i2b2 / SHRINE Platform

Shawn Murphy MD, Ph.D.
The National Center for Biomedical Computing entitled Informatics for Integrating Biology and the Bedside (i2b2), what is it?

- Software for explicitly organizing and transforming person-oriented clinical data to a way that is optimized for clinical genomics research
  - Allows integration of clinical data, trials data, and genotypic data
- A portable and extensible application framework
  - Software is built in a modular pattern that allows additions without disturbing core parts
  - Available as open source at https://www.i2b2.org
i2b2 Cell: The Canonical Software Module

HTTP XML
(minimum: RESTful)
An i2b2 Environment (the Hive) is built from i2b2 Cells

“Hive” of software services provided by i2b2 cells
I2b2 Software components are distributed as open source.
Implementations

CTSA's

- Boston University
- Case Western Reserve University (including Cleveland Clinic)
- Children's National Medical Center (GWU), Washington D.C.
- Duke University
- Emory University (including Morehouse School of Medicine and Georgia Tech)
- Harvard University (including Beth Israel Deaconess Medical Center, Brigham and Women's Hospital, Children's Hospital Boston, Dana Farber Cancer Center, Joslin Diabetes Center, Massachusetts General Hospital)
- Medical University of South Carolina
- Medical College of Wisconsin
- Oregon Health & Science University
- Penn State Milton S. Hershey Medical Center
- Tufts University
- University of Alabama at Birmingham
- University 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 Chicago
- University of Cincinnati (including Cincinnat Children's Hospital Medical Center)
- University of Colorado Denver (including Children's Hospital Colorado)
- University of Florida
- University of Kansas Medical Center
- University of Kentucky Research Foundation
- University of Massachusetts Medical School, Worcester
- University of Michigan
- University of Pennsylvania (including Children's Hospital of Philadelphia)
- University of Pittsburgh (including their Cancer Institute)
- University of Rochester School of Medicine and Dentistry
- University of Texas Health Sciences Center at Houston
- University of Texas Health Sciences Center at San Antonio
- University of Texas Medical Branch (Galveston)
- University of Texas Southwestern Medical Center at Dallas
- University of Utah
- University of Washington
- University of Wisconsin - Madison (including Marshfield Clinic)
- Virginia Commonwealth University
- Weill Cornell Medical College

Academic Health Centers (does not include AHCs that are part of a CTSA):

- Arizona State University
- City of Hope, Los Angeles
- Georgia Health Sciences University, Augusta
- Hartford Hospital, CN
- HealthShare Montana
- Massachusetts Veterans Epidemiology Research and Information Center (MAVERICK), Boston
- Nemours
- Phoenix Children's Hospital
- Regenstrief Institute
- Thomas Jefferson University
- University of Connecticut Health Center
- University of Missouri School of Medicine
- University of Tennessee Health Sciences Center
- Wake Forest University Baptist Medical Center

HMOs:

- Group Health Cooperative
- Kaiser Permanente

International:

- Georges Pompidou Hospital, Paris, France
- Hospital of the Free University of Brussels, Belgium
- Inserm U936, Rennes, France
- Institute for Data Technology and Informatics (IDT), NTNU, Norway
- Institute for Molecular Medicine Finland (FIMM)
- Karolinska Institute, Sweden
- Landspitali University Hospital, Reykjavik, Iceland
- Tokyo Medical and Dental University, Japan
- University of Bordeaux Segalen, France
- University of Erlangen-Nuremberg, Germany
- University of Goettingen, Goettingen, Germany
- University of Pavia, Pavia, Italy
- University of Seoul, Seoul, Korea

Companies:

- Johnson and Johnson (TransMART)
- GE Healthcare Clinical Data Services
Set of patients is selected through Enterprise Repository and data is gathered into a data mart

EDR

Selected patients

Data directly from EDR

Data from other sources

Data imported specifically for project

Automated Queries search for Patients and add Data

Project Specific Phenotypic Data
Privacy Levels in i2b2

**PHI-enabled User**
Client shows PHI to Key holders (limited to Entity)
Keys limit to specific users and patients

**Notes-enabled LDS User**
Client show LDS and scrubbed text
Fundamental data set is LDS (PHI encrypted)

**Limited-data-set User**
Client shows LDS (need DUA)
Fundamental data set is LDS (PHI encrypted)

**Obfuscated-data User**
Client only shows aggregated data
Fundamental data set is obfuscated

**Aggregated-data User**
Client only shows aggregated data
Fundamental data set is de-identified
Interrogation can occur through i2b2 web client
Data is available through the i2b2 Workbench
Builds complex “Custom Study” displays
Builds complex “Custom Study” displays
Launching SMART-i2b2 views
“SMART Enabling” i2b2
SMART Container in i2b2
The job of the SMART cell is to do most of the heavy lifting to convert SMART queries into serviceable i2b2 queries and convert i2b2 responses to SMART RDF responses that can be consumed by the SMART apps.
SMART editable EMR-Screen
Can We Trust the Phenotypes?

