

Apps To Enable AI In EHR Research



Butte Lab

Ben Glicksberg, PhD

Butte Lab

Bakar Computational Health Sciences Institute

University of California, San Francisco (UCSF)

UCSF Bakar Computational Health
Sciences Institute

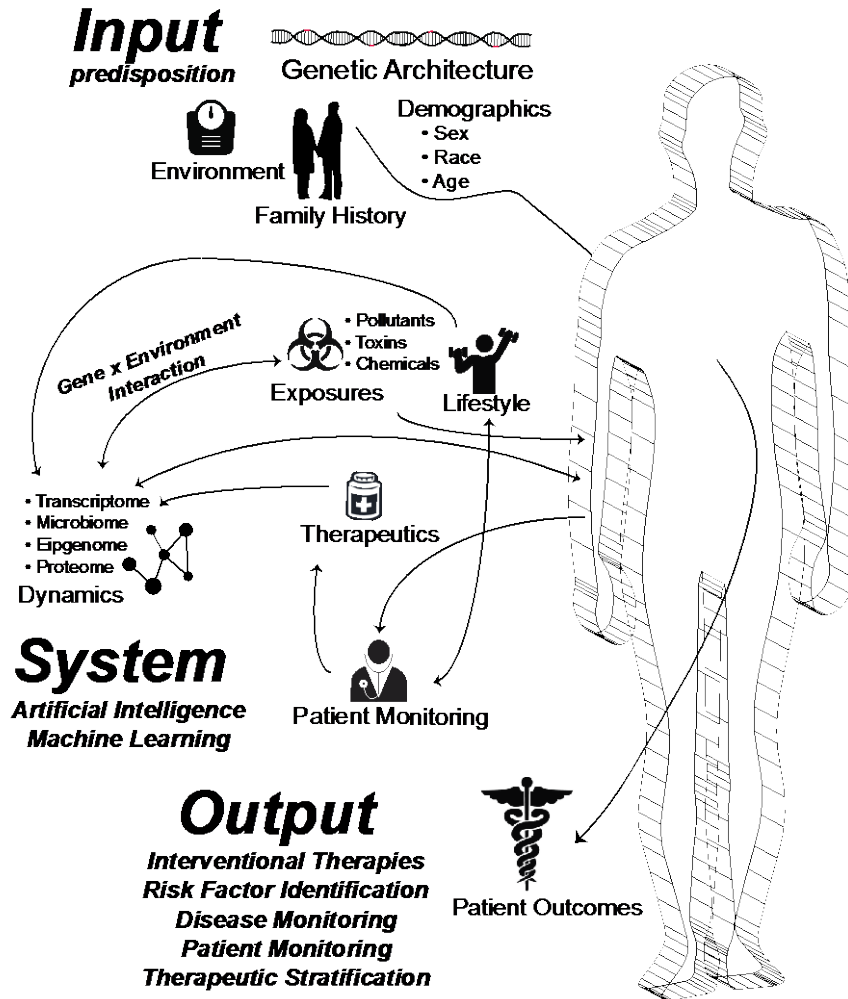
UCSF

University of California
San Francisco

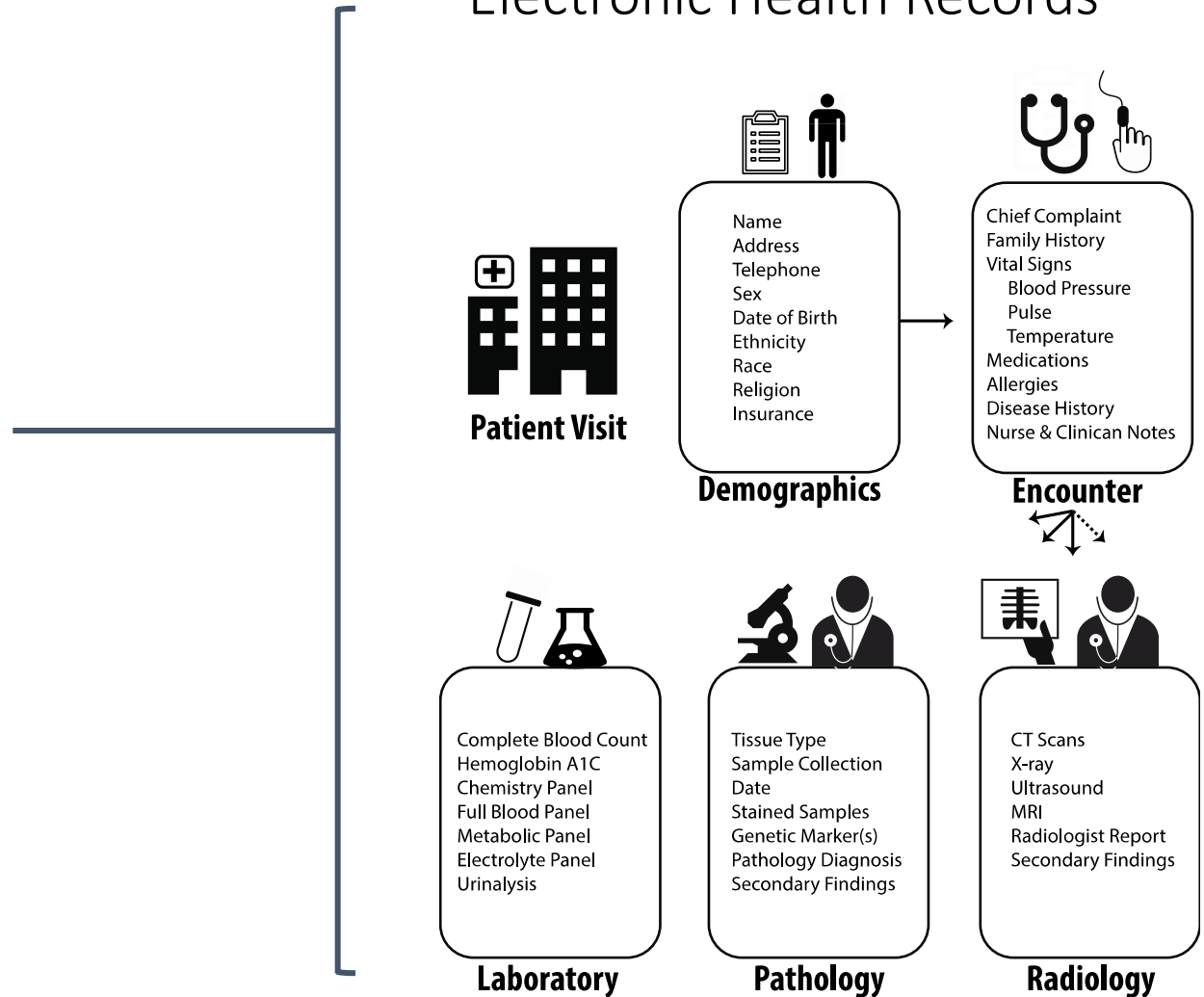
 @BenGlicksberg

Clinical Informatics in the era of big data

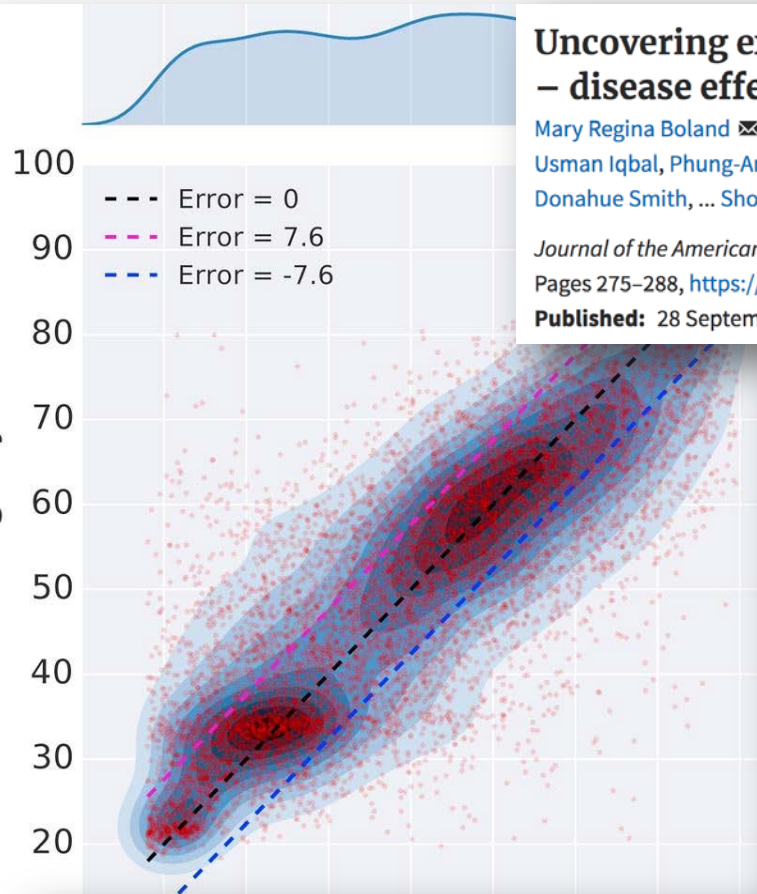
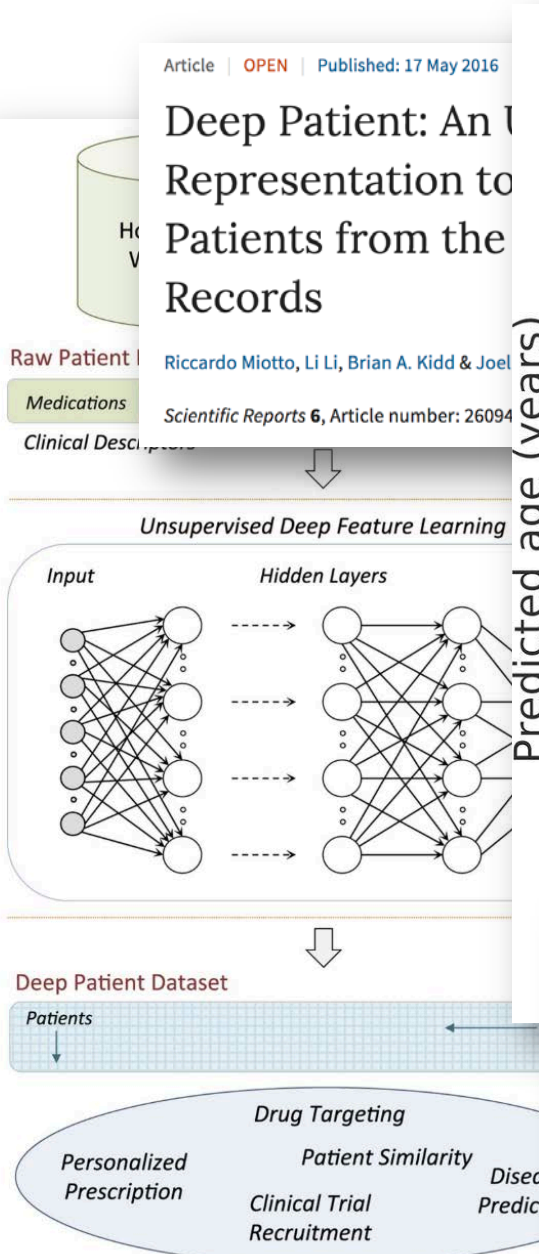
“The Quantified Self”



