

# **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
- Home
- Data Sources
- Search
- Concept Sets
- Cohort Definitions**
- Characterizations
- Cohort Pathways
- Incidence Rates
- Profiles
- Estimation
- Prediction
- Jobs
- Configuration
- Feedback

Cohort definitions

New Cohort

Column visibilityCopyCSVShow15▼entries

Filter:diabetes type 2

Showing 1 to 15 of 42 entries (filtered from 2,841 total entries)

Previous123Next

| Id      | Name   | Created                | Updated                | Author |
|---------|--|------------------------|------------------------|--------|
| 1769510 | <a href="#">Type 2 Diabetes Mellitus Medications</a>                                     | 10/25/2018 3:17:00 AM  | 10/25/2018 3:26:00 AM  |        |
| 1769380 | <a href="#">PheKB-Type 2 Diabetes (T2DM)-OMOP5-(GH+SS)</a>                               | 07/31/2018 2:15:00 PM  | 07/31/2018 2:15:00 PM  |        |
| 1769317 | <a href="#">Type2DiabetesTCAbcd</a>  | 07/12/2018 4:19:00 PM  | 07/12/2018 4:26:00 PM  |        |
| 1769314 | <a href="#">[C2Q]type 2 diabetes</a>   | 07/11/2018 3:27:00 PM  | 07/11/2018 3:28:00 PM  |        |
| 1662611 | <a href="#">PheKB-Type 2 Diabetes (T2DM)-GH</a>  | 12/04/2017 5:29:00 PM  | 06/09/2018 4:29:00 AM  |        |
| 1769058 | <a href="#">[C2Q]Type 2 diabetes mellitus</a>  | 05/25/2018 9:05:00 PM  | 05/25/2018 9:21:00 PM  |        |
| 98644   | <a href="#">[Target] Adults with Diabetes Mellitus Type 2</a>                            | 03/11/2017 8:48:00 PM  | 05/25/2018 12:35:00 PM |        |
| 1769011 | <a href="#">Adults with Diabetes Mellitus Type 2 - by visit - retain all events</a>      | 05/13/2018 12:01:00 PM | 05/13/2018 12:01:00 PM |        |
| 1768981 | <a href="#">[C2Q]type 2 diabetes insulin V1</a>  | 05/04/2018 4:01:00 PM  | 05/04/2018 4:03:00 PM  |        |
| 1768974 | <a href="#">Diabetes Type 2</a>  | 05/04/2018 9:12:00 AM  | 05/04/2018 11:51:00 AM |        |
| 1734996 | <a href="#">Type 2 Diabetes with Neuropathy</a>  | 03/15/2018 10:19:00 PM | 03/15/2018 10:25:00 PM |        |
| 917977  | <a href="#">EMIF demo: patients that have type 2 diabetes to show reporting function</a> | 07/16/2017 6:00:00 PM  | 07/16/2017 6:01:00 PM  |        |
| 917881  | <a href="#">EMIF demo: patients that have type 2 diabetes</a>                            | 07/16/2017 4:59:00 PM  | 07/16/2017 5:56:00 PM  |        |

Last Modified

2+ Weeks Ago (2798)

This Week (28)

Last Week (15)

Author

NULL (2736)

demo (105)

Apache 2.0

open source software

provided by

OHDSI

join the journey



# Atlas is a Sandbox

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

| Id     | Name   |
|--------|--|
| 922762 | <a href="#">Hypertension_test</a>                                |
| 226235 | <a href="#">Hyperglycemia -- Glucose Test &gt; 200 MSHN Test</a> |
| 146    | <a href="#">Hypertension Test</a>                                |
| 922763 | <a href="#">Hypertension_test2</a>                               |
| 922764 | <a href="#">Hypertension_test3</a>                               |
| 922765 | <a href="#">Hypertension_test3</a>                               |
| 922783 | <a href="#">Hypertension_test4</a>                               |
| 922784 | <a href="#">Hypertension_test4</a>                               |
| 922785 | <a href="#">Hypertension_test4</a>                               |



