INFORMATICS PROFESSIONALS. LEADING THE WAY.

Introduction to querying electronic health records for patient cohorts using the i2b2 data platform

I2b2 and the Query Tool

W16, Session 1: Introduction to i2b2

Jeffrey Klann, PhD

Harvard Medical School / Mass General Brigham

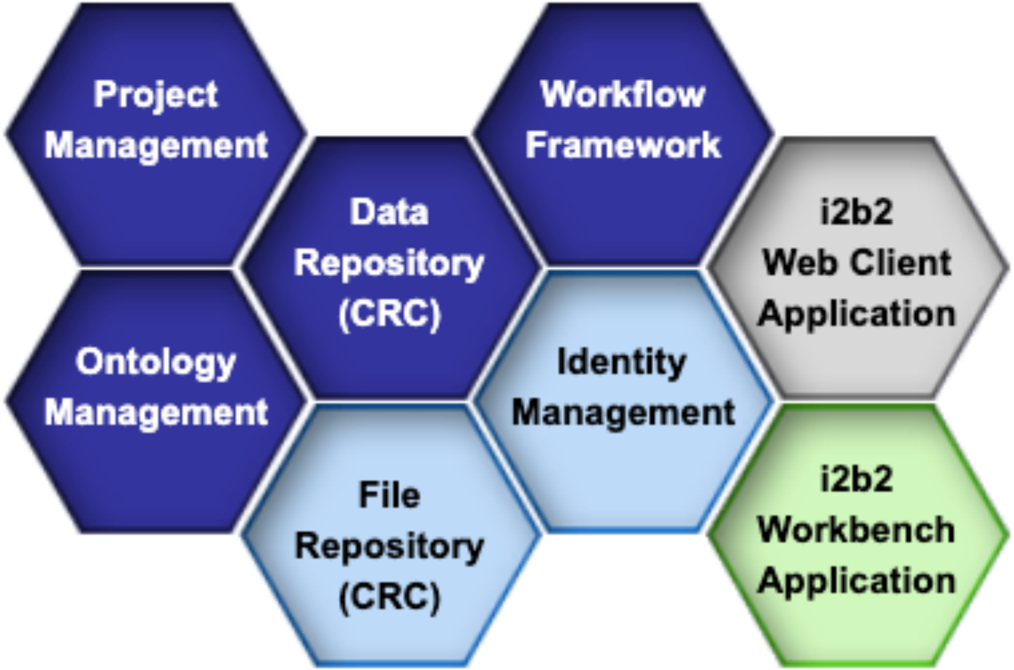
Twitter: @JeffKlann

#AMIA2020

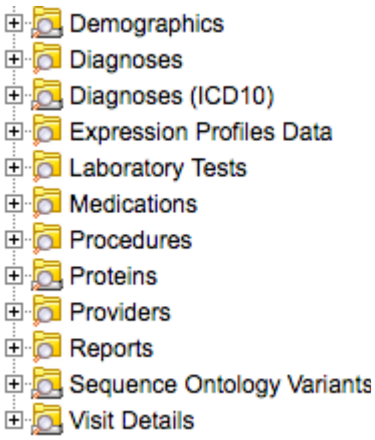
What is i2b2?

- A Clinical Data Warehousing and Analytics Platform
  - >200 installations worldwide
  - Active user community with many plugins and tools
  - Open-source (Mozilla license)
- Componentized, extensible architecture
- Curated biomedical ontologies
- Flexible star-schema data model
- Graphical query tool
- https://community.i2b2.org/

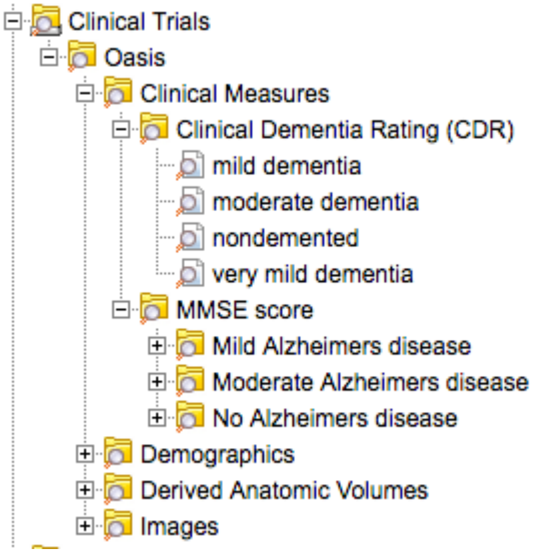
The i2b2 “hive”



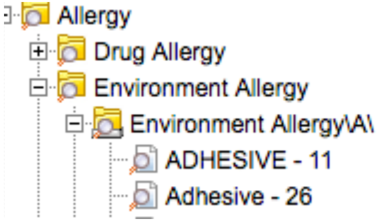
Many Ontologies



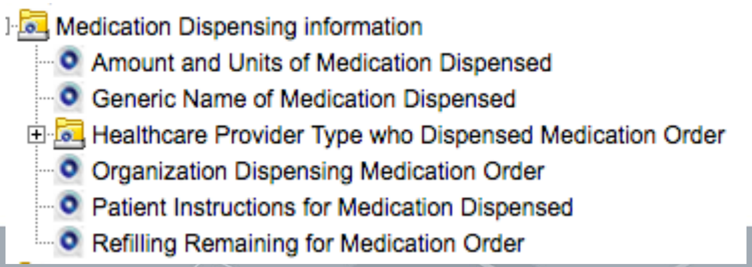
- Demographics
- Diagnoses
- Diagnoses (ICD10)
- Expression Profiles Data
- Laboratory Tests
- Medications
- Procedures
- Proteins
- Providers
- Reports
- Sequence Ontology Variants
- Visit Details



- Clinical Trials
  - Oasis
    - Clinical Measures
      - Clinical Dementia Rating (CDR)
        - mild dementia
        - moderate dementia
        - nondemented
        - very mild dementia
      - MMSE score
        - Mild Alzheimers disease
        - Moderate Alzheimers disease
        - No Alzheimers disease
    - Demographics
    - Derived Anatomic Volumes
    - Images

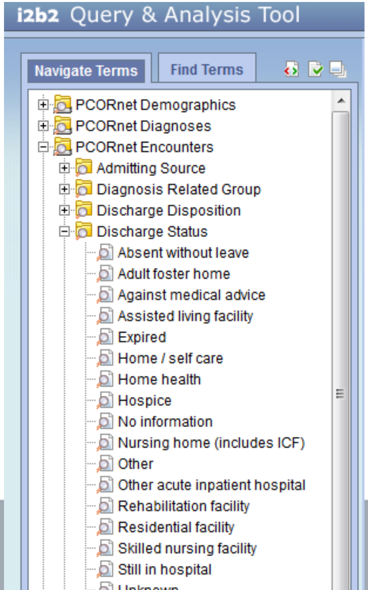


- Allergy
  - Drug Allergy
  - Environment Allergy
    - Environment Allergy\A\
      - ADHESIVE - 11
      - Adhesive - 26

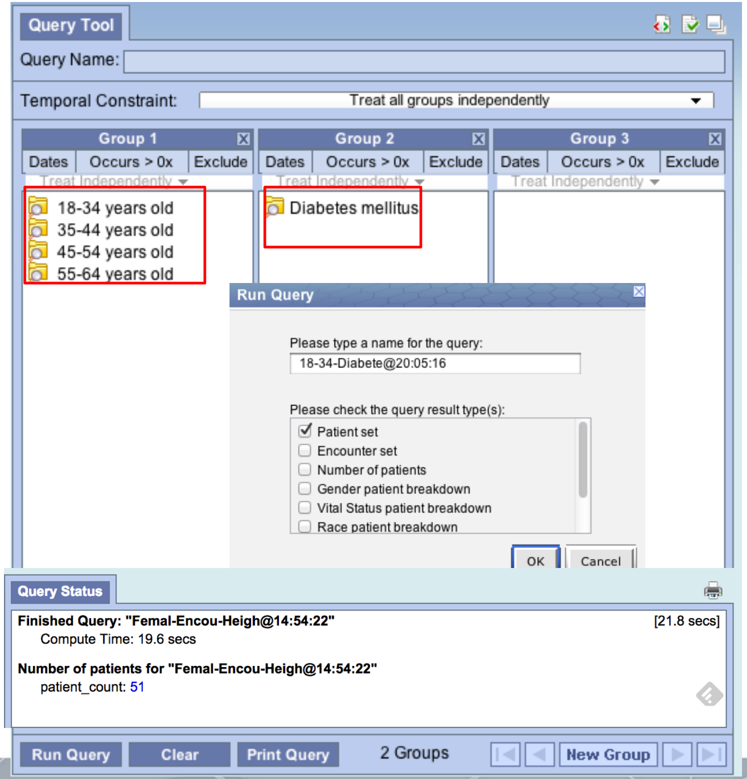


- Medication Dispensing information
  - Amount and Units of Medication Dispensed
  - Generic Name of Medication Dispensed
  - Healthcare Provider Type who Dispensed Medication Order
  - Organization Dispensing Medication Order
  - Patient Instructions for Medication Dispensed
  - Refilling Remaining for Medication Order

