Welcome to the Sentinel Innovation Center Webinar Series

The webinar will begin momentarily

- Please visit www.sentinelinitiative.org for recordings of past sessions and details on upcoming webinars.
- Note: closed-captioning for today's webinar will be available on the recording posted at the link above.

Sentinel

Portability and Scalability of Computable Phenotypes: Trade Offs and Trajectories

Jon Duke, MD
Director, Center for Health analytics and Informatics
Georgia Tech
August 24th 2020

Georgia Institute of Technology CHAI Georgia Tech Research Institute

Conflicts of Interest

- CHAI receives funding from organizations including FDA and CDC to conduct work that relates to the topics discussed in today's webinar
- No personal financial COIs

Overview

- Constructing phenotypes is a cornerstone activity in health data analytics, whether for research, quality, safety, cost or clinical objectives
- We will explore aspects of phenotype *portability* and *scalability,* with particular emphasis on recent policy, standards, and tools that can inform phenotyping decisions
- A key goal is to enable broader use of electronic health data beyond current commonly used sources

What's a phenotype?

- A phenotype is a logical definition of a clinical event, state, or or characteristic of interest
  - "Computable" indicates that the phenotype is machine-interpretable so that the query that it be run on a data source to find matching patients
- In safety contexts, phenotypes often define exposures (eg patients who received a given drug) and outcomes (eg patients who had a myocardial infarction)

Portability and Scalability

- Portability refers to the performance of a phenotype in generating consistent cohorts across diverse sites
  - ie, did you find the patients you are looking for
  - Reflects the design of the phenotype content and logic
- Scalability refers to the ability for a phenotype to be run on a wide range of sites and data sources
  - Reflects infrastructure, platform, adoption etc

Treatment patterns and burden of behavioral disturbances in patients with dementia in the United States: a claims database analysis

Myrlene Sanon Aigbogun[1*], Robert Stellhorn[1], Ann Hartry[2], Ross A. Baker[3] and Howard Fillit[4,5]

Patient population

Selected patients had at least one International Classification of Diseases, Ninth Edition (ICD-9) or ICD-9 clinical modification (ICD-9-CM) code for Dementia/Dementia related (ICD-9: 290.0, 290.10, 290.11, 290.12, 290.13, 290.20, 290.21, 290.3, 290.40, 290.41, 290.42, 290.43, 294.0, 294.8, 294.10, 294.20; ICD-9-CM: 331.11, 331.2, 331.7) and AD (ICD-9-CM: 331.0) in any of four diagnosis fields on outpatient claims or any of five diagnosis fields on inpatients claims. Of these, 398,128 patients were at least 65 years old, and 103,402 were continuously enrolled in the Medicare supplemental database and had continuous health plan enrollment with medical and pharmacy benefits for at least 6 months pre-index period (baseline) and 6 months post-index date. Existing BD diagnostic codes 294.11 and 294.21 were used as a proxy for agitation because ICD-9 diagnostic codes were not available to identify agitation. To identify patients with late-stage disease, patients were flagged with late-stage disease per Fillit et al. 2002 criteria [26]: presence of decubiti (707.00), malnutrition (260, 261, 262, 263.1, 263.2, 263.8, 263.9), and aspiration pneumonia (507.x).

Aigbogun MS, Stellhorn R, Hartry A, Baker RA, Fillit H. Treatment patterns and burden of behavioral disturbances in patients with dementia in the United States: a claims database analysis. BMC neurology. 2019 Dec 1;19(1):33.

PheKB

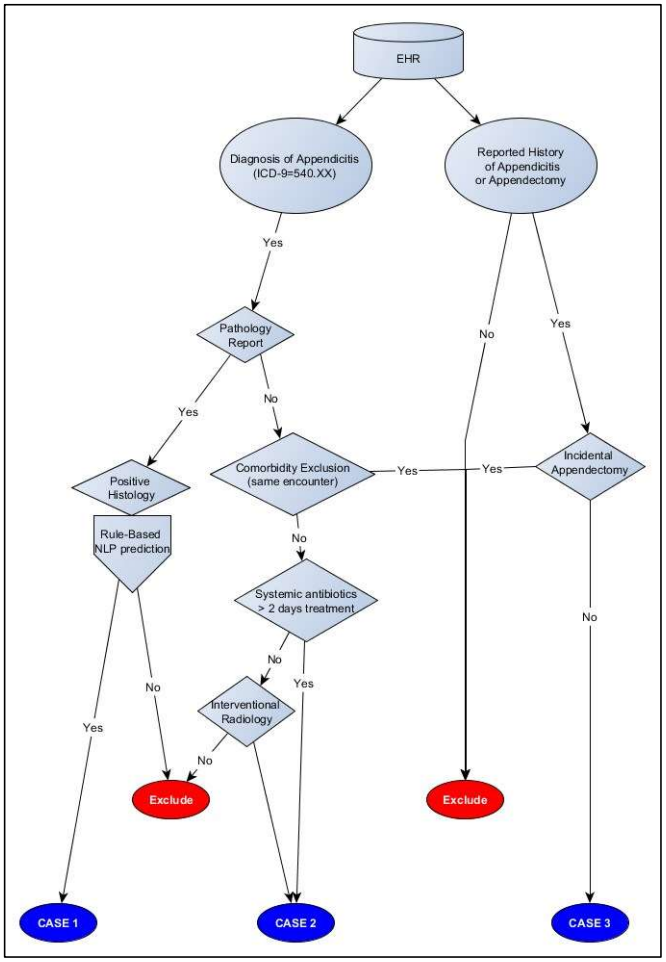


Table 1: Comorbidities for Appendicitis (Case Exclusion)

ICD9	Description
575.0	Acute cholecystitis
788.0	Right ureteric colic
633.0	Ectopic pregnancy
486	Pneumonia
533.5	Perforated peptic ulcer (without obstruction)
533.51	Perforated peptic ulcer (with obstruction)
599.0	Urinary tract infection
614.2	Salpingitis/
614.9	pelvic inflammatory disease
558.9	Gastroenteritis
560.9	Intestinal obstruction
590.80	pyelonephritis
590.10	pyelonephritis (acute pyelonephritis)
590.00	pyelonephritis (chronic pyelonephritis)
620.0	Ruptured ovarian follicle
250.1	Diabetic ketoacidosis (type 2 or unspecified)
250.11	Diabetic ketoacidosis (type 1; juvenile)
250.12	Diabetic ketoacidosis (type 2 uncontrolled)
250.13	Diabetic ketoacidosis (type 1 ; juvenile, uncontrolled)
577.0	Pancreatitis
620.2	Torted ovarian cyst
555.0	Terminal ileitis (small intestine)
555.1	Terminal ileitis (large intestine)
555.2	Terminal ileitis(small and large intestine)
560.0	Intussusception
277.1	Porphyria
751.0	Meckel's diverticulitis
780.96	Preherpetic pain (generalized pain)
562.11	Colonic/diverticulitis
543.9	appendicular diverticulitis (diverticula appendicular)
289.2	Mesenteric adenitis
924.9	Rectus sheath haematoma (hematoma) (no code for rectus sheath)

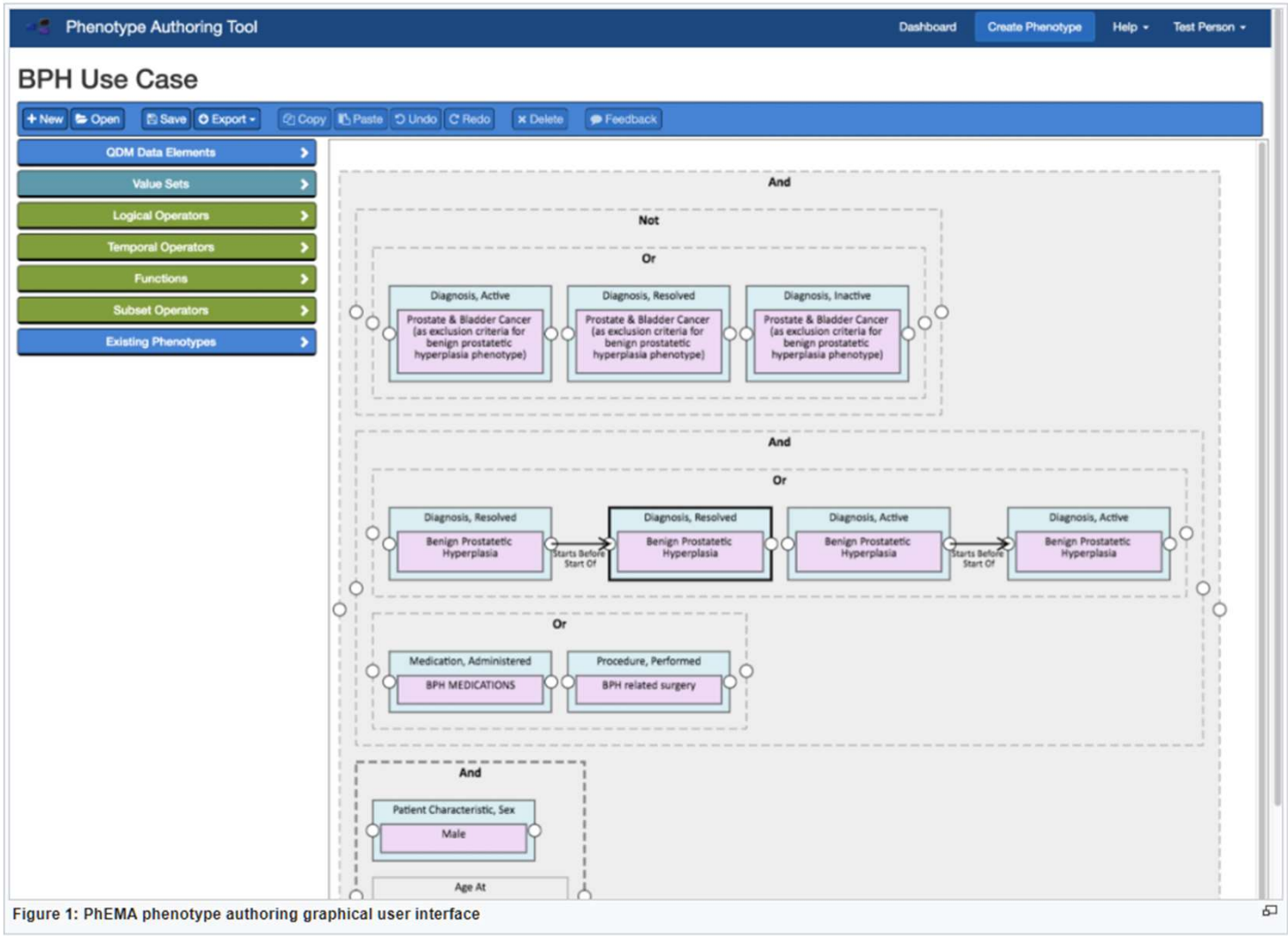
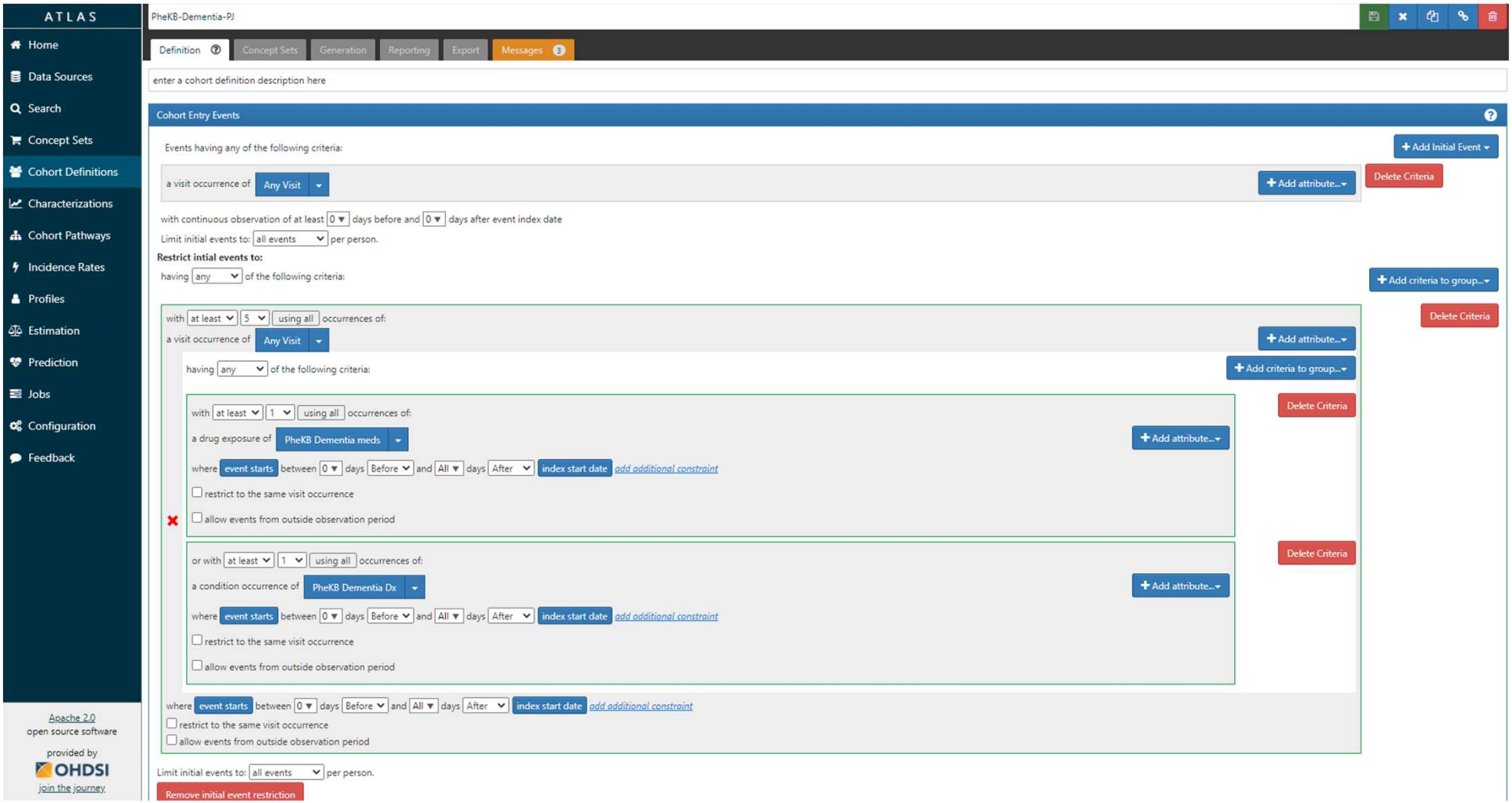
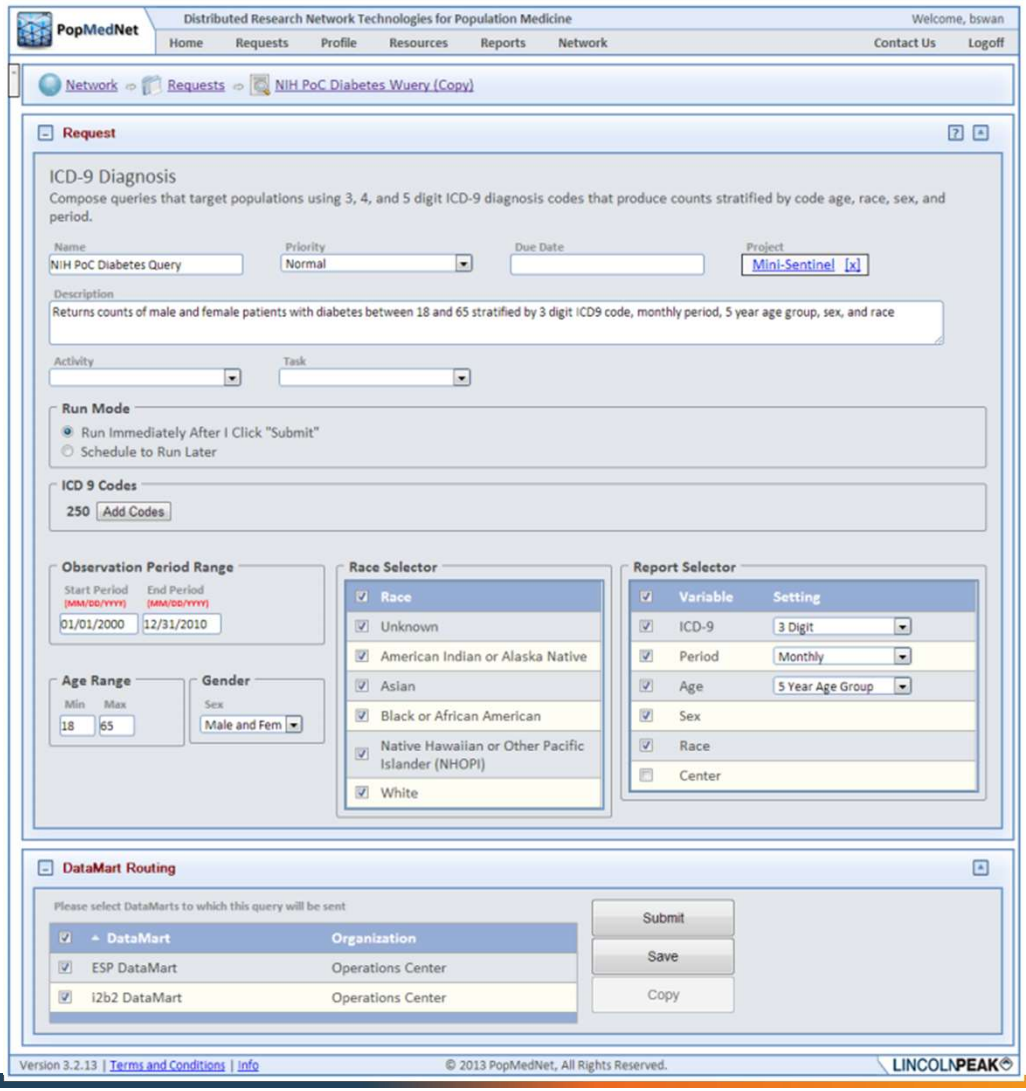