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



# Versioning is Desired

## 3) The community wants different phenotype versions.

| Id      | Name   |
|---------|--|
| 422     | <a href="#">eMERGE PheKB Type 2 Diabetes phenotype algorithm (Northwestern University)</a> Literal |
| 341     | <a href="#">eMERGE PheKB Type 2 Diabetes phenotype algorithm (Northwestern University)</a> oldV2   |
| 336     | <a href="#">eMERGE PheKB Type 2 Diabetes phenotype algorithm (Northwestern University)</a> oldV1   |
| 343     | <a href="#">eMERGE PheKB Type 2 Diabetes phenotype algorithm (Northwestern University)</a> oldV3   |
| 1198483 | <a href="#">beva_ch_mbc_v01</a>  |
| 112062  | <a href="#">Thia_Cohort_v2</a>   |
| 112061  | <a href="#">Sulf_Cohort_v2</a>   |



# Tagging is Desired

## 4) The community wants different ways to tag phenotypes.

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



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

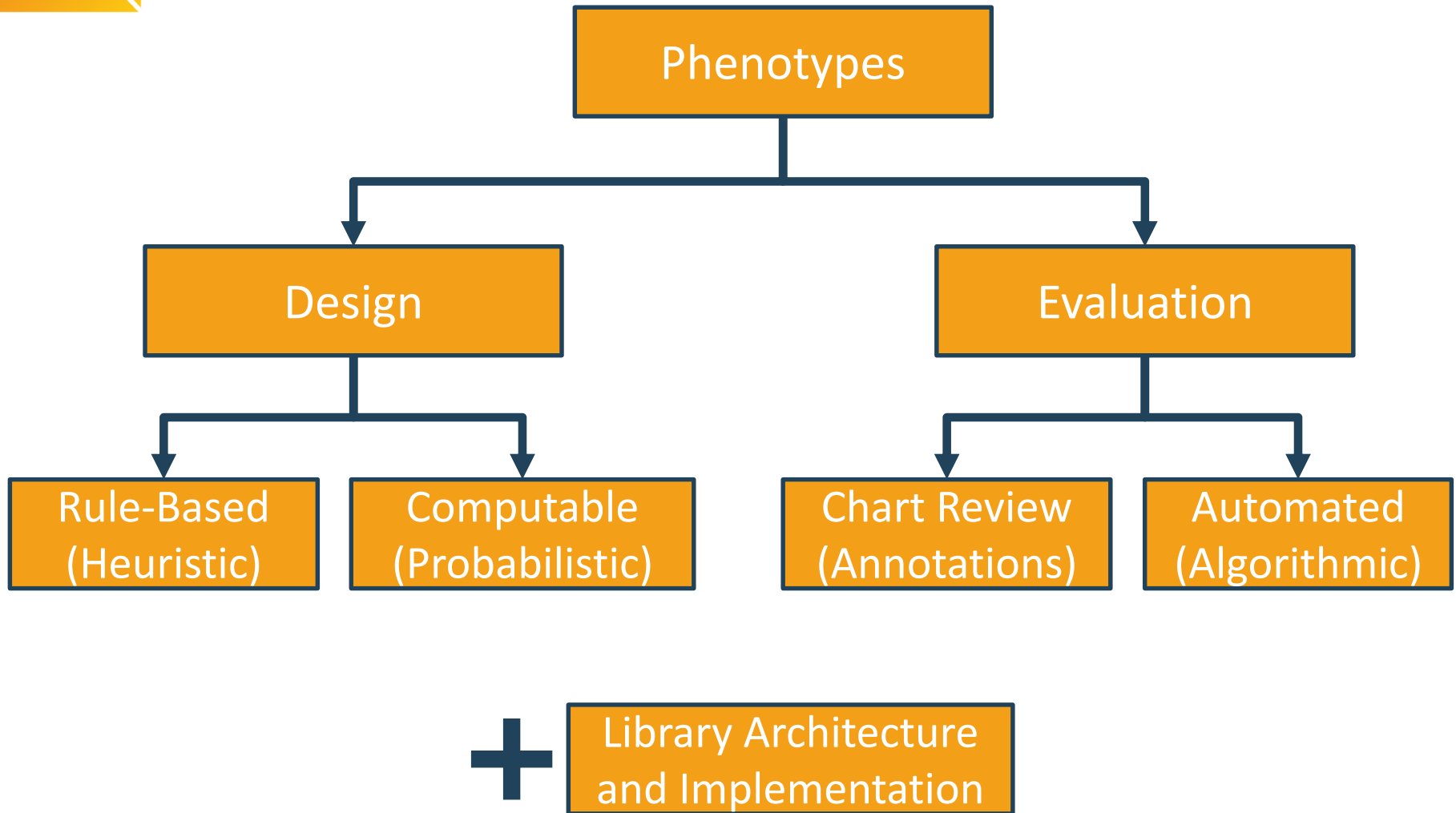


# What Makes a Phenotype a “Gold Standard”?

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



# Heuristic and Computational





# Heuristic Phenotype Design

[PHEKB] Type 2 Diabetes

Definition ?Concept SetsGenerationReportingExportMessages 3

This definition is based on the PheKB Diabetes definition found at <https://phekb.org/phenotype/type-2-diabetes-mellitus>

Cohort Entry Events?

Events having any of the following criteria:

+ Add Initial Event

a condition occurrence of [PHEKB] T2DM

+ Add attribute...

Delete Criteria

a drug exposure of [PHEKB] T2DM Medications

+ Add attribute...

Delete Criteria

a measurement of [PHEKB] HBA1c

+ Add attribute...

Delete Criteria

✗ with value as number Greater Than 6

✗ Unit is: ✗ percent

AddImport

a measurement of [PHEKB] Lab: Random Glucose

+ Add attribute...

Delete Criteria

✗ with an abnormal result

a measurement of [PHEKB] Lab: Fasting Glucos...

+ Add attribute...