The i2b2 Query Tool



- PCORnet Demographics
- PCORnet Diagnoses
- PCORnet Encounters
  - Admitting Source
  - Diagnosis Related Group
  - Discharge Disposition
  - Discharge Status
    - Absent without leave
    - Adult foster home
    - Against medical advice
    - Assisted living facility
    - Expired
    - Home / self care
    - Home health
    - Hospice
    - No information
    - Nursing home (includes ICF)
    - Other
    - Other acute inpatient hospital
    - Rehabilitation facility
    - Residential facility
    - Skilled nursing facility
    - Still in hospital
    - Unknown



Query Name:

Temporal Constraint: Treat all groups independently

Group 1	Group 2	Group 3
18-34 years old 35-44 years old 45-54 years old 55-64 years old	Diabetes mellitus	

Please type a name for the query:
18-34-Diabete@20:05:16

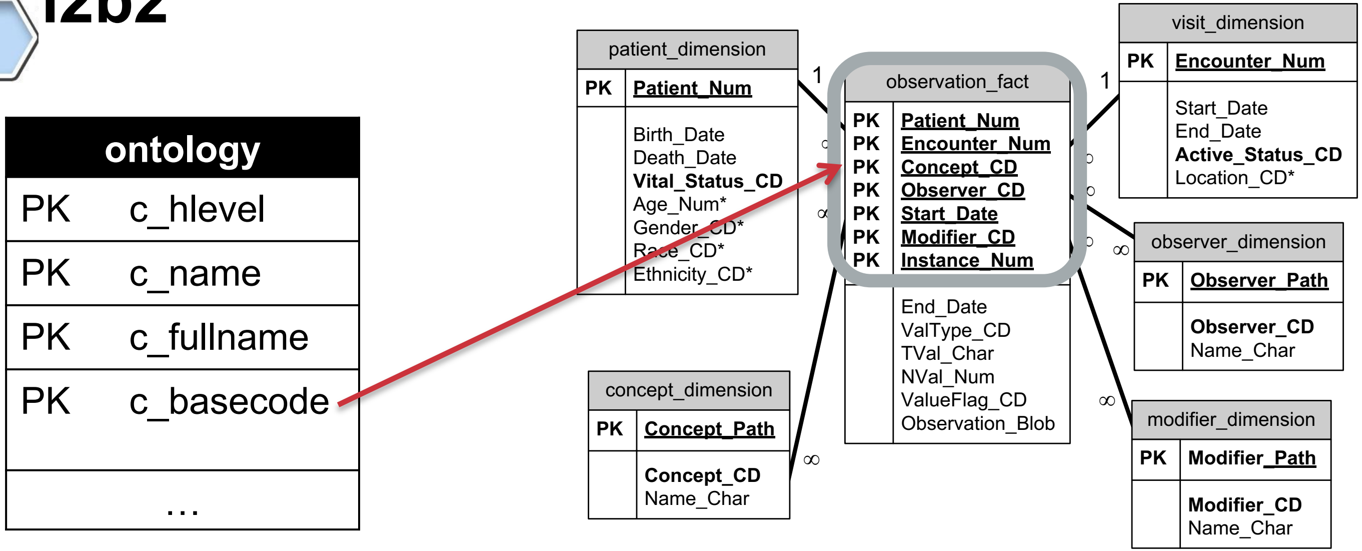
Please check the query result type(s):
☑ Patient set
□ Encounter set
□ Number of patients
□ Gender patient breakdown
□ Vital Status patient breakdown
□ Race patient breakdown

Finished Query: "Femal-Encou-Heigh@14:54:22" [21.8 secs]
Compute Time: 19.6 secs

Number of patients for "Femal-Encou-Heigh@14:54:22"
patient_count: 51

The i2b2 Data Model

i2b2



ontology	
PK	c_hlevel
PK	c_name
PK	c_fullname
PK	c_basecode
…	

patient_dimension	
PK	**Patient_Num**
	Birth_Date Death_Date **Vital_Status_CD** Age_Num* Gender_CD* Race_CD* Ethnicity_CD*

concept_dimension	
PK	**Concept_Path**
	Concept_CD Name_Char

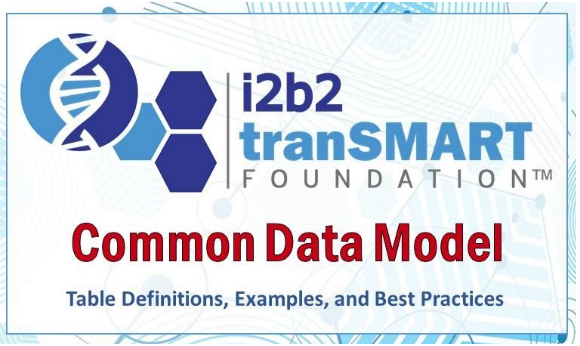
observation_fact	
PK	**Patient_Num**
PK	**Encounter_Num**
PK	**Concept_CD**
PK	**Observer_CD**
PK	**Start_Date**
PK	**Modifier_CD**
PK	**Instance_Num**
	End_Date ValType_CD TVal_Char NVal_Num ValueFlag_CD Observation_Blob

visit_dimension	
PK	**Encounter_Num**
	Start_Date End_Date **Active_Status_CD** Location_CD*

observer_dimension	
PK	**Observer_Path**
	Observer_CD Name_Char

modifier_dimension	
PK	**Modifier_Path**
	Modifier_CD Name_Char

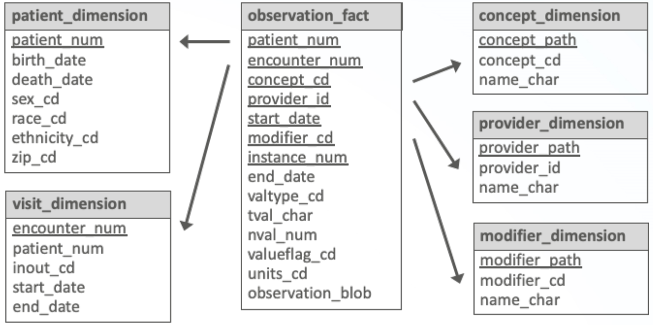
Common Data Model for i2b2 and tranSMART



i2b2 tranSMART FOUNDATION™

Common Data Model

Table Definitions, Examples, and Best Practices



patient_dimension
patient_num
birth_date
death_date
sex_cd
race_cd
ethnicity_cd
zip_cd

visit_dimension
encounter_num
patient_num
inout_cd
start_date
end_date

observation_fact
patient_num
encounter_num
concept_cd
provider_id
start_date
modifier_cd
instance_num
end_date
valtype_cd
tval_char
nval_num
valueflag_cd
units_cd
observation_blob

concept_dimension
concept_path
concept_cd
name_char

provider_dimension
provider_path
provider_id
name_char

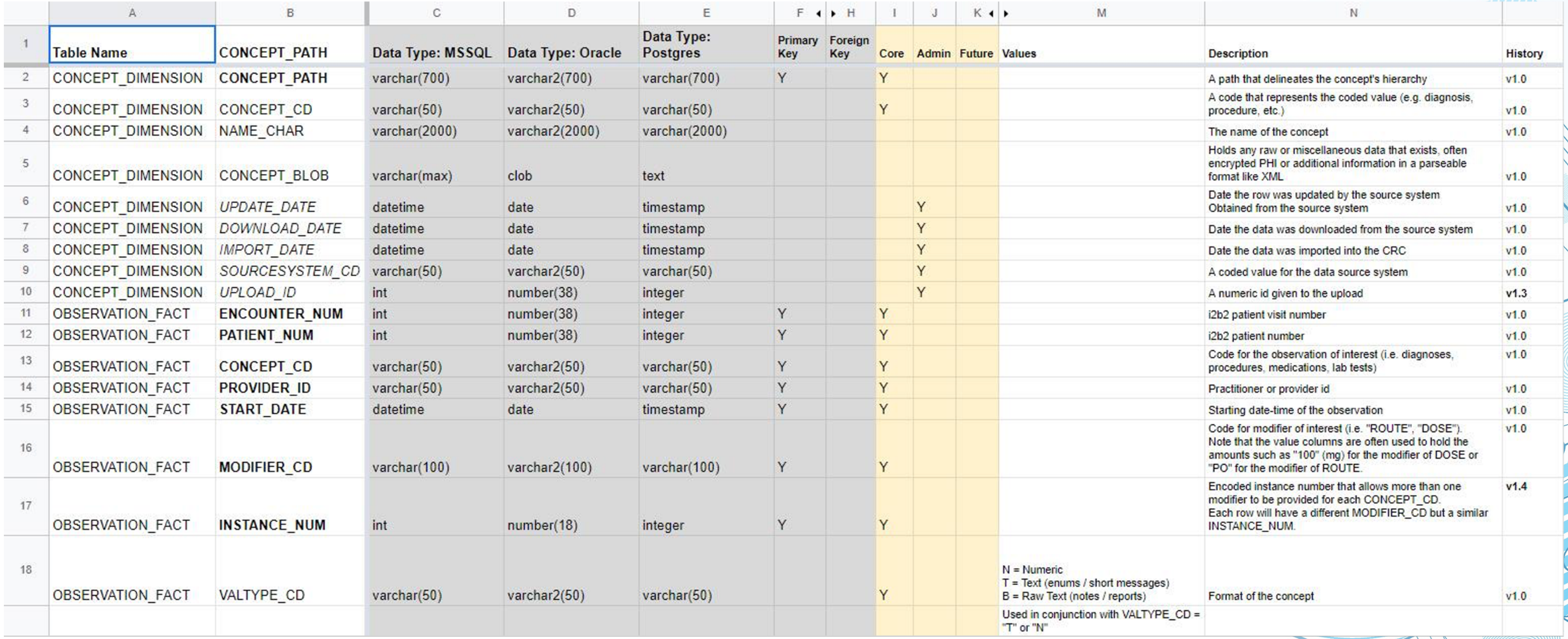
modifier_dimension
modifier_path
modifier_cd
name_char