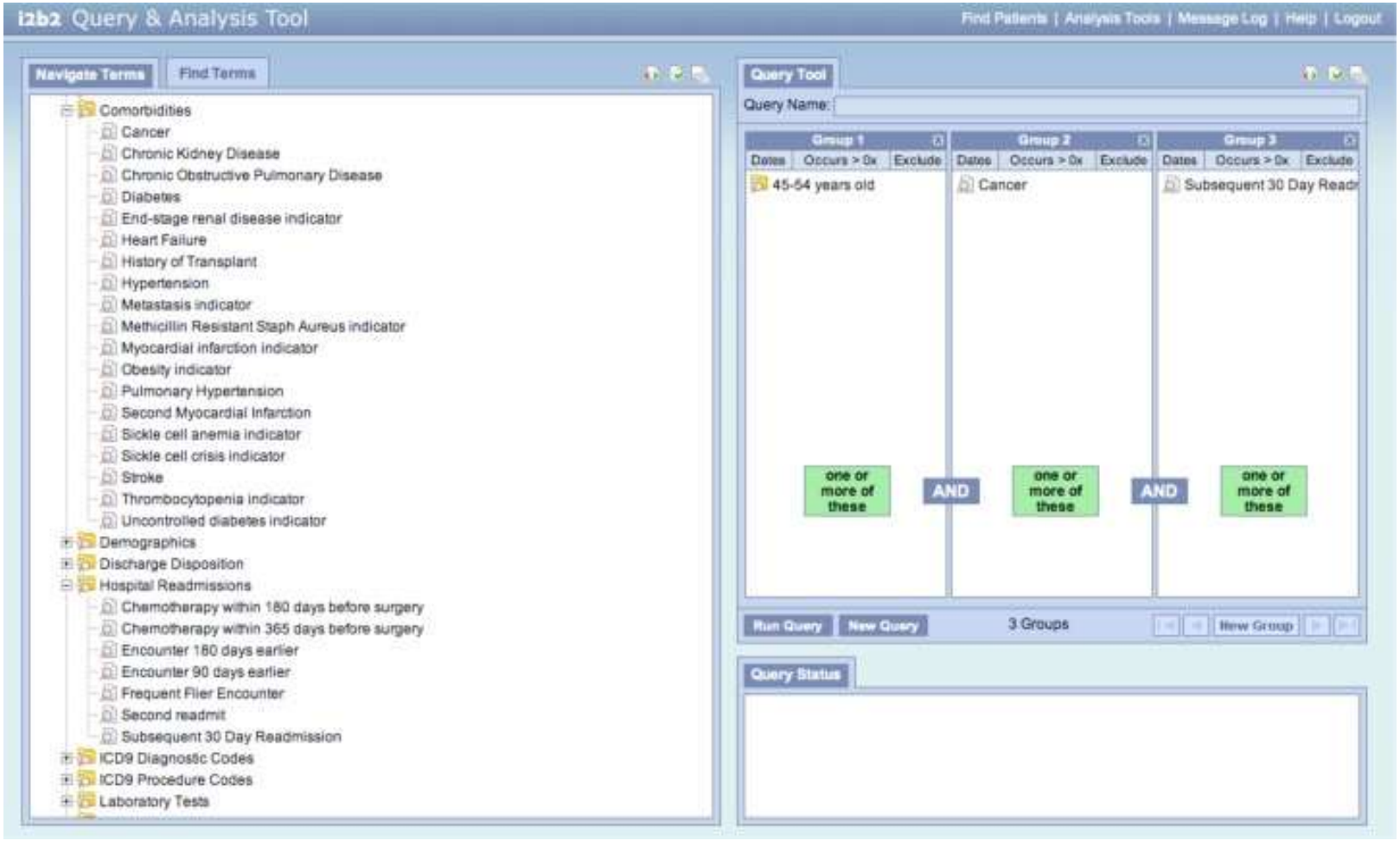
Figure 1: PhEMA phenotype authoring graphical user interface



http://www.ohdsi.org



https://www.popmednet.org/



https://www.i2b2.org/

Computable Phenotype Instantiations


```
{
 "ConceptSets": [
  {
   "id": 0,
   "name": "PheKB Dementia Dx",
   "expression": {
    "items": [
     {
      "concept": {
       "CONCEPT_ID": 374326,
       "CONCEPT_NAME": "Arteriosclerotic dementia with depression",
       "STANDARD_CONCEPT": "S",
       "STANDARD_CONCEPT_CAPTION": "Standard",
       "INVALID_REASON": "V",
       "INVALID_REASON_CAPTION": "Valid",
       "CONCEPT_CODE": "191466007",
       "DOMAIN_ID": "Condition",
       "VOCABULARY_ID": "SNOMED",
       "CONCEPT_CLASS_ID": "Clinical Finding"
      },
      "includeDescendants": true
     },
     {
      "concept": {
       "CONCEPT_ID": 4100252,
       "CONCEPT_NAME": "Arteriosclerotic dementia with paranoia",
       "STANDARD_CONCEPT": "S",
       "STANDARD_CONCEPT_CAPTION": "Standard",
       "INVALID_REASON": "V",
       "INVALID_REASON_CAPTION": "Valid",
       "CONCEPT_CODE": "191465006",
       "DOMAIN_ID": "Condition",
       "VOCABULARY_ID": "SNOMED",
       "CONCEPT_CLASS_ID": "Clinical Finding"
      },
      "includeDescendants": true
     },
     {
      "concept": {
       "CONCEPT_ID": 376094,
       "CONCEPT_NAME": "Arteriosclerotic dementia with delirium",
```


```
with t as (
 select  f.patient_num
from synpuf_omop.dbo.measurement_view f
where
f.concept_cd IN (select concept_cd from
synpuf_omop.dbo.concept_dimension   where concept_path LIKE
'\i2b2\Diagnoses\Endocrine disorders (240-259)\Other endocrine gland
diseases (250-259)\(250) Diabetes mellitus\%')
group by  f.patient_num
 )
insert into #global_temp_table (patient_num, panel_count)
select  t.patient_num, 0 as panel_count  from t
```

```
<panel>
  <item>
     <item_name>Diabetes mellitus</item_name>
     <item_key>\\OMOP_COND\i2b2\Diagnoses\Endocrine disorders (240-259)\Other endocrine gland
diseases (250-259)\(250) Diabetes mellitus\
     </item_key>
  </item>
</panel>

<panel>
  <item>
     <item_name>Female</item_name>
     <item_key>\\OMOP_DEMO\OMOP Demographics\Gender\(8532)Female\</item_key>
  </item>
</panel>

<panel>
  <item>
     <item_name>Black or African American</item_name>
     <item_key>\\OMOP_DEMO\OMOP Demographics\Race\(8516)Black or African American\</item_key>
  </item>
</panel>
```

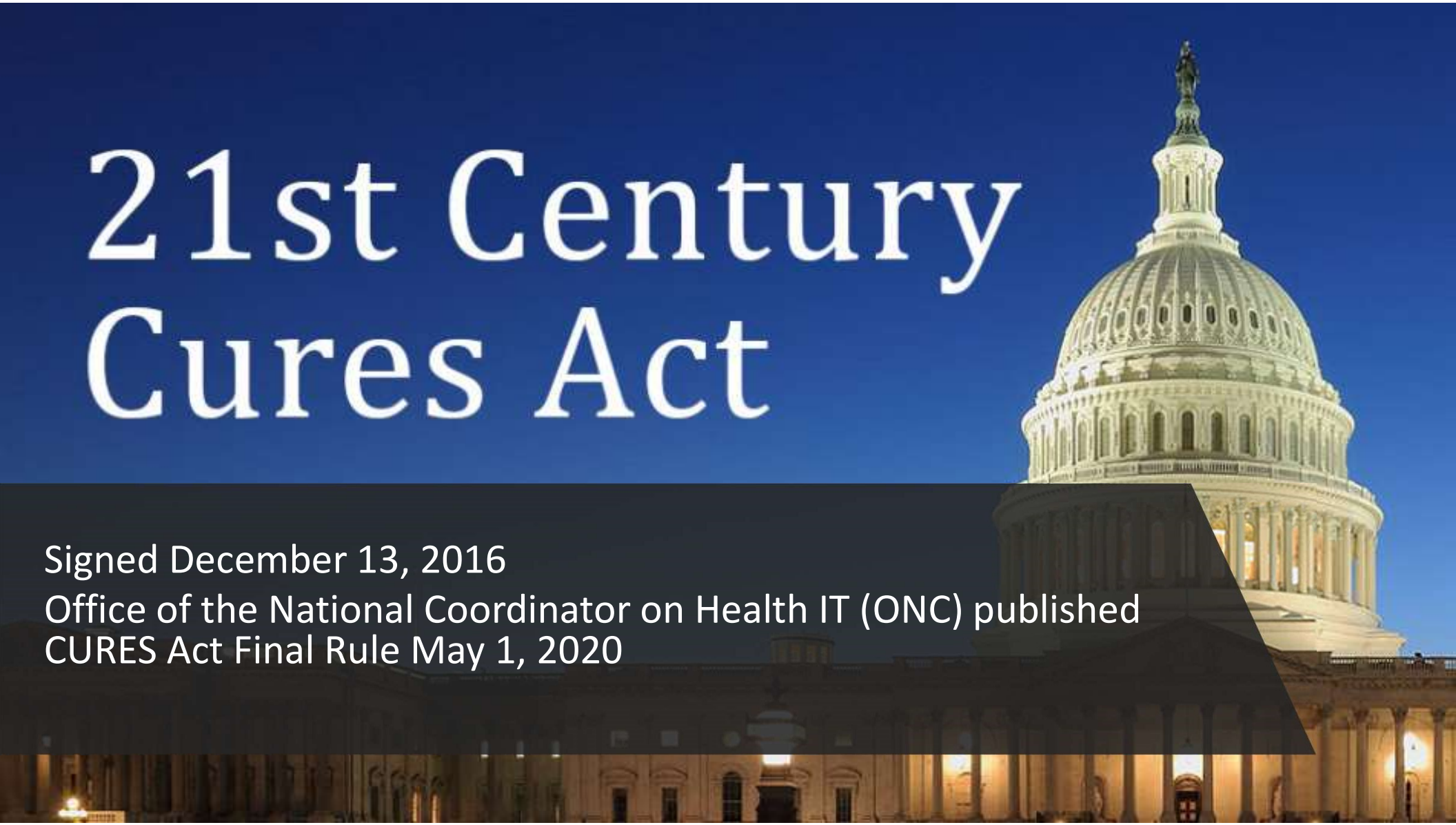
All the above examples are Research Data Models

Research Data Models

- Population level
- Designed to support analytics
- Handful of models
- Adoption varies nationally and internationally
  - Driven by funding and collaborations
  - Dominated by academic medical centers
- In the United States, no mandate regarding research data models

Transactional Data Models

- Patient level
- Designed to support clinical operations
- Hundreds of data models (EHRs + HL7 v2/3/FHIR)
- Wide array of local code systems
- Proprietary/bespoke formats limit utility for queries beyond a single site
- In the US, new regulatory policy has emerged with significant implications for transactional data models



Signed December 13, 2016

Office of the National Coordinator on Health IT (ONC) published CURES Act Final Rule May 1, 2020

The Big Stuff

- US Core Data for Interoperability (USCDI)
  - Establishes a set of data classes and elements that all Certified Health IT software must be able to export (by 2022/2023)
  - Specifies terminology(ies) for each data class
  - Specifies a process for updating the above based on stakeholder needs and IT burden

https://www.federalregister.gov/documents/2020/05/01/2020-07419/21st-century-cures-act-interoperability-information-blocking-and-the-onc-health-it-certification

USCDI v1 Classes and Elements

Allergies and Intolerances

Represents harmful or undesirable physiological response associated with exposure to a substance.

Substance (Drug Class)
Substance (Medication)
Reaction

Assessment and Plan of Treatment

Represents a health professional's conclusions and working assumptions that will guide treatment of the patient.

Assessment and Plan of Treatment

Care Team Member(s)

The specific person(s) who participate or are expected to participate in the care team.

Care Team Member(s)

Clinical Notes

Composed of both structured (i.e. obtained via pick-list and/or check the box) and unstructured (free text) data. A clinical note may include the history, Review of Systems (ROS), physical data, assessment, diagnosis, plan of care and evaluation of plan, patient teaching and other relevant data points.

Consultation Note
Discharge Summary Note
History & Physical
Imaging Narrative
Laboratory Report Narrative
Pathology Report Narrative
Procedure Note
Progress Note

Goals

An expressed desired health state to be achieved by a subject of care (or family/group) over a period of time or at a specific point of time

Patient's Goals

Health Concerns

Health related matter that is of interest, importance, or worry to someone who may be the patient, patient's family or patient's health care provider.

Health Concerns

Immunizations

Record of an administration of a vaccination or a record of a vaccination as reported by a patient, a clinician, or another party.

Immunizations

Laboratory

Tests
Values/Results

Medications

Medications

Patient Demographics

First Name
Last Name
Previous Name
Middle Name (including middle initial)
Suffix
Birth Sex
Date of Birth
Race
Ethnicity
Preferred Language
Current Address
Previous Address
Phone Number
Phone Number Type
Email Address

Problems

Information about a condition, diagnosis, or other event, situation, issue, or clinical concept that is documented.

Problems

Procedures

An activity that is performed with or on a patient as part of the provision of care.

Procedures

Provenance

The metadata, or extra information about data, that can help answer questions such as when and who created the data.

Author Time Stamp
Author Organization

Smoking Status

Classification of a patient's smoking behavior.

Smoking Status

Unique Device Identifier(s) for a Patient's Implantable Device(s)

A unique numeric or alphanumeric code that consists of a device identifier (DI) and a production identifier (PI).