Delete Criteria

✗ with an abnormal result

a specimen of [PHEKB] HBA1c

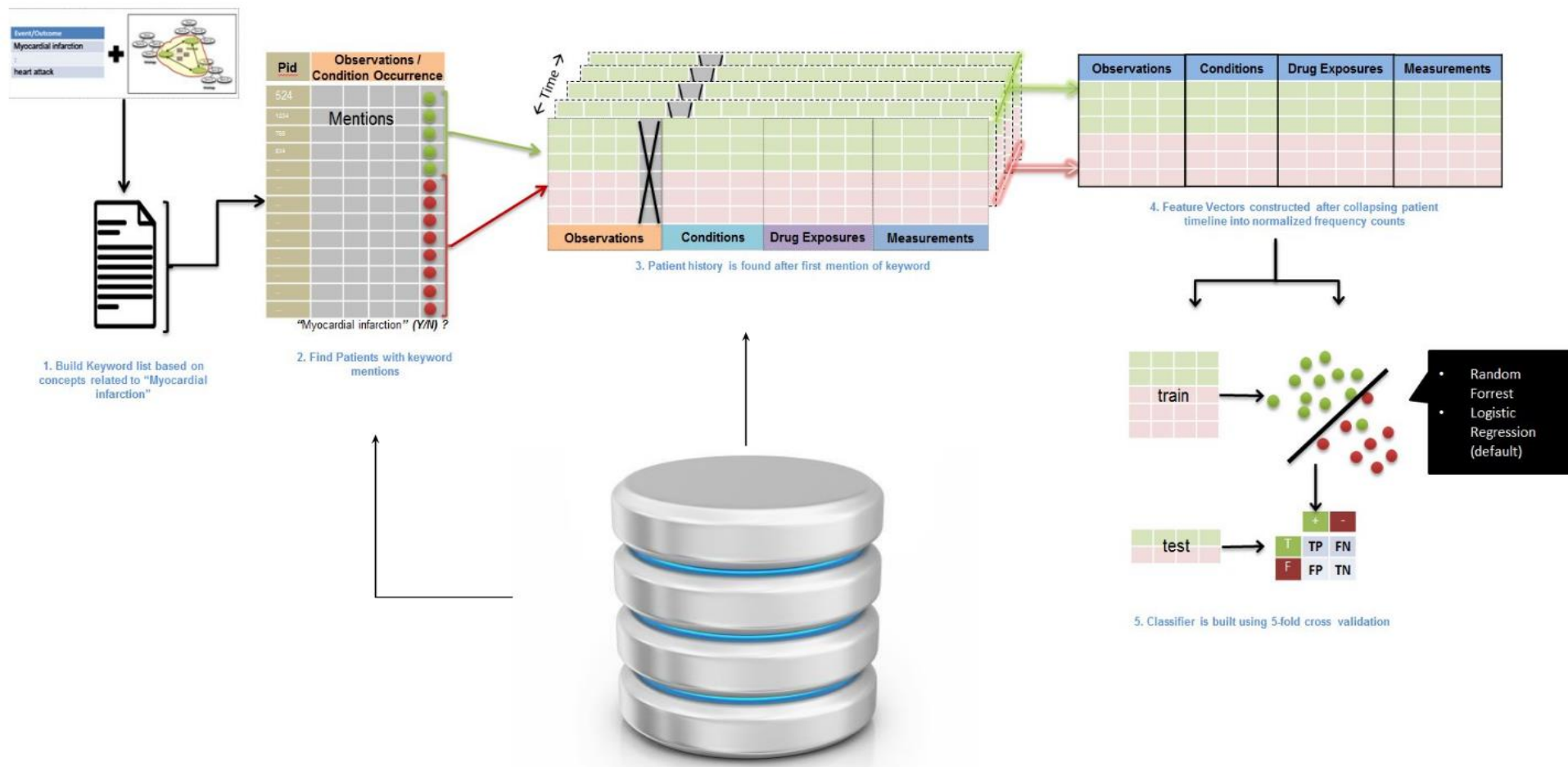
+ Add attribute...

Delete Criteria



# Computable Phenotype Design

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





# Chart Review / Evaluation

- Annotation Tool – Trey Schneider

← → ↺ localhost/#/profiles/OHDSI-CDMV5/15317/6

ATLAS

- Home
- Data Sources
- Search
- Concept Sets
- Cohort Definitions
- Incidence Rates
- Profiles
- Estimation
- Prediction
- Jobs
- Configuration
- Feedback

Profiles

OHDSI-CDMV5 15317

FEMALE | 6163 events | Age 305 at index | Cohort: death\_4

Annotations

Annotations Complete: 1 / 72

1. What medications was this subject taking at time-of-death?

☐ Tylenol ☐ Aspirin ☐ Advil

2. Has this subject died?

☐ YES ☐ NO

3. Any additional comments?

Submit

Column visibility Copy CSV Filter: Show 15 entries

Showing 1 to 15 of 6,163 entries

| Concept Id | Concept Name  | Domain      | Start Day | End Day |
|------------|---|-------------|-----------|---------|
| 3010421    | pH (acidity) in blood   | measurement | -1957     | -1957   |
| 3034426    | Prothrombin time - patient  | measurement | -1957     | -1957   |
| 3013466    | aPTT in Blood by Coagulation assay  | measurement | -1957     | -1957   |
| 3012030    | Erythrocyte mean corpuscular hemoglobin [Entitic mass] by Automated count | measurement | -1957     | -1957   |

Domain

- measurement (5110)
- drug (628)
- druga (134)
- condition (94)
- conditionera (84)
- procedure (72)
- visit (21)
- observation (19)
- death (1)

Apache 2.0 open source software provided by OHDSI join the journey.



# Chart Review / Evaluation

## Shiny App – Seng Chan You

### Phenotype Validator

random seed

sampling proportion (0 to 1.0)

Cohort Definition Id

Change the setting

214 th event

의뢰사유

의뢰사유

의뢰 (to)

Dept

주소 Free text

주소 Free text

진행병원 선정방법 (out)

진행병원 선정방법 (out)

Impression

Impression

Ischemic stroke

- Rt ICA stenosis 로 인한 증상 재발시 즉시 재내원 필요하며, 회복 불가능한 complication 발생 가능성을 재차 설명하였습니다.

