

Intro to the Integrated Data Repository (IDR)

Dr. Chris Harle

Becky Liao

Integrated Data Repository (IDR)

Chris Harle



- Professor, Health Outcomes and Biomedical Informatics
- Chief Research Information Officer, UF Health
 - Advancing the use of health system data and information technology in research

Becky Liao, IDR Analyst @ the CDS



- Hello! 😊
- Recent graduate from Johns Hopkins University with a B.S. in Applied Mathematics and Computer Science
- Joined IDR as a Data Management Analyst in February
- Working on bridging the gap between the UF Health campuses in JAX and GNV in terms of the medical record data in the IDR

What is the IDR?

- A large-scale clinical data warehouse (CDW)
- UF Health data
 - ~1.96 million patients, >1.6 billion facts
- Enables new research discoveries
as well as improvements in the quality
and safety of patient care



Why the IDR?

- Clinical care generates massive amounts of data, which can be valuable for research



- but data are stored in different places and formats without much thought for research



Why the IDR?



Solution: IDR analysts navigate these huge databases and using their knowledge to run the appropriate queries, provide researchers with the data needed for their studies.

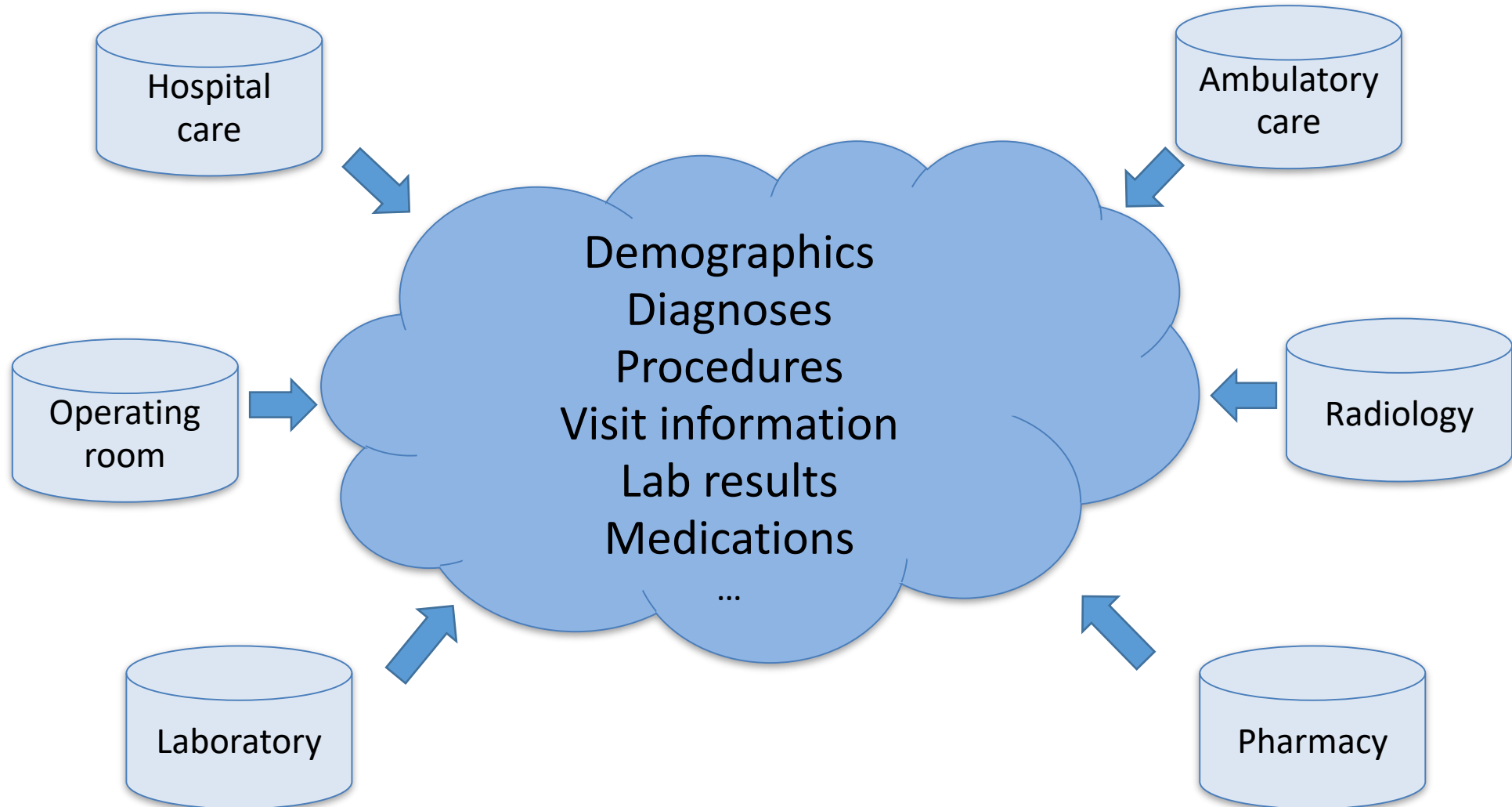


What *will* the IDR be?

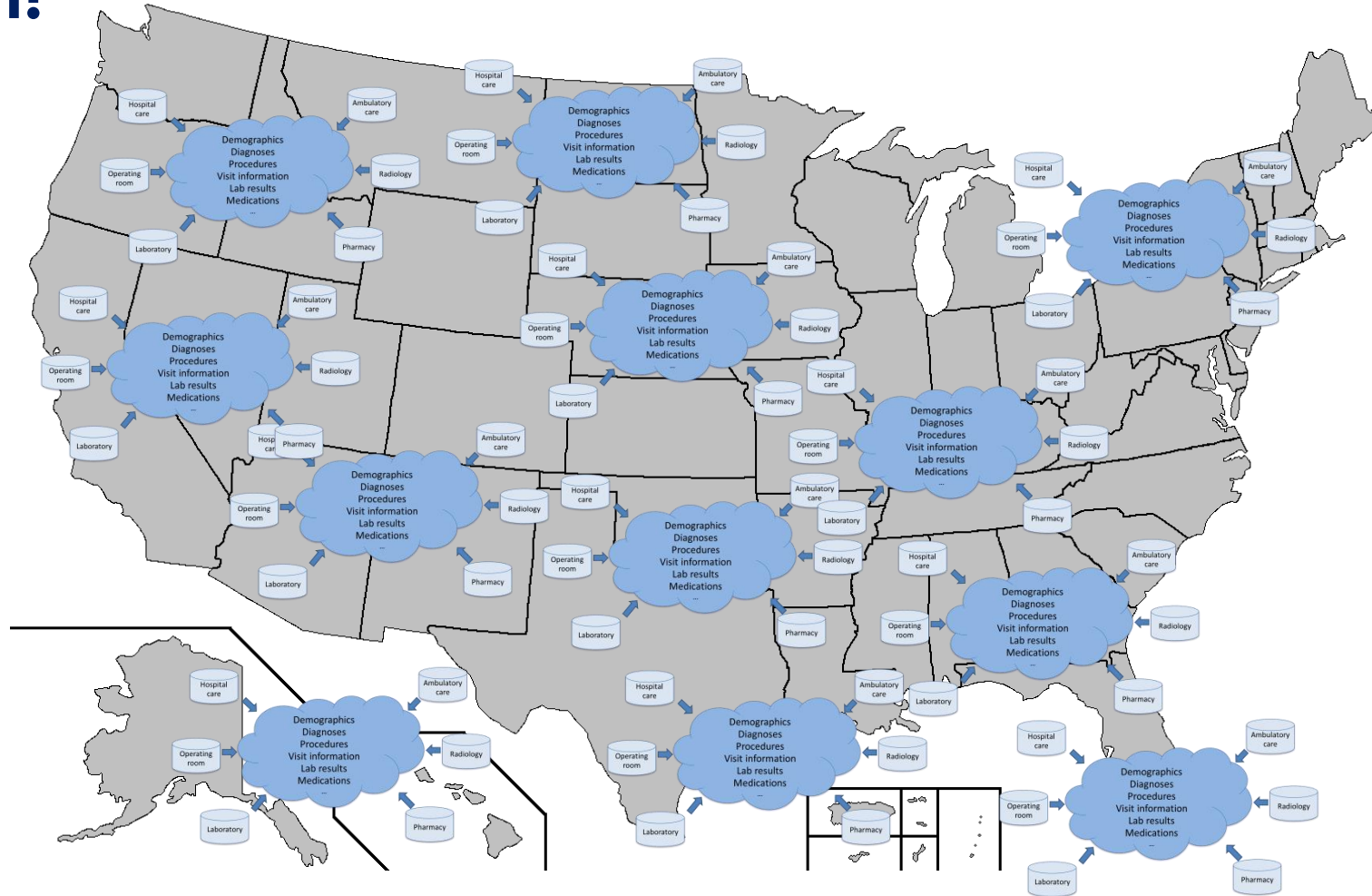
Vision for UF Health Integrated Data Repository