- **Unified Core Data Model for i2b2 and tranSMART**

- Core tables, fields, keys, indexes, constraints
- Detailed descriptions of each field
- Best practices and examples for different use cases

Common Data Model for i2b2 and tranSMART

Single schema, documented, MSSQL/Oracle/Postgres

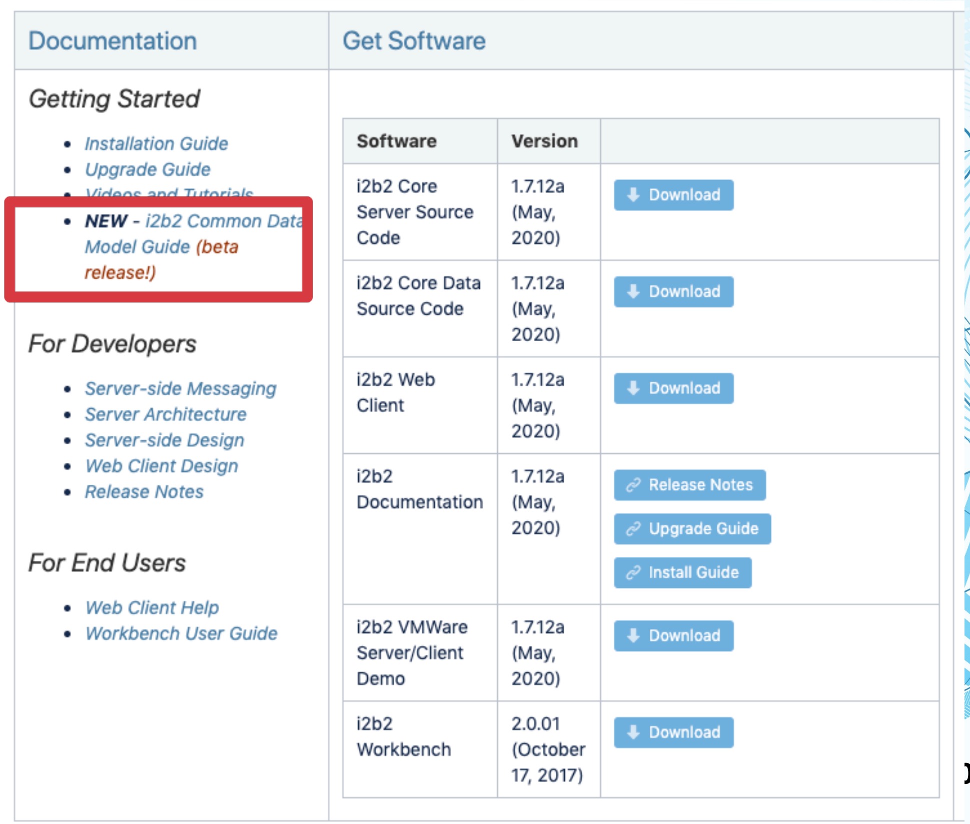


	A	B	C	D	E	F	H	I	J	K	M	N	
1	**Table Name**	**CONCEPT_PATH**	**Data Type: MSSQL**	**Data Type: Oracle**	**Data Type: Postgres**	**Primary Key**	**Foreign Key**	**Core**	**Admin**	**Future**	**Values**	**Description**	**History**
2	CONCEPT_DIMENSION	CONCEPT_PATH	varchar(700)	varchar2(700)	varchar(700)	Y		Y				A path that delineates the concept's hierarchy	v1.0
3	CONCEPT_DIMENSION	CONCEPT_CD	varchar(50)	varchar2(50)	varchar(50)			Y				A code that represents the coded value (e.g. diagnosis, procedure, etc.)	v1.0
4	CONCEPT_DIMENSION	NAME_CHAR	varchar(2000)	varchar2(2000)	varchar(2000)							The name of the concept	v1.0
5	CONCEPT_DIMENSION	CONCEPT_BLOB	varchar(max)	clob	text							Holds any raw or miscellaneous data that exists, often encrypted PHI or additional information in a parseable format like XML	v1.0
6	CONCEPT_DIMENSION	*UPDATE_DATE*	datetime	date	timestamp				Y			Date the row was updated by the source system Obtained from the source system	v1.0
7	CONCEPT_DIMENSION	*DOWNLOAD_DATE*	datetime	date	timestamp				Y			Date the data was downloaded from the source system	v1.0
8	CONCEPT_DIMENSION	*IMPORT_DATE*	datetime	date	timestamp				Y			Date the data was imported into the CRC	v1.0
9	CONCEPT_DIMENSION	*SOURCESYSTEM_CD*	varchar(50)	varchar2(50)	varchar(50)				Y			A coded value for the data source system	v1.0
10	CONCEPT_DIMENSION	*UPLOAD_ID*	int	number(38)	integer				Y			A numeric id given to the upload	**v1.3**
11	OBSERVATION_FACT	**ENCOUNTER_NUM**	int	number(38)	integer	Y		Y				i2b2 patient visit number	v1.0
12	OBSERVATION_FACT	**PATIENT_NUM**	int	number(38)	integer	Y		Y				i2b2 patient number	v1.0
13	OBSERVATION_FACT	**CONCEPT_CD**	varchar(50)	varchar2(50)	varchar(50)	Y		Y				Code for the observation of interest (i.e. diagnoses, procedures, medications, lab tests)	v1.0
14	OBSERVATION_FACT	**PROVIDER_ID**	varchar(50)	varchar2(50)	varchar(50)	Y		Y				Practitioner or provider id	v1.0
15	OBSERVATION_FACT	**START_DATE**	datetime	date	timestamp	Y		Y				Starting date-time of the observation	v1.0
16	OBSERVATION_FACT	**MODIFIER_CD**	varchar(100)	varchar2(100)	varchar(100)	Y		Y				Code for modifier of interest (i.e. "ROUTE", "DOSE"). Note that the value columns are often used to hold the amounts such as "100" (mg) for the modifier of DOSE or "PO" for the modifier of ROUTE.	v1.0
17	OBSERVATION_FACT	**INSTANCE_NUM**	int	number(18)	integer	Y		Y				Encoded instance number that allows more than one modifier to be provided for each CONCEPT_CD. Each row will have a different MODIFIER_CD but a similar INSTANCE_NUM.	**v1.4**
18	OBSERVATION_FACT	VALTYPE_CD	varchar(50)	varchar2(50)	varchar(50)			Y			N = Numeric T = Text (enums / short messages) B = Raw Text (notes / reports)	Format of the concept	v1.0
											Used in conjunction with VALTYPE_CD = "T" or "N"		

Single schema, documented, MSSQL/Oracle/Postgres

Common Data Model for i2b2 and tranSMART

- Draft version available for download!
- On https://community.i2b2.org/wiki/



Documentation	Get Software

Getting Started

- *Installation Guide*
- *Upgrade Guide*
- *Videos and Tutorials*
- ***NEW** - i2b2 Common Data Model Guide (beta release!)*

For Developers

- *Server-side Messaging*
- *Server Architecture*
- *Server-side Design*
- *Web Client Design*
- *Release Notes*