Unique Device Identifier(s) for a patient's implantable device(s)

Vital Signs

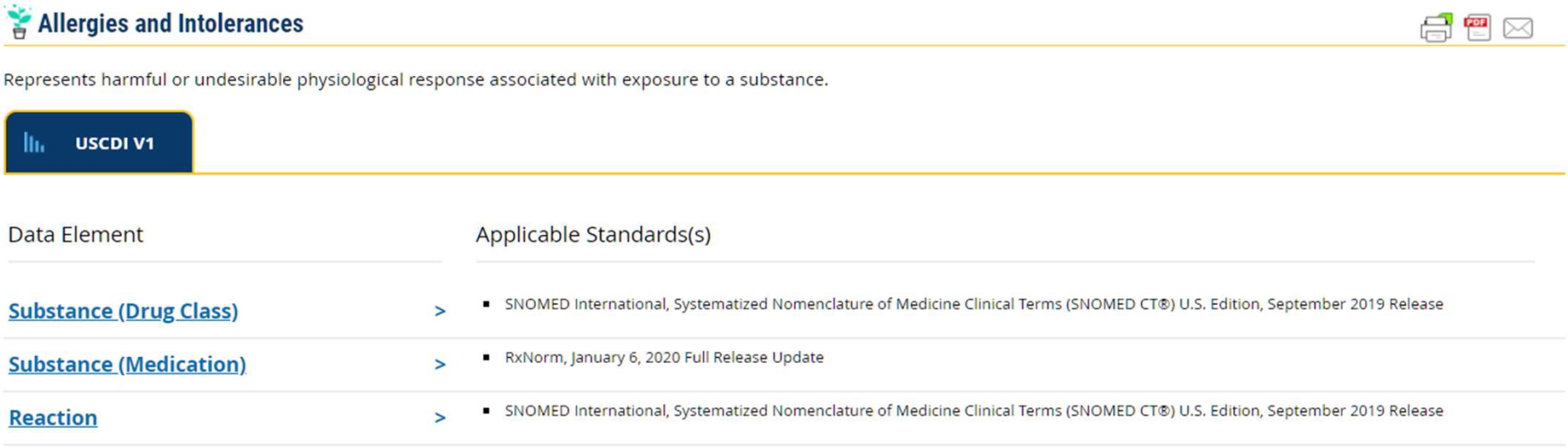
Physiologic measurements of a patient that indicate the status of the body's life sustaining functions.

Diastolic blood pressure
Systolic blood pressure
Body height
Body weight
Heart Rate
Respiratory rate
Body temperature
Pulse oximetry
Inhaled oxygen concentration
BMI Percentile (2 - 20 years)
Weight-for-length Percentile (Birth - 36 Months)
Head Occipital-frontal Circumference (Birth - 36 Months)

USCDI Standard Terminologies

Data Element	Applicable Standards(s)
Problems >	SNOMED International, Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT®) U.S. Edition, September 2019 Release
Tests >	▪ Logical Observation Identifiers Names and Codes (LOINC®) Database version 2.67
Medications >	▪ RxNorm, January 6, 2020 Full Release Update

USCDI Standard Terminologies



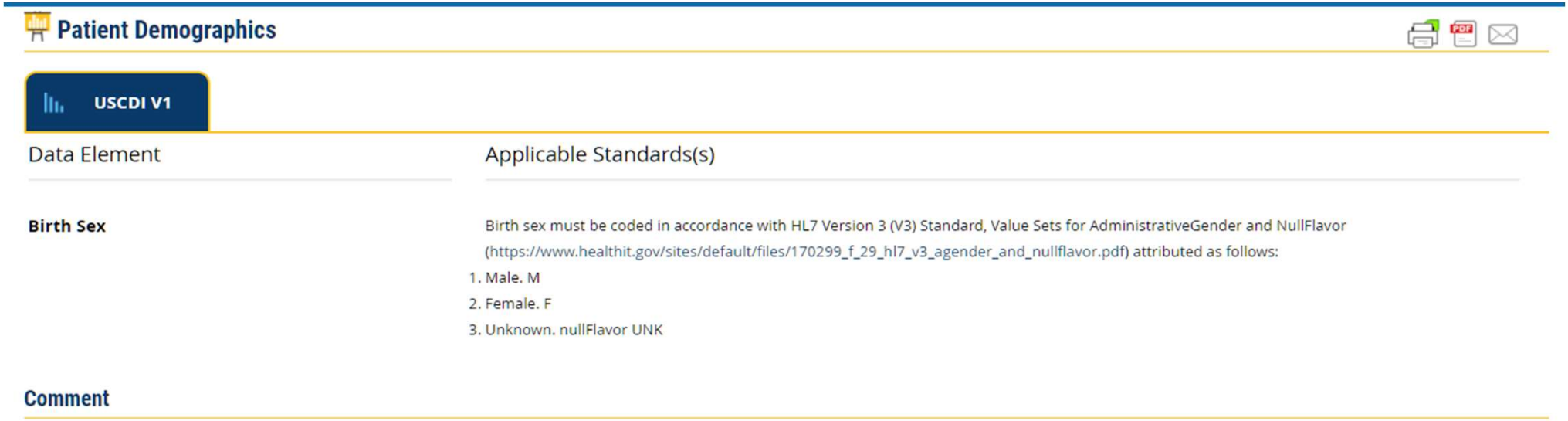
Allergies and Intolerances

Represents harmful or undesirable physiological response associated with exposure to a substance.

USCDI V1

Data Element	Applicable Standards(s)
Substance (Drug Class) >	▪ SNOMED International, Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT®) U.S. Edition, September 2019 Release
Substance (Medication) >	▪ RxNorm, January 6, 2020 Full Release Update
Reaction >	▪ SNOMED International, Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT®) U.S. Edition, September 2019 Release

USCDI Standard Terminologies



Patient Demographics

USCDI V1

Data Element	Applicable Standards(s)
Birth Sex	Birth sex must be coded in accordance with HL7 Version 3 (V3) Standard, Value Sets for AdministrativeGender and NullFlavor (https://www.healthit.gov/sites/default/files/170299_f_29_hl7_v3_agender_and_nullflavor.pdf) attributed as follows: 1. Male. M 2. Female. F 3. Unknown. nullFlavor UNK

Comment

USCDI Standard Terminologies

Clinical Notes

Composed of both structured (i.e. obtained via pick-list and/or check the box) and unstructured (free text) data. A clinical note may include the history, Review of Systems (ROS), physical data, assessment, diagnosis, plan of care and evaluation of plan, patient teaching and other relevant data points.

USCDI V1

Data Element	Applicable Standards(s)
Consultation Note >	▪ Consult Note (LOINC® code 11488-4)
Discharge Summary Note >	▪ Discharge Summary (LOINC® code 18842-5)
History & Physical >	▪ History and Physical Note (LOINC® code 34117-2)
Imaging Narrative >	▪ Diagnostic Imaging Study (LOINC® code 18748-4)
Laboratory Report Narrative >	
Pathology Report Narrative >	
Procedure Note >	▪ Procedure Note (LOINC® code 28570-0)
Progress Note >	▪ Progress Note (LOINC® code 11506-3)

Cures Act and FHIR

- USCDI specifies only content, not the means of transmission
- In terms of format, ONC specifies that API-based exchange of health data should be done via FHIR Release 4
- These two elements (USCDI and FHIR R4) are rolled together through the FHIR US Core profiles
  - Define which FHIR resources and elements satisfy the USCDI exchange requirements

https://www.hl7.org/fhir/us/core/

USCDI<->FHIR US Core

USCDI v1 Summary of Data Classes and Data Elements	US Core Profile	FHIR Resource
Assessment and Plan of Treatment	US Core CarePlan Profile	CarePlan
Care Team Members	US Core CareTeam Profile	CareTeam
Clinical Notes:		
Consultation Note	US Core DocumentReference Profile	DocumentReference
Discharge	US Core DocumentReference Profile	DocumentReference
Summary Note	US Core DocumentReference Profile	DocumentReference
History & Physical	US Core DocumentReference Profile,US Core DiagnosticReport Profile for Report and Note exchange	DocumentReference,DiagnosticReport
Imaging Narrative	US Core DocumentReference Profile,US Core DiagnosticReport Profile for Report and Note exchange	DocumentReference,DiagnosticReport
Laboratory Report Narrative	US Core DocumentReference Profile,US Core DiagnosticReport Profile for Report and Note exchange	DocumentReference,DiagnosticReport
Pathology Report Narrative	US Core DocumentReference Profile,US Core DiagnosticReport Profile for Report and Note exchange	DocumentReference,DiagnosticReport
Procedure Note	US Core DocumentReference Profile,US Core DiagnosticReport Profile for Report and Note exchange	DocumentReference,DiagnosticReport
Progress Note	US Core DocumentReference Profile	DocumentReference
Consultation Note	US Core DocumentReference Profile	DocumentReference
Goals:		
Patient Goals	US Core Goal Profile	Goal
Health Concerns	US Core Condition Profile	Condition
Immunizations	US Core Immunization Profile	Immunization
Laboratory:		
Tests	US Core Laboratory Result Observation Profile, US Core DiagnosticReport Profile for Laboratory Results Reporting	Observation, DiagnosticReport
Values/Results	US Core Laboratory Result Observation Profile, US Core DiagnosticReport Profile for Laboratory Results Reporting	Observation, DiagnosticReport
Medications:		
Medications	US Core Medication Profile, , US Core Medication Request Profile	Medication, MedicationStatement
Medication Allergies	US Core Allergies Profile	AllergyIntolerance
Patient Demographics:		
First Name	US Core Patient Profile	Patient.name.given
Last Name	US Core Patient Profile	Patient.name.family
Previous Name	US Core Patient Profile	Patient.name
Middle Name (including middle initial)	US Core Patient Profile	Patient.name.given
Suffix	US Core Patient Profile	Patient.name.suffix
Birth Sex	US Core Patient Profile	US Core Birth Sex Extension
Date of Birth	US Core Patient Profile	Patient.birthDate
Race	US Core Patient Profile	US Core Race Extension

https://www.hl7.org/fhir/us/core/general-guidance.html

3.3.1 StructureDefinition-us-core-observation-lab

Laboratory results are grouped and summarized using the DiagnosticReport resource which reference Observation resources. Each Observation resource represents an individual laboratory test and result value, a "nested" panel (such as a microbial susceptibility panel) which references other observations, or rarely a laboratory test with component result values. This profile sets minimum expectations for the Observation resource resource to record, search, and fetch laboratory test results associated with a patient. It identifies which core elements, extensions, vocabularies and value sets **SHALL** be present in the resource when using this profile.

Example Usage Scenarios:

The following are example usage scenarios for the US Core-Results profile:

- Query for lab results belonging to a Patient
- Record or update lab results belonging to a Patient

3.3.1.1 Mandatory and Must Support Data Elements

The following data-elements are mandatory (i.e data MUST be present) or must be supported if the data is present in the sending system (Must Support definition). They are presented below in a simple human-readable explanation. Profile specific guidance and examples are provided as well. The Formal Profile Definition below provides the formal summary, definitions, and terminology requirements.

Each Observation must have:

1. a status
2. a category code of 'laboratory'
3. a LOINC code, if available, which tells you what is being measured
4. a patient

Each Observation must support:

1. a time indicating when the measurement was taken
2. a result value or a reason why the data is absent
    - if the result value is a numeric quantity, a standard UCUM unit

Profile specific implementation guidance:

- Additional codes that translate or map to the Observation code or category codes are allowed. For example:
    - providing both a local code and LOINC code
    - providing a more specific category codes such as 'chemistry', SNOMED CT concept, or system specific codes in addition to the 'laboratory' category code.

FHIR US Core Laboratory Example


```
"status" : "final",
"category" : [
  {
    "coding" : [
      {
        "system" : "http://terminology.hl7.org/CodeSystem/observation-category",
        "code" : "laboratory",
        "display" : "Laboratory"
      }
    ],
    "text" : "Laboratory"
  }
],
"code" : {
  "coding" : [
    {
      "system" : "http://loinc.org",
      "code" : "1975-2",
      "display" : "Bilirub SerPl-mCnc"
    }
  ],
  "text" : "Bilirub SerPl-mCnc"
},
"subject" : {
  "reference" : "Patient/example",
  "display" : "Amy Shaw"
},
"effectiveDateTime" : "2005-07-07",
"valueQuantity" : {
  "value" : 8.6,
  "unit" : "mg/dL",
  "system" : "http://unitsofmeasure.org"
},
"referenceRange" : [
  {
    "low" : {
      "value" : 2.0,
      "unit" : "mg/dL",
      "system" : "http://unitsofmeasure.org",
      "code" : "mg/dL"
    },
    "high" : {
      "value" : 7.0,
      "unit" : "mg/dL",
      "system" : "http://unitsofmeasure.org",
      "code" : "mg/dL"
    },
    "appliesTo" : [
      {
        "coding" : [
          {
            "system" : "http://terminology.hl7.org/CodeSystem/referencerange-meaning",
            "code" : "normal",
            "display" : "Normal Range"
          }
        ],
        "text" : "Normal Range"
      }
    ]
  }
```

Population Level Data

- Cures Act mandates the support for both patient and population-level data export
- Does not specify the format for population-level export, but does mandate adoption of the “FHIR Bulk Data Access” (aka BulkFHIR, aka FlatFHIR) group-export functionality

Back to Phenotyping

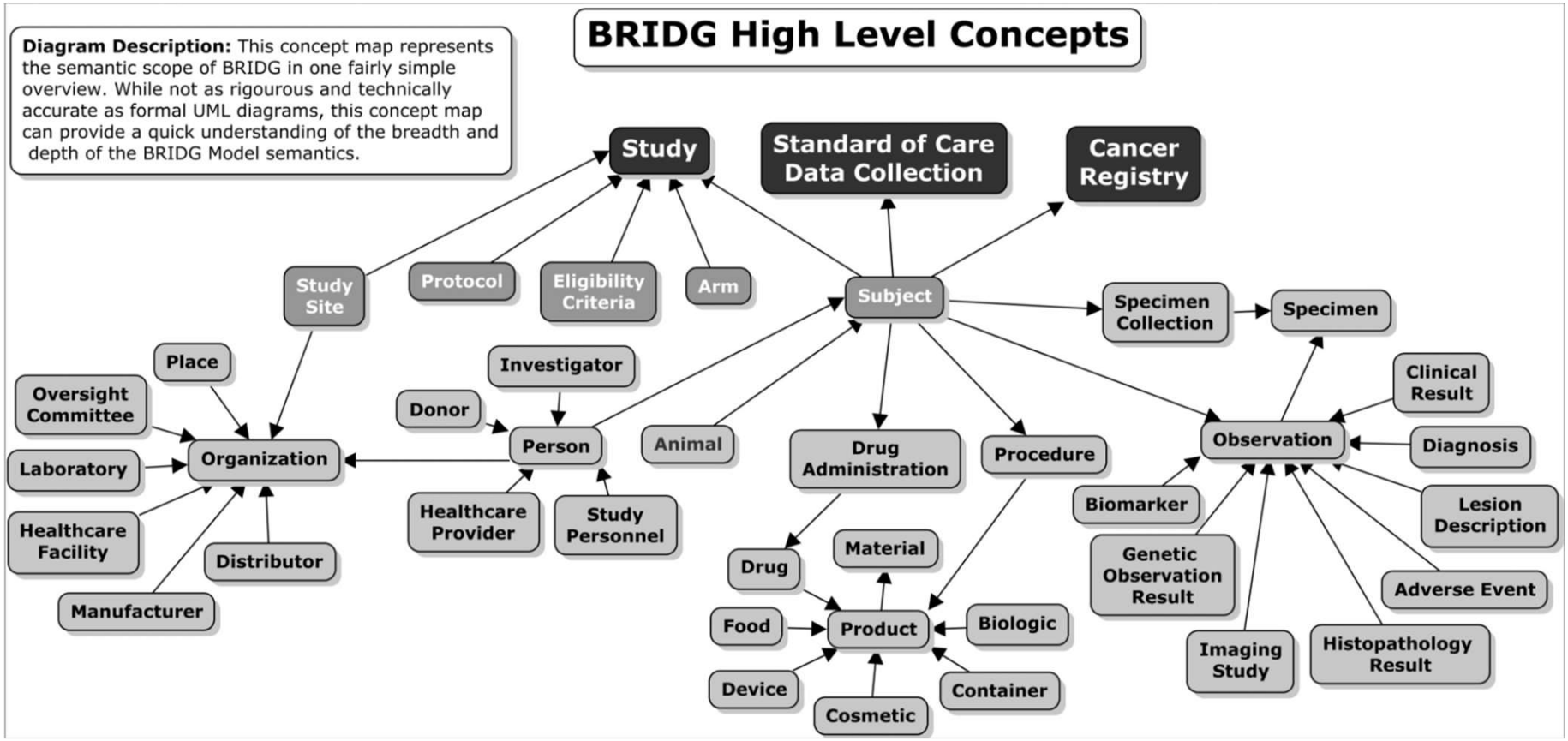
What makes a phenotype portable?

