

UF Health Office of Data Science & Research Implementation (ODSRI) IDR Research Data Services

Why an Integrated Data Repository (IDR) for research?

- Clinical care generates massive amounts of data
- Data are stored without much thought for research
- **People**, **Technology**, and **Process** needed to curate these data and deliver to researchers.

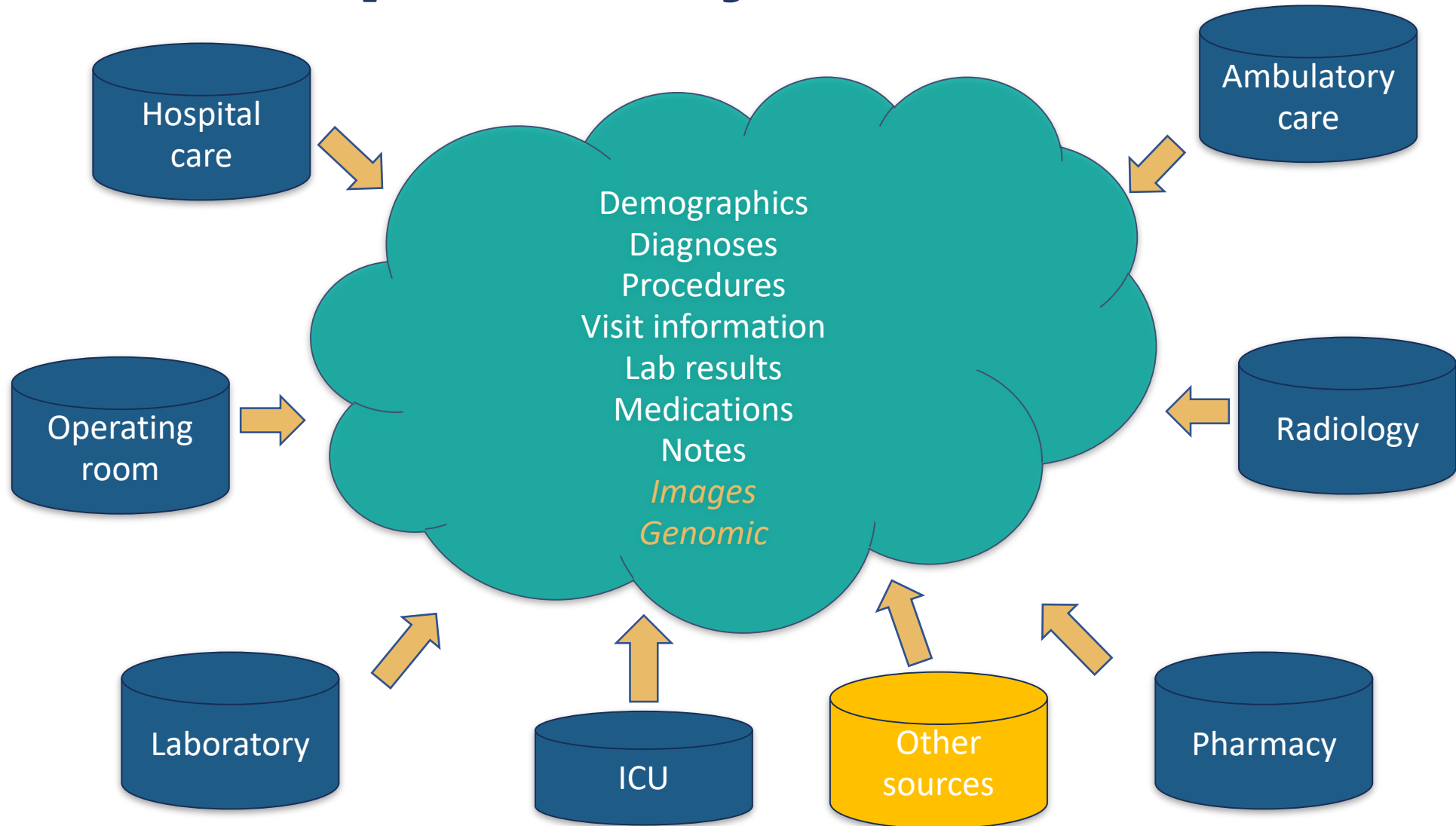


The Technology: UF Clinical Data Warehouse

- Large-scale clinical data warehouse (CDW)
- UF Health data
 - ~2.5 million patients
- Developed over decades by Shands and UF Health decision support services/IT analytics, and CTSI.
- Supports both Epic-generated and non-Epic generated data
- Supports clinical, operational, and research missions