Electronic Health Records

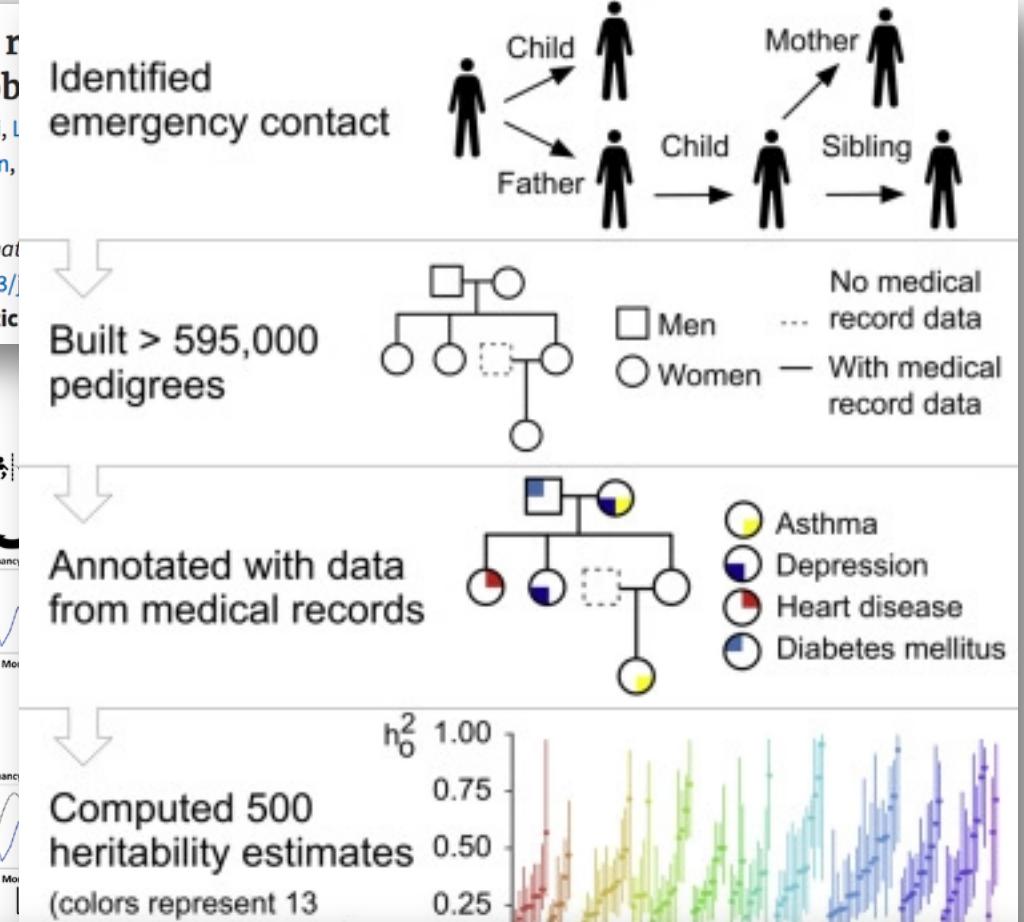


The power and diversity of EHR studies



Predicting age by mining electronic medical records: unsupervised deep learning characterizes differences between chronological and physiological age

Zichen Wang^a, Li Li^b, Benjamin S. Glicksberg^b, Ariel Israel^c, Joel T. Dudley^b, ... Show more
<https://doi.org/10.1016/j.jbi.2017.11.003>
 Under an Elsevier user license



Disease Heritability Inferred from Familial Relationships Reported in Medical Records

Fernanda C.G. Polubriaginof¹⁶ • Rami Vanguri¹⁶ • Kayla Quinnes¹⁶ • ... Joel Dudley • David K. Vawdrey¹⁷ • Nicholas P. Tatonetti^{17, 18} • Show all authors • Show footnotes

Published: May 17, 2018 • DOI: <https://doi.org/10.1016/j.cell.2018.04.032> • Check for updates


Towards a learning health system

nature
medicine

Comment | Published: 07 January 2019

A call for deep-learning healthcare

Beau Norgeot, Benjamin S. Glicksberg & Atul J. Butte 

Nature Medicine **25**, 14–15 (2019) | [Download Citation](#) 

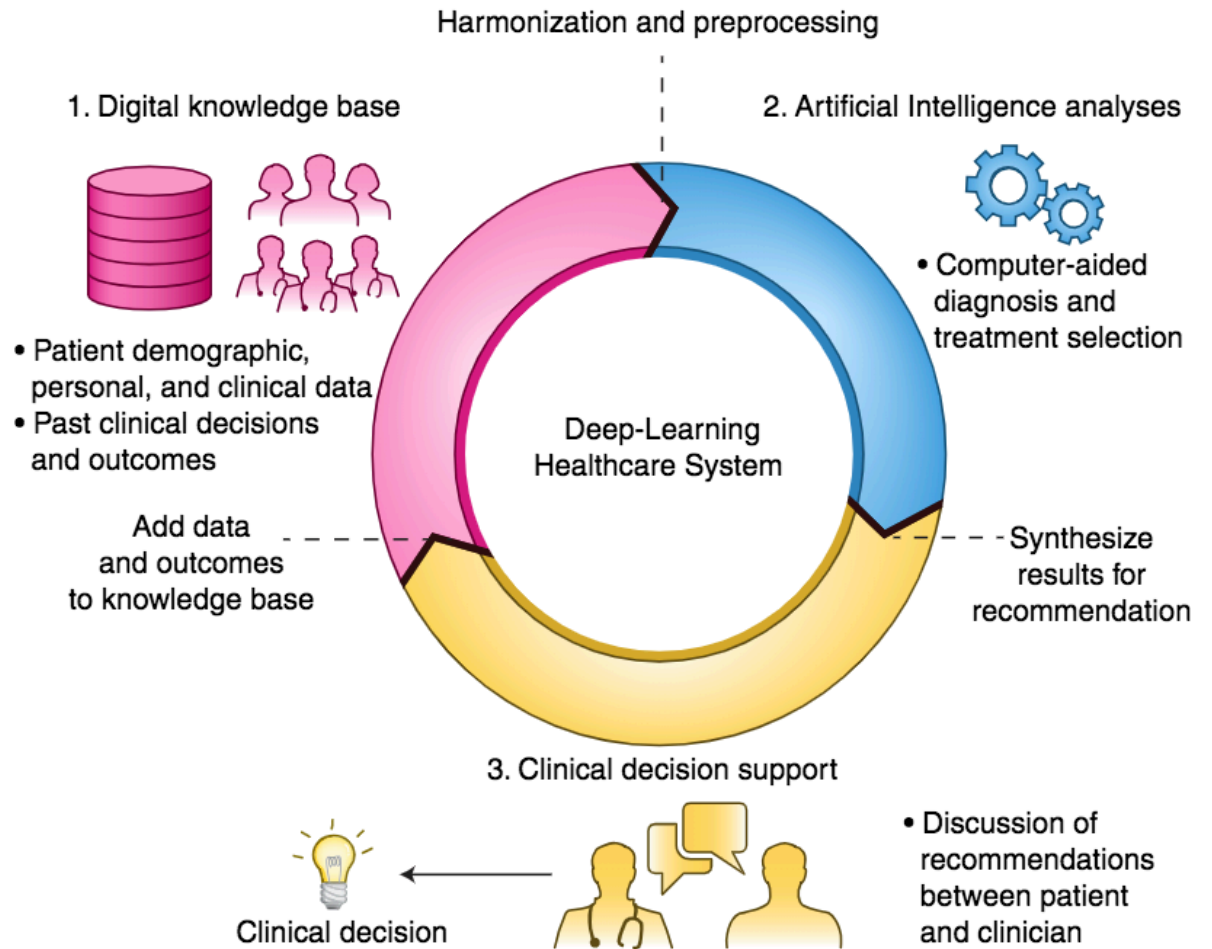


Fig. 1 | A deep-learning healthcare system. A schematic representation of a deep-learning healthcare system is shown.

Challenges of using EHR data for research

- EHRs are challenging to represent health state
 - heterogeneous
 - noisy
 - incomplete
 - structured / unstructured
 - redundant
 - subject to random errors
 - subject to systematic errors
 - ...*and so and so forth*

EHR barriers to entry

- Computational

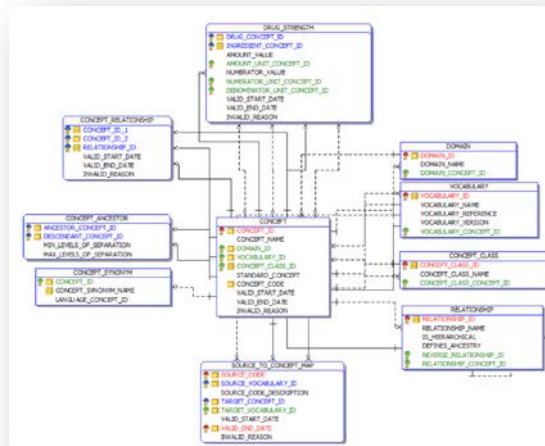
```
mysql> SELECT B.* FROM PATIENTS A INNER JOIN DIAGNOSES B ON A.Patient_ID = B.Patient_ID WHERE B.ICD10_Code = "I10" LIMIT 5;
```

Diagnosis_Start_Date	Diagnosis_Key	ICD9_Code	ICD10_Code	Diagnosis_Event_Key	Diagnosis_End_Date	Diagnosis_Hospital_Diagnosis	Diagnosis_Emergency_Department_Diagnosis	Diagnosis_Chronic	Diagnosis_Event_Type	Diagnosis_Name	Diagnosis_ID
NULL	142335	938025	481.9	I10	884819414932281	NULL	NULL	No	No	Hypertensive diseases (I10-I15)	Essential (primary) hypertension
NULL	142335	938025	481.9	I10	491569671779871	NULL	NULL	No	No	Hypertensive diseases (I10-I15)	Essential (primary) hypertension
NULL	142335	938025	481.9	I10	544386856588079	NULL	NULL	No	No	Hypertensive diseases (I10-I15)	Essential (primary) hypertension
NULL	142335	938025	481.9	I10	99908489035994	NULL	NULL	No	No	Hypertensive diseases (I10-I15)	Essential (primary) hypertension
NULL	142335	938025	481.9	I10	265244842588854	NULL	NULL	No	No	Hypertensive diseases (I10-I15)	Essential (primary) hypertension

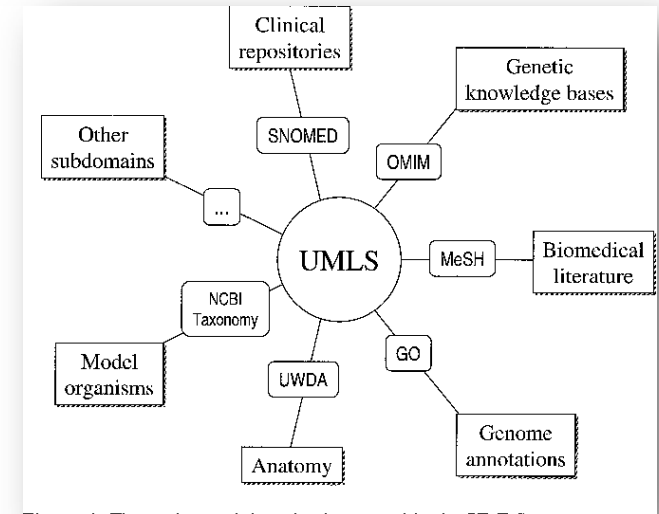
5 rows in set (0.03 sec)

- Domain knowledge:

- Structure

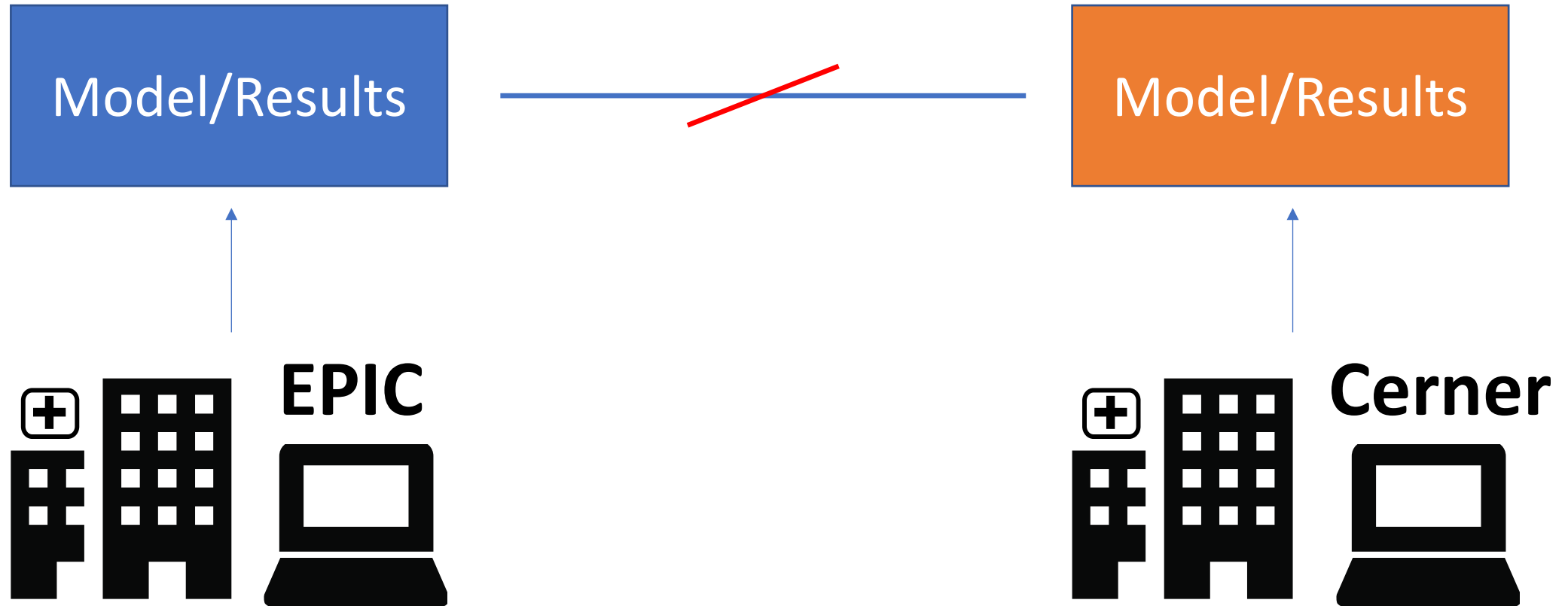


- Language



Bodenreider, O (2004): Medical Language System (UMLS) : integrating biomedical terminology

Cross-validation & replication in EHR research



OMOP common data model (CDM)



OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Language

Structure

Resources:

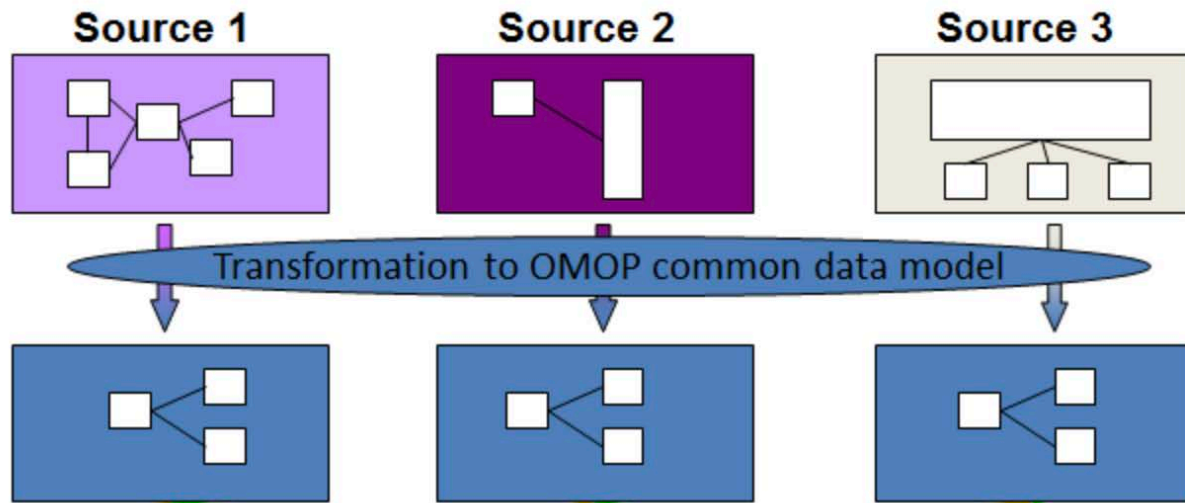
<https://www.ohdsi.org/>

<http://www.ohdsi.org/web/wiki/doku.php>

<http://forums.ohdsi.org/>

<https://github.com/OHDSI/>

(most documentation)



Analysis

Analysis
method

Analysis
results

CDM facilitates cross-validation and reproducibility

Scalable and accurate deep learning with electronic health records

Alvin Rajkomar, Eyal Oren, Jeffrey Dean

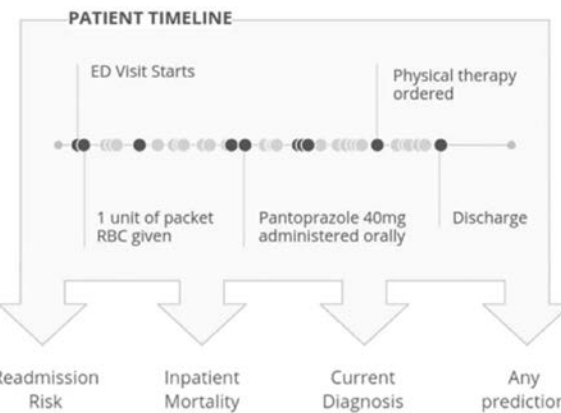
npj Digital Medicine 1, Article number: 18 (2018) | Download Citation

Systems collect and store health records in various formats in databases.

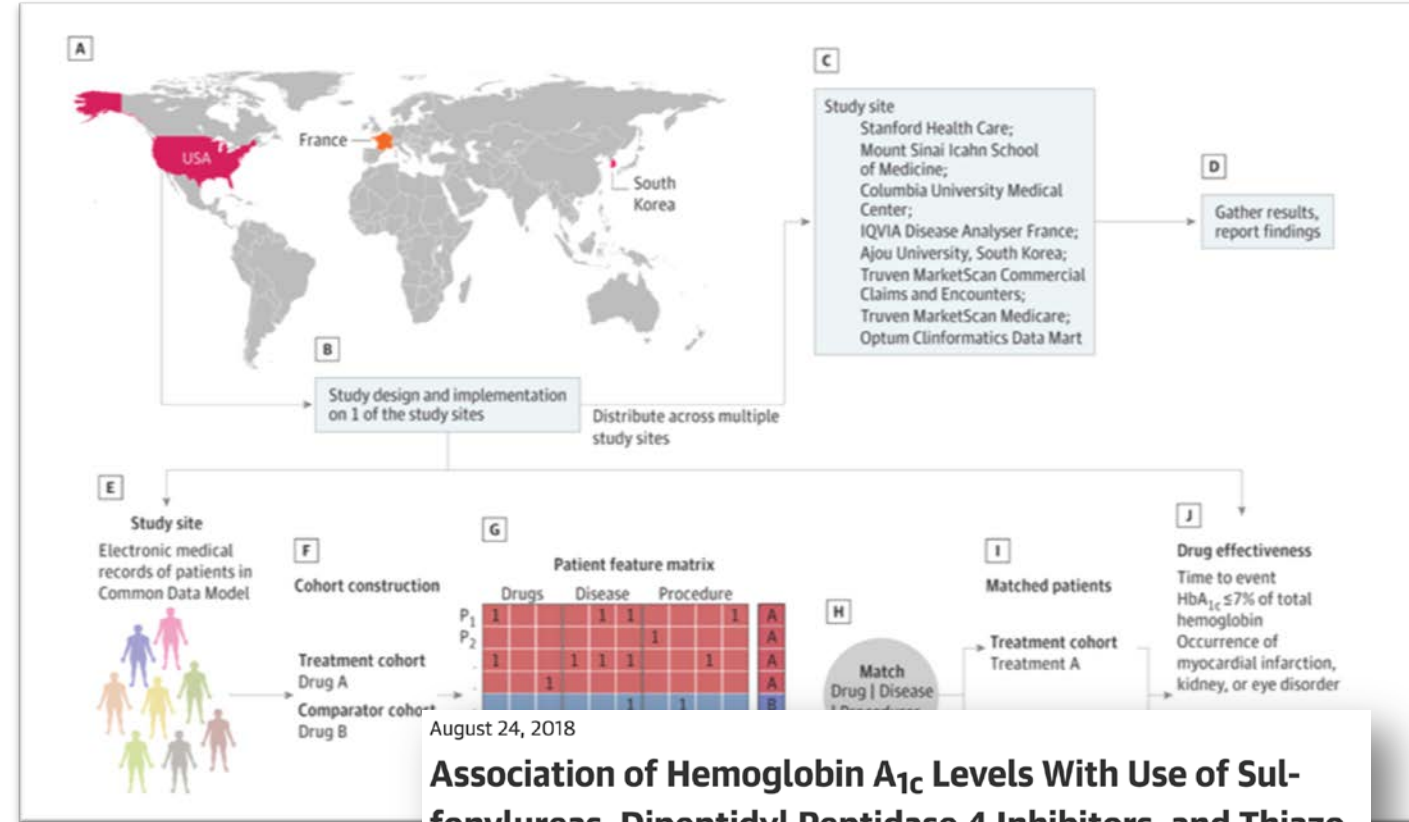
FHIR



2 All available data for each patient is converted to events recorded in containers based on the Fast Healthcare Interoperability Resource (FHIR) specification.



3 The FHIR resources are placed in temporal order, depicting all events recorded in the EHR (i.e. timeline). The deep learning model uses this full history to make each prediction.



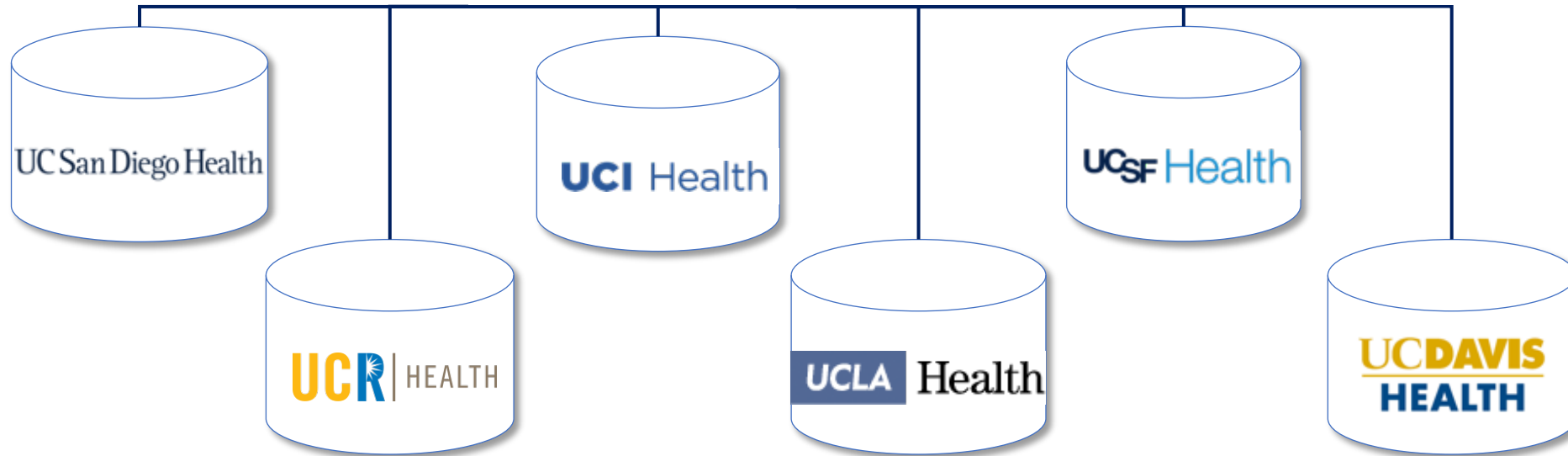
OMOP

Association of Hemoglobin A_{1c} Levels With Use of Sulfonylureas, Dipeptidyl Peptidase 4 Inhibitors, and Thiazolidinediones in Patients With Type 2 Diabetes Treated With Metformin