- Uses data types that can be found across systems
- Includes codes that can be found across systems
- Uses logic that can be applied across systems

What makes a phenotype portable?

- **Uses data types that can be found across systems**
- Includes codes that can be found across systems
- Uses logic that can be applied across systems

Data Types in Common Models



https://bridgmodel.nci.nih.gov/ https://cbiit.github.io/bridg-model/HTML/BRIDG5.3.1/

Data Types in Common Models

Common Data Model Harmonization

- Patient Centered Outcome Research Network (PCORNet) Common Data Model (CDM)
- Informatics for Integration Biology and Bedside (i2b2) model
- Observational Medical Outcomes Partnership (OMOP) model
- Federal Drug Administration's Sentinel model

FDA

NIH National Library of Medicine

NIH National Center for Advancing Translational Sciences

1	SENTINEL ID	PCORNet ID	PCORNet v4 ID	i2b2 ID	OMOP ID	CDMH CONCEPTUAL MAPPING	Tagged BRIDG Class Name	Tagged BRIDG Attribute Name
88	4.3.PROCEDURE.0	05.PROC.0	05.PROC.0	PROC.0	PROCEDURE_OCCURRENCE.0	PerformedProcedure	PerformedProcedure	
89		05.PROC.01	05.PROC.01		PROCEDURE_OCCURRENCE.1	PerformedProcedure.identifier(DSET<ID>).item(ID).identifier	Activity	identifier
90	4.3.PROCEDURE.1	05.PROC.02	05.PROC.02		PROCEDURE_OCCURRENCE.2	PerformedProcedure > Subject.identifier(ID).identifier	Subject	identifier
91	4.3.PROCEDURE.2	05.PROC.03	05.PROC.03		PROCEDURE_OCCURRENCE.10	PerformedProcedure > PerformedCompositionRelationship > PerformedEncounter.identifier.item(ID).identifier	Activity	identifier
92	4.3.PROCEDURE.3	05.PROC.05	05.PROC.05			PerformedProcedure > PerformedCompositionRelationship > PerformedEncounter.dateRange(IVL<TS.DATETIME>).low	PerformedActivity	dateRange
93	4.3.PROCEDURE.4	05.PROC.06	05.PROC.06		PROCEDURE_OCCURRENCE.9	PerformedProcedure > Performer > HealthcareProvider.identifier(DSET<ID>).item(ID).identifier	HealthcareProvider	identifier
94	4.3.PROCEDURE.5	05.PROC.04	05.PROC.04			PerformedProcedure > PerformedCompositionRelationship > PerformedEncounter > DefinedSubjectActivityGroup.nameCode	DefinedActivity	nameCode
95		05.PROC.07	05.PROC.07		PROCEDURE_OCCURRENCE.4	PerformedProcedure.dateRange(IVL<TS.DATETIME>).high	PerformedActivity	dateRange
96				PROC.4	PROCEDURE_OCCURRENCE.5	PerformedProcedure.dateRange(IVL<TS.DATETIME>).high	PerformedActivity	dateRange

https://www.healthit.gov/topic/scientific-initiatives/pcor/common-data-model-harmonization-cdm

CDMs to FHIR US core

PCORNet
OMOP
i2b2
Sentinel

OMOP Data Element	FHIR Data Element	FHIR Resource/Profile/Extension
Table : Person		
person_id	Patient.identifier	us-core-patient
provider_id	Patient.generalPractitioner	Patient
care_site_id	BodySite.patient	BodySite
gender_concept_id	Patient.gender	us-core-patient
year_of_birth	Patient.birthDate	us-core-patient
month_of_birth	Patient.birthDate	us-core-patient
day_of_birth	Patient.birthDate	us-core-patient
birth_datetime	Patient.birthDate	us-core-patient
race_concept_id	Patient.extension: us-core-race	us-core-patient
ethnicity_concept_id	Patient.extension: us-core-ethnicity	us-core-patient
location_id	Patient.address	Patient
Table : VISIT_OCCURRENCE		
visit_occurence_id	Encounter.id	us-core-encounter
care_site_id	Encounter.location.location.identifier	us-core-encounter, us-core-location
admitting_source_concept_id	Encounter.hospitalization.admitSource or Encounter.hospitalization.origin(location).type	us-core-encounter, us-core-location
discharge_to_concept_id	Encounter.location.location.type	us-core-encounter,us-core-location
preceding_visit_occurence	Encounter.partOf	us-core-encounter
person_id	Encounter.subject	us-core-encounter
visit_concept_id	Encounter.type	us-core-encounter
visit_start_date	Encounter.period	us-core-encounter
visit_start_datetime	Encounter.period	us-core-encounter
visit_end_date	Encounter.period	us-core-encounter
visit_end_datetime	Encounter.period	us-core-encounter
visit_type_concept_id	Encounter.extension (Proposed Name: source-data-type : CodeableConcept)	us-core-encounter
Table : CARE_SITE		
care_site_id	Location.id	us-core-location
care_site_name	Location.name	us-core-location
place_of_service_concept_id	Location.type	us-core-location
location_id	Location.address	us-core-location
Table : CONDITION_OCCURRENCE		
condition_occurrence_id	Condition.id	us-core-condition
provider_id	Condition.asserter	us-core-condition
visit_occurrence_id	Condition.encounter	us-core-condition
condition_status_concept_id	Condition.clinicalStatus	us-core-condition
person_id	Condition.subject	us-core-condition
condition_concept_id	Condition.code	us-core-condition

http://build.fhir.org/ig/HL7/cdmh/profiles.html

Playing the Classics vs Skating to the Puck

- Ensuring the portability of phenotypes at present still requires focusing on the classic data types
  - Conditions
  - Medications
  - Procedures
  - Observations / Labs (presence of)
  - Encounters
- But adoption of USCDI and FHIR Core expands the potential to incorporate elements such as vital signs, laboratory results, and clinical notes from EHR sources

What makes a phenotype portable?

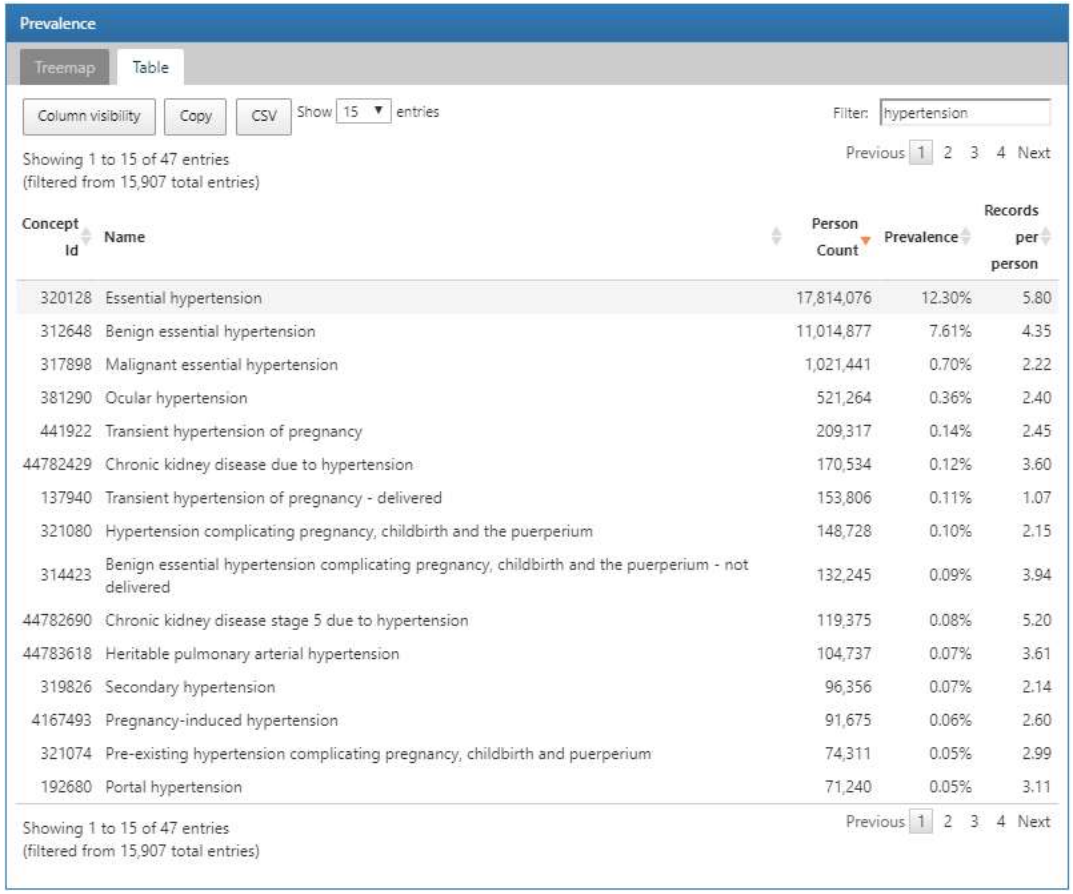
- Uses data types that can be found across systems
- **Includes codes that can be found across systems**
- Uses logic that can be applied across systems

How do we know what codes are in what system?

- New cohort characterization tools greatly expand our ability to understand the availability and impact of code selection
- Specifically we are seeking to understand what codes are actually used in cohort generation and how that differs from site to site

Spotlight: OHDSI Characterization Tools

Older OHDSI tools let you look at counts of codes for a given dataset or a generated cohort



Prevalence

Treemap | Table

Column visibility | Copy | CSV | Show 15 entries | Filter: hypertension

Showing 1 to 15 of 47 entries
(filtered from 15,907 total entries)

Previous 1 2 3 4 Next

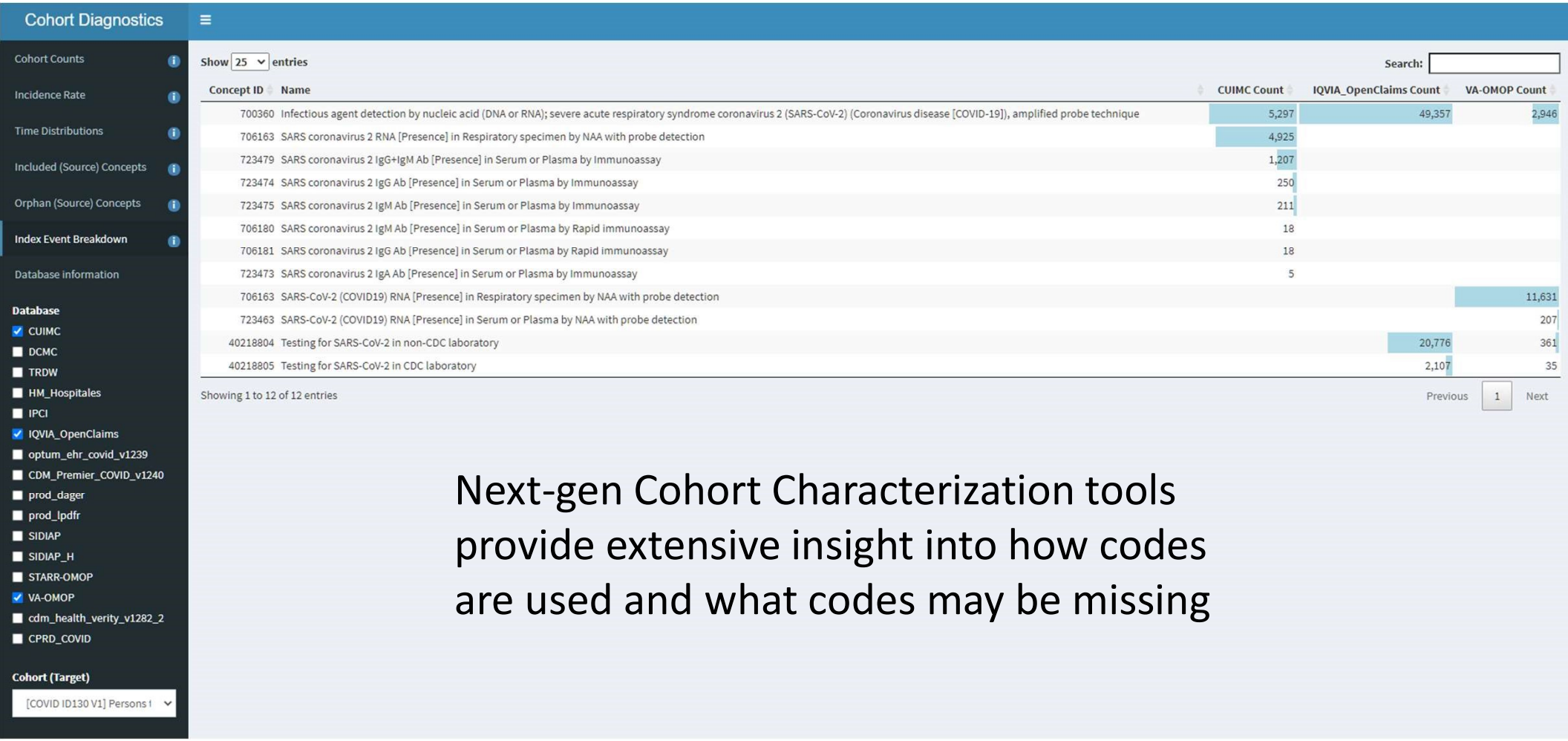
Concept Id	Name	Person Count	Prevalence	Records per person
320128	Essential hypertension	17,814,076	12.30%	5.80
312648	Benign essential hypertension	11,014,877	7.61%	4.35
317898	Malignant essential hypertension	1,021,441	0.70%	2.22
381290	Ocular hypertension	521,264	0.36%	2.40
441922	Transient hypertension of pregnancy	209,317	0.14%	2.45
44782429	Chronic kidney disease due to hypertension	170,534	0.12%	3.60
137940	Transient hypertension of pregnancy - delivered	153,806	0.11%	1.07
321080	Hypertension complicating pregnancy, childbirth and the puerperium	148,728	0.10%	2.15
314423	Benign essential hypertension complicating pregnancy, childbirth and the puerperium - not delivered	132,245	0.09%	3.94
44782690	Chronic kidney disease stage 5 due to hypertension	119,375	0.08%	5.20
44783618	Heritable pulmonary arterial hypertension	104,737	0.07%	3.61
319826	Secondary hypertension	96,356	0.07%	2.14
4167493	Pregnancy-induced hypertension	91,675	0.06%	2.60
321074	Pre-existing hypertension complicating pregnancy, childbirth and puerperium	74,311	0.05%	2.99
192680	Portal hypertension	71,240	0.05%	3.11

Showing 1 to 15 of 47 entries
(filtered from 15,907 total entries)