- Enterprise-wide support and collaboration
- Patient-oriented data in a single secure place
- A single vocabulary to simplify the use of clinical information for research

Vision for UF Health Integrated Data Repository



Expansion!

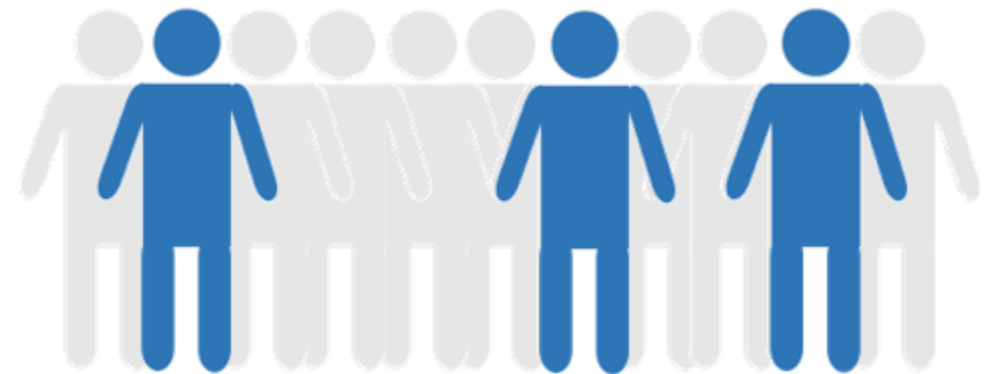


Services offered by IDR

1. Cohort discovery
2. Patient data for retrospective studies
3. Consent2Share
 - Patient recruitment for research studies

1. Cohort discovery

- Establishing study feasibility
 - Determine the number of patients matching study criteria (e.g. patient demographics, ICD and CPT codes, etc.)
- A key first step in study design!



i2b2 <https://i2b2.idr.ahc.ufl.edu/>

- **Informatics for Integrating Biology & the Bedside (i2b2)**
- A NIH-funded data exploration and cohort discovery tool
- Provides HIPAA-compliant, query de-identified data via a web application
- Returns an obfuscated number of unique patients that meets query criteria
 - Data-sets are IRB-approved – no need to submit a proposal to query i2b2
- For UFHealth, there are two instances: GNV and JAX

What data are NOT available?

1. Patient identifiers
2. Clinician identifiers
3. & any HIPAA identifiers EXCEPT dates

Data loaded quarterly!