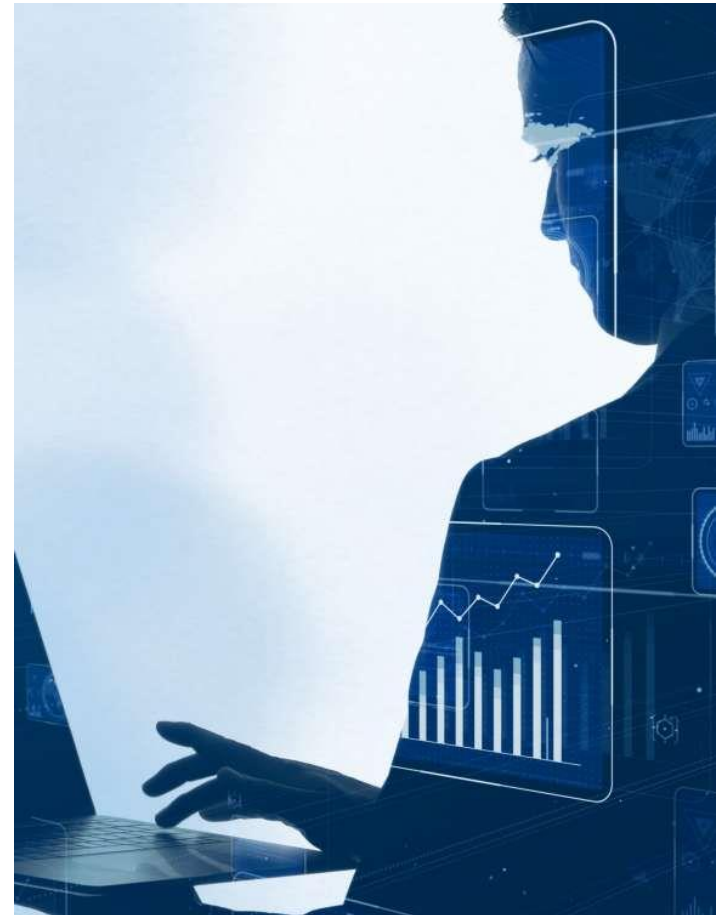


Data sources: Epic and beyond



The Process: IDR Research Services

1. Consultation
2. Feasibility and Cohort discovery
3. Recruitment Support (C2S)
4. Research data
5. OMOP/ATLAS support



1. Consultation

- What is your research objective?
- What data are you hoping to collect?
- What types of data are available that the IDR can provide?



2. Feasibility and cohort discovery

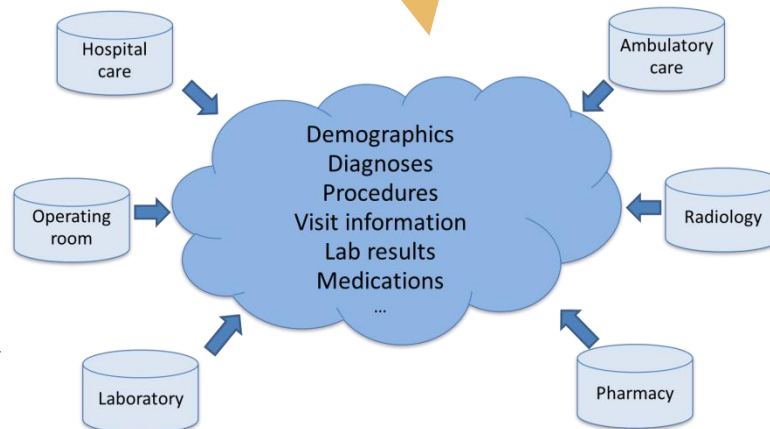
i2b2

How many children were diagnosed with diabetes and had a hemoglobin count > 7% in the same encounter?

The screenshot shows the i2b2 Query & Analysis Tool interface. The 'Query Name' is 'diabetes_A1c7_child'. The 'Temporal Constraint' is 'Selected groups occur in the same financial encounter'. The query is structured with three groups: Group 1 (E06-E13 Diabetes mellitus [54055 patients]), Group 2 (Hemoglobin A1c/Hemoglobin total in Blood (LOINC: 4548-4) [76866 patients] > 7%), and Group 3 (Child (< 18 years old) at visit). The results show 'Number of patients: 48 ± 3 For Query "diabetes_A1c7_child"'. The interface also includes a 'Navigate Terms' pane on the left and a 'Previous Queries' pane at the bottom.

N = 48 patients

Integrated Data Repository



3. Recruitment support

- Consent2Share (C2S) Research Contact Registry
 - Voluntary database of patients consented to be contacted for research opportunities
 - Must be approved as C2S study by IRB
 - In addition to C2S data requests, IRB research services manages deletions from C2S program.
 - Can provide patient name, basic demographics, and contact information
- Patient lists per provider “clinical relationship”
- Coordination with care teams for automated, real-time recruitment (MyChart messages)