Previous 1 2 3 4 Next

https://ohdsi.github.io/TheBookOfOhdsi/Characterization.html

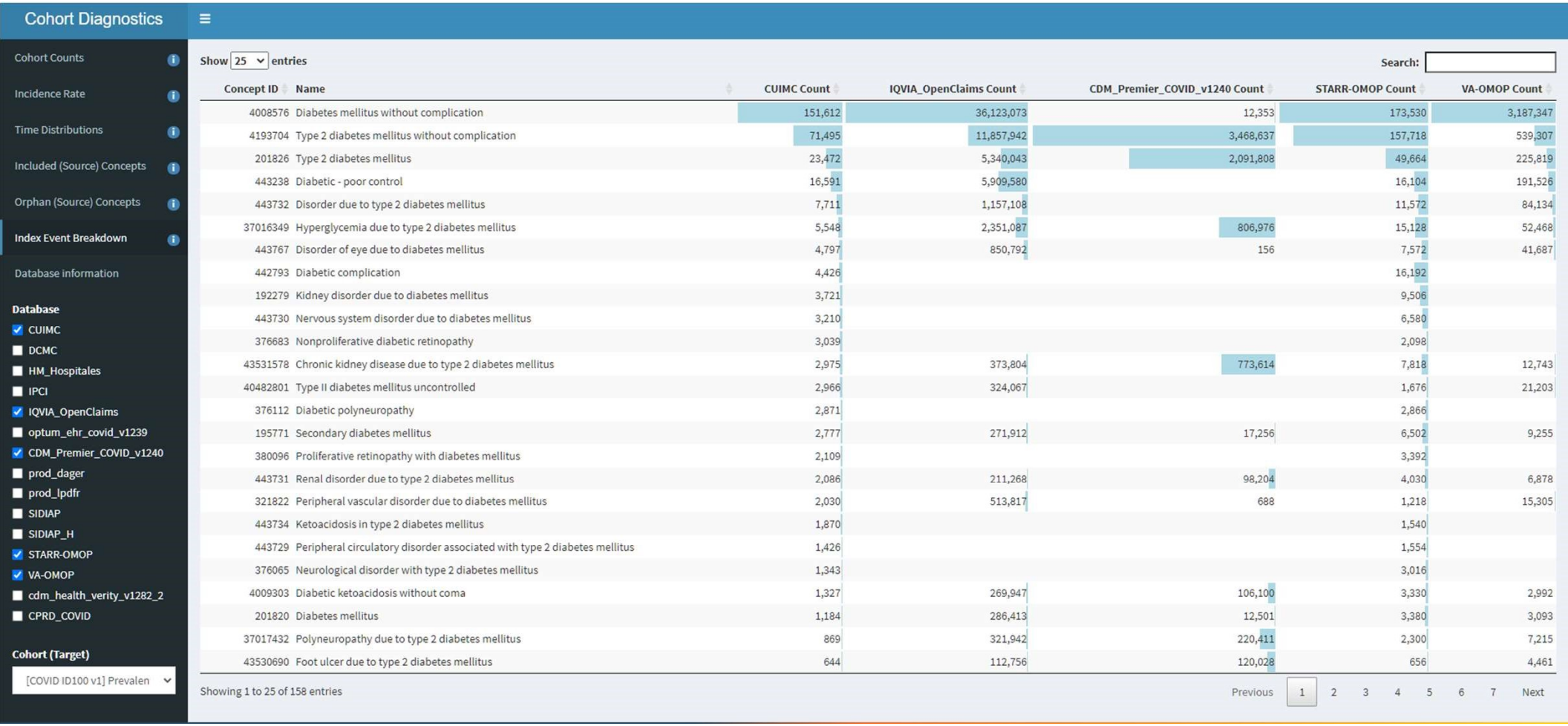
Spotlight: OHDSI Characterization Tools



Next-gen Cohort Characterization tools provide extensive insight into how codes are used and what codes may be missing

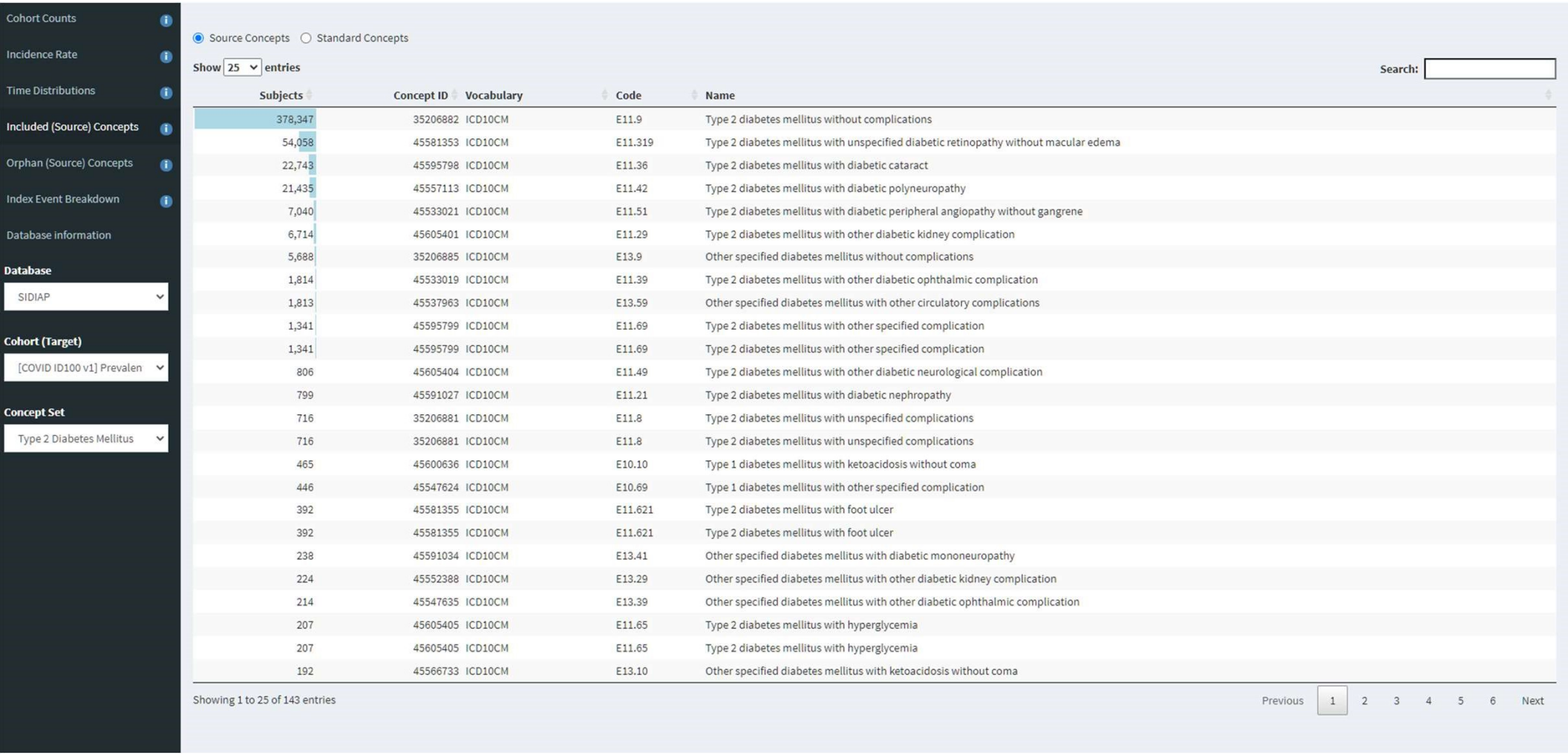
https://data.ohdsi.org/Covid19CharacterizationCharybdisDiagCovid/

How are diabetes patients detected at different sites?



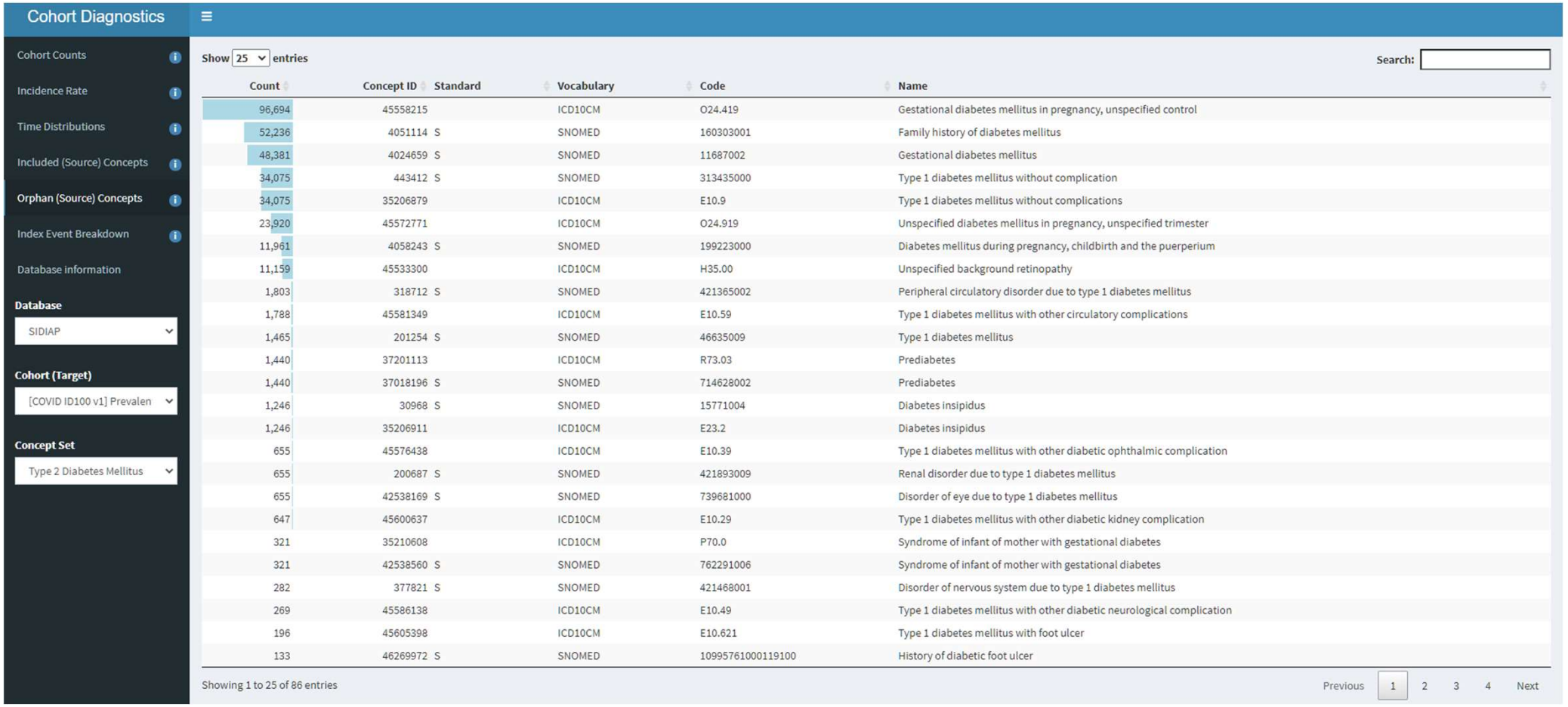
https://data.ohdsi.org/Covid19CharacterizationCharybdisDiagStrata/

What codes included in the definition are found?



Subjects	Concept ID	Vocabulary	Code	Name
378,347	35206882	ICD10CM	E11.9	Type 2 diabetes mellitus without complications
54,058	45581353	ICD10CM	E11.319	Type 2 diabetes mellitus with unspecified diabetic retinopathy without macular edema
22,743	45595798	ICD10CM	E11.36	Type 2 diabetes mellitus with diabetic cataract
21,435	45557113	ICD10CM	E11.42	Type 2 diabetes mellitus with diabetic polyneuropathy
7,040	45533021	ICD10CM	E11.51	Type 2 diabetes mellitus with diabetic peripheral angiopathy without gangrene
6,714	45605401	ICD10CM	E11.29	Type 2 diabetes mellitus with other diabetic kidney complication
5,688	35206885	ICD10CM	E13.9	Other specified diabetes mellitus without complications
1,814	45533019	ICD10CM	E11.39	Type 2 diabetes mellitus with other diabetic ophthalmic complication
1,813	45537963	ICD10CM	E13.59	Other specified diabetes mellitus with other circulatory complications
1,341	45595799	ICD10CM	E11.69	Type 2 diabetes mellitus with other specified complication
1,341	45595799	ICD10CM	E11.69	Type 2 diabetes mellitus with other specified complication
806	45605404	ICD10CM	E11.49	Type 2 diabetes mellitus with other diabetic neurological complication
799	45591027	ICD10CM	E11.21	Type 2 diabetes mellitus with diabetic nephropathy
716	35206881	ICD10CM	E11.8	Type 2 diabetes mellitus with unspecified complications
716	35206881	ICD10CM	E11.8	Type 2 diabetes mellitus with unspecified complications
465	45600636	ICD10CM	E10.10	Type 1 diabetes mellitus with ketoacidosis without coma
446	45547624	ICD10CM	E10.69	Type 1 diabetes mellitus with other specified complication
392	45581355	ICD10CM	E11.621	Type 2 diabetes mellitus with foot ulcer
392	45581355	ICD10CM	E11.621	Type 2 diabetes mellitus with foot ulcer
238	45591034	ICD10CM	E13.41	Other specified diabetes mellitus with diabetic mononeuropathy
224	45552388	ICD10CM	E13.29	Other specified diabetes mellitus with other diabetic kidney complication
214	45547635	ICD10CM	E13.39	Other specified diabetes mellitus with other diabetic ophthalmic complication
207	45605405	ICD10CM	E11.65	Type 2 diabetes mellitus with hyperglycemia
207	45605405	ICD10CM	E11.65	Type 2 diabetes mellitus with hyperglycemia
192	45566733	ICD10CM	E13.10	Other specified diabetes mellitus with ketoacidosis without coma

What potentially relevant codes were not in the definition?



https://data.ohdsi.org/Covid19CharacterizationCharybdisDiagStrata/

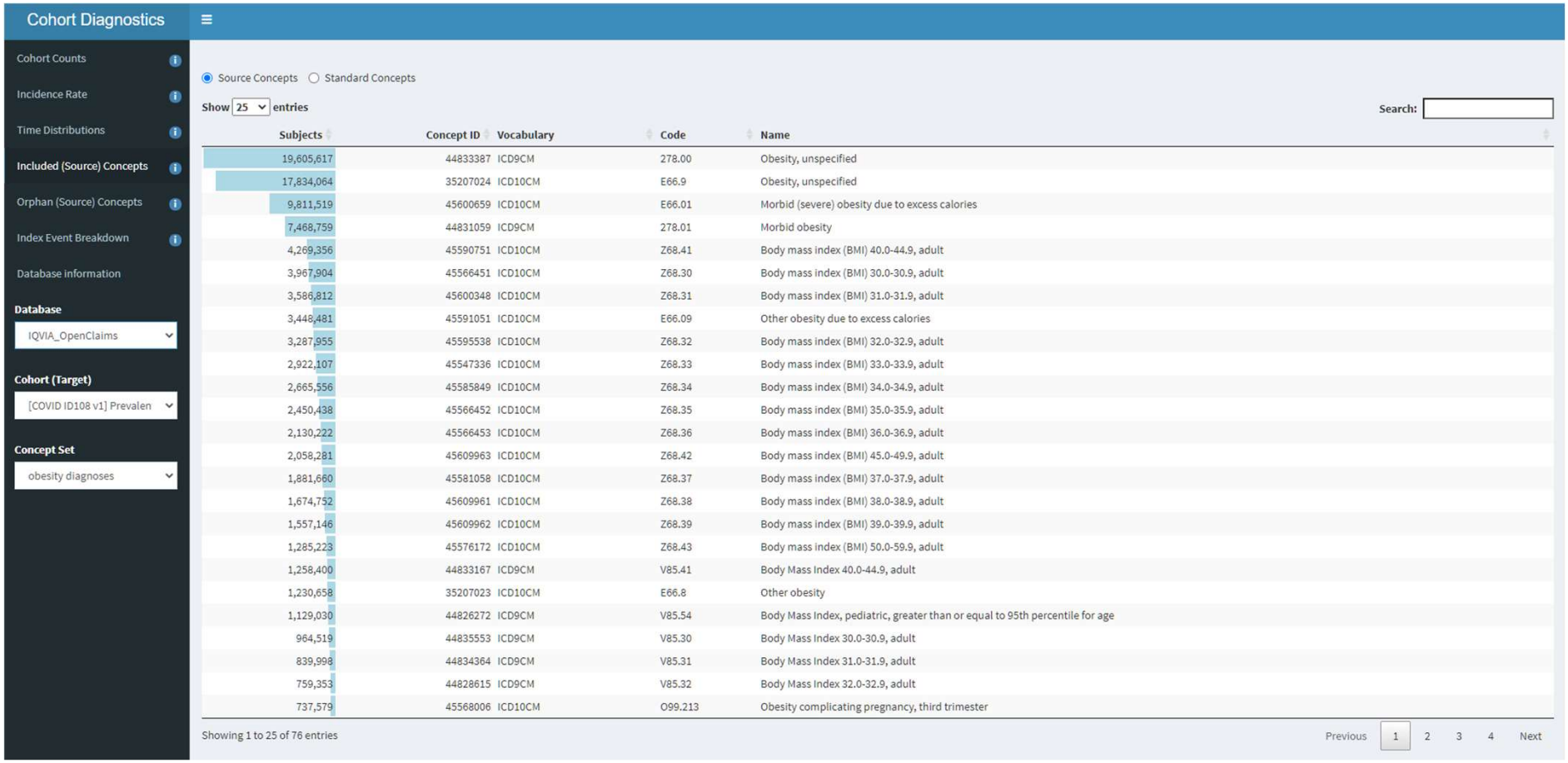
What about Obesity?