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

i2b2

Project: Cohort JAX (03/2013-09/2019) User: Rebecca Liao Find Patients | Analysis Tools | Message Log | Help | Logout

Query Name: diabetes_A1c7_child

Temporal Constraint: Selected groups occur in the same financial encounter

Group 1: Occurs > 0x Exclude Dates Occurs in Same Encounter
E06-E13 Diabetes mellitus [54055 patients]

Group 2: Occurs > 0x Exclude Dates Occurs in Same Encounter
Hemoglobin A1c/Hemoglobin total in Blood [LOINC: 4548-4] [76866 patients] > 7%

Group 3: Occurs > 0x Exclude Dates Occurs in Same Encounter
Child (< 18 years old) at visit

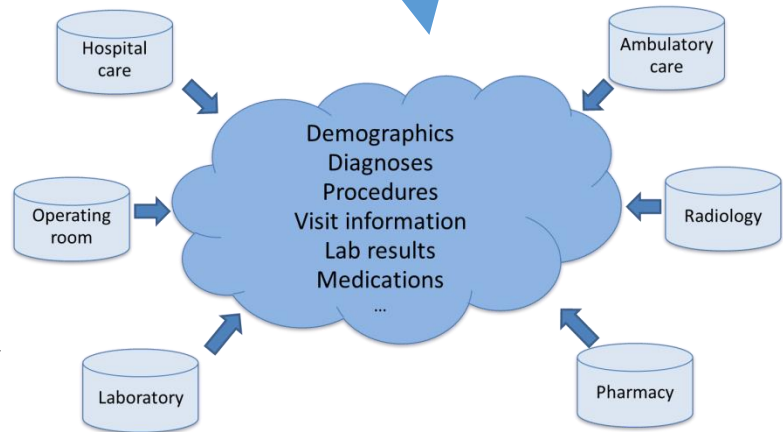
one of more of these AND one of more of these AND one of more of these

3 Groups

Number of patients
48 ± 3
For Query "diabetes_A1c7_child"

N = 48 patients

Integrated Data Repository



Navigate Terms

Find

- causes [130848 patients]
 - V00-Y99 External causes of morbidity [103721 patients]
 - Z00-Z99 Factors influencing health status and contact with health services [383110 patients]
- ICD-9-CM [437654 patients]
- Findings
- Immunizations
- Lab Results
- Medications
- Personalized Medicine
- Procedures
- Visit Details

Workplace

ria0001

Previous Queries

Find

- diabetes_A1c7_child [2-27-2020] [ria0001]
- diabetes_A1c7_child [2-27-2020] [ria0001]

Query Tool

Query Name: diabetes_A1c7_child

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		
<ul style="list-style-type: none"> E08-E13 Diabetes mellitus [54055 patients] 250 Diabetes mellitus [41116 patients] 			<ul style="list-style-type: none"> Hemoglobin A1c/Hemoglobin.total in Blood (LOINC: 4548-4) [76866 patients] > 7 % 			<ul style="list-style-type: none"> Child (< 18 years old) at visit 		