4. Research data

- IRB approval required
- Line-level (individual-level) data
- Support observational studies
- May reduce chart review burden
- Wide variation in complexity and effort to complete



patientID	date	diag_code	diag_icd_type	diag_desc
P1	12/7/2005	784	ICD9	HEADACHE
P1	11/12/2012	V14.7	ICD9	Itx-vaccine allergy
P1	11/28/2012	305.1	ICD9	Tobacco use disorder
P2	4/12/2011	V76.51	ICD9	COLON CA SCREENING
P6	3/7/2014	345.9	ICD9	Epilep NOS w/o intr epil
P17	3/19/2004	401.9	ICD9	HYPERTENSION NOS
P9	12/14/2006	434.9	ICD9	CEREB ART OCCL S INFARCT
P2	7/25/2008	599	ICD9	URINARY TRACT INF NOS
P3	5/15/2008	401.1	ICD9	BENIGN HYPERTENSION
P45	7/8/2013	V58.63	ICD9	Lng use antipite/thrombct
P1	11/9/2004	V76.13	ICD9	'U'RFN MAMMOGRAPHY NRC

PatientID	hospital	encounter_date	sex	race	ethnicity
P1	UFHP CLINIC	9/11/2018	MALE	WHITE	HISPANIC
P5	UFHP CLINIC	9/11/2018	FEMALE	WHITE	NOT HISPANIC
P17	UFHP CLINIC	9/13/2018	FEMALE	BLACK	NOT HISPANIC
P3	UF	9/13/2018	MALE	WHITE	NOT HISPANIC
P4	UFHP CLINIC	9/13/2018	FEMALE	WHITE	HISPANIC
P6	UFHP CLINIC	9/18/2018	MALE	ASIAN	NOT HISPANIC
P23	UF	9/19/2018	FEMALE	BLACK	NOT HISPANIC
P2	UFHP CLINIC	9/19/2018	FEMALE	WHITE	NOT HISPANIC
P6	UFHP CLINIC	9/19/2018	MALE	ASIAN	NOT HISPANIC

PatientID	lab_result	lab_unit	normal_low	normal_high	lab date	lab_name
P1		0.02 x10E3/uL	0.03	0.46	9/19/2018 10:22	EOSINOPHILS ABSOLUTE COUNT
P1		11 ng/mL			9/19/2018 10:22	DELTA-9-THC-CARBOXYLIC ACID URINE CONFIRMATION
P2		3.3 mmol/L	3.3	5.1	9/19/2018 10:22	POTASSIUM
P3		None Detected			9/19/2018 10:22	BENZODIAZEPINE SCREEN, URINE
P4		0.5 %	0	2	9/19/2018 10:22	BASOS
P17		0.61 x10E3/uL	0.2	0.7	9/19/2018 10:22	MONOCYTES ABSOLUTE COUNT
P23		10 IU/L	0	35	9/19/2018 10:22	ALT (SGPT)
P3		11 mmol/L	8	16	9/19/2018 10:22	ANION GAP
P12		3.07 x10E6/uL	4	5.2	9/19/2018 10:22	RED BLOOD COUNT
P1		34 seconds	25	38	9/19/2018 10:22	PARTIAL THROMBOPLASTIN TIME

The Process: Levels of Service

IDR analysts can guide in matching research teams to appropriate data and tools

1. Free services
 - Consultations, i2b2, OMOP/ATLAS support
2. Low cost services
 - Recruitment Support (C2S) or simple MRN Only requests
3. More costly and time intensive services
 - Complex data request with detailed elements

INTRO TO i2b2

<https://i2b2.idr.ufhealth.org>



**Clinical and Translational
Science Institute**

Integrated Data Repository Research Services

UNIVERSITY of FLORIDA

Claire Layton

Data Management Analyst II
Integrated Data Repository (IDR)

Outline

- What is i2b2?
- How can I obtain access to i2b2?
- How can i2b2 help me with research questions?
- i2b2 demonstration

i2b2

- **Informatics for Integrating Biology & the Bedside (i2b2)**
- An NIH-funded tool built at Harvard
- Feasibility and cohort discovery tool
- Provides HIPAA-compliant, limited data
 - Returns an obfuscated number of unique patients that meet query criteria
- **Self-serve tool!**

What typical questions can i2b2 answer?

- How many men with hypertension diagnoses had at least one inpatient encounter between January and December 2018?
- How many women ages 45-65 had successful breast cancer surgery?
- How many children were diagnosed with diabetes and had Hemoglobin A1c greater than 7% in the same financial encounter?