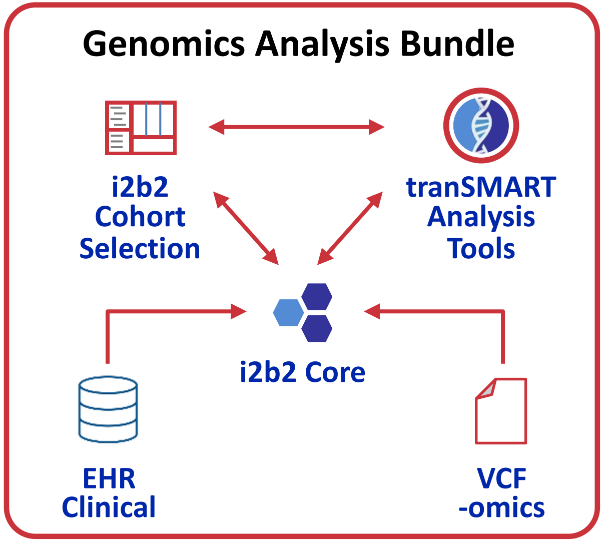
For End Users

- *Web Client Help*
- *Workbench User Guide*

Software	Version	
i2b2 Core Server Source Code	1.7.12a (May, 2020)	Download
i2b2 Core Data Source Code	1.7.12a (May, 2020)	Download
i2b2 Web Client	1.7.12a (May, 2020)	Download
i2b2 Documentation	1.7.12a (May, 2020)	Release Notes Upgrade Guide Install Guide
i2b2 VMWare Server/Client Demo	1.7.12a (May, 2020)	Download
i2b2 Workbench	2.0.01 (October 17, 2017)	Download

I2b2/tranSMART Bundle

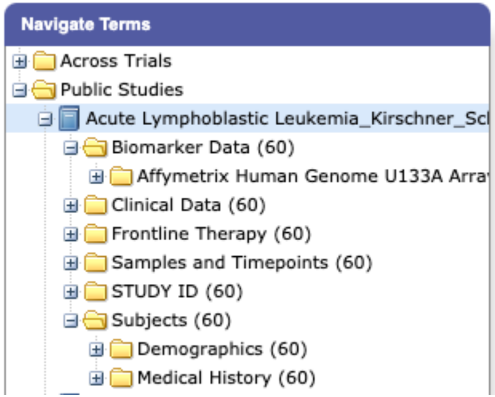
- Run I2b2 and tranSMART on the same database!
- Detailed documentation with step-by-step instructions.



i2b2 tranSMART FOUNDATION™

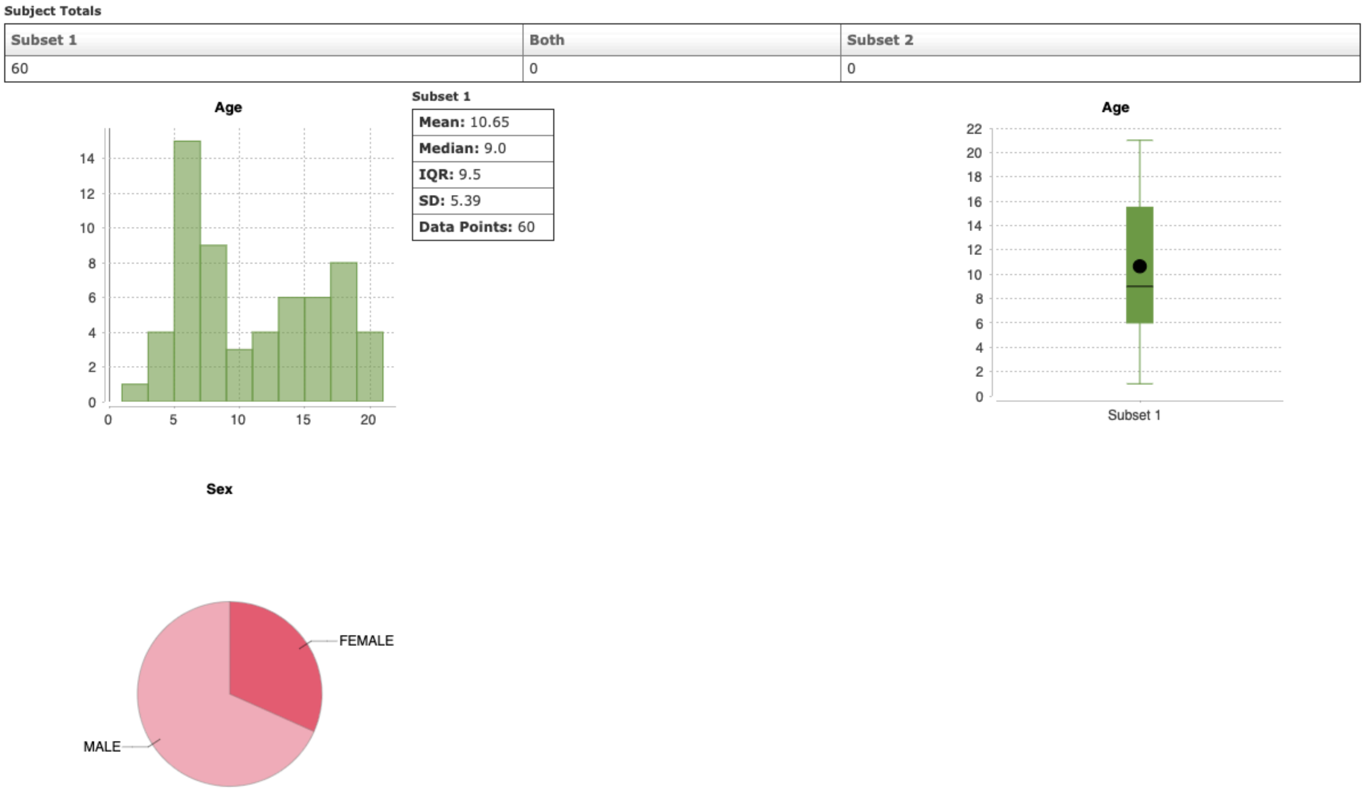
HARVARD SYMPOSIUM 2020

tranSMART



Navigate Terms

- Across Trials
- Public Studies
  - Acute Lymphoblastic Leukemia_Kirschner_Scl
    - Biomarker Data (60)
      - Affymetrix Human Genome U133A Arra
    - Clinical Data (60)
    - Frontline Therapy (60)
    - Samples and Timepoints (60)
    - STUDY ID (60)
    - Subjects (60)
      - Demographics (60)
      - Medical History (60)



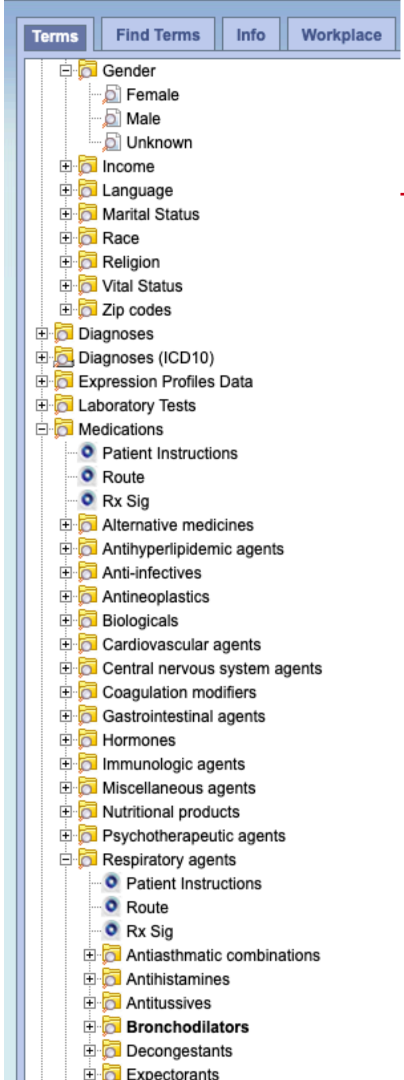
Subject Totals

Subset 1	Both	Subset 2
60	0	0

Subset 1

Mean: 10.65	
Median: 9.0	
IQR: 9.5	
SD: 5.39	
Data Points: 60	

I2b2 and tranSMART on the same database



- Gender
  - Female
  - Male
  - Unknown
- Income
- Language
- Marital Status
- Race
- Religion
- Vital Status
- Zip codes

Diagnoses

Diagnoses (ICD10)