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

Run Query

Clear

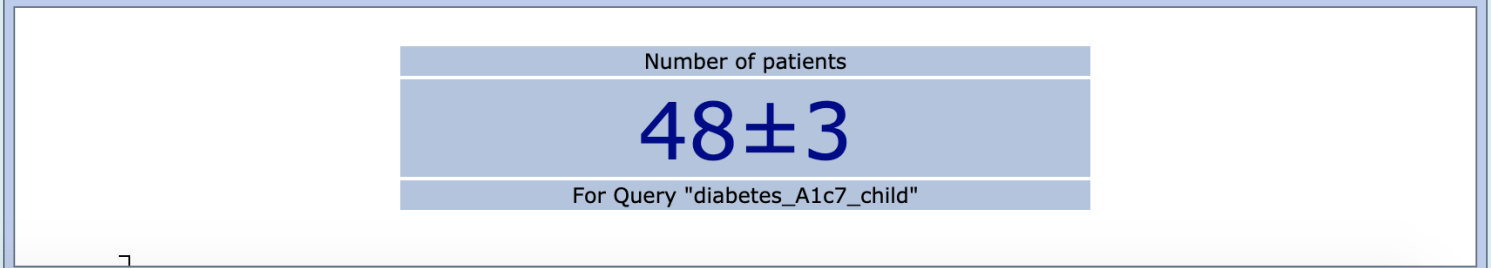
3 Groups

Navigation buttons: back, forward, New Group

Show Query Status

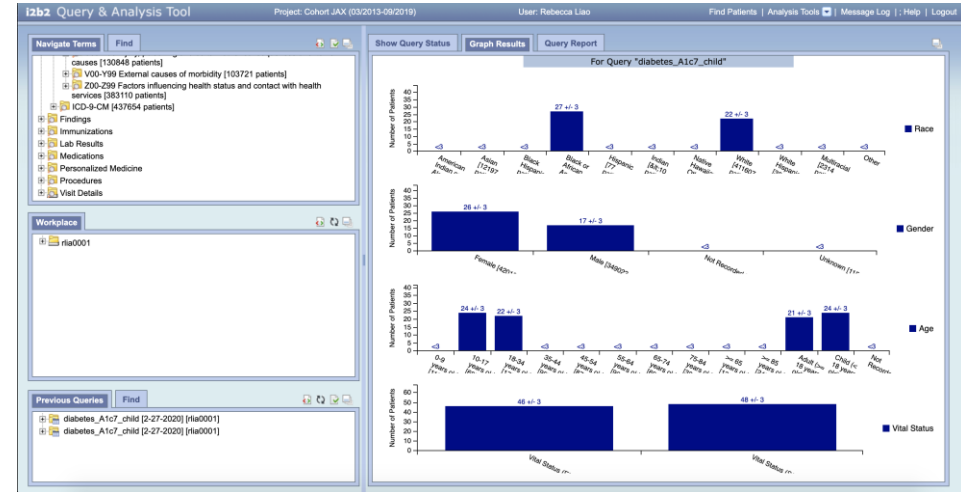
Graph Results

Query Report



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

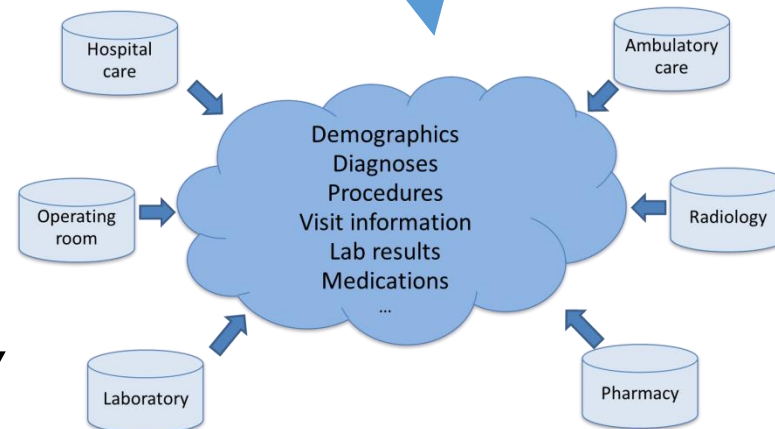
i2b2



Female: 26 N = 48
Male: 17 patients

Black/African American 27
White: 22
Other: <3

Integrated Data Repository



i2b2 instances available at UF



125 MILLION PATIENTS
41 SITES CONNECTED
AND GROWING.



Connected to ACT:

- Boston University
- Children's National
- Columbia University
- Duke University
- Emory Univ./Morehouse Univ.
- Harvard University
- Indiana University
- Johns Hopkins University
- Mayo Clinic
- Medical College of Wisconsin
- Medical University of South Carolina
- New York University
- Northwestern University
- Ohio State University
- Oregon Health & Science University
- Pennsylvania State University
- Stanford University
- University of Alabama at Birmingham
- U. of Arkansas for Medical Sciences
- University of California, Davis