Analysis From the Observational Health Data Sciences and Informatics Initiative

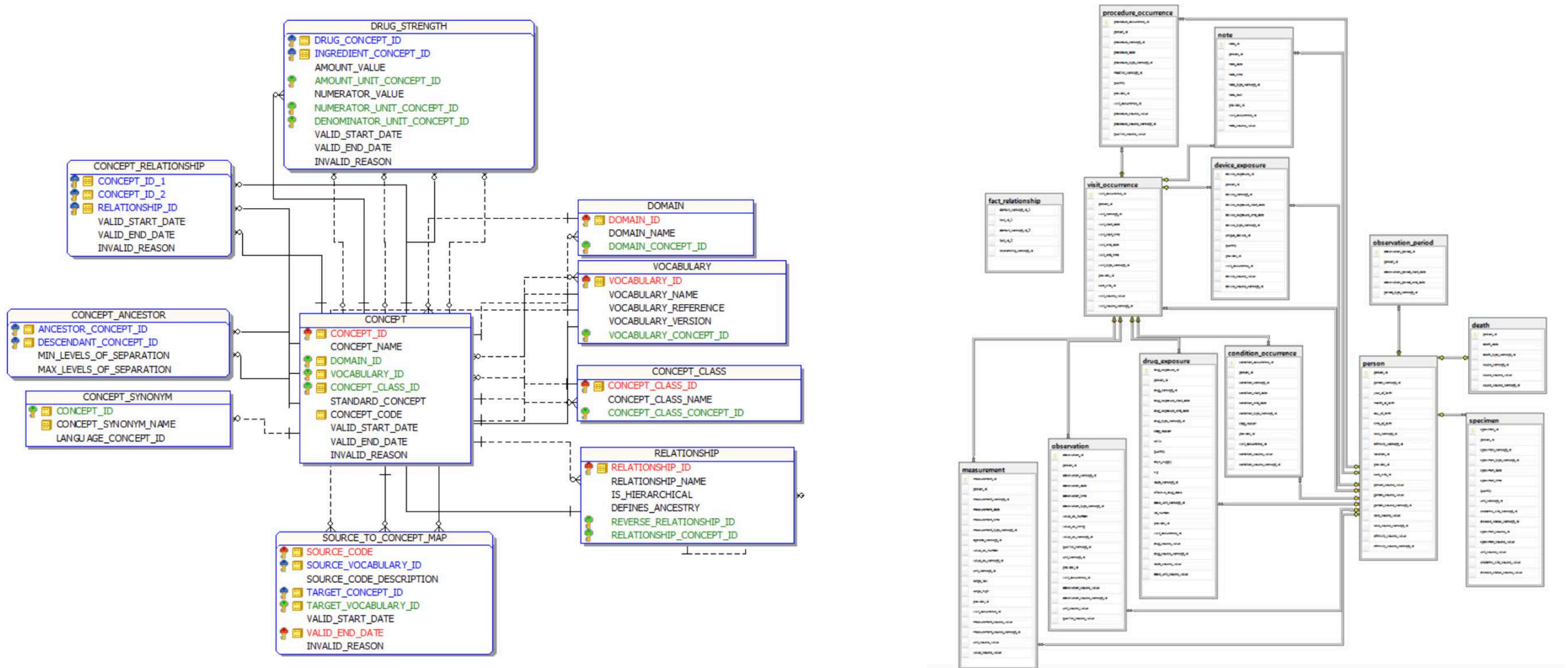
Rohit Vashisht, PhD^{1,2}; Kenneth Jung, PhD^{1,2}; Alejandro Schuler, MS^{1,2}; et al
 > Author Affiliations | Article Information
 JAMA Netw Open. 2018;1(4):e181755. doi:10.1001/jamanetworkopen.2018.1755

OMOP CDM across the UC system



The OMOP system is efficient but complicated

- OMOP still requires extensive domain and computational expertise



OHDSI has developed powerful, advanced tools



Observational Health

<http://ohdsi.org>

Repositories 133

People 5

Projects 0

Find a repository...

Type: All

Language: All

PatientLevelPrediction

An R package for performing patient level prediction.

R ★ 48 🍷 31 Updated 2 days ago

BrokenAdaptiveRidge

R ★ 1 🍷 2 Updated 2 days ago

Open-Source Software

Observational Data Management – tools and processes to standardize the structure and content of healthcare data in preparation for observational analyses, including:

- [ATHENA standardized vocabularies](#)
- Common data model and standardized [vocabularies specifications](#)
- [Extract, transform, and load](#) (ETL) design, development, and testing
- Database profiling and [data quality assessment](#)

Clinical Characterization – descriptive analyses to support disease natural history and quality improvement, including:

- Cohort definition and phenotype evaluation
- Patient record profiling
- Study feasibility assessment
- Population summarization and comparison

Population-Level Estimation – epidemiologic designs for estimating average treatment effects for medical product safety surveillance and comparative effectiveness, including:

- Comparative cohort analysis
- Self-controlled case series
- Self-controlled cohort

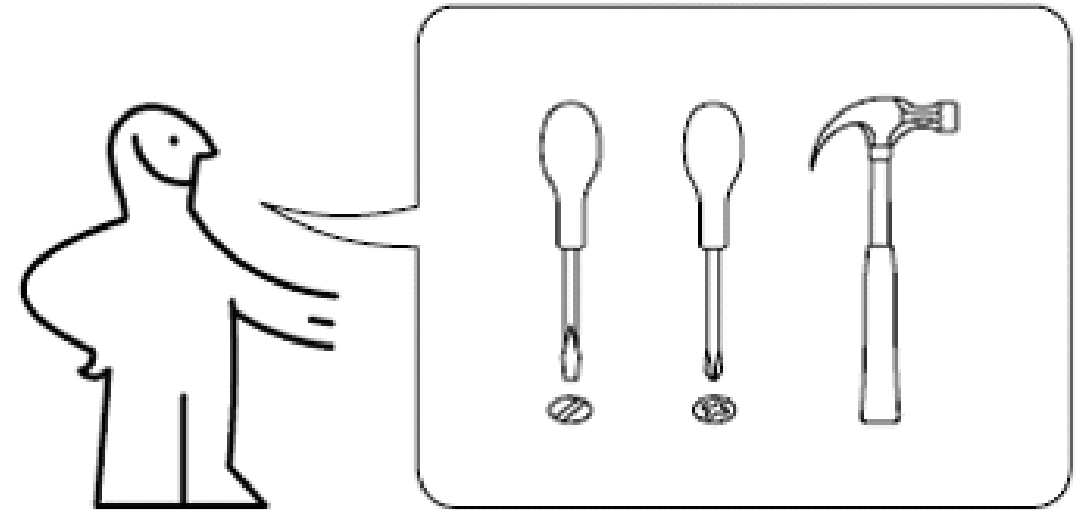
Patient-level prediction – machine learning methods for precision medicine and disease interception, including:

- Regularized regression
- Random forest
- k-nearest neighbors

<https://github.com/OHDSI>

<https://www.ohdsi.org/analytic-tools/>

...that are sometimes *too* advanced for most tasks



<http://remembar.me/wp-content/uploads/2018/07/garage-pegboard-organization-interior-furniture-full-image-for-tool-storage-special-tools-and-ideas.jpg>

https://www.ikea.com/ms/en_CA/customer_service/assembly_instructions/assembly_instructions1.html

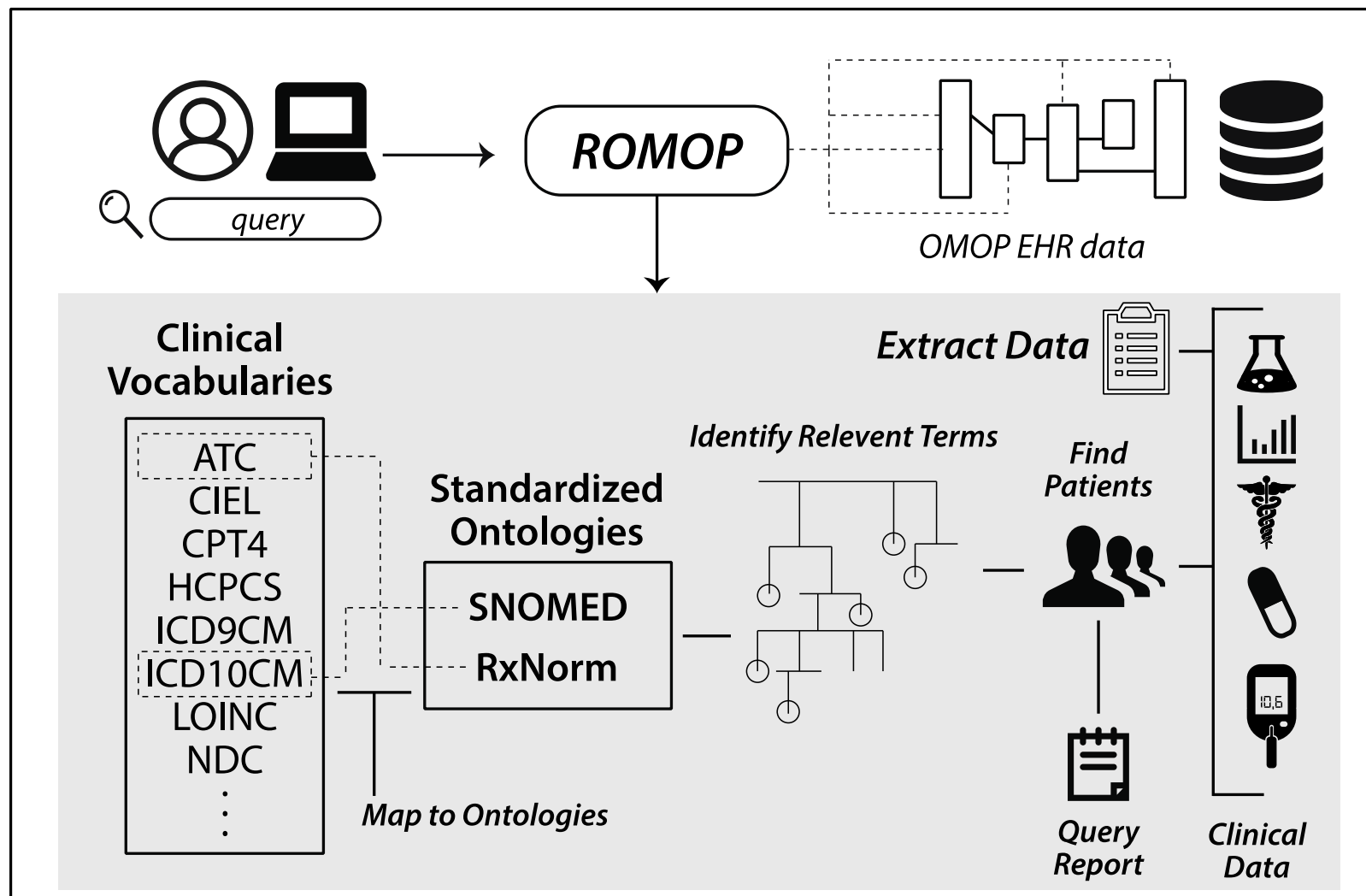
ROMOP

a light-weight R package for interfacing with
OMOP-formatted Electronic Health Record data

Glicksberg et al. *JAMIA Open* (ooy059)