Concept ID	Name	CUIMC Count	IQVIA_OpenClaims Count	CDM_Premier_COVID_v1240 Count	STARR-OMOP Count	VA-OMOP Count
3025315	Body weight	488,882		341	346,855	2,926,259
3038553	Body mass index (BMI) [Ratio]	485,396			340,567	
3027492	Dry body weight Measured	152,568				
433736	Obesity	106,801	31,355,725	2,209,239	76,399	3,312,894
434005	Morbid obesity	42,115	9,764,068	1,270,926	25,091	249,938
42872398	Maternal obesity complicating pregnancy, childbirth and the puerperium, antepartum	4,890	1,299,918	93,910	3,870	926
439893	Maternal obesity syndrome	3,123	401,402	3,793	641	1,464
3013762	Body weight Measured	3,011				<5
4060985	Body mass index 30+ - obesity	2,026	13,374,128	1,486,249	5,525	412,896
438731	Localized adiposity	1,640	239,261	10,621	2,185	9,182
4256640	Body mass index 40+ - severely obese	1,292	3,608,905	972,849	2,459	56,019
40481140	Childhood obesity	688	973,161		4,721	73
3023166	Body weight Stated	201			103	
4100857	Extreme obesity with alveolar hypoventilation	122	129,478	39,878	652	3,171
4097996	Drug-induced obesity	112	25,632	1,737	97	641
380500	Hypertrophy of fat pad of knee	65	35,156		168	1,935
4029277	Fat pad syndrome	60	46,048	1,983	184	1,344
45757112	Obesity in mother complicating childbirth	38	144,328	114,934	102	928
4171147	Hypertrophic obesity				<10	
4217557	Simple obesity		1,325		305	
37018860	Severe obesity				341	

Showing 1 to 21 of 21 entries

Previous 1 Next

Diagnosis Codes found in Claims



Subjects	Concept ID	Vocabulary	Code	Name
19,605,617	44833387	ICD9CM	278.00	Obesity, unspecified
17,834,064	35207024	ICD10CM	E66.9	Obesity, unspecified
9,811,519	45600659	ICD10CM	E66.01	Morbid (severe) obesity due to excess calories
7,468,759	44831059	ICD9CM	278.01	Morbid obesity
4,269,356	45590751	ICD10CM	Z68.41	Body mass index (BMI) 40.0-44.9, adult
3,967,904	45566451	ICD10CM	Z68.30	Body mass index (BMI) 30.0-30.9, adult
3,586,812	45600348	ICD10CM	Z68.31	Body mass index (BMI) 31.0-31.9, adult
3,448,481	45591051	ICD10CM	E66.09	Other obesity due to excess calories
3,287,955	45595538	ICD10CM	Z68.32	Body mass index (BMI) 32.0-32.9, adult
2,922,107	45547336	ICD10CM	Z68.33	Body mass index (BMI) 33.0-33.9, adult
2,665,556	45585849	ICD10CM	Z68.34	Body mass index (BMI) 34.0-34.9, adult
2,450,438	45566452	ICD10CM	Z68.35	Body mass index (BMI) 35.0-35.9, adult
2,130,222	45566453	ICD10CM	Z68.36	Body mass index (BMI) 36.0-36.9, adult
2,058,281	45609963	ICD10CM	Z68.42	Body mass index (BMI) 45.0-49.9, adult
1,881,660	45581058	ICD10CM	Z68.37	Body mass index (BMI) 37.0-37.9, adult
1,674,752	45609961	ICD10CM	Z68.38	Body mass index (BMI) 38.0-38.9, adult
1,557,146	45609962	ICD10CM	Z68.39	Body mass index (BMI) 39.0-39.9, adult
1,285,223	45576172	ICD10CM	Z68.43	Body mass index (BMI) 50.0-59.9, adult
1,258,400	44833167	ICD9CM	V85.41	Body Mass Index 40.0-44.9, adult
1,230,658	35207023	ICD10CM	E66.8	Other obesity
1,129,030	44826272	ICD9CM	V85.54	Body Mass Index, pediatric, greater than or equal to 95th percentile for age
964,519	44835553	ICD9CM	V85.30	Body Mass Index 30.0-30.9, adult
839,998	44834364	ICD9CM	V85.31	Body Mass Index 31.0-31.9, adult
759,353	44828615	ICD9CM	V85.32	Body Mass Index 32.0-32.9, adult
737,579	45568006	ICD10CM	O99.213	Obesity complicating pregnancy, third trimester

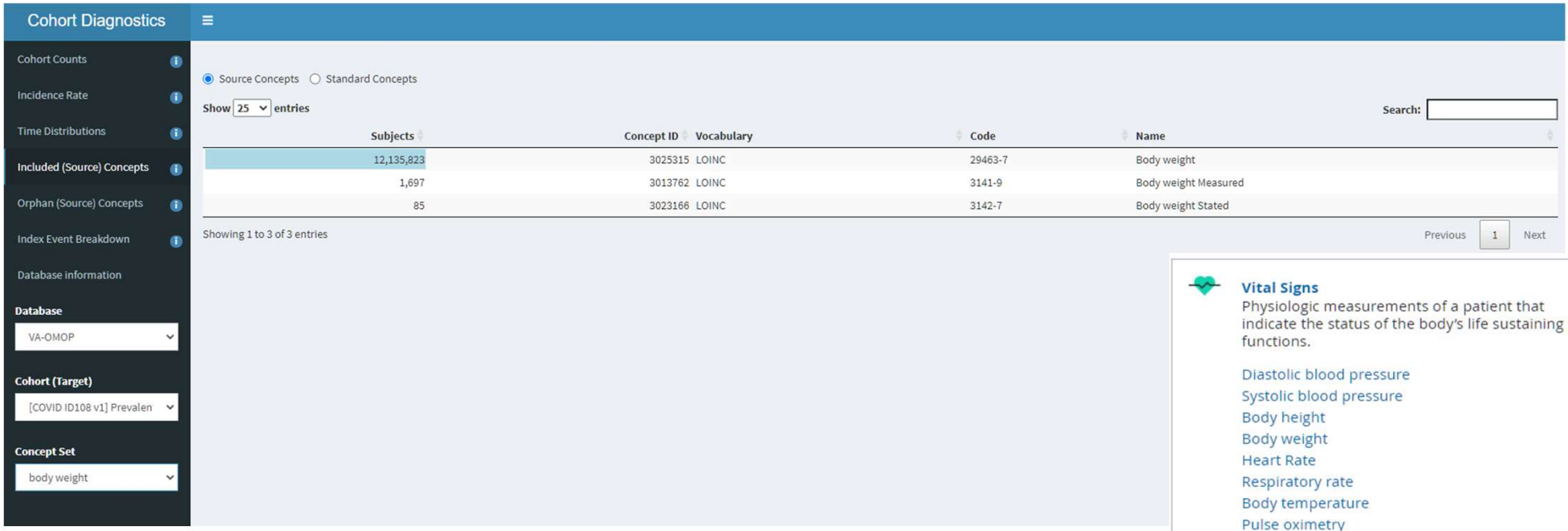
Missing codes, some missing for a reason

Count	Concept ID	Standard	Vocabulary	Code	Name
7,474	2100973	S	CPT4	00797	Anesthesia for intraperitoneal procedures in upper abdomen including laparoscopy; gastric restrictive procedure for morbid obesity
588	44819525		ICD9CM	278.8	Other hyperalimentation
244	2109000	S	CPT4	43843	Gastric restrictive procedure, without gastric bypass, for morbid obesity; other than vertical-banded gastroplasty
157	2109003	S	CPT4	43847	Gastric restrictive procedure, with gastric bypass for morbid obesity; with small intestine reconstruction to limit absorption
154	2102458	S	CPT4	15821	Blepharoplasty, lower eyelid; with extensive herniated fat pad
125	2109002	S	CPT4	43846	Gastric restrictive procedure, with gastric bypass for morbid obesity; with short limb (150 cm or less) Roux-en-Y gastroenterostomy
101	4149383	S	SNOMED	268551005	Obesity screening
101	44828608		ICD9CM	V77.8	Screening for obesity
85	2108468	S	CPT4	38520	Biopsy or excision of lymph node(s); open, deep cervical node(s) with excision scalene fat pad
79	2109004	S	CPT4	43848	Revision, open, of gastric restrictive procedure for morbid obesity, other than adjustable gastric restrictive device (separate procedure)
50	40664664	S	HCPCS	G0447	Face-to-face behavioral counseling for obesity, 15 minutes
47	2108999	S	CPT4	43842	Gastric restrictive procedure, without gastric bypass, for morbid obesity; vertical-banded gastroplasty
8	2102494	S	CPT4	15838	Excision, excessive skin and subcutaneous tissue (includes lipectomy); submental fat pad
5	2101698	S	CPT4	0156T	Laparoscopy, surgical; revision or removal of gastric stimulation electrodes, lesser curvature (ie, morbid obesity)

Showing 1 to 14 of 14 entries

Previous 1 Next

Weight Codes found in EHR



What makes a phenotype portable?

- Uses data types that can be found across systems
- Includes codes that can be found across systems
- **Uses logic that can be applied across systems**

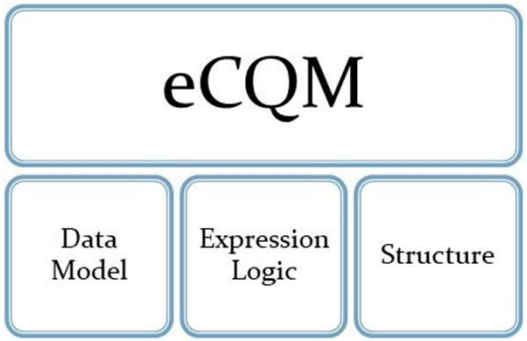
Standards for required logic not well-defined

- Unlike data elements and terminologies, there is not a defining set of logical querying capabilities specified, eg
  - Temporal logic (+/- relative to index event)
  - Count of occurrences
  - Groupings (any/all, and or not)
- But hope can be found in the CMS electronic clinical quality measures (eCQMs) that are being transitioned to broad interoperability standards

What is an eCQM?

- Clinical Quality Measures are required reporting for all providers and hospitals that participate in Medicare/Medicaid
  - Performance on these measures is tied to reimbursement
- Each CQM is defined by numerator and denominator cohorts
- An eCQM is a computable phenotype representing these cohorts to ensure consistency in the reporting processing
- They include a data model, code sets, and logic

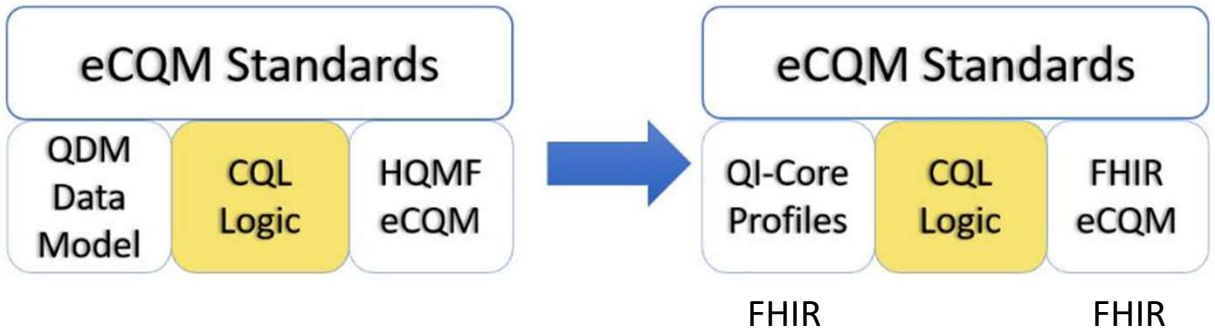
eCQMs' Logic is Expressed Using CQL (Clinical Quality Language)



Data Model: How to describe the patient's medical record data needed to calculate the measure

Expression Logic: How to calculate the result and evaluate the performance

Structure: The container and sections describing measure metadata, numerator, denominator, exclusions, exceptions



Clinical Quality Language (CQL)

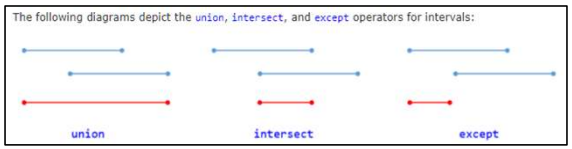
- HL7's CQL aims to be both human readable and machine interpretable
- Leverages both embedded and NLM VSAC value sets
- Has extensive date manipulation logic
- Gets compiled into a query execution mechanism such as FHIR[1]

```
[Encounter: "Ambulatory/ED Visit"] E
  with [Condition: "Acute Pharyngitis"] P
    such that P.onsetDateTime during E.period
      and P.abatementDate after end of E.period
```

```
Concept {
  Code '66071002' from "SNOMED-CT",
  Code 'B18.1' from "ICD-10-CM"
} display 'Type B viral hepatitis'
```

```
valueset "Acute Pharyngitis": 'urn:oid:2.16.840.1.113883.3.464.1003.102.12.1011'
valueset "Acute Tonsillitis": 'urn:oid:2.16.840.1.113883.3.464.1003.102.12.1012'
valueset "Ambulatory/ED Visit": 'urn:oid:2.16.840.1.113883.3.464.1003.101.12.1061'
```

```
X starts 3 days before start Y
```



1. https://github.com/DBCG/cql_engine

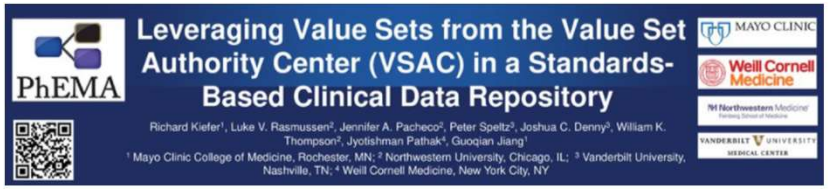
```
valueset "Flutter Diagnosis": '2.16.840.1.113883.3.117.1.7.1.202'
valueset "Warfarin Medication": '2.16.840.1.113883.3.117.1.7.1.232'
valueset "Face-to-face Encounter": '2.16.840.1.113883.3.464.1003.101.12.1048'
valueset "Office Visit": '2.16.840.1.113883.3.464.1003.101.12.1001'
```