Data availability in i2b2

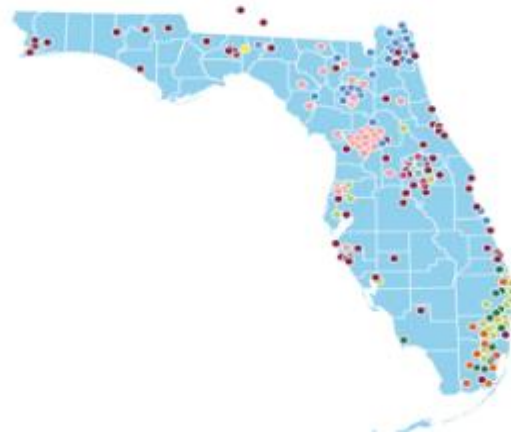
- Data guide
 - <https://idr.ufhealth.org/research-services/feasibility-cohort-discovery/uf-health-i2b2-training-tools/>
 - Click on 'The i2b2 Data Guide'
- Date range: 6/1/2011 – Today-ish
 - Data refreshed quarterly, is current up to the last quarter
- Data is represented as closely as possible to the original source

i2b2 instances available at UF

ACT
The ACT Network
Powered by MCAT® CTR Program

OneFlorida
Clinical Research Consortium

UFHealth
UNIVERSITY OF FLORIDA HEALTH



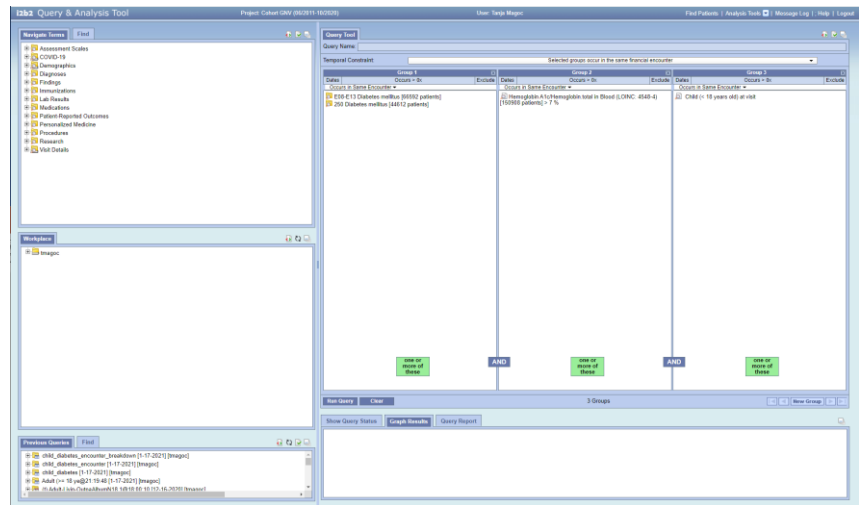
i2b2 instance	Geography
UFHealth	Gainesville, Jacksonville
OneFlorida	Florida
ACT Network	Nationwide

Access i2b2

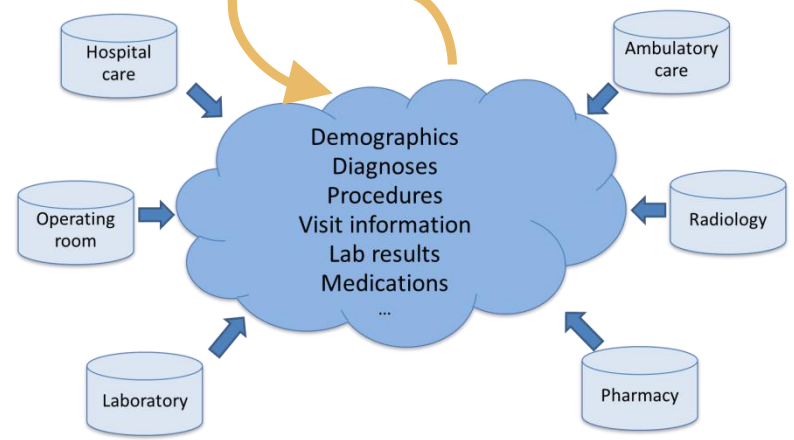
- Request access:
 - <https://idr.ufhealth.org/research-services/data-request-form/>
 - Select 'i2b2 Registration'
- View training material:
 - <https://idr.ufhealth.org/research-services/feasibility-cohort-discovery/uf-health-i2b2-training-tools/>
- I2B2 Support:
 - i2b2support@ahc.ufl.edu
 - IRBDataRequest@ahc.ufl.edu

How many children were diagnosed with diabetes and had a hemoglobin A1c count > 7% in the same encounter?

N = 1,622 patients



Integrated Data Repository



i2b2 Live Demo

How many children were diagnosed with diabetes and had Hemoglobin a1c greater than 7% in the same financial encounter?