- University of California, Irvine
- University of California, Los Angeles
- University of California, San Diego
- University of California, San Francisco
- Univ of Cincinnati/Cincinnati Children's
- Univ of Colo/Children's Hosp. Colorado
- University of Florida
- University of Illinois-Chicago
- University of Kansas
- University of Kentucky
- University of Minnesota
- University of North Carolina at Chapel Hill
- University of Pittsburgh
- University of Southern California
- UTHealth Houston
- UT Health San Antonio
- UT Southwestern
- Vanderbilt University Medical Center
- Virginia Commonwealth Univ.
- Washington University in St. Louis
- Weill Cornell Medicine

Staging for ACT:

- Case Western University
- Dartmouth College
- Scripps Research / Scripps Health
- Tufts University
- University at Buffalo
- University of Massachusetts
- University of Miami
- University of Michigan
- University of New Mexico
- University of Rochester
- University of Texas Medical Branch
- University of Utah
- University of Virginia
- University of Washington
- University of Wisconsin-Madison
- Wake Forest University



- o Florida Hospital (AdventHealth)
- o Orlando Health
- o Tallahassee Memorial Inpatient
- o Tallahassee Memorial Outpatient
- o UF Health
- o University of Miami
- o Nicklaus Children's Hospital
- o Capital Health Plan
- o Florida Medicaid

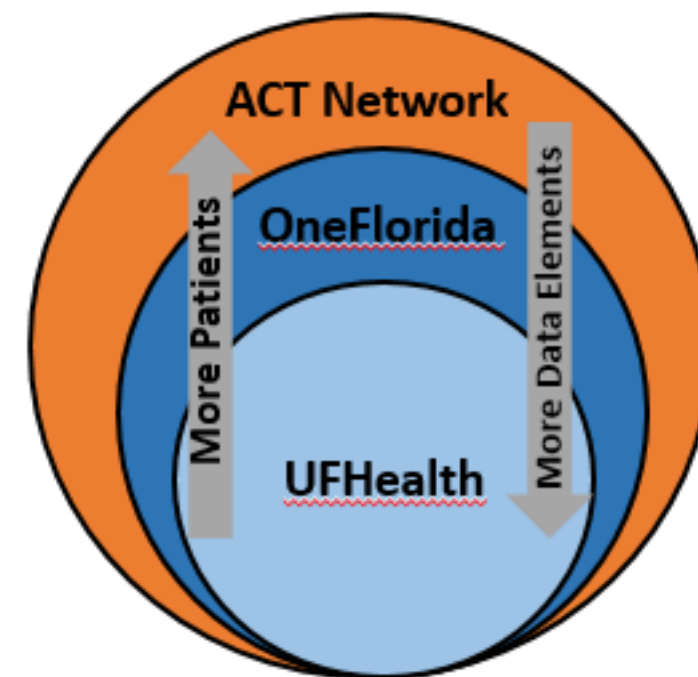


i2b2 instance	Geography
UFHealth	Gainesville, Jacksonville
OneFlorida	Florida
ACT Network	Nationwide

i2b2 instances available at UF

i2b2 instance	Patients	Facts	Geography
UFHealth – GNV	1.27M	1.26B	Gainesville
UFHealth – JAX	0.77 M	0.37B	Jacksonville
OneFlorida	13.8M	18.4B	Florida
ACT Network	125M	>50B	Nationwide

Data as of January 1, 2020.



i2b2 resources

UFHealth i2b2

- Project overview: <https://idr.ufhealth.org/i2b2/>
- Training: <http://i2b2.idr.ufhealth.org/education/>, <https://idr.ufhealth.org/i2b2/i2b2-training-tools/>
- Request access: <https://idr.ufhealth.org/i2b2/i2b2-registration/>
- i2b2 support: i2b2support@ahc.ufl.edu