Code System	Code System Version	Code System OID	Code	Description
SNOMED CT®			5370000	Atrial flutter (disorder)
SNOMED CT®			49436004	Atrial fibrillation (disorder)
SNOMED CT®			440059007	Persistent atrial fibrillation (disorder)
SNOMED CT®			440028005	Permanent atrial fibrillation (disorder)
SNOMED CT®			429218009	History of maze procedure for atrial fibrillation (situation)
SNOMED CT®			428076002	History of atrial flutter (situation)
SNOMED CT®			427665004	Paroxysmal atrial flutter (disorder)
SNOMED CT®			427665004	Paroxysmal atrial fibrillation (disorder)
SNOMED CT®			426814001	Transient cerebral ischemia due to atrial fibrillation (disorder)
SNOMED CT®			426749004	Chronic atrial fibrillation (disorder)
SNOMED CT®			425615007	Chronic atrial flutter (disorder)
SNOMED CT®			425615007	Chronic atrial fibrillation (disorder)
SNOMED CT®			314208002	Rapid atrial fibrillation (disorder)
SNOMED CT®			312442005	History of - atrial fibrillation (situation)
SNOMED CT®			300996004	Controlled atrial fibrillation (disorder)
SNOMED CT®			282825002	Paroxysmal atrial fibrillation (disorder)
SNOMED CT®			233911009	Non-rheumatic atrial fibrillation (disorder)
SNOMED CT®			233910005	Lone atrial fibrillation (disorder)
SNOMED CT®			195082009	Atrial fibrillation and flutter NOS (disorder)
SNOMED CT®			195080001	Atrial fibrillation and flutter (disorder)
SNOMED CT®			164890007	Electrocardiogram: atrial flutter (finding)
SNOMED CT®			164889003	Electrocardiogram: atrial fibrillation (finding)

```
define "ActiveWarfarinDuringLookback":
  "WarfarinMedications" M where M."effectiveTimePeriod" overlaps "LookbackInterval"

define "WarfarinUsageIntervals":
  collapse
    "ActiveWarfarinDuringLookback" M
      return M."effectiveTimePeriod" intersect "LookbackInterval"

define "WarfarinUsage": Sum("WarfarinUsageIntervals" I return duration in days of I)
```

```
valueset "Flutter Diagnosis": '2.16.840.1.113883.3.117.1.7.1.202'
valueset "Warfarin Medication": '2.16.840.1.113883.3.117.1.7.1.232'
valueset "Face-to-face Encounter": '2.16.840.1.113883.3.464.1003.101.12.1048'
valueset "Office Visit": '2.16.840.1.113883.3.464.1003.101.12.1001'
valueset "Valvular Heart Disease": '2.16.840.1.113883.3.464.1003.104.12.1017'
valueset "INR Lab Result": '2.16.840.1.113883.3.117.1.7.1.213'

parameter MeasurementPeriod default Interval[DateTime(2013, 1, 1, 0, 0, 0, 0), DateTime(2014, 1, 1, 0, 0, 0, 0))

context Patient

define "FlutterDiagnoses": ["Condition": "Flutter Diagnosis"]
define "WarfarinMedications": ["MedicationAdministration": "Warfarin Medication"]
define "FaceToFaceEncounters": ["Encounter": "Face-to-face Encounter"]
define "OfficeVisitEncounters": ["Encounter": "Office Visit"]
define "ValvularHeartDiseaseDiagnoses": ["Condition": "Valvular Heart Disease"]
define "INRLabResults": ["Observation": "INR Lab Result"]

define "InDemographic":
  AgeInYearsAt(start of MeasurementPeriod) >= 18

define "InpatientEncounters": "FaceToFaceEncounters" union "OfficeVisitEncounters"
define "ActiveFlutterDiagnoses": "FlutterDiagnoses" F where Interval[F."onsetDateTime", F."abatementDate"] overlaps before MeasurementPeriod
define "ActiveValvularHeartDiseaseDiagnoses": "ValvularHeartDiseaseDiagnoses" D where Interval[D."onsetDateTime", D."abatementDate"] overlaps before MeasurementPeriod

define "LookbackInterval": Interval[start of MeasurementPeriod - 200 days, start of MeasurementPeriod]

define "ActiveWarfarinDuringLookback":
  "WarfarinMedications" M where M."effectiveTimePeriod" overlaps "LookbackInterval"

define "WarfarinUsageIntervals":
  collapse
    "ActiveWarfarinDuringLookback" M
      return M."effectiveTimePeriod" intersect "LookbackInterval"

define "WarfarinUsage": Sum("WarfarinUsageIntervals" I return duration in days of I)
```

Sharing code sets with VSAC across platforms



FHIR® Terminology Service for VSAC Resources

The FHIR Terminology Service for VSAC Resources is a RESTful API service for accessing the VSAC value sets and supported code systems.

VSAC SVS API

The VSAC SVS API is based on the IHE SVS Technical Framework, section 2.2.21 Sharing Value Set Integration Profile (SVS), and the IHE SVS XML Schema

emerge network
ELECTRONIC MEDICAL RECORDS & GENOMICS

README.md

VSAC to OMOP

This is a command-line utility that takes a Value Set Authority Center (VSAC) OID for a particular value set, and using the NLM's FHIR API, retrieves the value set definition including codes. It then communicates using the OHDSI WebAPI to create the value set (as a concept set) and populate it with all codes that it is able to find.

Configuration

Currently the program uses a properties file found at `config/config.properties`. This will be changed in the future to use command line options. For now, you should set the following options within that file:

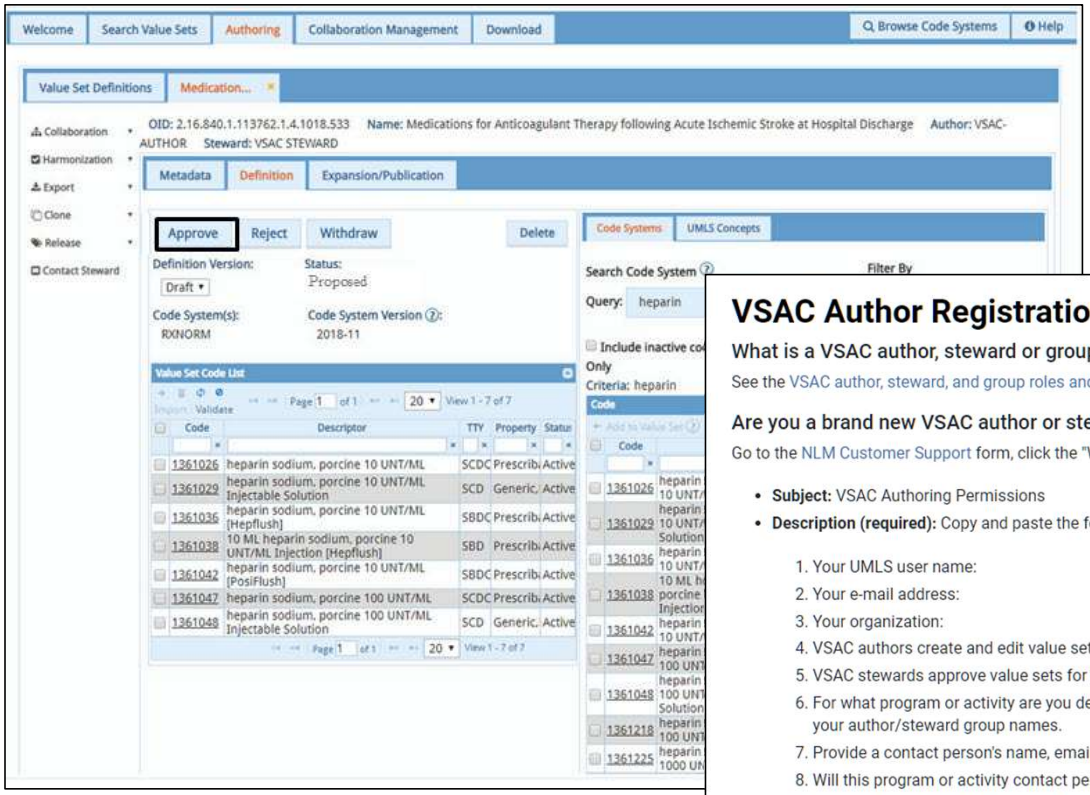
```
UMLS_USER=username                          -- Your UMLS username
UMLS_PASS=password                          -- Your UMLS password
FHIR_URL=cts.nlm.nih.gov/fhir/              -- Note the lack of http(s) in front

VS_NAME=Acute Myocardial Infarction         -- The name of the value set you wish to load
VS_OIDS=2.16.840.1.113883.3.67.1.101.1.317  -- The VSAC OID of the value set you wish to load
VS_VER=20170609                             -- The Definition Version (found from the VSAC UI) of the value s

OMOP_BASE_URL=http://api.ohdsi.org/WebAPI/  -- The base WebAPI URL of the OHDSI instance you want to load to
OMOP_SOURCE=SYNPUF5PCT                      -- The data source in OHDSI
OMOP_CREATE=true                            -- If the value set should actually be created or not.
```

Base URI: https://cts.nlm.nih.gov/fhir

NOTICE:
- The VSAC implementation of FHIR cannot retrieve intensional value sets. Use the SVS API to retrieve intensional value sets.
- VSAC value set identifiers must be specified as a VSAC value set OID. Value set names are not unique in VSAC.

FHIR Resource	FHIR Operation	Examples
ValueSet		https://cts.nlm.nih.gov/fhir/ValueSet/2.16.840.1.113883.3.464.1003.113.11.1090
	$expand	https://cts.nlm.nih.gov/fhir/ValueSet/2.16.840.1.113883.3.464.1003.113.11.1090/$expand
		https://cts.nlm.nih.gov/fhir/ValueSet/2.16.840.1.113883.3.464.1003.113.11.1090/$expand?filter=Ankylos
	$validate-code	https://cts.nlm.nih.gov/fhir/ValueSet/2.16.840.1.113883.3.464.1003.113.11.1090/$validate-code?system=http://hl7.org/fhir/sid/icd-
CodeSystem	$lookup	https://cts.nlm.nih.gov/fhir/CodeSystem/$lookup?system=http://loinc.org&code=1963-8
		https://cts.nlm.nih.gov/fhir/CodeSystem/$lookup?system=http://loinc.org&code=1963-8&version=2.56
		https://cts.nlm.nih.gov/fhir/CodeSystem/$lookup?system=http://loinc.org&code=1963-8&date=20150501
	$subsumes	https://cts.nlm.nih.gov/fhir/CodeSystem/$subsumes?system=http://snomed.info/sct&codeA=29857009&codeB=10000006
		https://cts.nlm.nih.gov/fhir/CodeSystem/$subsumes?system=http://hl7.org/fhir/sid/icd-10-cm&version=2018&codeA=A01.01&code

Sharing code sets with VSAC across platforms



VSAC Author Registration

What is a VSAC author, steward or group?

See the VSAC author, steward, and group roles and permissions page for details about these roles.

Are you a brand new VSAC author or steward and need to create new groups?

Go to the NLM Customer Support form, click the "Write to the help desk" button, and type the following text for these fields:

- **Subject:** VSAC Authoring Permissions
- **Description (required):** Copy and paste the following questions into the description field of the e-mail form, and provide your answers in the same field.

1. Your UMLS user name:
2. Your e-mail address:
3. Your organization:
4. VSAC authors create and edit value sets. Do you require value set authoring permissions? - Yes/No
5. VSAC stewards approve value sets for publication. Do you require steward permissions? - Yes/No
6. For what program or activity are you developing value sets? Include the URL for your program or activity. Note that the program or activity name will be reflected in your author/steward group names.
7. Provide a contact person's name, email, and phone # for the program or activity noted in #6:
8. Will this program or activity contact person be acting as a Steward, providing direct VSAC oversight to your authoring work in the VSAC?
9. If you answered YES to #8, please ensure the contact person (Steward) also requests VSAC Authoring access. When your Steward applies for VSAC Authoring permissions, they should list you as working with them as an Author, a Steward, or both, depending on your working arrangement.

Do you need to join an existing group?

Current group members can add you to the group. See the Group Management page for instructions.

Logic Beyond Structured Data

- Many of the most critical data elements for determining cohort inclusion come from unstructured data
- As noted, USCDI mandates that systems be able to output a core set of 8 common and valuable clinical note types
- But how to cull data from clinical notes while conforming to the broader phenotype structure?

Many Valid Approaches

- Using NLP to prepopulate structured fields
  - Eg OMOP Note table → NLP Derived fields in structured tables
- Using NLP in real-time queries to augment structured feature definition
  - Eg “Stroke” = {Stroke Codeset} OR {[occlusion OR ischemia] on Head CT}
- Using NLP in real-time queries to create unstructured features
  - Eg “> 2cm lung mass”= MeasurementFinder ({mass OR lesion} on Chest XR/CT) value >20 mm

In all cases, you must be able to integrate the NLP derived features back into the logical pipeline for structured data. @GT we use ClarityNLP to hybridize structured features (via CQL) and unstructured features (via NLPQL).

http://github.com/claritynlp

He J, Mark L, Hilton C, Martin J, Baker J, Duke J, Hui SL, Li X, Dexter P. A comparison of structured data query methods versus natural language processing to identify metastatic melanoma cases from electronic health records. International Journal of Computational Medicine and Healthcare. 2019;1(1):101-11.

How can you assess portability?

- In the absence of markedly different baseline populations, a portable phenotype should create comparable patient populations across sites
- That is, a group of patients who have a similar clinical condition should look more or less the same

Spotlight: OHDSI Cohort Characterization

Cohort	CUIMC		IQVIA_OpenClaims		CDM_Premier_COVID_v1240		STARR-OMOP		VA-OMOP	
	Entries	Subjects	Entries	Subjects	Entries	Subjects	Entries	Subjects	Entries	Subjects
[COVID ID100 v1] Prevalent Type 2 Diabetes Mellitus	229,196	229,196	55,299,457	55,299,457	5,294,334	5,294,334	236,742	118,371	3,692,023	3,692,023
[COVID ID101 v1] Prevalent hypertension	598,064	598,064	122,784,203	122,784,203	12,336,370	12,336,370	621,640	310,820	8,033,128	8,033,128
[COVID ID102 v1] Prevalent chronic kidney disease	61,060	61,060	16,310,729	16,310,729	407,057	407,057	40,109	40,109	1,334,789	1,334,789
[COVID ID103 v1] Prevalent end stage renal disease	25,136	25,136	4,245,352	4,245,352	302,356	302,356	13,262	13,262	306,344	306,344
[COVID ID104 v1] Prevalent heart disease	607,680	607,680	87,858,660	87,858,660	6,742,394	6,742,394	270,169	270,169	5,737,199	5,737,199
[COVID ID105 v1] Prevalent malignant neoplasm excluding non-melanoma skin cancer	422,372	422,372	40,027,369	40,027,369	3,122,450	3,122,450	183,139	183,139	3,033,303	3,033,303
[COVID ID106 v1] Prevalent Human immunodeficiency virus infection	16,417	16,417	1,349,258	1,349,258	35,096	35,096	2,259	2,259	53,978	53,978
[COVID ID107 V1] Prevalent Hepatitis C	21,250	21,250	3,658,562	3,658,562	285,341	285,341	12,835	12,835	463,484	463,484
[COVID ID108 v1] Prevalent obesity	431,821	431,821	53,718,965	53,718,965	3,617,717	3,617,717	357,490	357,490	3,499,354	3,499,354
[COVID ID109 v1] Prevalent Dementia	52,645	52,645	10,684,811	10,684,811	774,074	774,074	13,901	13,901	851,901	851,901
[COVID ID106 v1] Prevalent tuberculosis	642	642	15,649	15,649	1,114	1,114	345	345	5,380	5,380
[COVID ID118 v1] Prevalent Autoimmune condition	200,518	200,518	36,453,532	36,453,532	1,500,146	1,500,146	92,620	92,620	1,728,469	1,728,469
[COVID ID119 V1] Prevalent chronic obstructive pulmonary disease (COPD) without asthma	119,993	119,993	31,326,149	31,326,149	2,222,462	2,222,462	94,796	47,398	2,970,936	2,970,936
[COVID ID120 V1] Prevalent Asthma without COPD	253,686	253,686	55,525,684	55,525,684	2,477,682	2,477,682	246,340	123,170	733,011	733,011
[COVID ID125 V1] Prevalent pre-existing condition of COVID risk factor	1,055,017	1,055,017	138,902,076	138,902,076	12,031,423	12,031,423	482,561	482,561	8,011,205	8,011,205
[COVID ID199 V1] Pregnant women	312,562	178,818	39,893,561	23,813,763	1,114,145	1,083,716	85,085	61,338	60,250	42,098
[COVID ID200 v1] Flu-like symptom episodes	532,684	307,207	182,421,964	96,488,751	9,099,499	7,786,607	459,834	256,291	3,571,173	1,869,676
[COVID ID203 v1] Prevalent chronic kidney disease broad	125,098	125,098	25,275,744	25,275,744	2,267,496	2,267,496	72,042	72,042	1,925,292	1,925,292
[COVID ID204 v1] Prevalent end stage renal disease broad	56,902	56,902	6,655,206	6,655,206	374,252	374,252	25,563	25,563	735,648	735,648
[COVID ID121 v1] Prevalent Human immunodeficiency virus infection broad	23,187	23,187	1,796,443	1,796,443	141,194	141,194	3,010	3,010	66,111	66,111
[COVID ID102 v1] Prevalent tuberculosis broad	15,078	15,078	504,836	504,836	5,682	5,682	2,348	2,348	41,746	41,746