Gold Standard Validation Study

- Evaluate case and control algorithms compared to gold standard of diagnostic interview by expert clinician
- Recruit cases and controls as defined by informatics algorithm
- Interview by clinicians blinded to ascertainment group

Jordan Smoller MD, ScD and team
Natural Language Processing
Train classification algorithms

1. Over 300 words/phrases (features) were identified using chart review

2. Important features were selected for model using adaptive LASSO shrinkage

Tianxi Cai PhD and team
Refining diagnosis of Depression

<table>
<thead>
<tr>
<th>Clinical Status</th>
<th>Model</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>Precision</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depressed</td>
<td>Billing Codes</td>
<td>0.95</td>
<td>0.09 (0.03)</td>
<td>0.57 (0.14)</td>
<td>0.54 (0.02)</td>
</tr>
<tr>
<td>Depressed</td>
<td>NLP</td>
<td>0.95</td>
<td>0.42 (0.05)</td>
<td>0.78 (0.02)</td>
<td>0.88 (0.02)</td>
</tr>
<tr>
<td>Depressed</td>
<td>NLP + Billing Codes</td>
<td>0.95</td>
<td>0.39 (0.06)</td>
<td>0.78 (0.02)</td>
<td>0.87 (0.02)</td>
</tr>
<tr>
<td>Well</td>
<td>Billing Codes</td>
<td>0.95</td>
<td>0.06 (0.02)</td>
<td>0.26 (0.27)</td>
<td>0.55 (0.03)</td>
</tr>
<tr>
<td>Well</td>
<td>NLP</td>
<td>0.95</td>
<td>0.37 (0.06)</td>
<td>0.86 (0.02)</td>
<td>0.85 (0.02)</td>
</tr>
<tr>
<td>Well</td>
<td>NLP + Billing Codes</td>
<td>0.95</td>
<td>0.39 (0.07)</td>
<td>0.85 (0.02)</td>
<td>0.86 (0.02)</td>
</tr>
</tbody>
</table>

Specificity: 95%
AUC > 85%
Use Case: QT interval and antidepressant use

QT interval and antidepressant use: a cross sectional study of electronic health records

BMJ 2013;346:f288 doi:

Victor M Castro team lead1, Caitlin C Clements clinical research coordinator2,3, Shawn N Murphy associate professor of neurology4, Vivian S Gainer team lead1, Maurizio Fava Slater Family professor of psychiatry4, Jeffrey B Weilburg assistant professor of psychiatry5, Jane L Erb assistant professor of psychiatry4, Susanne E Churchill executive director, i2b2 National Center for Biomedical Computing7, Isaac S Kohane director, i2b2 National Center for Biomedical Computing6, Dan V Iosifescu associate professor of psychiatry9, Jordan W Smoller associate professor of psychiatry2, Roy H Perlis associate professor of psychiatry9

Partners EMR

~4 M

Antidepressant / methadone datamart
241,308

Adult patients prescribed an anti-depressant or methadone and no prior history of Torsade de pointes

202,911
patients had no EKG in the exposure window EXCLUDED (Table S1)

38,397
Main Study Cohort
patients prescribed 1+ type of antidepressant or methadone with a follow-up EKG 14-90 days after prescription

467
Paired Dose Cohort
Patients prescribed escalating doses of the same antidepressant (or methadone) with a follow-up EKG for each prescription
Use Case: QT interval and antidepressant use

- **FDA warning 2011 for Celexa**
  
  Safety Announcement: [8-24-2011] ”should no longer be used at doses greater than 40 mg per day because it can cause abnormal changes in the electrical activity of the heart.”