Expression Profiles Data

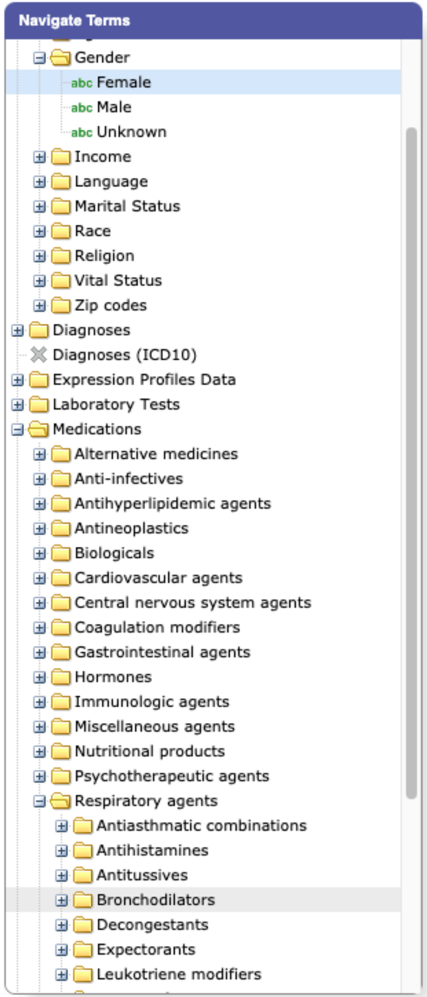
Laboratory Tests

Medications

- Patient Instructions
- Route
- Rx Sig
- Alternative medicines
- Antihyperlipidemic agents
- Anti-infectives
- Antineoplastics
- Biologicals
- Cardiovascular agents
- Central nervous system agents
- Coagulation modifiers
- Gastrointestinal agents
- Hormones
- Immunologic agents
- Miscellaneous agents
- Nutritional products
- Psychotherapeutic agents
- Respiratory agents
  - Patient Instructions
  - Route
  - Rx Sig
  - Antiasthmatic combinations
  - Antihistamines
  - Antitussives
  - **Bronchodilators**
  - Decongestants
  - Expectorants

i2b2 Core

tranSMART



- Gender
  - abc Female
  - abc Male
  - abc Unknown
- Income
- Language
- Marital Status
- Race
- Religion
- Vital Status
- Zip codes

Diagnoses

Diagnoses (ICD10)

Expression Profiles Data

Laboratory Tests

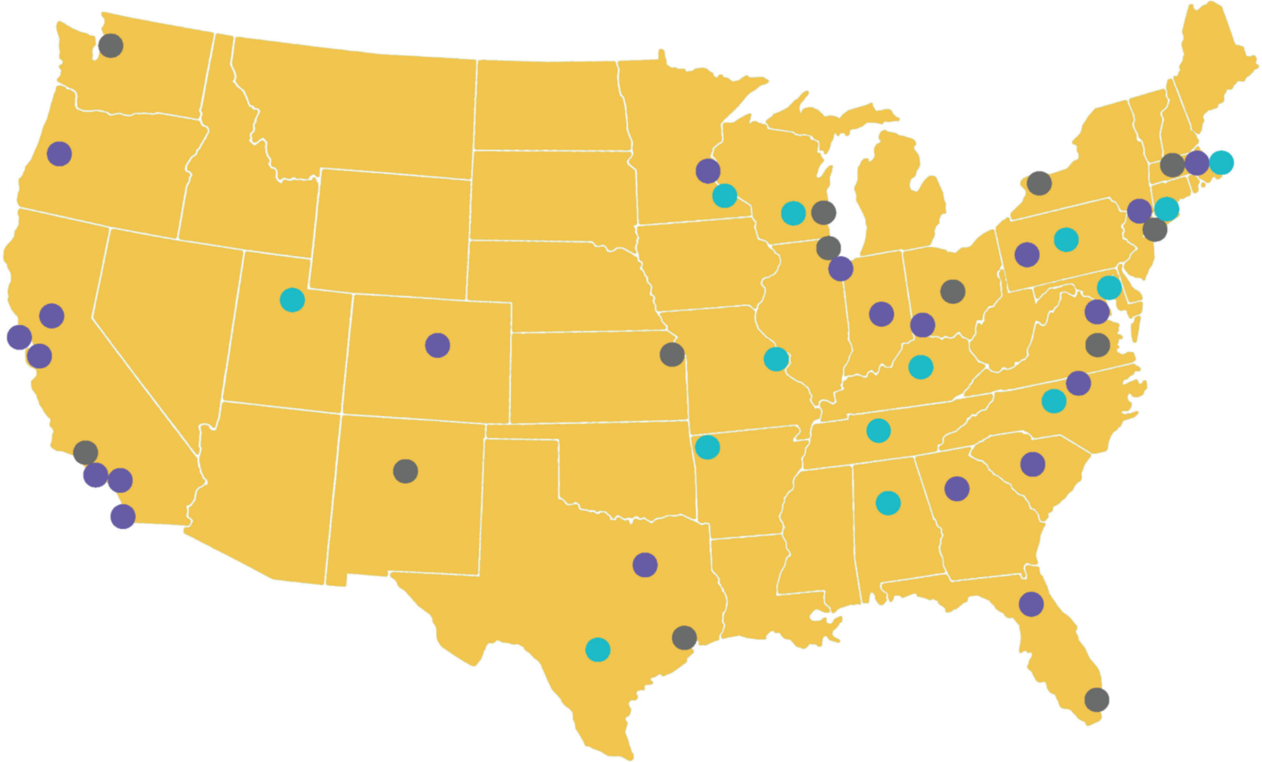
Medications

- Alternative medicines
- Anti-infectives
- Antihyperlipidemic agents
- Antineoplastics
- Biologicals
- Cardiovascular agents
- Central nervous system agents
- Coagulation modifiers
- Gastrointestinal agents
- Hormones
- Immunologic agents
- Miscellaneous agents
- Nutritional products
- Psychotherapeutic agents
- Respiratory agents
  - Antiasthmatic combinations
  - Antihistamines
  - Antitussives
  - Bronchodilators
  - Decongestants
  - Expectorants
  - Leukotriene modifiers

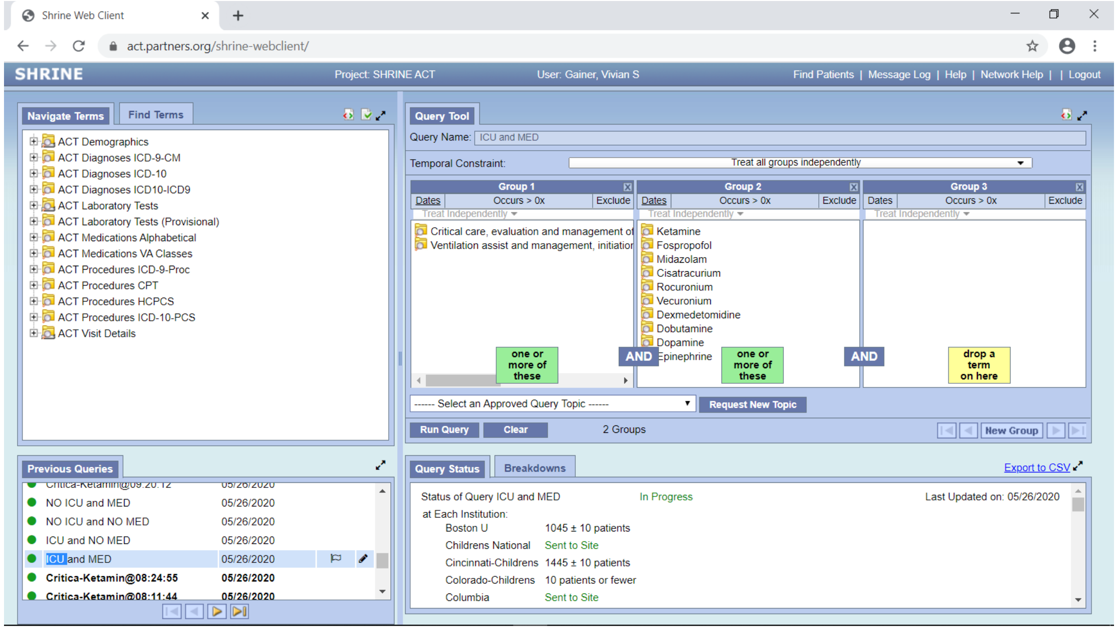
i2b2 Community Projects

https://community.i2b2.org/

Accrual to Clinical Trials (ACT) Data Network



ACT Research Network



Shrine Web Client

act.partners.org/shrine-webclient/

SHRINE   Project: SHRINE ACT   User: Gainer, Vivian S   Find Patients | Message Log | Help | Network Help | | Logout

Navigate Terms | Find Terms

- ACT Demographics
- ACT Diagnoses ICD-9-CM
- ACT Diagnoses ICD-10
- ACT Diagnoses ICD10-ICD9
- ACT Laboratory Tests
- ACT Laboratory Tests (Provisional)
- ACT Medications Alphabetical
- ACT Medications VA Classes
- ACT Procedures ICD-9-Proc
- ACT Procedures CPT
- ACT Procedures HCPCS
- ACT Procedures ICD-10-PCS
- ACT Visit Details