Showing 1 to 21 of 21 entries

Previous 1 Next

‘Asthma’ Incidence Rates Across Sites

‘Heart Disease’ Incidence Rates Across Sites

Comorbidities for `Influenza` Patients

Covariate Name	IBM_CCAE	OPTUM_EXTENDED_DOD	OPTUM_PANTHER
	Proportion	Proportion	Proportion
Medical history: General			
Acute respiratory disease	34.4%	33.7%	18.5%
Attention deficit hyperactivity disorder	1.5%	1.2%	0.8%
Chronic liver disease	3.3%	3.9%	3.0%
Chronic obstructive lung disease	4.5%	11.6%	9.5%
Crohn's disease	1.5%	2.0%	1.3%
Dementia	0.2%	1.2%	0.9%
Depressive disorder	13.6%	18.9%	15.7%
Diabetes mellitus	13.5%	20.9%	15.0%
Gastroesophageal reflux disease	13.7%	21.9%	18.4%
Gastrointestinal hemorrhage	3.3%	4.4%	2.3%
Human immunodeficiency virus infection	0.2%	0.2%	0.2%
Hyperlipidemia	30.6%	45.9%	31.2%
Hypertensive disorder	34.8%	51.4%	38.4%
Lesion of liver	0.8%	1.3%	1.0%
Obesity	9.5%	12.9%	12.3%
Osteoarthritis	44.4%	56.4%	37.0%
Pneumonia	4.0%	6.2%	5.1%
Psoriasis	7.7%	7.3%	3.8%
Renal impairment	2.8%	9.8%	6.9%
Rheumatoid arthritis	85.7%	85.7%	83.4%
Schizophrenia	0.1%	0.1%	0.2%
Ulcerative colitis	1.5%	1.9%	1.2%
Urinary tract infectious disease	10.7%	14.1%	6.4%
Viral hepatitis C	1.1%	1.7%	1.6%
Visual system disorder	28.8%	42.9%	17.0%
Medical history: Cardiovascular disease			
Atrial fibrillation	1.3%	4.4%	4.0%
Cerebrovascular disease	2.5%	5.6%	3.1%
Coronary arteriosclerosis	4.6%	10.7%	8.5%
Heart disease	15.0%	26.4%	20.5%
Heart failure	2.0%	6.0%	4.5%

Medication Use for `Influenza` Patients

Medication use			
Agents acting on the renin-angiotensin system	24.9%	33.5%	31.5%
Antibacterials for systemic use	66.9%	66.3%	42.8%
Antidepressants	36.5%	36.4%	36.4%
Antiepileptics	20.9%	23.8%	25.2%
Antiinflammatory and antirheumatic products	65.8%	59.4%	55.9%
Antineoplastic agents	39.6%	40.5%	40.2%
Antipsoriatics	1.1%	1.0%	0.7%
Antithrombotic agents	7.4%	11.7%	32.0%
Beta blocking agents	16.1%	23.2%	27.0%
Calcium channel blockers	11.4%	17.0%	16.8%
Diuretics	24.4%	30.0%	28.5%
Drugs for acid related disorders	33.5%	35.8%	46.0%
Drugs for obstructive airway diseases	29.7%	30.4%	36.6%
Drugs used in diabetes	10.7%	14.6%	15.8%
Immunosuppressants	54.2%	54.9%	57.8%
Lipid modifying agents	23.3%	33.1%	31.7%
Opioids	40.7%	43.3%	41.0%
Psycholeptics	33.7%	32.4%	35.2%
Psychostimulants, agents used for adhd and nootropics	5.7%	3.9%	4.4%

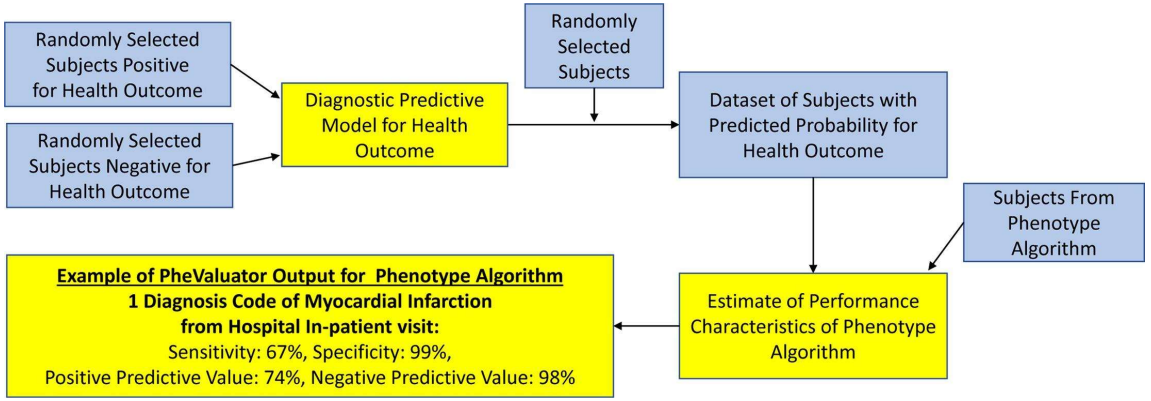
Machine Learning Based Tools

ELSEVIER

Journal of Biomedical Informatics
Volume 97, September 2019, 103258

PheValuator: Development and evaluation of a phenotype algorithm evaluator

Joel N. Swerdel [a, b], George Hripcsak [b, c], Patrick B. Ryan [a, b, c]



So how do we make phenotypes portable

- Use data types that can be found across systems
  - Look across models and at emerging mandated standards
  - Leverage the required data types
- Include codes that can be found across systems
  - Leverage cohort and dataset characterization tools
  - Publish concept sets on NLM VSAC for all to use
- Use logic that can be applied across systems
  - Avoid exotic logical operators
  - CQL capabilities can serve as a guide

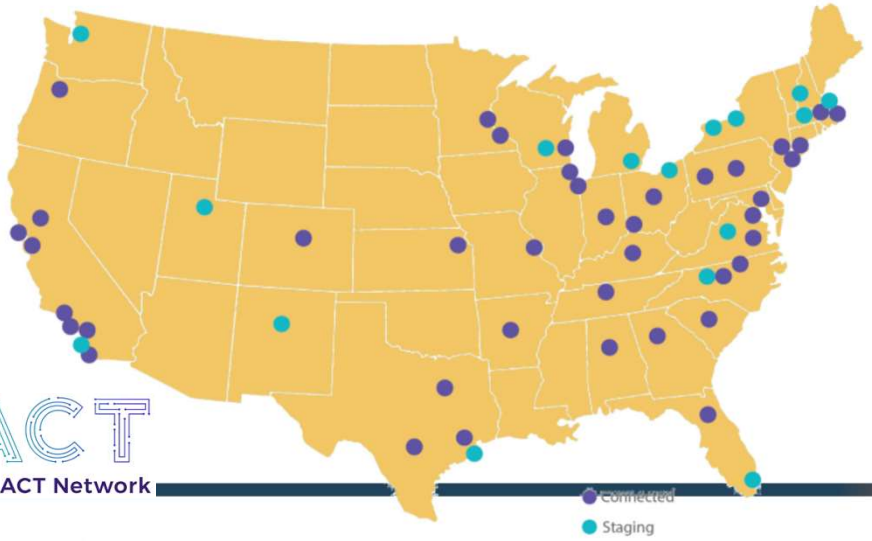
What makes a phenotype scalable?

- Meets all the criteria for portability and…
- There exists actual machinery to run the phenotype at a large number of sites

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

PCORnet Coverage Map

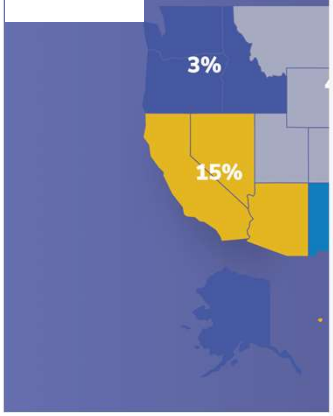
This map depicts the number of PCORI-funded Patient-Powered Research Networks (PPRNs) or Clinical Data Research Networks (CDRNs) that have coverage in each state.



ACT
The ACT Network

Sentinel

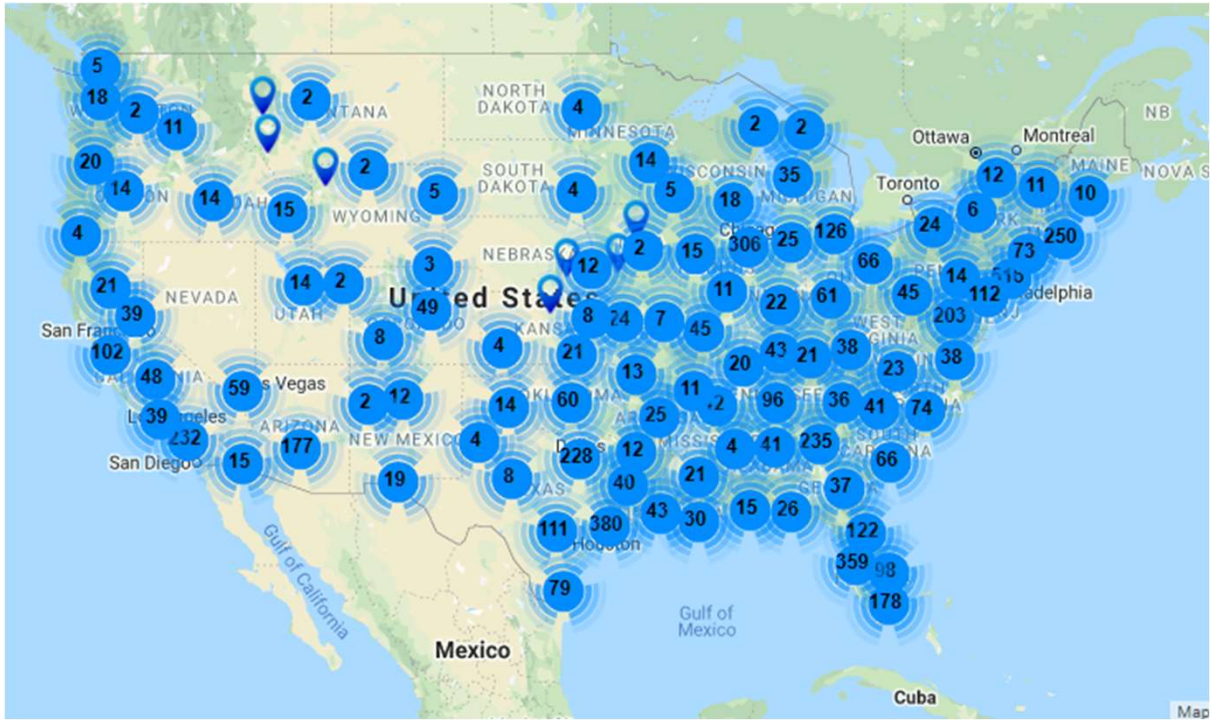
Member Distribution of the Sentinel Distributed Database by Geographical Region



1. Aetna, a CVS Health company
2. Blue Cross Blue Shield of Massachusetts
3. Duke University School of Medicine: Department of Population Health Sciences (Medicare Fee-for-Service data)
4. Harvard Pilgrim Health Care Institute
5. HCA Healthcare
6. HealthCore, Inc. (Anthem, Inc. data)
7. HealthPartners Institute
8. Humana, Inc.
9. Kaiser Permanente Colorado Institute for Health Research
10. Kaiser Permanente Integrated Health Care Research Hawaii
11. Kaiser Foundation Health Plan of the Mid-Atlantic States, Inc.
12. Kaiser Permanente Northern California, Division of Research
13. Kaiser Permanente Northwest Center for Health Research
14. Kaiser Permanente Washington Health Research Institute
15. Marshfield Clinic Research Institute
16. Optum (OptumInsight Life Sciences Inc. and Optum Labs®)
17. Vanderbilt University Medical Center, Department of Health Policy (Tennessee Medicaid data)

FHIR

1upHealth



Da Vinci 2020 Multi-Stakeholder Membership

Payers using CQL and FHIR

PROVIDERS

YAMA Medical Group
Cedars-Sinai
*RUSH
Sutter Health
MultiCare Connected Care
Providence St. Joseph Health
OHSU
Texas Health Resources
UNC Health Care
Weill Cornell Medicine
UC Davis Health

INDUSTRY PARTNERS

HIMSS
*HL7 International
NCQA Measuring quality. Improving health care.

EHRs

*Pulse8
athenahealth
*Cerner
*Epic
healow | Insights

PAYERS

*Anthem
*BlueCross BlueShield Association
*BlueCross BlueShield of Alabama
*Blue Cross of Idaho
Blue Cross Blue Shield Blue Care Network of Michigan
*of Tennessee
*CAMBIA
*CMS
*Cigna
CVS Health
CENTENE Corporation
GUIDEWELL
*HCSC Health Care Service Corporation
*Humana
*Independence
*UnitedHealthcare

VENDORS

casenet
*cognosante
*edifecs
infor
InterSystems Health | Business | Government
juxly
mcg
*OPTUM
*surescripts
*ZeOmega

DEPLOYMENT

Availity
MiHIN
CHANGE HEALTHCARE

For current membership: http://www.hl7.org/about/davinci/members.cfm

https://confluence.hl7.org/display/DVP/Da+Vinci

*Indicates a founding member of the Da Vinci Project.
Organization shown in primary Da Vinci role, Many members participate across categories.

How to improve Phenotype Scalability?

- It is not about phenotype design (ie portability)
- We need to make our phenotypes runnable in more places
- So yes continue to expand individual platform adoption
  - simplify CDM ETL, code deployment, add value to participating organizations, publications, etc
- But we **must take advantage of the transitions already underway in health systems and interoperability**

Keeping Both the Baby and the Bathwater

- When building phenotypes for your network, begin looking at a parallel pipeline to export logic using CQL, concept sets using VSAC codesets
- As FHIR mandates come into place and CQL becomes standard for health system payer interactions and reporting, you will have phenotypes that can piggyback on these technologies
- You can thus leverage a far broader range of sites as data partners (perhaps in an “Extended” tier) in addition to your native platform adopters

Caveats

- The aforementioned regulations are not internationally adopted at present and thus this pathway is US-centric
- Performance of patient-level technologies for population-level analytics will take time to catch up
- Culling data from a broader range of sites will introduce even more variability in populations and data quality that will need to be addressed

Questions?

jon.duke@gatech.edu