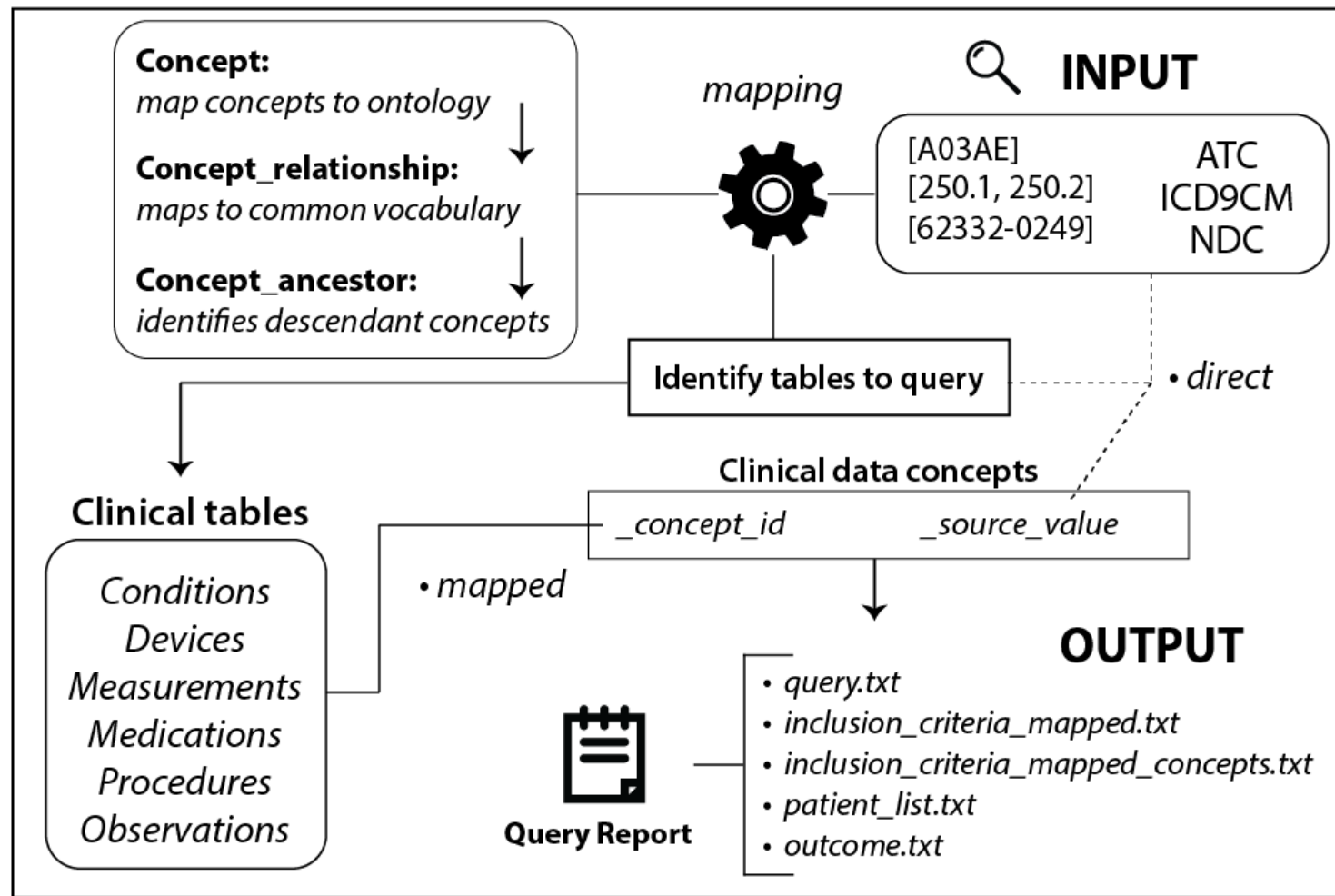
Goals of ROMOP

1. Automatically connect to OMOP EHR relational database
2. Enable non-technical experts to easily pull data into R-object
3. Facilitate follow-up analyses



What can ROMOP do?

1. Explore CDM fields
2. Generate population statistics
3. Search for patients:
 - Any vocabulary
 - Inclusion/Exclusion criteria
 - Flexible search strategies (e.g., and vs. or)
4. Retrieve all relevant data for patients:
 - Demographics
 - Encounters
 - Clinical
5. Automatically map concepts to ontologies
6. Export search report



Public sandbox server: interactive tutorial

<http://romop.ucsf.edu>

- 1MM patients from CMS synthesized clinical dataset (DE-SymPUF)

- Package:

<https://github.com/BenGlicksberg/ROMOP>

ROMOP Sandbox Tutorial

Benjamin S. Glicksberg
Butte Lab
Bakar Computational Health Sciences Institute
University of California, San Francisco
2018

ROMOP
Initialization
Data exploration
Finding cohort/patients
Extracting clinical data
Start Over

ROMOP

ROMOP is a flexible, light-weight R package for interfacing with Electronic Health Record (EHR) data in the [Observational Health Data Sciences and Informatics \(OHDSI\) OMOP Common Data Model](#). This sandbox server is set up for individuals without access to OMOP-formatted EHR data. This resource will also provide an interactive tutorial.

- For a detailed description of the OMOP common data model, please visit this [helpful wiki](#).

Project Information

- For the open-source package, visit <https://github.com/BenGlicksberg/ROMOP>.
- We provide detailed documentation in the [Readme file](#).
- For the manuscript, please click here.

Data and Server Information

The Centers for Medicare and Medicaid Services (CMS) have released a synthetic clinical dataset ([DE-SymPUF](#)) in the public domain with the aim of being reflective of the patient population but containing no protected health information. The OHDSI group has undertaken the task of converting these data into the [OMOP CDM format](#). Users are certainly able to set up this configuration on their own system following the instructions on the GitHub page. We obtained all data files from the [OHDSI FTP server](#) (accessed June 17th, 2018) and created the CDM (DDL and indexes) according to their [official instructions](#), but modified for MySQL. For space considerations, we only uploaded one million rows of each of the data files. The sandbox server is a Rshiny server running as an Elastic Compute Cloud (EC2) instance on Amazon Web Services (AWS) querying a MySQL database server (AWS Aurora MySQL).

Who We Are

- [Butte Lab](#)
- [Bakar Computational Health Sciences Institute \(BCHSI\)](#)
- [University of California, San Francisco \(UCSF\)](#)

Contact

For questions, comments, errors, bug reports, or issues, please contact: benjamin.glicksberg@ucsf.edu
For general correspondence, please contact: atul.butte@ucsf.edu

Next Topic

Data and CDM exploration

ROMOP Sandbox Tutorial

Benjamin S. Glicksberg

Butte Lab

Bakar Computational Health Sciences Institute

University of California, San Francisco

2018

ROMOP

Initialization

Data exploration

Finding cohort/patients

Extracting clinical data

Start Over

Data exploration

- ✓ Explore data types in the data ontology

For those unfamiliar with OMOP structure, this function details relevant vocabularies per clinical domain: Condition, Observation, Measurement, Device, Procedure, Drug.

Show data types:

```
Code Start Over Run Code  
1 showDataTypes()  
2  
3
```

domain_id	vocabulary_id
<chr>	<chr>
Condition	ICD10CM
Condition	SNOMED
Condition	ICD9CM
Device	SNOMED
Device	HCPCS
Device	NDC
Device	SPL
Drug	NDFRT
Drug	RxNorm
Drug	SNOMED

1-10 of 35 rows [Previous](#) [1](#) [2](#) [3](#) [4](#) [Next](#)

Define cohorts/Find patients

ROMOP Sandbox Tutorial

Benjamin S. Glicksberg

Butte Lab

Bakar Computational Health Sciences Institute

University of California, San Francisco

2018

ROMOP

Initialization

Data exploration

Finding cohort/patients

Extracting clinical data

Start Over

Finding cohort/patients

ROMOP has a straight-forward yet flexible ways to search for patients that takes advantage of the underlying OMOP CDM structure. If the "mapped" option is selected, searching for a broad code like ATC level 3 code A05A ("Bile Therapies"), or even a specific term code like RxNorm code 1544460 for idelalisib, will automatically identify and query for all bottom-level (e.g., idelalisib 150 MG Delayed Release Oral Tablet) codes contained underneath that seed concept. This works by ROMOP first mapping the initial search criteria to a standard concept (SNOMED or RxNorm) and finding all descendants underneath it. This function allows for incorporation of multiple vocabulary types (e.g., ATC and LOINC codes) and codes simultaneously and can support both inclusion and exclusion criteria, if desired. The user can also set the strategy of dealing with criteria, namely either union (i.e., or) or intersection (i.e., and) requirements.

Find all "Type 2 Diabetes Mellitus" patients using ICD10 code (E11):

```
Code Start Over Run Code  
1 patient_list <- findPatients(strategy_in="mapped", vocabulary_in = "ICD10CM", codes_in = "E11")  
2  
3
```

```
[1] "5378 patients found that meet the inclusion criteria."
```

Find all patients prescribed with any "Serotonin receptor antagonists" using ATC code (A03AE):

```
Code Start Over Run Code  
1 patient_list <- findPatients(strategy_in="mapped", vocabulary_in = "ATC", codes_in = "A03AE")  
2  
3
```

```
[1] "96 patients found that meet the inclusion criteria."
```

Find all patients with "Other anxiety disorders" using ICD10 code (F31), but not prescribed with "Clonazepam" using RxNorm code (2598):

```
Code Start Over Run Code  
1 patient_list <- findPatients(strategy_in="mapped", vocabulary_in = "ICD10CM", codes_in = "F31", strategy_out="mapped",  
2  
3
```

```
[1] "268 overlapping patients excluded from the original inclusion input based on the exclusion criteria."
```

```
[1] "2057 patients found that meet the inclusion criteria."
```

[Previous Topic](#)

[Next Topic](#)

ROMOP Sandbox Tutorial

Benjamin S. Glicksberg

Butte Lab

Bakar Computational Health Sciences Institute

University of California, San Francisco

2018

ROMOP

Initialization

Data exploration

Finding cohort/patients

Extracting clinical data

Start Over

Extract Data

✓ Retrieve clinical data for pre-defined cohort

Retrieve clinical data for patient ids found from the findPatients function:

Clinical data can also be retrieved for a patient list that is defined using the findPatients function.

```
Code Start Over Run Code
1 patient_list <- findPatients(strategy_in="mapped", vocabulary_in = "ATC", codes_in = "A03AE")
2
3 ptClinicalData <- getClinicalData(patient_list, declare=FALSE)
4
5 head(ptClinicalData$Condition)
```

```
[1] "96 patients found that meet the inclusion criteria."
```

condition_concept_vocabulary	condition_concept_code	condition_concept_name
SNOMED	40257000	Contusion of shoulder region
SNOMED	40257000	Contusion of shoulder region
SNOMED	35678005	Multiple joint pain
SNOMED	44465007	Sprain of ankle
SNOMED	95210003	Plasma cell leukemia
SNOMED	11437003	Contusion of back

6 rows | 5-7 of 12 columns

As mentioned, the clinical data are stored as a list of data.tables in the ptClinicalData object.

Summarize cohort

ROMOP Sandbox Tutorial

Benjamin S. Glicksberg
Butte Lab
Bakar Computational Health Sciences Institute
University of California, San Francisco

2018

ROMOP

Initialization

Data exploration

Finding cohort/patients

Extracting clinical data

Start Over

✓ Summarize demographic information of clinical cohort

ROMOP provides a function to quickly summarize the demographic information for a cohort of interest.

