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

- Hospital care
- Ambulatory care
- Operating room
- Radiology
- Laboratory
- Pharmacy
- Demographics
- Diagnoses
- Procedures
- Visit information
- Lab results
- Medications
- …
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!
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?

N = 48 patients
How many children were diagnosed with diabetes and had a hemoglobin count > 7% in the same encounter?

N = 48 patients

Female: 26
Male: 17
Black/African American: 27
White: 22
Other: <3
i2b2 instances available at UF

- Florida Hospital (AdventHealth)
- Orlando Health
- Tallahassee Memorial Inpatient
- Tallahassee Memorial Outpatient
- UF Health
- University of Miami
- Nicklaus Children’s Hospital
- Capital Health Plan
- Florida Medicaid
i2b2 instances available at UF

<table>
<thead>
<tr>
<th>i2b2 instance</th>
<th>Patients</th>
<th>Facts</th>
<th>Geography</th>
</tr>
</thead>
<tbody>
<tr>
<td>UFHealth – GNV</td>
<td>1.27M</td>
<td>1.26B</td>
<td>Gainesville</td>
</tr>
<tr>
<td>UFHealth – JAX</td>
<td>0.77 M</td>
<td>0.37B</td>
<td>Jacksonville</td>
</tr>
<tr>
<td>OneFlorida</td>
<td>13.8M</td>
<td>18.4B</td>
<td>Florida</td>
</tr>
<tr>
<td>ACT Network</td>
<td>125M</td>
<td>&gt;50B</td>
<td>Nationwide</td>
</tr>
</tbody>
</table>

Data as of January 1, 2020.
i2b2 resources

**UFHealth i2b2**
- Project overview: [https://idr.ufhealth.org/i2b2/](https://idr.ufhealth.org/i2b2/)
- Request access: [https://idr.ufhealth.org/i2b2/i2b2-registration/](https://idr.ufhealth.org/i2b2/i2b2-registration/)
- i2b2 support: i2b2support@ahc.ufl.edu

**OneFlorida i2b2**
- Project overview: [https://onefloridaconsortium.org/front-door/onfli2b2/](https://onefloridaconsortium.org/front-door/onfli2b2/)
- Training: [https://onefloridaconsortium.org/front-door/onfli2b2/](https://onefloridaconsortium.org/front-door/onfli2b2/), go to "Training" section
- Request access: [https://onefloridaconsortium.org/i2b2/webform/i2b2.html](https://onefloridaconsortium.org/i2b2/webform/i2b2.html)
- i2b2 support: i2b2support@onefloridaconsortium.org

**NCATS ACT i2b2**
- Project overview: [http://www.actnetwork.us/FloridaCTSI](http://www.actnetwork.us/FloridaCTSI)
- Training: [http://www.actnetwork.us/FloridaCTSI](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/](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
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

<table>
<thead>
<tr>
<th>Need an approved IRB</th>
<th>Need an approved IRB for OneFlorida</th>
<th>Need an approved IRB at each participating institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact IDR Data Team <a href="mailto:IRBDataRequest@ahc.ufl.edu">IRBDataRequest@ahc.ufl.edu</a></td>
<td>Contact OneFlorida Front Door Team <a href="mailto:OneFloridaOperations@health.ufl.edu">OneFloridaOperations@health.ufl.edu</a></td>
<td>Submit an online data request form <a href="https://idr.ufhealth.org/services/analyst-data-support-services/idr-data-request-form/">https://idr.ufhealth.org/services/analyst-data-support-services/idr-data-request-form/</a> to get UF data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Contact IDR Data Team <a href="mailto:IRBDataRequest@ahc.ufl.edu">IRBDataRequest@ahc.ufl.edu</a></td>
</tr>
</tbody>
</table>
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
  - >= 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
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!