OneFlorida i2b2

- Project overview: <https://onefloridaconsortium.org/front-door/onfli2b2/>
- Training: <https://onefloridaconsortium.org/front-door/onfli2b2/>, go to "Training" section
- Request access: <https://onefloridaconsortium.org/i2b2/webform/i2b2.html>
- i2b2 support: i2b2support@onefloridaconsortium.org

NCATS ACT i2b2

- Project overview: <http://www.actnetwork.us/FloridaCTSI>
- Training: <http://www.actnetwork.us/FloridaCTSI>, click on "How-To Resources"
- Request access: <https://idr.ufhealth.org/services/act-i2b2-cohort-discovery/ncats-act-i2b2-registration/>
- i2b2 support: ACT-I2B2Support@ahc.ufl.edu

2. Retrospective patient data

- Reliable
- Secure
- Protects patients' privacy



patientID	date	diag_code	diag_icd_type	diag_desc
P1	12/7/2005	784	ICD9	HEADACHE
P3	11/12/2012	V14.7	ICD9	Itc-vaccine allergy
P1	11/28/2012	305.1	ICD9	Tobacco use disorder
P2	4/12/2013	V76.51	ICD9	COLON CA SCREENING
P6	3/7/2014	345.9	ICD9	Epilep NOS w/o Intra-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 antipla/thromboc
P1	11/4/2004	V76.12	ICD9	M/RRFN MAMMOGRAPHY NFI

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

What data are available?

Clinical data from 3/1/2013 – 12/31/2019

- **HIPAA identifiers (MRN, Encounter #, Dates)**
- **Patient demographics** (age, gender, race, vital status, language, zip code, county, payer, and more)
- **Visit types and locations**
- **Diagnoses** (ICD-9-CM, ICD-10-CM)
- **Procedures** (ICD-9-CM, ICD-10-CM, CPT)
- **Select laboratory results**
- **Medications dispensed in hospital or prescribed in ambulatory**
- **Select other clinical assessments** (pain scores, BMI, health literacy)
- **Consent to research re-contact (select patients only)**
- *...And more!*

Accessing Patient Data



- Need an approved IRB
- Submit an online data request form <https://idr.ufhealth.org/services/analyst-data-support-services/idr-data-request-form/>
- Contact IDR Data Team IRBDataRequest@ahc.ufl.edu

- Need an approved IRB for OneFlorida
- Submit an online application form <https://onefloridaconsortium.org/front-door/research-infrastructure-utilization-application/>
- Contact OneFlorida Front Door Team OneFloridaOperations@health.ufl.edu

- Need an approved IRB at each participating institution
 - Need to find a collaborator at an institution of interest
- Submit an online data request form <https://idr.ufhealth.org/services/analyst-data-support-services/idr-data-request-form/> to get UF data
- Contact IDR Data Team IRBDataRequest@ahc.ufl.edu

The Data Request Form

IDR Data Request Form

IDR Data Request

Form to collect pertinent information and requirements to service data requests.

PI Full Name *

Name as reflected on Project, IRB or Protocol

PI Email Address *

PI Email Address as listed on IRB.

Enter Email

Confirm Email

Point of Contact Full Name *

Point of Contact for additional information.

Inclusion Criterion *

List all your inclusion criterion. Please include ICD, CPT, LOINC, RxNorm codes. ****PLEASE DO NOT INCLUDE ANY PHI****

For Example:

- Timeline
- Age
- Geography: UF JAX, UF GNV study, or both
- Specific diagnosis codes
- Specific procedure codes
- Medication taken
- Specific lab result evaluation
- & more!

Data Elements *

List all data elements you are requesting from the IDR for your study or protocol. **PLEASE DO NOT INCLUDE ANY PHI**

For example

- Most common -
 - Diagnoses, lab values, medications
- MRN
- Encounter #
- Name
- Demographics
- etc.

Example of Good Data Request

Inclusion criteria:

- All adult patients (age ≥ 18 at the time of the procedure) who underwent thoracic and thoracoabdominal aortic repair (CPT codes: 33875,33877,33880,33881,34841,34842,34843,34844,34845,34846,34847,34848,35091,35092) in the period 1/1/2016-1/1/2020.