Summarize demographic information for patient ids found from the findPatients function:

```
Code Start Over Run Code
```

```
1 patient_list <- findPatients(strategy_in="mapped", vocabulary_in = "ATC", codes_in = "A03AE")
2
3 ptDemo <- getDemographics(patient_list, declare=FALSE)
4
5 summarizeDemographics(ptDemo)
```

```
[1] "96 patients found that meet the inclusion criteria."
```

```
# of patients: 96
```

```
Mean age: 79.375
```

```
Median age: 82.5
```

```
STD age: 14.145
```

```
Status breakdown:
```

```
      Status  n proportion
1:   Alive  94 0.97916667
2: Deceased   2 0.02083333
```

```
Gender breakdown:
```

```
      Gender  n proportion
1: FEMALE  61 0.6354167
2:   MALE  35 0.3645833
```

```
Race breakdown:
```

```
      Race  n proportion
1: Black or African American  7 0.07291667
2:                Unknown    9 0.09375000
3:                White   80 0.83333333
```

```
Ethnicity breakdown:
```

```
      Ethnicity  n proportion
1: Hispanic or Latino  5 0.05208333
2: Not Hispanic or Latino 91 0.94791667
```

PatientExploreR

dynamic visualization of clinical history in OMOP
format

Glicksberg et al. (in revision)

No flexible application exists

ATLAS warfarin-new user

Profiles

1PCT 2

MALE | 501 events | Age 65 at index

conditioners

drugs

measurements

procedures

visits

0 100 200 300 400 500 600 700

Column visibility Copy CSV Show 15 entries Filter:

Showing 1 to 15 of 323 entries

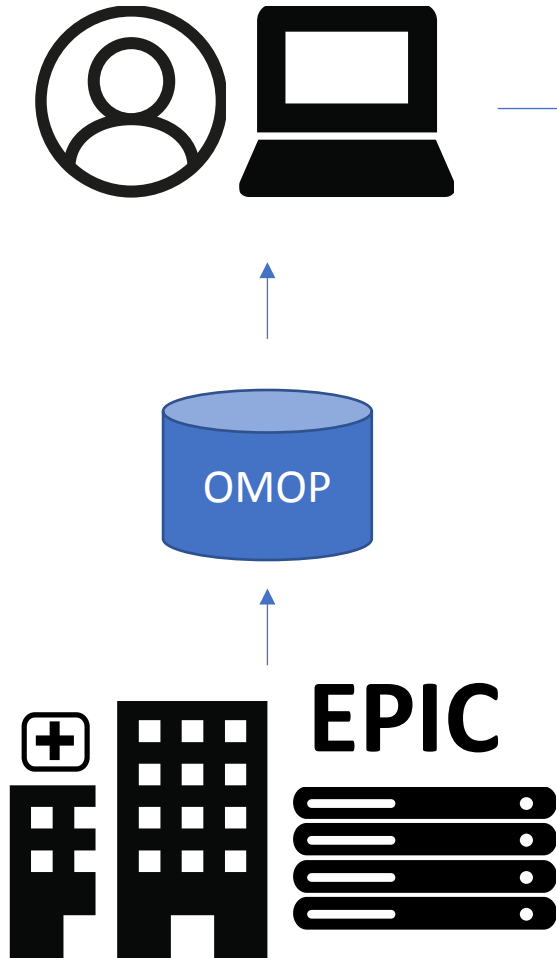
Concept Id	Concept Name	Domain	Start Day	End Day
439947	Open fracture of upper end of fibula	conditionera	0	0
434502	Closed fracture of phalanx of foot	conditionera	0	0
43531028	Mononeuropathy of lower limb	conditionera	0	0
0	No matching concept	visit	0	0
2212392	Measurement of glycosylated hemoglobin (HbA1C)	measurement	0	0
2105893	Application of short leg cast (below knee to toes)	procedure	0	0
27674	Nausea and vomiting	conditionera	4	4
200219	Abdominal pain	conditionera	4	4
317898	Malignant essential hypertension	conditionera	4	4
40490986	Tenderness of right lower quadrant of abdomen	conditionera	4	4
0	No matching concept	visit	4	4
2211359	Radiologic examination, chest; single view, frontal	procedure	4	4

Domain

- condition (143)
- conditionera (119)
- procedure (112)
- visit (44)
- measurement (32)
- observation (19)
- drug (16)
- drugera (16)

Apache 2.0 open source software provided by OHDSI join the journey

Goals



PatientExploreR: dynamic visualization of clinical history

This application allows for flexible searching and extracts patient-level interactive and dynamic reports and visualization of clinical data

User ID
glicksbergb

Password
.....

LOGIN LOGOUT



Please log-in with your credentials.

Successfully logged in.

? First time user? Check out the [Help](#) page or start the [Tutorial](#)



Patient Finder

Identify a patient to explore: query the EMR for all patients with data a concept or concepts of interest. Can search by Diagnosis, Medication, Procedure, and Lab related concepts. Can further filter patients by demographic features (e.g., age range, self-reported race).



Overall Report

Generate overall report of a selected patient's clinical history: this report will provide a chronological history of all events of all data modalities (e.g., diseases, medications). Can filter by event type for more focused displays.



Encounter Timeline

Interact and explore a selected patient's clinical encounter timeline: investigate clinical events by encounter. Selecting an encounter in the timeline will detail all associated clinical events. Can filter by encounter (e.g., Appointment) and visit (e.g., Screening) types.



Data Explorer

Explore patterns of clinical events over time: for a selected patient, can view all data measured for categorical (diseases, medications, procedures) and numeric (labs, vital signs, and flowsheet) types over time. Categorical variables displayed in a timeline and can be filtered for what is shown. Numeric variables are displayed as a timeseries which the user can interact with. Targeted view provides an in-depth graph of one variable at a time while the Multiplex view allows for simultaneous and linked exploration of multiple variables.

Who We Are

Public Sandbox Server

<http://patientexplorer.ucsf.edu>

- Synthesized data (no PHI) from CMS
- 1 million patients
- OMOP format
- Open to the public

Code: <https://github.com/BenGlicksberg/PatientExploreR>

The screenshot shows the PatientExploreR Sandbox Server interface. At the top, there is a navigation bar with links: PatientExploreR, Home, Patient Finder, Overall Report, Encounter Timeline, Data Explorer, and More. The main heading is "PatientExploreR Sandbox Server". Below this, a paragraph states: "PatientExploreR interfaces with a relational database of EHR data in the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). This application produces patient-level interactive and dynamic reports and visualization of clinical data, without requiring programming skills." A prominent message reads: "All patient data are synthesized and contain no Protected Health Information". Three icons are displayed: a question mark for "Help", an information icon for "About", and a download icon for "Download App". A instruction says: "To begin: click [Load Credentials](#), then [Login](#)".

The main content area is titled "Please log-in below:" and contains a form with the following fields:

- User ID (with a copy icon)
- Password (with a copy icon)
- Host
- Database (pre-filled with "aws_omop_synpuf")
- Driver (a dropdown menu currently showing "MYSQL")
- Port (pre-filled with "3306")

Below the form are two buttons: "SAVE CREDENTIALS" and "LOAD CREDENTIALS". At the bottom of the form area, there is a breadcrumb trail: "/srv/shiny-server/patientexplorer/" and a "..." menu icon. At the very bottom, there are "LOGOUT" and "LOGIN" buttons.

On the right side of the interface, there is a sidebar with a heading "First time user? Check out the [Help](#) page." Below this are four menu items, each with an icon and a description:

- Patient Finder** (magnifying glass icon): Search for a patient directly or identify a cohort: query the EHR for a certain patient or find all patients that meet any criteria concept available from the CDM of any modality (e.g., Condition, Procedure). Cohorts can be further filtered by demographic features (e.g., age range, self-reported race), visualized, and exported.
- Overall Report** (document icon): Generate overall report of a selected patient's clinical history: this report will provide a chronological history of all events of all data modalities (e.g., Observations, Medications). Can filter by specific concepts and export.
- Encounter Timeline** (mouse cursor icon): Interact and explore a selected patient's clinical encounter and visit timeline: investigate and visualize clinical events by visit occurrence. Selecting a visit in the interactive timeline will detail all associated clinical events. Can filter by visit (e.g., Outpatient) and admitting/discharge types.
- Data Explorer** (wrench icon): Explore patterns of clinical events over time: for a selected patient, can view all data measured for categorical (e.g., Medications, Devices) and numeric (e.g., Measurement, Observation) types over time. Categorical variables displayed in a timeline and can be filtered for what is shown. Numeric variables are displayed as a timeseries which the user can interact with. Targeted view provides an in-depth graph of one variable at a time while the Multiplex view allows for simultaneous and linked exploration of multiple variables.

Patient Finder

Search for patients directly or based on clinical criteria (e.g., Condition ICD-10CM code). By selecting 'Criteria', all available ontologies will be displayed per modality which the user can use for searching. This will load demographic information for matching patients to allow for further refining.

Search Mode:

- Search by Patient
 Search by Criteria

Criteria (select from table):

Select Domain: Select Vocabulary: Select Concept Class:

Search:

concept_code	concept_name	domain_id	vocabulary_id	concept_class_id
K51.4	Inflammatory polyps of colon	Condition	ICD10CM	4-char nonbill code
K51.414	Inflammatory polyps of colon with abscess	Condition	ICD10CM	6-char billing code
K51.41	Inflammatory polyps of colon with complications	Condition	ICD10CM	5-char nonbill code
K51.413	Inflammatory polyps of colon with fistula	Condition	ICD10CM	6-char billing code
K51.412	Inflammatory polyps of colon with intestinal obstruction	Condition	ICD10CM	6-char billing code

Showing 1 to 5 of 64 entries (filtered from 93,463 total entries) Previous ... Next

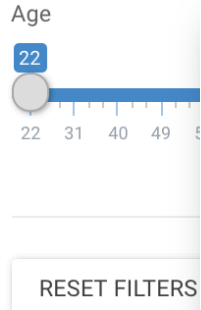
Selected Criteria:

vocabulary	term
ICD10CM	K51.4
ICD10CM	K51.414
ICD10CM	K51.41
ICD10CM	K51.413
ICD10CM	K51.412

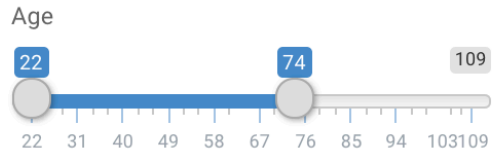
Search Type: or
Search Strategy: Mapped Direct

Showing 1 to 5 of 64 entries Previous ... Next

Filter Cohort:



Filter Cohort:



Gender: MALE

Status: ALIVE

Race: 3 ITEMS SELECTED

Ethnicity: 2 ITEMS SELECTED

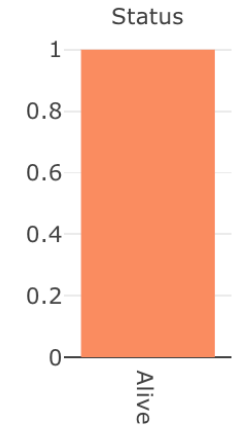
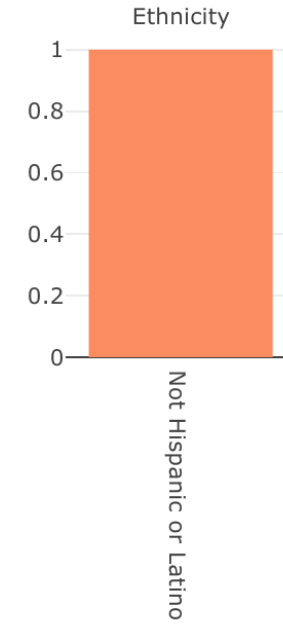
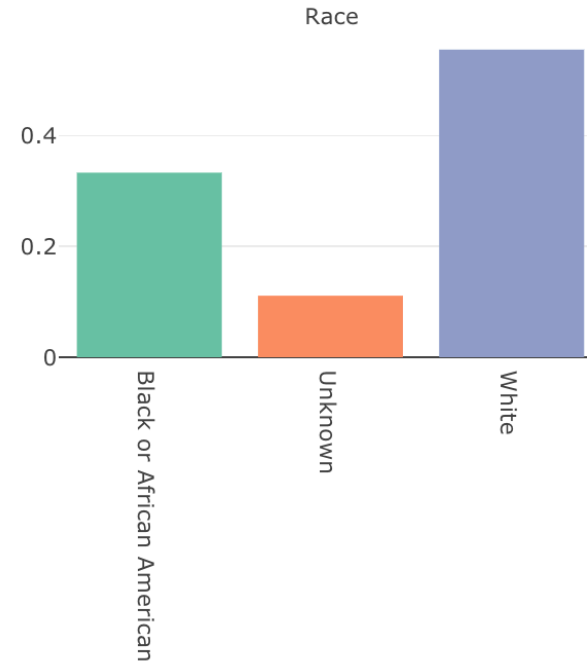
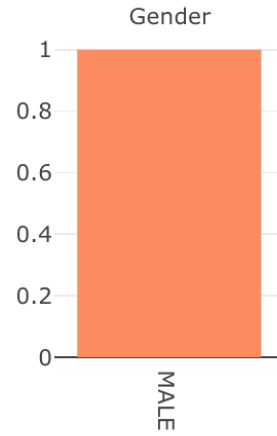
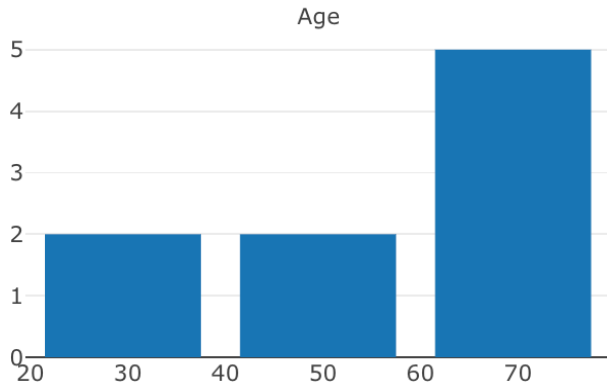
RESET FILTERS