Query Tool

Query Name: ICU and MED

Temporal Constraint: Treat all groups independently

Group 1	Group 2	Group 3
Dates · Occurs > 0x · Exclude	Dates · Occurs > 0x · Exclude	Dates · Occurs > 0x · Exclude
Treat Independently	Treat Independently	Treat Independently
Critical care, evaluation and management of Ventilation assist and management, initiation	Ketamine Fospropofol Midazolam Cisatracurium Rocuronium Vecuronium Dexmedetomidine Dobutamine Dopamine Epinephrine	
one or more of these	AND · one or more of these	AND · drop a term on here

------ Select an Approved Query Topic ------   Request New Topic

Run Query   Clear   2 Groups   New Group

Previous Queries

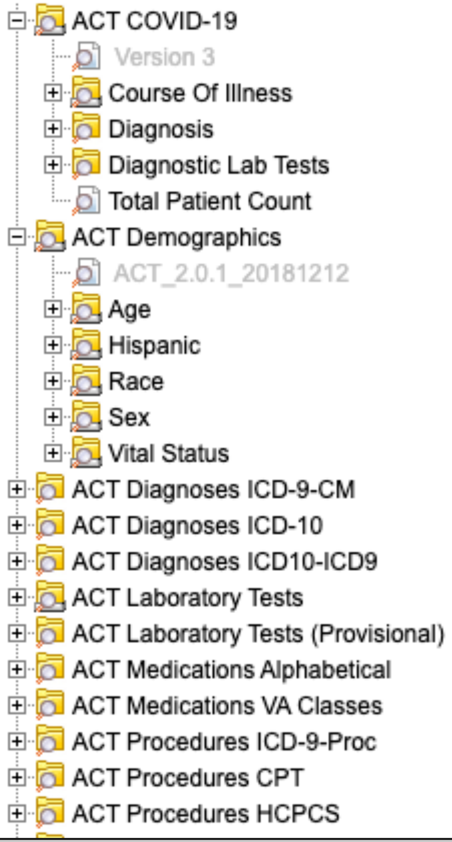
Critica-Ketamin@09:20:12	05/26/2020
NO ICU and MED	05/26/2020
NO ICU and NO MED	05/26/2020
ICU and NO MED	05/26/2020
ICU and MED	05/26/2020
Critica-Ketamin@08:24:55	**05/26/2020**
Critica-Ketamin@08:11:44	**05/26/2020**

Query Status | Breakdowns   Export to CSV

Status of Query ICU and MED   In Progress   Last Updated on: 05/26/2020

at Each Institution:

Boston U	1045 ± 10 patients
Childrens National	Sent to Site
Cincinnati-Childrens	1445 ± 10 patients
Colorado-Childrens	10 patients or fewer
Columbia	Sent to Site

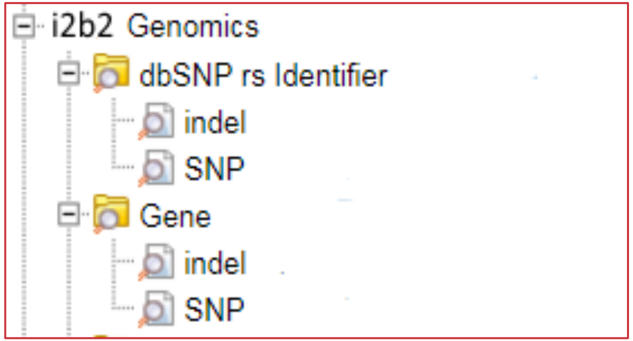


- ACT COVID-19
  - Version 3
  - Course Of Illness
  - Diagnosis
  - Diagnostic Lab Tests
  - Total Patient Count
- ACT Demographics
  - ACT_2.0.1_20181212
  - Age
  - Hispanic
  - Race
  - Sex
  - Vital Status
- ACT Diagnoses ICD-9-CM
- ACT Diagnoses ICD-10
- ACT Diagnoses ICD10-ICD9
- ACT Laboratory Tests
- ACT Laboratory Tests (Provisional)
- ACT Medications Alphabetical
- ACT Medications VA Classes
- ACT Procedures ICD-9-Proc
- ACT Procedures CPT
- ACT Procedures HCPCS

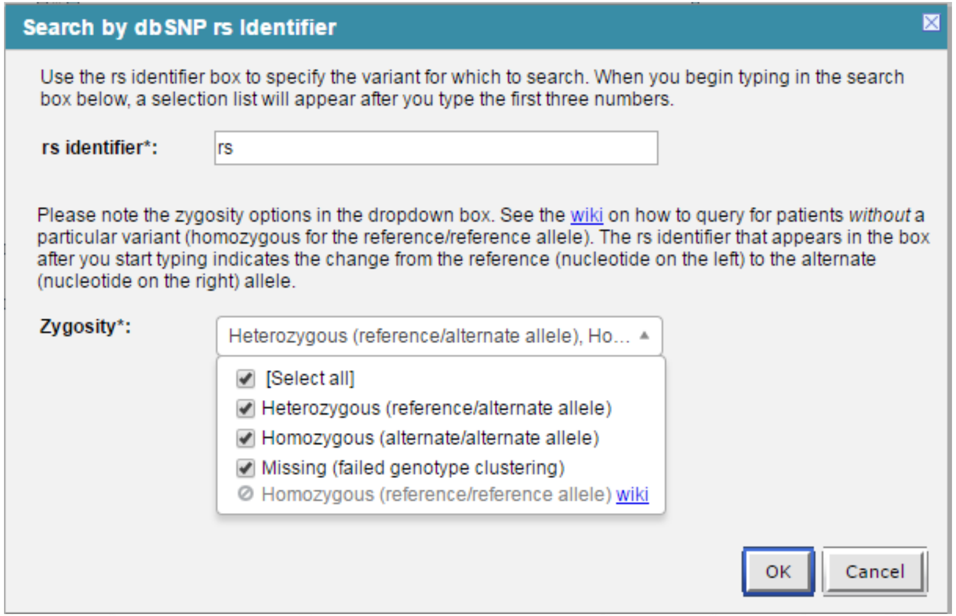
Query for sites with adequate data

Curated, mapped ontology

Genomic Data in i2b2



- i2b2 Genomics
  - dbSNP rs Identifier
    - indel
    - SNP
  - Gene
    - indel
    - SNP



Search by dbSNP rs Identifier

Use the rs identifier box to specify the variant for which to search. When you begin typing in the search box below, a selection list will appear after you type the first three numbers.

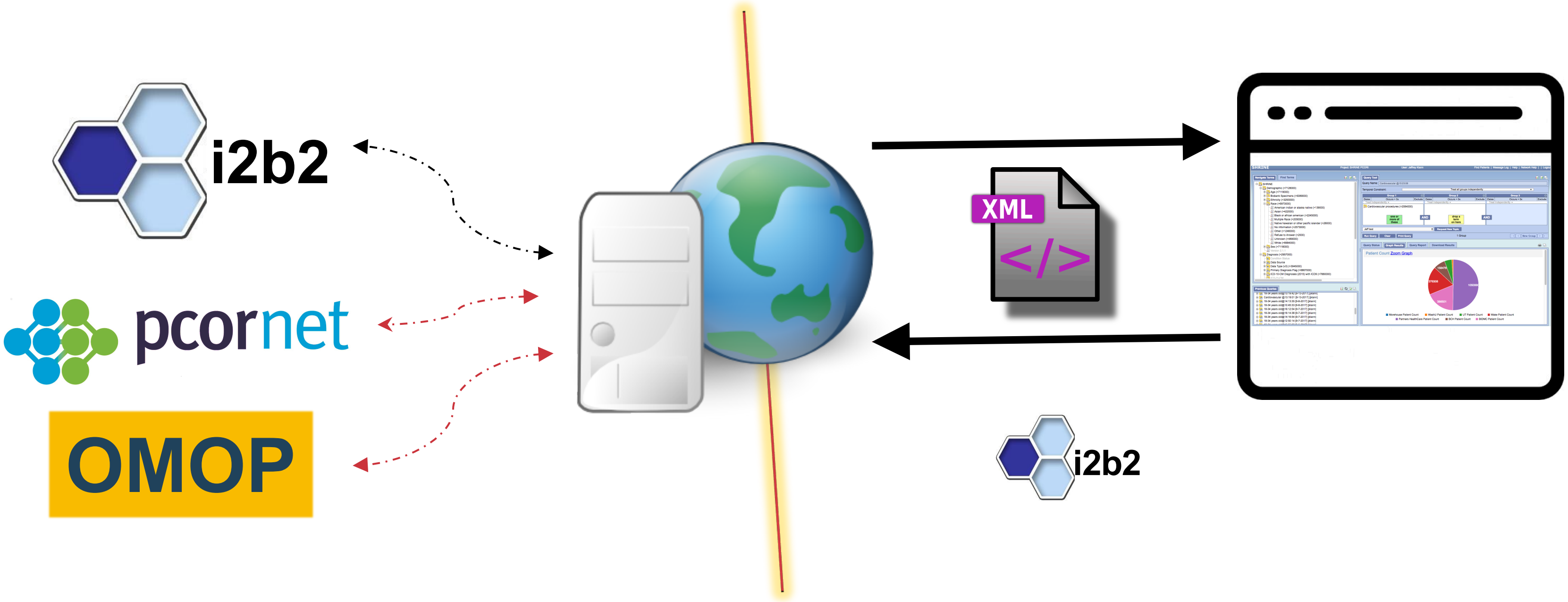
rs identifier*: rs

Please note the zygosity options in the dropdown box. See the wiki on how to query for patients *without* a particular variant (homozygous for the reference/reference allele). The rs identifier that appears in the box after you start typing indicates the change from the reference (nucleotide on the left) to the alternate (nucleotide on the right) allele.