← Visualize Discharge Notes

### The result of validation

- ☒ valid
- ☐ invalid
- ☐ inconclusive

← Check Validity

### Additional comment

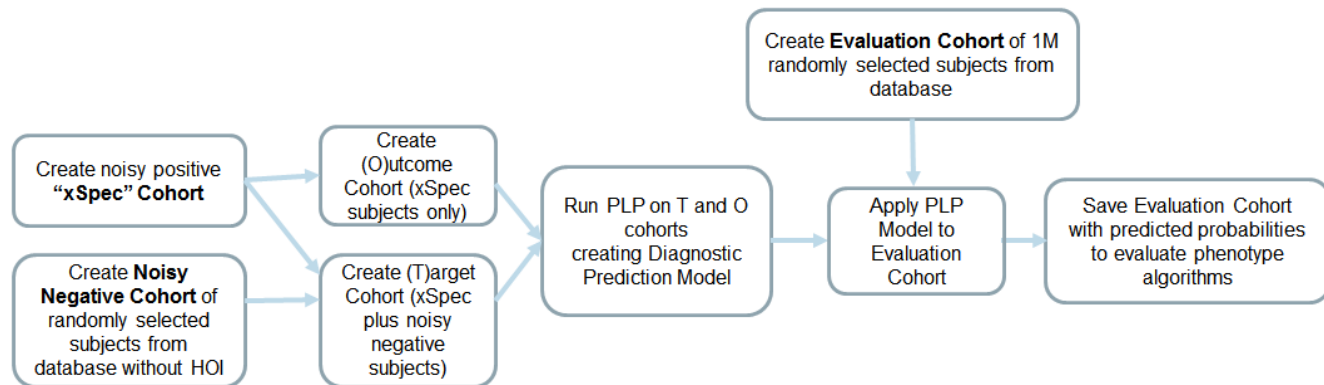
Show

Enter the result of validation

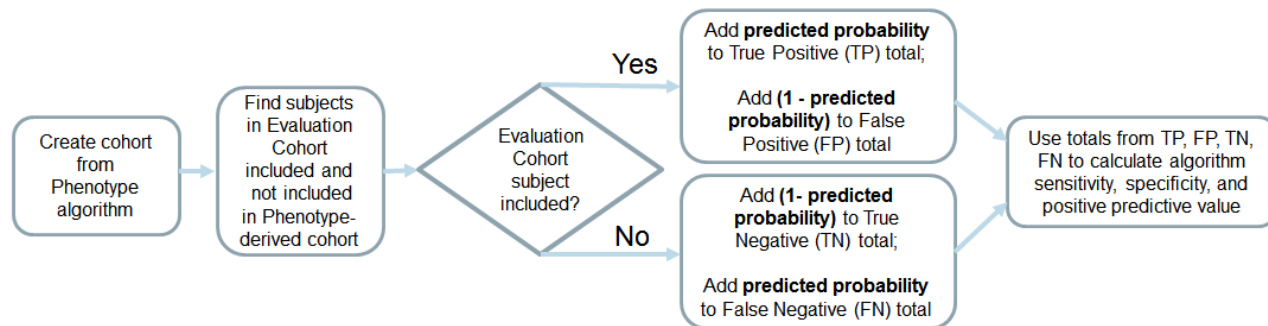


# Automated Evaluation Phenotype Evaluator (PheValuator – Joel Swerdel)

## 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 =  $\text{Sum TP} / (\text{Sum TP} + \text{Sum FN})$   
Specificity =  $\text{Sum TN} / (\text{Sum TN} + \text{Sum FP})$   
PPV =  $\text{Sum TP} / (\text{Sum TP} + \text{Sum FP})$



# 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 16<sup>th</sup>, 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](mailto: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!