<https://i2b2.idr.ufhealth.org/>

I2B2 Interface

i2b2 Query & Analysis Tool Project: Cohort GNV (06/2011-06/2022) User: Lei McDevitt Find Patients | Analysis Tools | Message Log | Help | Logout

Terms Find Trm Info Workplace Queries Find Qry

- Assessment Scales
- COVID-19
- Demographics
- Diagnoses
- Findings
- Immunizations
- Lab Results
- Medications
- Patient-Reported Outcomes
- Personalized Medicine
- Procedures
- Research
- Visit Details

Query Tool

Query Name: _____

Query Timing: 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		

Query build Tool

drop a term on here

Run Query Clear 0 Groups New Group

Show Query Status Graph Results Query Report

Query Status/Results/Reports

i2b2 - Navigate Terms

Navigate Terms Find

- Assessment Scales
- COVID-19
- Demographics
- Diagnoses
 - ICD-10-CM [1066108 patients]**
 - ICD-9-CM [692828 patients]
 - AHRQ Comorbidities (Inpatient and Observation visit types only) [152777 patients]
- Findings
- Immunizations
- Lab Results
- Medications
- Patient-Reported Outcomes
- Personalized Medicine



Navigate Terms Find

- Assessment Scales
- COVID-19
- Demographics
- Diagnoses
 - ICD-10-CM [1066108 patients]
 - Present on Admission
 - Priority
 - Source
 - A00-B99 Certain infectious and parasitic diseases [179189 patients]
 - C00-D49 Neoplasms [299354 patients]
 - D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism [111313 patients]
 - E00-E89 Endocrine, nutritional and metabolic diseases [252308 patients]**
 - F01-F99 Mental, Behavioral and Neurodevelopmental disorders [237747 patients]
 - G00-G99 Diseases of the nervous system [237511 patients]
 - H00-H59 Diseases of the eye and adnexa [108060 patients]



Navigate Terms Find

- Assessment Scales
- COVID-19
- Demographics
- Diagnoses
 - ICD-10-CM [1066108 patients]
 - Present on Admission
 - Priority
 - Source
 - A00-B99 Certain infectious and parasitic diseases [179189 patients]
 - C00-D49 Neoplasms [299354 patients]
 - D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism [111313 patients]
 - E00-E89 Endocrine, nutritional and metabolic diseases [252308 patients]
 - Present on Admission
 - Priority
 - Source
 - E00-E07 Disorders of thyroid gland [59160 patients]
 - E08-E13 Diabetes mellitus [66592 patients]**
 - E15-E16 Other disorders of

i2b2 - Navigate Terms

Query Name: diabete@12:34:49

Query Timing: Treat all groups independently

Group 1	Group 2	Group 3
<p>Dates Occurs > 0x Exclude</p> <p>Treat Independently</p> <ul style="list-style-type: none">E08-E13 Diabetes mellitus [86376 patients]250 Diabetes mellitus [44912 patients]	<p>Dates Occurs > 0x Exclude</p> <p>Treat Independently</p>	<p>Dates Occurs > 0x Exclude</p> <p>Treat Independently</p>

one or more of these AND drop a term on here

Run Query Clear 1 Group New Group

Show Query Status Graph Results Query Report

Number of patients
107963±3
For Query diabete@12:34:49"

i2b2 - Navigate Terms

Navigate Terms

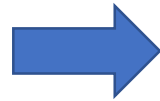
Find

Search by Names Search by Codes

Containing hemoglobin a1c

Find Any Category

- Hemoglobin A1c (LOINC: LP16413-4) [152825 patients]
- Hemoglobin A1c/Hemoglobin.total (LOINC: LP286151-8) [152825 patients]
- Hemoglobin A1c/Hemoglobin.total | Blood | Hematology and Cell counts (LOINC: LP393563-4) [152825 patients]
- Hemoglobin A1c/Hemoglobin.total in Blood (LOINC: 4548-4) [150908 patients]**
- Hemoglobin A1c/Hemoglobin.total in Blood by calculation (LOINC: 17855-8) [5910 patients]



Choose value of - (Test:Hgb A1c MFr Bld)

Searches by Lab values can be constrained by the high/low flag set by the performing laboratory, or by the values themselves.

No Value

By Flag

By Value

Please select operator:
GREATER THAN (>)

Please enter a value:
7

Units = %

OK Cancel

i2b2 – Query Tool

- Terms
 - Find Trm
 - Info
 - Workplace
 - Queries
 - Find Qry
- Assessment Scales
 - Braden Scale [346620 patients]
 - Morse Fall Risk Scale [418480 patients]
 - REALM Adult Health Literacy [152683 patients]
 - COVID-19
 - Demographics
 - Age
 - 0-9 years old [120859 patients]
 - 10-17 years old [114984 patients]
 - 18-34 years old [374174 patients]
 - 35-44 years old [180657 patients]
 - 45-54 years old [185280 patients]
 - 55-64 years old [268839 patients]
 - 65-74 years old [276623 patients]
 - 75-84 years old [188295 patients]
 - >= 65 years old [554480 patients]
 - >= 85 years old [89562 patients]
 - Adult (>= 18 years old)
 - Child (< 18 years old)
 - Not Recorded
 - County
 - Death Date [50745 patients]
 - Ethnicity
 - Gender
 - History of Cancer Diagnosis
 - Language
 - Marital Status
 - MyChart Status
 - Payer
 - Race
 - Vital Status
 - Zip Codes
 - Known axiUm Patient [108002 patients]
 - Known Healthstreet Patient [3785 patients]
 - Known Tumor Registry Patient [46023 patients]
 - Diagnoses
 - ICD-10-CM [1405909 patients]
 - Present on Admission