Zygosity*: Heterozygous (reference/alternate allele), Ho...

- ☑ [Select all]
- ☑ Heterozygous (reference/alternate allele)
- ☑ Homozygous (alternate/alternate allele)
- ☑ Missing (failed genotype clustering)
- ⊘ Homozygous (reference/reference allele) wiki

Supports Multiple Data Models



1.7.13: Community Contributions?

- Develop a new feature!
  - Submit a GitHub pull request
  - Attest to the DCO (Developer Certificate of Origin)
  - **Ideas:** Import/export, client usability, new authentication methods, refactoring, better demo data

- Clean up the Community Projects

- Improve the Documentation

Synthea/COVID/ACT Demodata

- Synthea synthetic patient data
  - 1.5 million patients

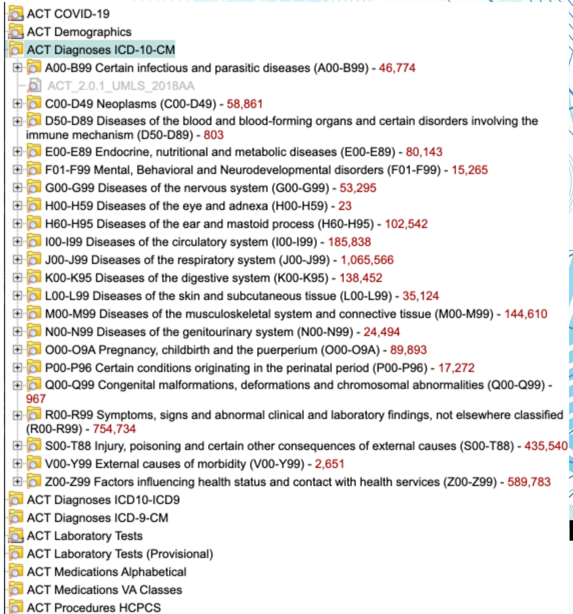
SYNTHEA

ACT

The ACT Network

Powered by NCATS CTSA Program

- Mapped to ACT Ontology
  - Demographics, PX, RX, DX, Labs
- Enhanced with COVID Lab Values
  - 20,000 patients



ACT COVID-19
ACT Demographics
ACT Diagnoses ICD-10-CM
A00-B99 Certain infectious and parasitic diseases (A00-B99) - 46,774
ACT_2.0.1_UMLS_2018AA
C00-D49 Neoplasms (C00-D49) - 58,861
D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism (D50-D89) - 803
E00-E89 Endocrine, nutritional and metabolic diseases (E00-E89) - 80,143
F01-F99 Mental, Behavioral and Neurodevelopmental disorders (F01-F99) - 15,265
G00-G99 Diseases of the nervous system (G00-G99) - 53,295
H00-H59 Diseases of the eye and adnexa (H00-H59) - 23
H60-H95 Diseases of the ear and mastoid process (H60-H95) - 102,542
I00-I99 Diseases of the circulatory system (I00-I99) - 185,838
J00-J99 Diseases of the respiratory system (J00-J99) - 1,065,566
K00-K95 Diseases of the digestive system (K00-K95) - 138,452
L00-L99 Diseases of the skin and subcutaneous tissue (L00-L99) - 35,124
M00-M99 Diseases of the musculoskeletal system and connective tissue (M00-M99) - 144,610
N00-N99 Diseases of the genitourinary system (N00-N99) - 24,494
O00-O9A Pregnancy, childbirth and the puerperium (O00-O9A) - 89,893
P00-P96 Certain conditions originating in the perinatal period (P00-P96) - 17,272
Q00-Q99 Congenital malformations, deformations and chromosomal abnormalities (Q00-Q99) - 967
R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99) - 754,734
S00-T88 Injury, poisoning and certain other consequences of external causes (S00-T88) - 435,540
V00-Y99 External causes of morbidity (V00-Y99) - 2,651
Z00-Z99 Factors influencing health status and contact with health services (Z00-Z99) - 589,783
ACT Diagnoses ICD10-ICD9
ACT Diagnoses ICD-9-CM
ACT Laboratory Tests
ACT Laboratory Tests (Provisional)
ACT Medications Alphabetical
ACT Medications VA Classes
ACT Procedures HCPCS

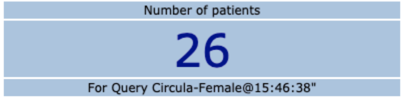
i2b2
tranSMART
FOUNDATION™

M 2020

Excercises

Exercise Set 1: Query

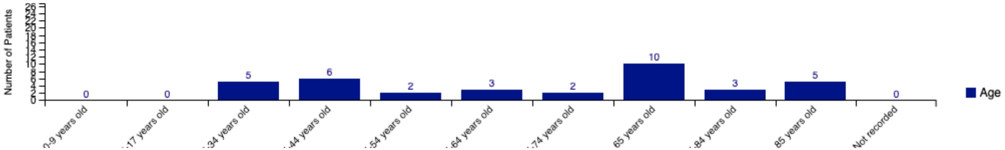
1a) Using the ontology browser and find terms, find how many females had circulatory system disease and create a bar chart by age.



Number of patients

26

For Query Circula-Female@15:46:38"



1b) How many patients had circulatory system disease and a CPK test >=400 u/l. (Answer: 5 patients.)

1c) Using the query from 1a, find the number of patients with a diagnosis before 11/2005. (Answer: 20).

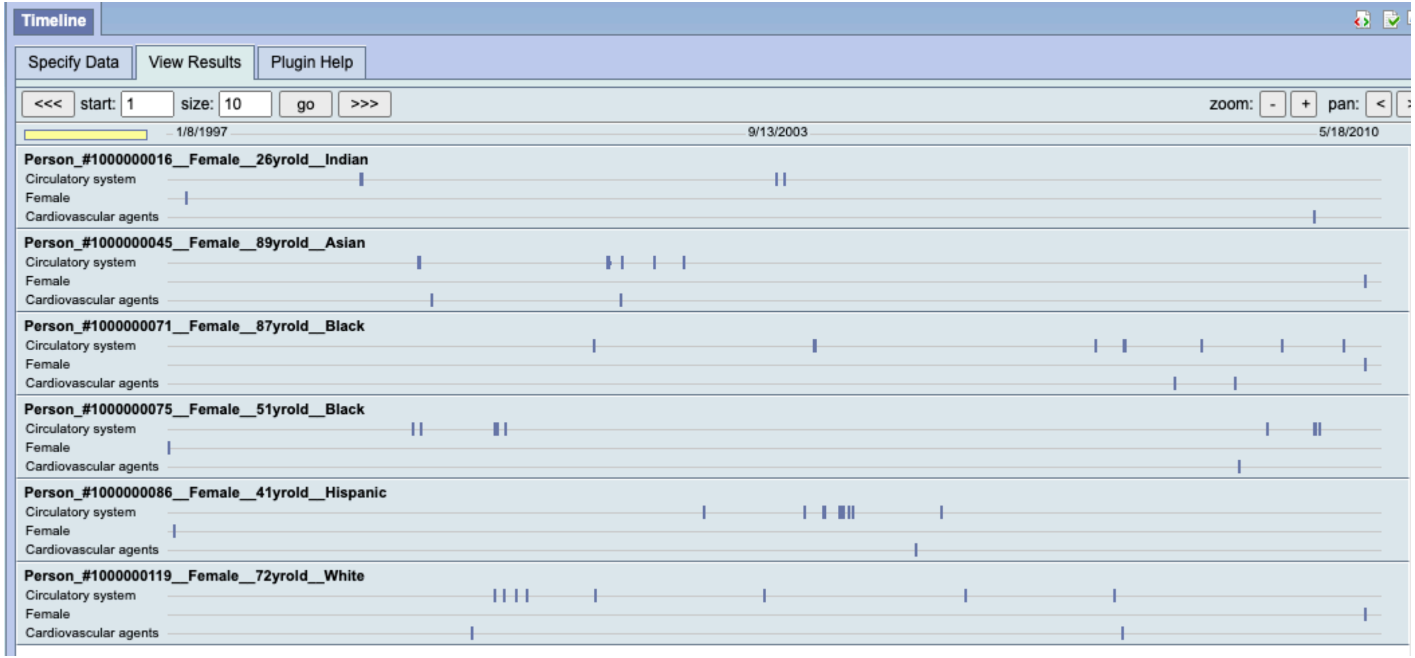
Exercise 1 cont'd

1d) Repeat 1c, but now find patients with three or more occurrences of circulatory system disease. (Answer: 13)

1e) How many patients from 1d used a cardiovascular agent medication? (Answer: 6)

Exercise 1 cont’d

1f) Now view these patients as a timeline.



