Clinical Data acquisition and management

David A. Dorr, MD, MS

Data after Dark
1/2016 OHSU
## Clinical data: Four major points (revisited)

<table>
<thead>
<tr>
<th>Point</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical data has <strong>potential</strong> to transform research across the translational continuum</td>
<td>Multiple opportunities to transform knowledge discovery and generate value</td>
</tr>
<tr>
<td>Clinical data is <strong>available</strong>, here at OHSU and elsewhere</td>
<td>Chief Clinical Research Informatics Officer; OCTRI - Research data warehouse; RedCAP; EHR</td>
</tr>
<tr>
<td>Clinical data is <strong>incomplete, inaccurate, and messy</strong></td>
<td>Computable phenotype variation (duke); EHR data source variation</td>
</tr>
<tr>
<td>You can be <strong>part of the solution</strong> by sharing data, recording metadata, and being responsible</td>
<td>Up to you ... just do it!</td>
</tr>
</tbody>
</table>
Clinical Data across the ecosystem

Figure 1. Vision of the Health IT Ecosystem

Source: Office of the National Coordinator
Clinical data is available, here at OHSU and elsewhere

Current
• OCTRI Data management (links provided)
• E.g., Epic Data for research
• ACC
• **Cloud**: Resource Identification Initiative; NIH; many other sources

Leading towards the future
* Tim Burdick, Chief Clinical Research Informatics Officer
Epic Data for Research Purposes

OCTRI maintains a repository of Epic data from patients and research subjects. This valuable resource is known as the Research Data Warehouse (RDW). The RDW provides access to over 2.5 million patient records with 20 million patient encounter records and more than 40,000 data points. Researchers work with OCTRI staff to obtain answers to a variety of research purposes.

- Cohort discovery (counters) for preparatory to research purposes
- Deidentified data sets for epidemiological studies
- Data sets that identify potential study participants or gather retrospective data on current study participants

The Research Data Warehouse (RDW) contains an enormous amount of linked clinical data and can be queried to answer a broad range of research questions. OCTRI staff can act as your guide through this complex data.
How much data are in the RDW?

- Epic Ambulatory since 2005
- Epic Inpatient and ED since May 2008
- >2.7 million patients
- >600,000 patients with at least 1 “visit”
- >10.3 million “visits”

- Epic data is complex!
  - Clarity contains over 8,700 data tables and over 75,000 data elements

Cohort Discovery Tool
What data are available?

- **Demographic information**
  - gender, race, ethnicity, preferred language, employment status
- **Provider, insurance coverage**
- **Birth history**
  - date of birth, gestation, birth weight
- **Vitals**
  - blood pressure, weight, height, pain ratings
- **Contact and census information**
  - Addresses, phone, census information
- **Lab orders and results**
- **Diagnoses, problem lists, medical history (ICD-9*)**

- **Hospital encounter information**
- **Medication lists and orders**
- **Ambulatory encounter information**
  - clinic date, provider, department/clinic, PCP, chief complaint, cancel reason
- **Procedures**
- **Surgeries**
- **Insurance coverage for patient encounter**
  - benefit plan, insurance class, co-pay

* ICD-10 coming soon!
Accessing the RDW

• Data can be released to investigators in three ways
  • **Counts** - no IRB required.
  • **De-identified** - IRB determination of “non-human subjects research”
  • **Fully identifiable** - requires IRB approval
The ACC is a service center and charges for services provided. Please send all inquiries to: acc@ohsu.edu.

**Mission Statement**
The mission of the Advanced Computing Center (ACC) at OHSU is to meet the computing needs of the OHSU research, academic and administrative community by providing a flexible and scalable set of advanced computing services that augment and supplement no-cost core services provided to the OHSU Enterprise by the Information Technology Group.

**Equipment**
The Advanced Computing Center's technology platforms are currently housed in 2000 square feet of the West Campus Data Center that provides redundant and emergency power, HVAC and multiple layers of security. Core Equipment comprises a 1000-processor compute resource with over 3 terabytes of memory as well as hundreds of terabytes of storage and backup capacity.

The Advanced Computing Center's network infrastructure consists of high-speed switches delivering Gigabit Ethernet to all hosted resources. This exclusively switched architecture is designed to eliminate the network bandwidth contention common in research environments, while allowing reliable high and low bandwidth connections between OHSU researchers, the internet, and OHSU's connection to Internet2.