Query Tool

Query Name: E08-E-Hemog- Chil@13:34:30

Query Timing: 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		
E08-E13 Diabetes mellitus [86376 patients]			Hemoglobin A1c/Hemoglobin.total in Blood (LOINC: 4548-4) [196785 patients] > 7 %			Child (< 18 years old) at visit		
250 Diabetes mellitus [44912 patients]								

one or more of these AND one or more of these AND one or more of these

Run Query Clear 3 Groups New Group

Show Query Status Graph Results Query Report

Number of patients

1791±3

For Query E08-E-Hemog- Chil@13:34:30"

i2b2 - Query Tool

i2b2 Query & Analysis Tool Project: Cohort GNV (06/2011-10/2020) User: Tanja Magoc Find Patients | Analysis Tools | Message Log | Help | Logout

Navigate Terms Find

- Present on Admission
- Priority
- Source
- E00-E07 Disorders of thyroid gland [59100 patients]
- E08-E13 Diabetes mellitus [66592 patients]**
- E15-E16 Other disorders of glucose regulation and pancreatic internal secretion [6620 patients]

Workplace

- tmagoc

Query Tool

Query Name: _____

Temporal Constraint:

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		

Run Query Clear 0 Groups New Group

i2b2 - Query Tool

i2b2 Query & Analysis Tool Project: Cohort GNV (06/2011-10/2020) User: Tanja Magoc Find Patients | Analysis Tools | Message Log | Help | Logout

Navigate Terms Find

- Present on Admission
- Priority
- Source
- E00-E07 Disorders of thyroid gland [59160 patients]
- E08-E13 Diabetes mellitus [66592 patients]**
- E15-E16 Other disorders of glucose regulation and pancreatic internal secretion [6620 patients]

Workplace

- tmagoc

Query Tool

Query Name: _____

Temporal Constraint:

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		
E08-E13 Diabetes mellitus [66592 patients]								

Run Query Clear 1 Group New Group

i2b2 - Query Tool

i2b2 Query & Analysis Tool Project: Cohort GNV (06/2011-10/2020) User: Tanja Magoc Find Patients | Analysis Tools | Message Log | ; Help | Logon

Navigate Terms Find

- Visit Details
 - Age at visit
 - Adult (>= 18 years old) at visit
 - Child (< 18 years old) at visit**
 - Visit at age 9 or less
 - Visit at ages 10-17 years old
 - Visit at ages 18-34 years old
 - Visit at ages 35-44 years old
 - Visit at ages 45-54 years old

Workplace

- tmagoc

Query Tool

Query Name: child_diabetes_encounter_breakdown

Temporal Constraint: Selected groups occur in the same financial encounter

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Occurs in Same Encounter			Occurs in Same Encounter			Occurs in Same Encounter		
E08-E13 Diabetes mellitus [66592 patients] 250 Diabetes mellitus [44612 patients]			Hemoglobin A1c/Hemoglobin total in Blood (LOINC: 4548-4) [150908 patients] > 7 %			Child (< 18 years old) at visit		