Person_#1000000016__Female__26yrold__Indian
Circulatory system
Female
Cardiovascular agents

Person_#1000000045__Female__89yrold__Asian
Circulatory system
Female
Cardiovascular agents

Person_#1000000071__Female__87yrold__Black
Circulatory system
Female
Cardiovascular agents

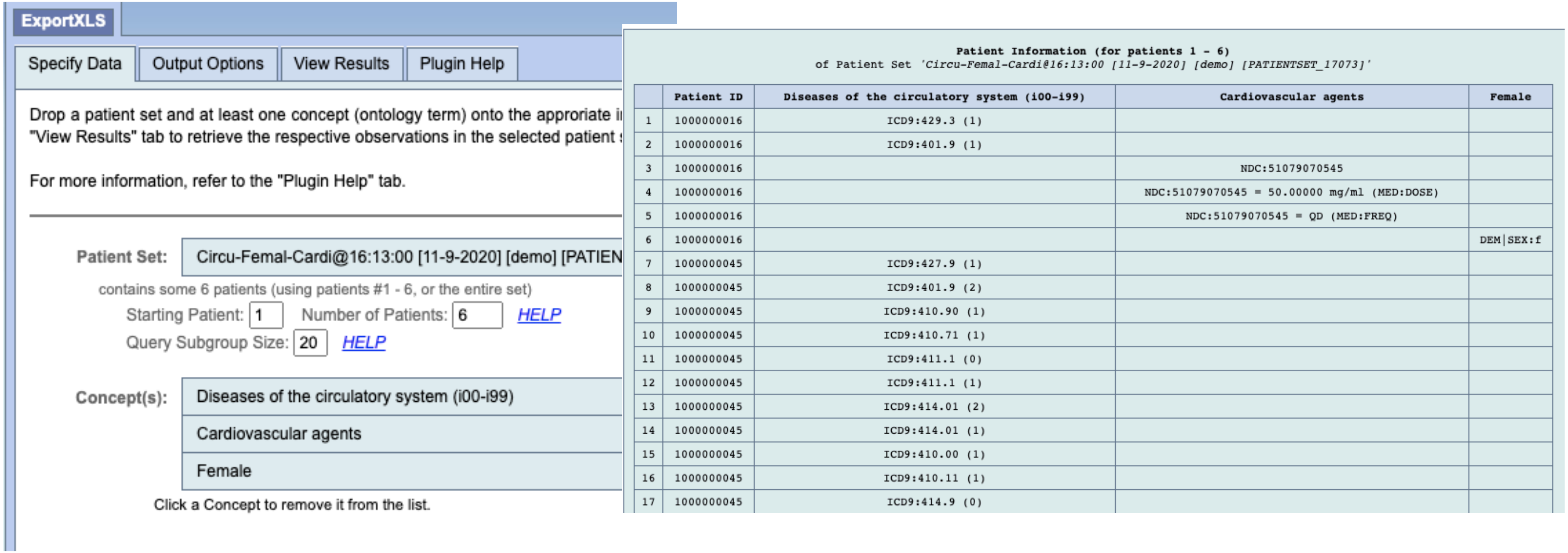
Person_#1000000075__Female__51yrold__Black
Circulatory system
Female
Cardiovascular agents

Person_#1000000086__Female__41yrold__Hispanic
Circulatory system
Female
Cardiovascular agents

Person_#1000000119__Female__72yrold__White
Circulatory system
Female
Cardiovascular agents

Exercise 2: Data Export

Export the previous query using the ExportXLS plugin.



Drop a patient set and at least one concept (ontology term) onto the approriate i
"View Results" tab to retrieve the respective observations in the selected patient s

For more information, refer to the "Plugin Help" tab.

Patient Set: Circu-Femal-Cardi@16:13:00 [11-9-2020] [demo] [PATIEN

contains some 6 patients (using patients #1 - 6, or the entire set)

Starting Patient: 1 Number of Patients: 6 HELP

Query Subgroup Size: 20 HELP

Concept(s): Diseases of the circulatory system (i00-i99)
Cardiovascular agents
Female

Click a Concept to remove it from the list.

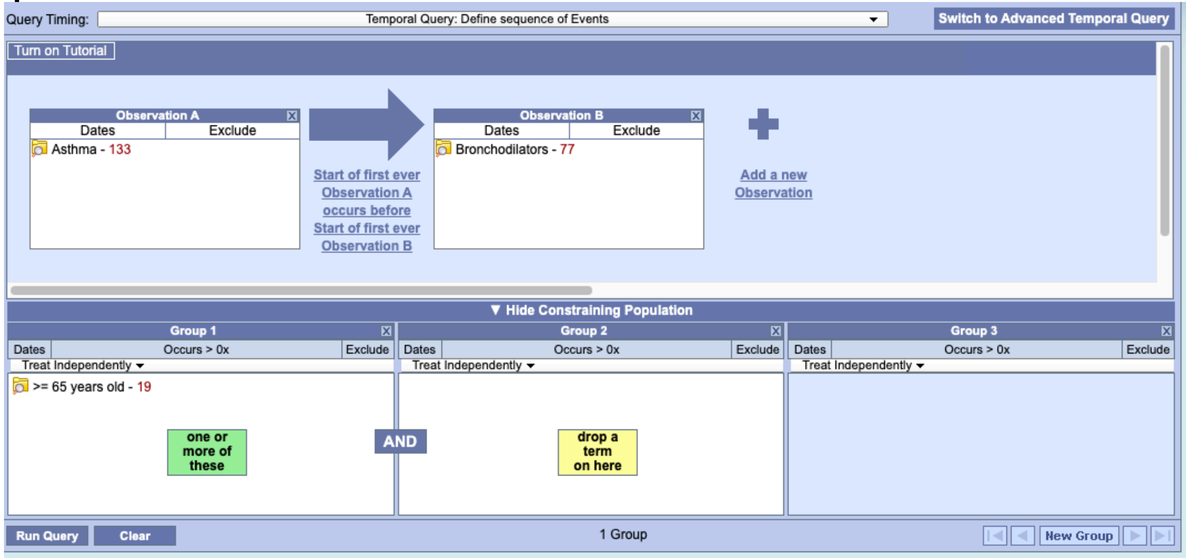
Patient Information (for patients 1 - 6)
of Patient Set *'Circu-Femal-Cardi@16:13:00 [11-9-2020] [demo] [PATIENTSET_17073]'*

	Patient ID	Diseases of the circulatory system (i00-i99)	Cardiovascular agents	Female
1	1000000016	ICD9:429.3 (1)		
2	1000000016	ICD9:401.9 (1)		
3	1000000016		NDC:51079070545	
4	1000000016		NDC:51079070545 = 50.00000 mg/ml (MED:DOSE)	
5	1000000016		NDC:51079070545 = QD (MED:FREQ)	
6	1000000016			DEM|SEX:f
7	1000000045	ICD9:427.9 (1)		
8	1000000045	ICD9:401.9 (2)		
9	1000000045	ICD9:410.90 (1)		
10	1000000045	ICD9:410.71 (1)		
11	1000000045	ICD9:411.1 (0)		
12	1000000045	ICD9:411.1 (1)		
13	1000000045	ICD9:414.01 (2)		
14	1000000045	ICD9:414.01 (1)		
15	1000000045	ICD9:410.00 (1)		
16	1000000045	ICD9:410.11 (1)		
17	1000000045	ICD9:414.9 (0)		

Exercise 2: Temporal query

Among patients over 65, how many have a diagnosis of asthma prior to their first ever prescription of bronchodilators?

Answer: 6



Query Timing: Temporal Query: Define sequence of Events

Observation A — Asthma - 133

Start of first ever Observation A occurs before Start of first ever Observation B

Observation B — Bronchodilators - 77

Group 1 — Occurs > 0x — Treat Independently — >= 65 years old - 19

Group 2 — Occurs > 0x — Treat Independently

Group 3 — Occurs > 0x — Treat Independently

1 Group

Now view this as a timeline as well.