- **But, did NOT include Lexapro (which is active ingredient of Celexa [s-enantiomer])**

- **Shown to be true with RPDR-derived data set with >38,000 EKGs obtained within 14 – 90 day window after medication initiated**

<table>
<thead>
<tr>
<th>Anti-depressant</th>
<th>Adjusted model†</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Citalopram (Celexa)</td>
<td>2.85</td>
<td>0.004</td>
</tr>
<tr>
<td>Escitalopram (Lexapro)</td>
<td>3.80</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Fluoxetine (Prozac)</td>
<td>1.44</td>
<td>0.150</td>
</tr>
<tr>
<td>Paroxetine (Paxil)</td>
<td>0.07</td>
<td>0.943</td>
</tr>
<tr>
<td>Sertraline (Zoloft)</td>
<td>0.87</td>
<td>0.383</td>
</tr>
<tr>
<td>Other anti-depressants</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amitriptyline</td>
<td>4.10</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Bupropion</td>
<td>-2.15</td>
<td>0.032</td>
</tr>
<tr>
<td>Duloxetine</td>
<td>0.60</td>
<td>0.547</td>
</tr>
<tr>
<td>Mirtazapine</td>
<td>-1.46</td>
<td>0.145</td>
</tr>
<tr>
<td>Nortriptyline</td>
<td>1.23</td>
<td>0.219</td>
</tr>
<tr>
<td>Venlafaxine</td>
<td>1.15</td>
<td>0.251</td>
</tr>
<tr>
<td>previously known prolonger</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methadone</td>
<td>5.32</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

† Adjusted for age, gender, race, type of insurance, history of major depression, history of myocardial infarction and Charlson comorbidity score.
Use Case: QT interval and antidepressant use

Mean QTc (ms)

Selectove serotonin reuptake inhibitor or methadone

<table>
<thead>
<tr>
<th>Drug</th>
<th>Dose (mg)</th>
<th>QTc (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citalopram</td>
<td>10-50</td>
<td>420-450</td>
</tr>
<tr>
<td>Escitalopram</td>
<td>10-50</td>
<td>430-460</td>
</tr>
<tr>
<td>Fluoxetine</td>
<td>10-80</td>
<td>440-470</td>
</tr>
<tr>
<td>Paroxetine</td>
<td>10-60</td>
<td>450-480</td>
</tr>
<tr>
<td>Sertraline</td>
<td>10-200</td>
<td>460-500</td>
</tr>
<tr>
<td>Methadone</td>
<td>5-50</td>
<td>470-510</td>
</tr>
</tbody>
</table>

Other antidepressant or methadone

<table>
<thead>
<tr>
<th>Drug</th>
<th>Dose (mg)</th>
<th>QTc (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amitriptyline</td>
<td>10-300</td>
<td>420-450</td>
</tr>
<tr>
<td>Bupropion</td>
<td>5-75</td>
<td>430-460</td>
</tr>
<tr>
<td>Duloxetine</td>
<td>10-60</td>
<td>440-470</td>
</tr>
<tr>
<td>Mirtazapine</td>
<td>15-100</td>
<td>450-480</td>
</tr>
<tr>
<td>Nortriptyline</td>
<td>10-50</td>
<td>460-490</td>
</tr>
<tr>
<td>Venlafaxine</td>
<td>37.5-250</td>
<td>470-510</td>
</tr>
<tr>
<td>Methadone</td>
<td>5-50</td>
<td>480-520</td>
</tr>
</tbody>
</table>

* Dose a significant predictor of QTc in fully adjusted linear models at α=0.05
† QTc at specified dose is significantly different from that at prior dose in fully adjusted linear models at α=0.05

Mean (SD) corrected QT (QTc) interval recorded on electrocardiogram 14–90 days after prescription of antidepressant or methadone, by drug dose.
Integration Fundamentally enabled by Data Model

- **Integration** of data from distributed and differently structured databases in order to perform comprehensive analyses.

- **Separation** of data used for research from daily operational or transactional data.

- **Standardization** of a model across systems.

- **Ease** of use by end-users.
i2b2 Data Model = Star Schema

**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

**modifier_dimension**
- **PK**: Modifier_Path
- Modifier_CD
- Name_Char
In i2b2, a fact is an observation on a patient.

Examples of FACTS:

- Diagnoses
- Procedures
- Health History
- Genetic Data
- Lab Data
- Provider Data
- Demographics Data

An observation is not necessarily the same thing as an event
i2b2 Dimension Tables

- Dimension tables contain descriptive information about facts.