Run Query Clear 3 Groups New Group

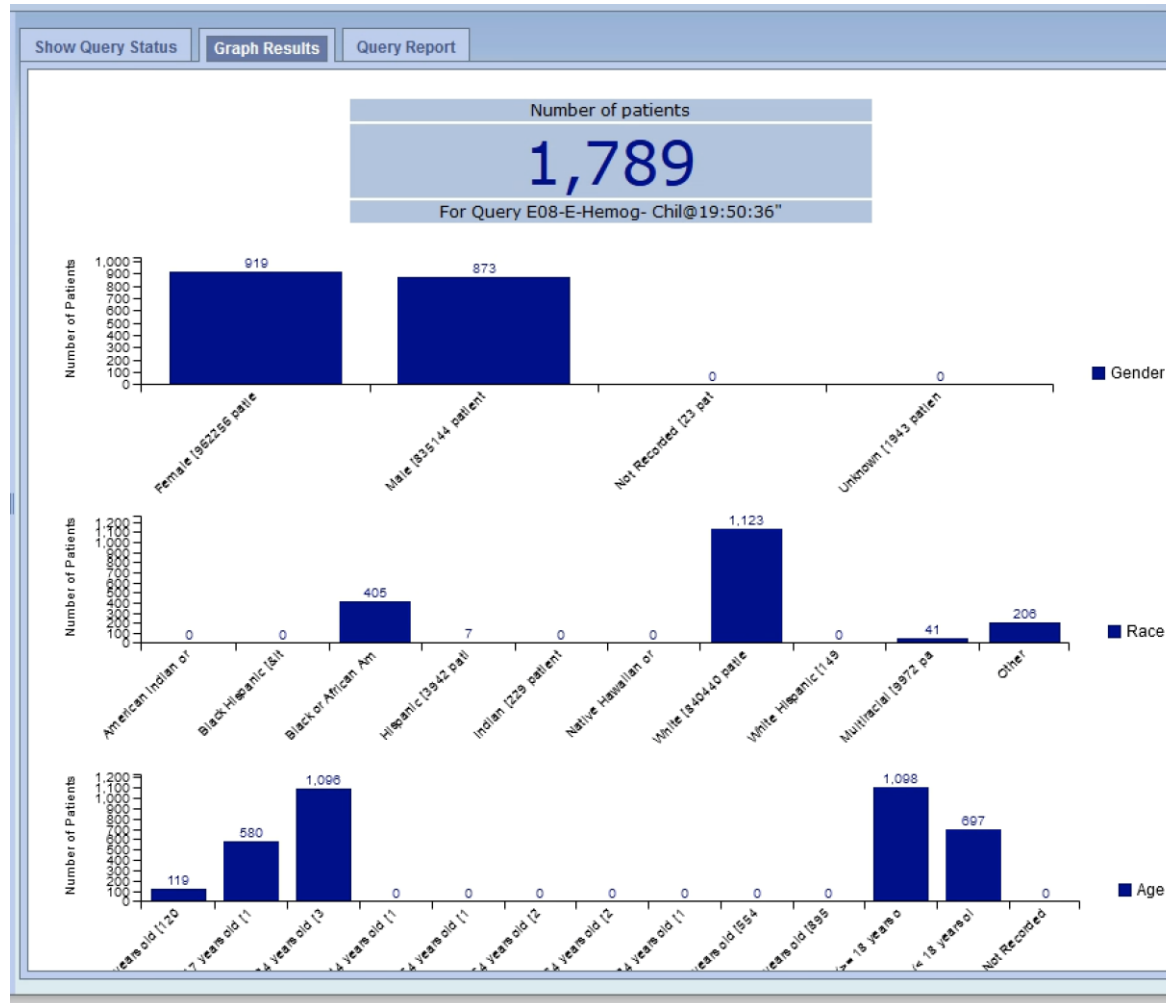
I2B2 -Results

The screenshot displays the I2B2 Query Tool interface. On the left, the 'Terms' panel shows a hierarchical tree of categories including Assessment Scales, COVID-19, Demographics, Diagnoses, Findings, Immunizations, Lab Results, Medications, Patient-Reported Outcomes, Personalized Medicine, Procedures, Research, and Visit Details. The 'Workplace' panel below it contains the user 'leimcdevitt'. The 'Queries' panel at the bottom left lists several queries, with the top one being 'E08-E-Hemog- Chil@16:18:30 [8-30-2022] [leimcdevitt]'. The main 'Query Tool' window shows a query named 'E08-E-Hemog- Chil@16:18:30' with a timing of 'Non-Temporal Query: Treat Independently'. The query is configured with three groups:

Group 1	Group 2	Group 3
E08-E13 Diabetes mellitus [86376 patients] 250 Diabetes mellitus [44912 patients]	Hemoglobin A1c/Hemoglobin.total in Blood (LOINC: 4548-4) [196785 patients] > 7 %	Child (< 18 years old) at visit

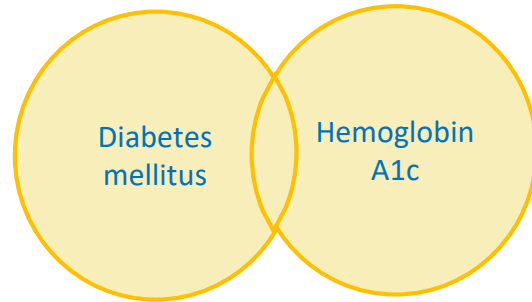
Each group is set to 'Treat Independently' and 'Occurs > 0x'. The groups are connected by 'AND' operators. Below the query configuration, there are three green boxes labeled 'one or more of these'. At the bottom of the query tool, there are buttons for 'Run Query', 'Clear', and 'New Group', along with a status indicator '3 Groups'. The results panel at the bottom right shows a summary: 'Number of patients' is '1790±3' for the query 'E08-E-Hemog- Chil@16:18:30'. This result is circled in green.

