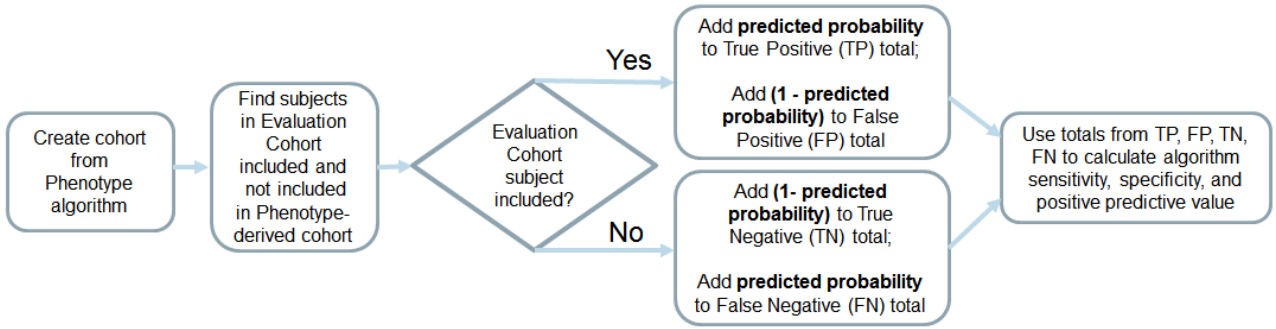
← Check Validity

Automated Evaluation
Phenotype Evaluator
(PheValuator – Joel Swerdel)

Evaluation Cohort from Diagnostic Predictive Model



Evaluate Phenotype Algorithm



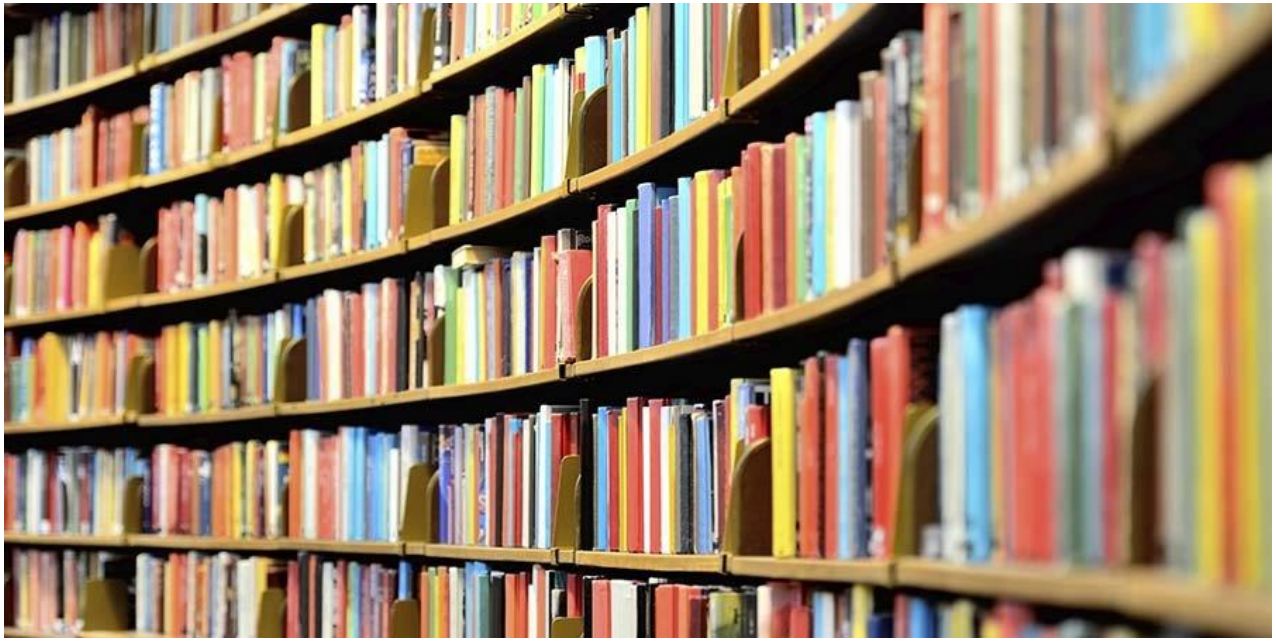
Example: Subject 1
Predicted Probability of HOI = 0.9 and **included** in Phenotype algorithm cohort:
0.9 added to TP ("90% correct case")
1 - 0.9 = 0.1 added to FP ("10% incorrect case")

Example: Subject 2
Predicted Probability of HOI = 0.1 and **not included** in Phenotype algorithm cohort:
1 - 0.1 = 0.9 added to TN ("90% correct non-case")
0.1 added to FN ("10% incorrect non-case")

Calculations:
Sensitivity = Sum TP / (Sum TP + Sum FN)
Specificity = Sum TN / (Sum TN + Sum FP)
PPV = Sum TP / (Sum TP + Sum FP)

Library Architecture / Implementation

- Where does this information reside?
- How does the library become populated?
- Who are the librarians?



What needs to be done?

- For 2019, we're aiming to have a GSPL to debut by the time of the next OHDSI Symposium (September 16th, 2019).

- We need to establish best phenotype design and evaluation practices and agree on an implementation framework for the library.

- This will require iterations of documentation, discussion, peer review, and literature review until we converge to agreement on what the best practices should be.

Current Project Leads

Subtask	Lead
1) Design: Rule-Based (Heuristic) Phenotypes	TBD
2) Design: Computable (Probabilistic) Phenotypes	Juan Banda
3) Evaluation: Manual Chart Review / Annotations	Seng Chan You
4) Evaluation: Automated / Algorithmic	Joel Swerdel
5) Library Architecture and Implementation	TBD

How to get involved?

Thanks to our current list of 11 WG members!

- Aaron Potvien
- Andrew Williams
- Joel Swerdel
- Jon Duke
- Juan Banda
- Kristin Kostka
- Matt Spotnitz
- Sarah Seager
- Seng Chan You
- Trey Schneider
- Vojtech Huser

Forums:
http://forums.ohdsi.org/t/requirements-development-for-the-ohdsi-gold-standard-phenotype-library/4876

Wiki:
http://www.ohdsi.org/web/wiki/doku.php?id=projects:workgroups:gold-library-wg

My e-mail:
Aaron.Potvien@gtri.gatech.edu

Next meeting is 10am ET tomorrow!

Summary

- OHDSI needs a way to **organize and disseminate** its **validated peer-reviewed phenotypes**.
- As a community, we are well-positioned to establish **phenotype design and evaluation best practices**.
- With a Gold Standard Phenotype Library, the **community** would **gain the ability** to **locate high-quality phenotypes** for study and research considerations.
- Those who have such phenotypes will be able to **submit them** to the **library for review**.
- We have a plan, and we're off to a **great start** for 2019, so **please join in!**

Thank you!