- In i2b2 there are five dimension tables

  - concept_dimension
  - provider_dimension
  - visit_dimension
  - patient_dimension
  - modifier_dimension
Ontology Flexibility

SNOMED Clinical Terms

- CORE Subset of SNOMED CT
- SNOMED Anatomy
- SNOMED Clinical Findings
  - Description: The Clinical Finding subtree of SNOMED CT
  - Definition: Class subtree of ClinicalFinding
  - Ontology ID: 2018
  - Definition Language: Manual

<table>
<thead>
<tr>
<th>VERSION</th>
<th>BASE VERSION</th>
<th>CREATED</th>
<th>CREATED BY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>2009_07_31</td>
<td>07/16/2010</td>
<td>Tania Tudorache, <a href="mailto:tudorache@stanford.edu">tudorache@stanford.edu</a></td>
</tr>
<tr>
<td>1.1</td>
<td>2009_07_31</td>
<td>03/23/2010</td>
<td>Tania Tudorache, <a href="mailto:tudorache@stanford.edu">tudorache@stanford.edu</a></td>
</tr>
<tr>
<td>1.0</td>
<td>2009_01_31</td>
<td>09/09/2009</td>
<td>Tania Tudorache, <a href="mailto:tudorache@stanford.edu">tudorache@stanford.edu</a></td>
</tr>
</tbody>
</table>

- SNOMED Ethnic Group
Query Flexibility
Constrain observations into the same encounter
Defining Temporal Relationships
Viewing Temporal Relationships in Time Align
Shared Health Research Information Network (SHRINE)
Current Shared Health Research Information Network (SHRINE) at Harvard
Quantity of Clinical Data

- 10 years (2001-2011)
- 5 hospitals
- 6 million total patients
- >1 billion medical observations
Distributed Query System

1. Certify investigator
2. Compose Query
3. Broadcast Query
4. Verify user, adapt for local CRC
5. I2B2 CRC Query (no changes)
6. Send results to aggregator
7. Aggregate Results

CHB PM
CHB Composer
CHB Broadcaster Aggregator
Partners Responder Adapter
CHB Responder Adapter
BIDMC Responder Adapter
CRC
CRC
CRC

PARTNERS HEALTHCARE
Search Interface
Search By Medication and Diagnoses

<table>
<thead>
<tr>
<th>Bipolar disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atypical depressive disorder</td>
</tr>
<tr>
<td>Atypical manic disorder</td>
</tr>
<tr>
<td>Bipolar I disorder, most recent episode (or current)</td>
</tr>
<tr>
<td>Bipolar affective disorder, depressed, in full remission</td>
</tr>
<tr>
<td>Bipolar affective disorder, depressed, in partial remission</td>
</tr>
<tr>
<td>Bipolar affective disorder, depressed, mild degree</td>
</tr>
<tr>
<td>Bipolar affective disorder, depressed, moderate degree</td>
</tr>
<tr>
<td>Bipolar affective disorder, depressed, severe degree</td>
</tr>
<tr>
<td>Bipolar affective disorder, depressed, unspecified</td>
</tr>
<tr>
<td>Bipolar affective disorder, manic, in full remission</td>
</tr>
<tr>
<td>Bipolar affective disorder, manic, in partial or incomplete remission</td>
</tr>
<tr>
<td>Bipolar affective disorder, manic, mild degree</td>
</tr>
<tr>
<td>Bipolar affective disorder, manic, moderate degree</td>
</tr>
<tr>
<td>Bipolar affective disorder, manic, severe degree</td>
</tr>
</tbody>
</table>

**Query Tool**

**Query Name:** Chlorpr-Bipolar@03:30:52

**Group 1**
- **Dates:**
- **Occurs > 0x:**
- **Exclude:**
  - Chlorpromazine
  - Chlorpromazine hydrochloride

**Group 2**
- **Dates:**
- **Occurs > 0x:**
- **Exclude:**
  - Bipolar disorders
  - Atypical depressive disorder
  - Atypical manic disorder
  - Bipolar I disorder
  - Bipolar affective disorder

**Group 3**
- **Dates:**
- **Occurs > 0x:**
- **Exclude:**

**Query Status**

**Finished Query:** "Chlorpr-Bipolar@03:30:52"
- Partners - 446 ±3 patients
- CHB - 22 ±3 patients
- BIDMC - 142 ±3 patients
- AGGREGATED - 609 ±9 patients

FINISHED [221.6 secs]
Paths this will lead us to ...

- Allow recruitment across several sites with SHRINE
  - Find cohort
  - Recruit doctors at other institutions for a clinical trial
  - Create a project for the study
  - Present patients one by one using SHRINE data + contact demographics at each site
References


I2b2, SHRINE, and SMART Information and Software on the Web

i2b2 Homepage (https://www.i2b2.org)
i2b2 Software (https://www.i2b2.org/software)
i2b2 Community Site (https://community.i2b2.org)
SHRINE at Harvard (http://shrine.catalyst.harvard.edu)
SHRINE Software:
(https://open.med.harvard.edu/display/SHRINE/Software)
SMART i2b2 Homepage (http://smarti2b2.org)
SMART Platforms Homepage (http://smartplatforms.org)