i2b2 - Results

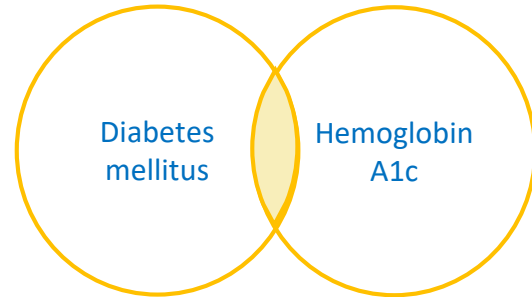


i2b2 - Query Logic

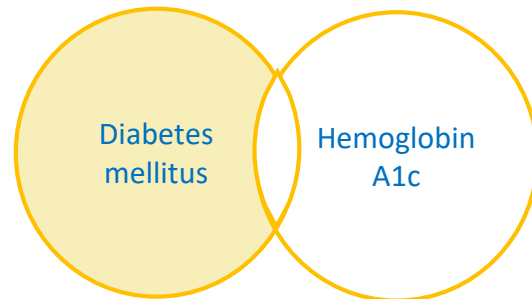
OR



AND



NOT



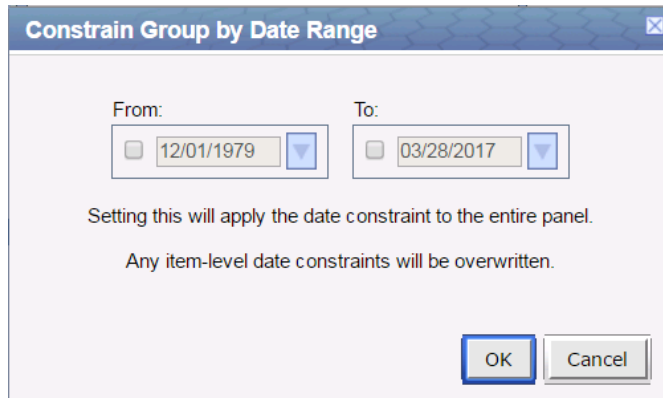
Group 1			Group 2		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently		
E08-E13 Diabetes mellitus [66592 patients]					
Hemoglobin A1c/Hemoglobin.total in Blood (LOINC: 4548-4) [150908 patients] > 7 %					

Group 1			Group 2		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently		
E08-E13 Diabetes mellitus [66592 patients]			Hemoglobin A1c/Hemoglobin.total in Blood (LOINC: 4548-4) [150908 patients] > 7 %		

Group 1			Group 2		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently		
E08-E13 Diabetes mellitus [66592 patients]			NOT Hemoglobin A1c/Hemoglobin.total in Blood (LOINC: 4548-4) [150908 patients] > 7 %		

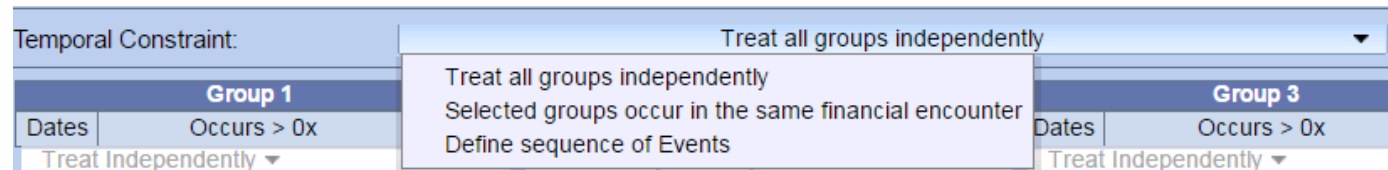
Additional filters

- Date range



The screenshot shows a dialog box titled "Constrain Group by Date Range". It contains two date input fields: "From:" with the date "12/01/1979" and "To:" with the date "03/28/2017". Below the fields, there is a message: "Setting this will apply the date constraint to the entire panel. Any item-level date constraints will be overwritten." At the bottom right, there are "OK" and "Cancel" buttons.

- Temporal constraints:
 - Treat all groups independently
 - Groups occur in the same financial encounter
 - Sequence of events



The screenshot shows a table with two columns, "Group 1" and "Group 3". Each column has a "Dates" header and a value "Occurs > 0x". Below the table, there is a dropdown menu for "Temporal Constraint:" with the following options: "Treat all groups independently", "Selected groups occur in the same financial encounter", and "Define sequence of Events". The "Treat all groups independently" option is currently selected.

Sequence of event

- Research interest:
 - Diagnosed with asthma: ICD 10 J45 → Define population
 - Treated with **SULFONAMIDE**: RXNORM 691843 → Event 1
 - Had adverse reaction to **SULFONAMIDE**: ICD10 T37.0 → Event 2

Sequence: event 1 occurred ≥ 30 days before event 2

Dx with Asthma

Treated w/
SULFONAMIDE

Adverse Reaction to
SULFONAMIDE

Accessing Patient Line Level Data



- Need an approved IRB
- Submit an online data request form <https://idr.ufhealth.org/services/analyst-data-support-services/idr-data-request-form/>
- Select IRB-Approved Line-level Data Elements and in body of request form, indicate that you have already identified your cohort in i2b2
 - Enter the exact name of your i2b2 query
- Contact IDR Data Team IRBDataRequest@ahc.ufl.edu for assistance or questions
- Or contact me directly claire.layton@ufl.edu

i2b2 QUESTIONS?