Data elements:

- MRN
- Gender, Race, Ethnicity, Age at the time of the procedure
- Date of the thoracic and thoracoabdominal aortic repair
- Admit date and Discharge date
- All CPT codes during the encounter of aortic repair
- All diagnosis codes during the encounter of aortic repair
- Survival rate - 30 days, 90 days, current

Example of Good Data Request

- The **inclusion criteria** is clear.
 - The researcher provided all codes that we need to consider.
 - In addition, they provided the age and the dates to consider.
- The **data elements** are clear.
 - It specifies that age, dates, CPTs and diagnosis codes are relevant to the encounter when the aortic repair happened rather than at any time in a patient's life.

Types of IRB-Approved Data Requests

- Re-identification for an i2b2 query
- Multi-dimensional data requests
- C2S

Data Request Logistics

- We **STRONGLY** recommend you to use i2b2 to identify your cohort before submitting a data request
- Make sure your IRB is approved
 - Submit your data request once it is!
- Iterative communication
 - More likely than not, we will follow up on your data request with questions
- Once we have finished your data request, we will send you the Excel sheets containing your data

3. Consent2Share (C2S)



- Patient recruitment for research studies
- IDR contains the C2S Research Contact Registry
 - A voluntary database of patients interested in hearing about research opportunities
- 65,829 patients with signed C2S by Jan. 1, 2020
 - <18 years of age: 10,196
 - \geq 18 years of age: 55,633

IDR Data Request Best Practices

- Obtain IRB approval... but
 - Is the study feasible?
 - Are the data available?
 - Are the data discrete?
 - Are the data needs well specified?
 - Sometimes broad data requests are good... sometimes they are bad
- Talk to us! ... Meet with IDR team
 - Often helpful to bring research plan before submitting IRB
 - Work closely with us to specify data needs

IDR Data Request Best Practices

- Data are as close as possible to their original source
 - (i.e., minimal deletions, deduplications, or edits)
- Trainees should have close faculty involvement
- Well-formulated research questions and study designs
- Research teams should have
 - Study design expertise
 - Clinical expertise
 - Data expertise

IDR Data Request Common Pitfalls

- Poor data request specifications
 - Unclear what is being requested
 - Inconsistency between IRB and data request
- Too many data elements requested
 - 50+ hours of analyst time
- Research emergencies
- Poor communication
 - Daily emails
 - Email non-response

IDR Successes

Annals of Surgery. 269(4):652–662, APRIL 2019

DOI: 10.1097/SLA.0000000000002706, PMID: 29489489

Issn Print: 0003-4932

Publication Date: April 2019



 Print

***MySurgeryRisk*: Development and Validation of a Machine-learning Risk Algorithm for Major Complications and Death After Surgery**

Azra Bihorac;Tezcan Ozrazgat-Baslanti;Ashkan Ebadi;Amir Motaei;Mohcine Madkour;Panagote Pardalos;Gloria Lipori;William Hogan;Philip Efron;Frederick Moore;Lyle Moldawer;Daisy Wang;Charles Hobson;Parisa Rashidi;Xiaolin Li;Petar Momcilovic;

IDR Successes

Common inpatient hypoglycemia phenotypes identified from an automated electronic health record–based prediction model

Yoonyoung Choi, B.S.Pharm., Ph.D., M.S, Ben Staley, Pharm.D., BCPS,
Rene Soria-Saucedo, M.D., M.P.H., Ph.D, Carl Henriksen, M.S,
Amy Rosenberg, Pharm.D., BCPS, Almut G Winterstein, B.S.Pharm., Ph.D., FISPE ✉

American Journal of Health-System Pharmacy, Volume 76, Issue 3, 1 February 2019,
Pages 166–174, <https://doi.org/10.1093/ajhp/zxy017>

Published: 25 January 2019

THANK
YOU!