**Services**
The ACC provides a tiered menu of service offerings, including consulting, hosting, backups and storage. These services are designed with the needs of research in mind, facilitating remote collaboration and co-located computing capabilities.

We offer virtual server environments, applications for research computing systems, secure and public

**Location**
OHSU
3181 SW Sam Jackson Park Road, Portland OR 97239
### OHSU MyChart for Recruitment pilot

Dennis McCarty (PI) – Public Health; 1 clinic, 2 MDs – Dept Medicine

<table>
<thead>
<tr>
<th>Enrollment by sex</th>
<th>MyChart msg sent (n)</th>
<th>MyChart msg read (%)</th>
<th>REDCap survey completed (%)</th>
<th>Patients who opted out from MyChart for research (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females</td>
<td>124</td>
<td>54%</td>
<td>36%</td>
<td>0</td>
</tr>
<tr>
<td>Males</td>
<td>124</td>
<td>53%</td>
<td>28%</td>
<td>0</td>
</tr>
<tr>
<td>Both</td>
<td>248</td>
<td>54%</td>
<td>32%</td>
<td>0</td>
</tr>
</tbody>
</table>

- Patients read MyChart messages identified as research.
- Patients see MyChart as a reasonable channel for conducting research.
- MyChart is a viable method for recruiting for a clinical research survey.

Slide from Tim Burdick; Help from OCTRI (NCATS UL1TR000128); ITG Epic Research Team; OHSU Office of Clinical Research Informatics
Cloud resources

1. You *read an article*, are *interested in expanding or reproducing its results*, and *want to know how*
   
   1. *Resource Identification Initiative – in progress*
   2. Clinical data study request - [https://www.clinicalstudydatarequest.com/](https://www.clinicalstudydatarequest.com/)

2. You have a novel idea that needs certain clinical data
   
   1. Federal sources, e.g.,
   2. Collaboratives, like eMERGE or PCORnet
• 29 partner networks (including one lead by OCHIN named ADVANCE)
• 60 million covered lives
Clinical data is **incomplete, inaccurate, and messy**

Duke – PCORNet
EHR data extraction to diagnose diabetes
3536 (14.4%) only diagnosis codes
1129 (4.6%) only medications
4203 (17.1%) only labs

**Figure 1** Overlap of diabetes cohorts identified from different categories of phenotype eligibility criteria; n=24,520 patients identified by criteria from any of the three categories.
EHR data is messy: Gold standard for diagnosis

Source: Martin, Dorr et al. unpublished data
### Does this messiness matter?

<table>
<thead>
<tr>
<th>Outcome / Risk score</th>
<th>EHR data Source</th>
<th>Med History</th>
<th>Problem List</th>
<th>Combined</th>
<th>Gold Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost &gt; 20k</td>
<td>Encounter</td>
<td>0.626</td>
<td>0.621</td>
<td>0.608</td>
<td>0.630</td>
</tr>
<tr>
<td>ACA ALL</td>
<td></td>
<td>0.618</td>
<td>0.630</td>
<td>0.593</td>
<td>0.625</td>
</tr>
<tr>
<td>Mod. Charlson</td>
<td></td>
<td>0.636</td>
<td>0.617</td>
<td>0.601</td>
<td>0.638</td>
</tr>
<tr>
<td>HCC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.631</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>0.627</td>
<td>0.623</td>
<td>0.601</td>
<td></td>
</tr>
</tbody>
</table>

Source: Martin, Dorr et al. unpublished data
Being Part of the Solution

• Use **standards**
  – Computable phenotypes (don’t just make up your own stuff)
  – Record metadata about your data, your processing, your analysis

• **Manage** your data carefully – get professional help

• **Validate** your approach
  – Test against known standards, measure a subset in two different ways, etc.

• **Share** all you can – your data, your approach
Links

http://skynet.ohsu.edu/~pedersbj/ShortCourse/Jan2016Schedule.html
http://www.ohsu.edu/xd/research/centers-institutes/octri/resources/octri-research-services/data-consult.cfm
http://www.ohsu.edu/xd/research/centers-institutes/octri/resources/octri-research-services/rdw.cfm
http://www.ohsu.edu/xd/research/centers-institutes/octri/resources/octri-research-services/cohort-discovery.cfm