EXPORT COHORT

HIDE PLOTS

Selected Patient ID:

SEARCH



Automatically generated clinical history

Overall Report: 9000000

Background:

Status: Alive
Age: 22
Age of Death: NA
Ethnicity: Not Hispanic or Latino
Race: Unknown
Gender: MALE

Clinical Summary:

Earliest encounter: 2017-01-17
Most recent encounter: 2017-07-28
unique encounter types: 1
Encounters: 7
Outpatient encounters: 7
Inpatient encounters: 0

observations: 3
unique observation concepts: 3
conditions: 5
unique condition concepts: 4
procedures: 0
unique procedure concepts: 0
medication prescriptions: 3
unique medication concepts: 2
measurements: 40
unique measurement concepts: 6
devices: 0
unique device concepts: 0

Select data modalities to include:

Data Modalities

4 ITEMS SELECTED

EXPORT REPORT

Observations

3 ITEMS SELECTED

Conditions

4 ITEMS SELECTED

Procedures

NOTHING SELECTED

Medications

2 ITEMS SELECTED

Measurements

6 ITEMS SELECTED

Devices

NOTHING SELECTED

Show 10 entries

Search:

Date	Type	Event	Value
2017-01-17	Observation	Contraceptive use behavior	
2017-01-17	Observation	Drug injection behavior	
2017-01-17	Measurement	Hematocrit	39
2017-01-17	Measurement	Calprotectin [Mass/mass] in Stool	70
2017-01-17	Measurement	C reactive protein [Mass/volume] in Serum or Plasma	0.5
2017-01-17	Measurement	Erythrocyte sedimentation rate	3
2017-01-17	Measurement	Creatinine serum/plasma	0.7
2017-01-17	Measurement	Albumin serum/plasma	4
2017-06-15	Condition	Keloid scar	
2017-07-01	Observation	Tobacco use and exposure	1

Showing 1 to 10 of 51 entries

Previous

1

2

3

4

5

6

Next

Encounters Timeline: 9000000

Plot Encounters:

- None
- Visit Types
- Admitting Concepts
- Discharge Concepts

Visit Types

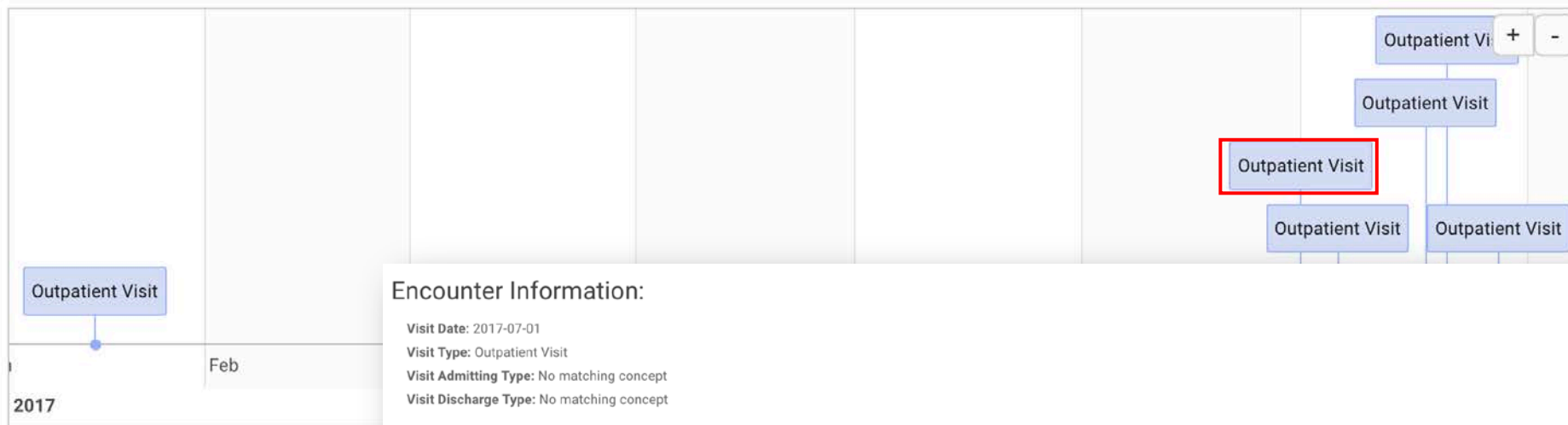
OUTPATIENT VISIT

Admitting Concept Type

NO MATCHING CONCEPT

Discharge Concept Type

NO MATCHING CONCEPT



Encounter Information:

Visit Date: 2017-07-01
Visit Type: Outpatient Visit
Visit Admitting Type: No matching concept
Visit Discharge Type: No matching concept

[Conditions](#) [Devices](#) [Measurements](#) [Medications](#) [Observations](#) [Procedures](#)

Show 10 entries

Search:

condition_concept_name	condition_type	condition_status_type	condition_concept_vocabulary	condition_concept_code	condition_source_vocabulary	condition_source_code	condition_start_date	condition_end_date
Allergic rhinitis	Primary Condition		SNOMED	61582004			2017-07-01	2017-07-10

Showing 1 to 1 of 1 entries

Previous

1

Next

Data Explorer: 9000000

Data Explorer Mode:

- Targeted
- Multiplex
- Multiplex Timeline

Explore all clinical events over the patient's history. The user can explore both categorical (Conditions, Medications, Procedures, or Devices) or numeric (Measurement or Observation) data. For categorical data, the events are visualized in an interactive timeline and the user can select which events to show. Further, diseases may be explored at different levels (Disease Name, ICD 9 or 10). For numeric data types, the events (e.g., WBC for Labs) are displayed as a table with # of measurements recorded. The user can select an event of interest which will display as an interactive timeseries plot.

Conditions Devices Measurements Medications Procedures Observations

View Type:

- Event
- Range

Conditions

4 ITEMS SELECTED



Visit Occurrence ID for Condition: 9000002

Condition Window: 2017-07-06 to 2017-07-21

Condition Status Type: NA

Condition Standardized Name Selected: Ulcerative colitis

Condition Standardized Vocabulary: SNOMED

Condition Standardized Vocabulary Code: 64766004

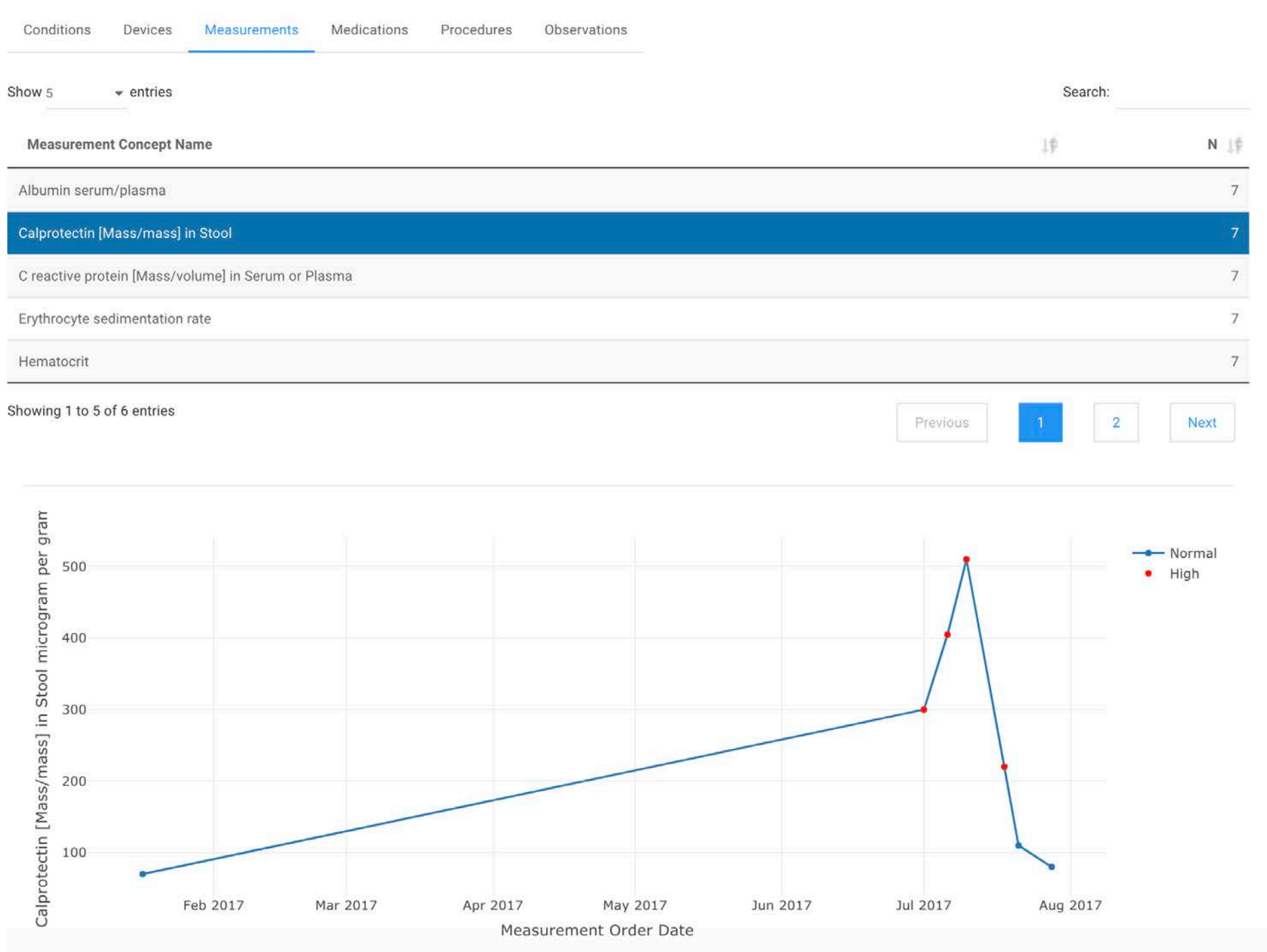
Condition Source Value: NA

Condition Source Vocabulary: NA

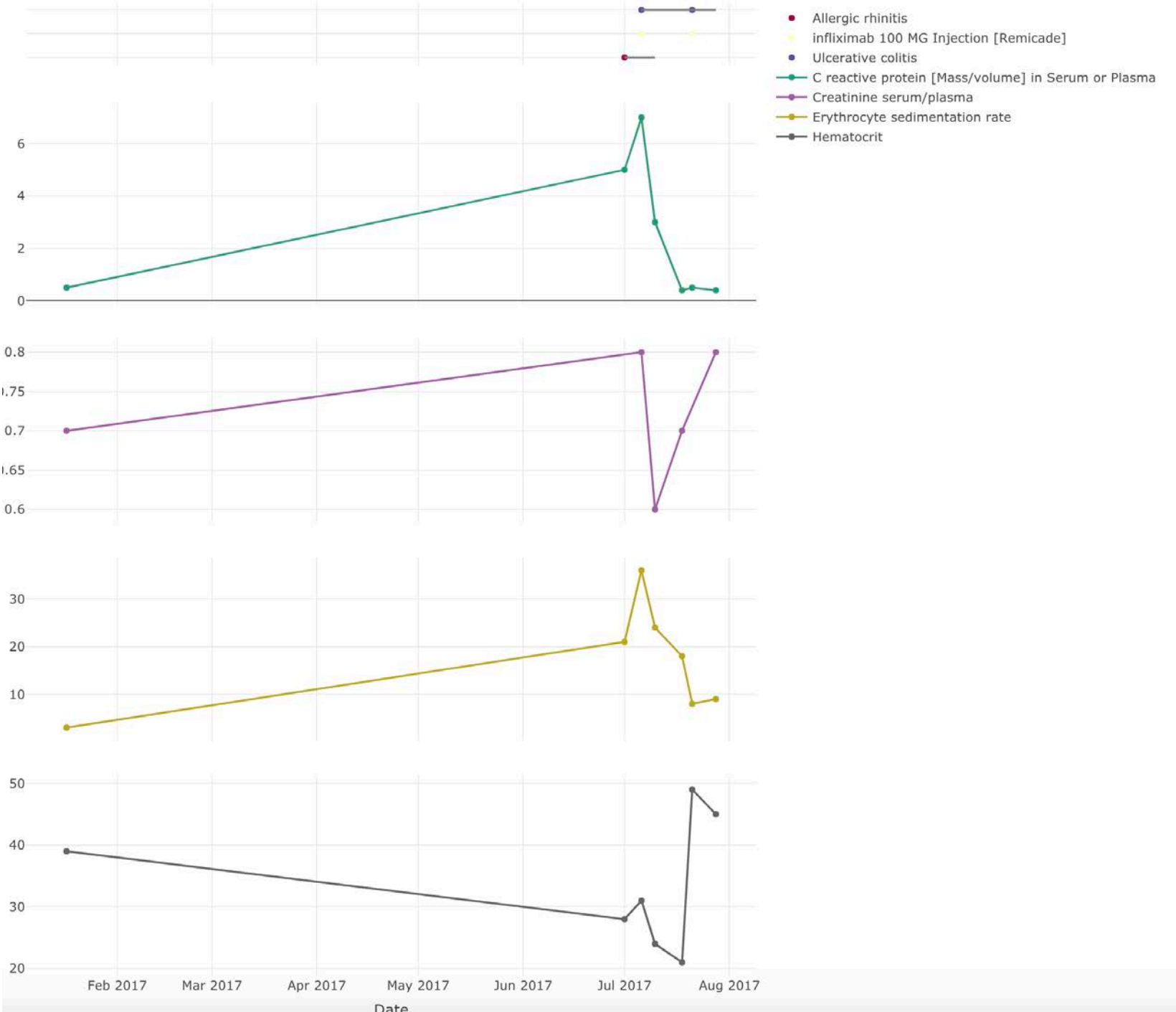
Condition Source Vocabulary Code: NA

Explore Trends in Data/ Outcomes (targeted)

Explore Trends in Data/ Outcomes (numeric; targeted)



Explore Trends in Data/ Outcomes (multiplex)



Explore Trends in Data/ Outcomes (multiplex timeline)

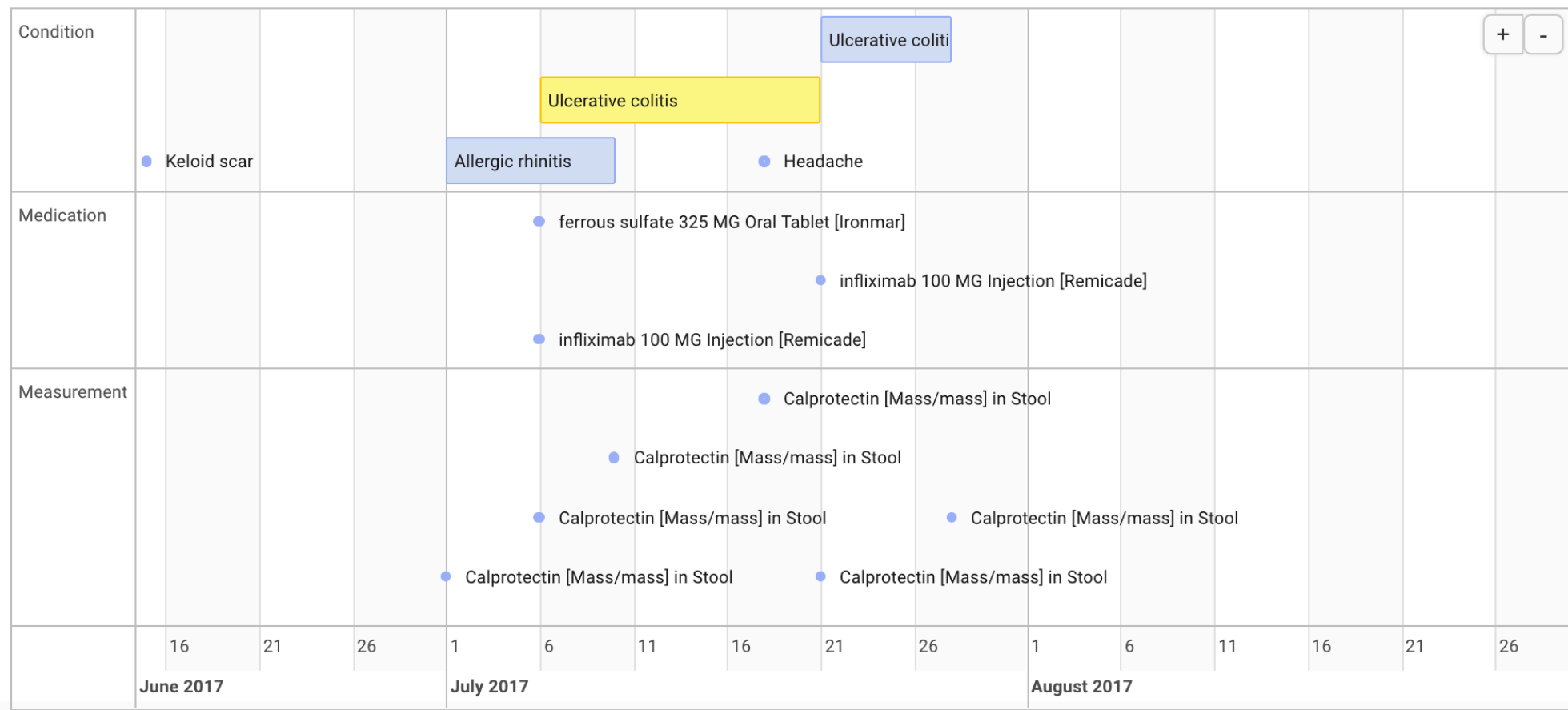
View Type:

View Type:

- Event
- Range

Selected Data Info:

Modality: Condition
Concept: Ulcerative colitis
Window: 2017-07-06 to 2017-07-21
Value: NA



How might these tools enable AI-based EHR research?

How well can we predict...

- Risk for disease
- Disease onset
- Symptom severity
- Treatment response
- Medication adverse events
- Ideal dose of medication
- Symptom flares
- Length of stay in hospital

Time Aggregation and Model Interpretation for Deep Multivariate Longitudinal Patient Outcome Forecasting Systems in Chronic Ambulatory Care

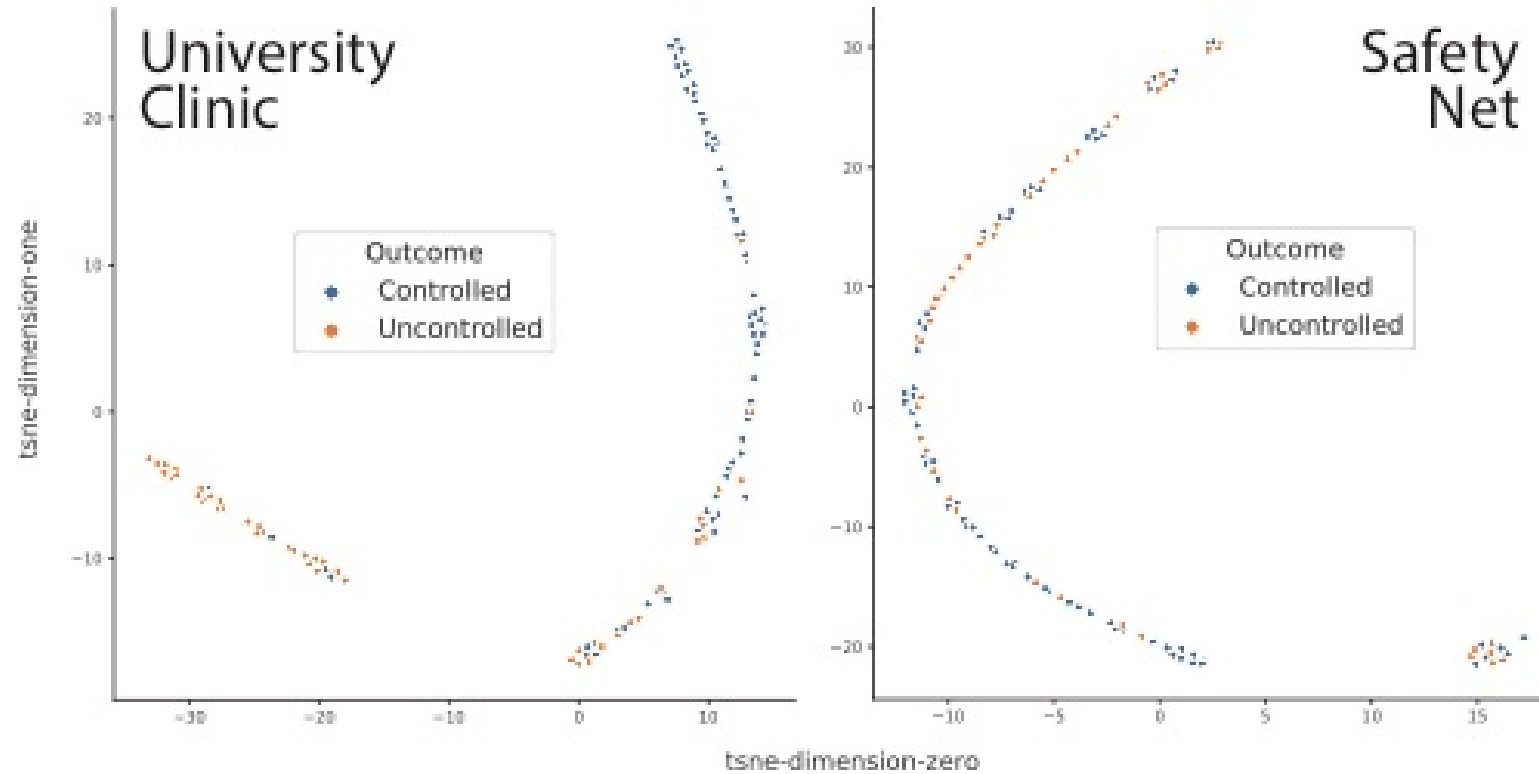
Beau Norgeot¹

Dmytro Lituiev¹

Benjamin S. Glicksberg¹

Atul J. Butte^{1*}

¹ Bakar Computational Health Sciences Institute, University of California, San Francisco
*atul.butte@ucsf.edu



More representation/data = better reflection of dx

UC DAVIS
UNIVERSITY OF CALIFORNIA



UCSF
University of California
San Francisco

UCLA
University of California, Los Angeles

UC San Diego

...perform here?

UNIVERSITY OF CALIFORNIA
UC RIVERSIDE



UC IRVINE

- 1. Precision medicine:** finding similar patients to go beyond treating doctor's, clinic's, department's, hospital's, or even institution's expertise.
- 2. Disease representation in EHR:** electronic phenotyping algorithms might not be fully generalizable. Building as a "meta" signature will be more robust
- 3. Prediction:** training and testing models across multiple institutions, alone and in conjunction, will enable identifying ideal strategies
- 4. Multi-omic factors:** incorporating genetics and environmental data (e.g., pollution) can help pinpoint etiology and discern GxE interactions

Butte Lab

Boris Oskotsky, PhD

Vivek Rudrapatna, MD, PhD

Debajyoti Datta, MD, PhD

Beau Norgeot

Nadav Rappoport, PhD

Ted Goldstein, PhD

Atul Butte, MD, PhD



Dana Ludwig, MD

Remi Frazier

Nelson Lee

Rick Larsen

Acknowledgements



Eugenia Rutenberg, MBA

Angelo Pelonero

Angela Rizk-Jackson, PhD

Sharat Israni, PhD



Nicholas Giangreco

Phyllis Thangaraj

Nicholas Tatonetti, PhD



*Graduate School of
Biomedical Sciences*

Li Li, MD

Khader Shameer, PhD

Riccardo Miotto, PhD

Kipp Johnson

Marcus Badgeley

Mark Shervey

Rong Chen, PhD

Joel Dudley, PhD



Community and Developers

