Drug Repurposing Roadmap that Optimizes Patient Impact through Collaboration

Nov 15, 2022
An Introduction

• “Smitty” Heavner
  - Pronouns: He/They
• Registered Nurse
  - 10 years, specialized in ED
• PhD in Evaluation Science
• Scientific Director
  - CURE Drug Repurposing Collaboratory
• Adjunct Faculty
  - Clemson University – Public Health Sciences
  - University of South Carolina School of Medicine Greenville
Disclosures

• CDRC is a public-private partnership with FDA and NIH National Center for Advancing Translational Science

• Critical Path Institute is supported by the Food and Drug Administration (FDA) of the U.S. Department of Health and Human Services (HHS) and is 54.2% funded by the FDA/HHS, totaling $13,239,950, and 45.8% funded by non-government source(s), totaling $11,196,634.

• The views expressed are those of the presenter and do not necessarily represent the official views of, nor an endorsement by, FDA/HHS NIH or the U.S. Government.
Outline

• CDRC Overview
• Drug Repurposing
• Real-World Data
• The Edge Tool
  - Use Case: COVID-19
• Real-World Evidence
• Future Directions
**Goal**: Facilitate a drug’s demonstrated utility in clinical practice, harnessing a combination of traditional trials and real-world evidence.
Drug Repurposing

• Significant portion of conditions have no FDA approved treatments
• Traditional drug development and labeling is slow and expensive.
• Repurposing may be needed in diseases which are:
  - rapidly emerging
  - extremely rare
  - impacting vulnerable groups
  - treated by standardized guidelines
The best use case for CURE ID is for diseases where clinical trials are not or cannot be conducted.

In some instances (e.g., start of outbreak), it may be helpful to capture data that could later be used to inform the design and conduct of RCTs.
Real-World Data

• Explore clinical practice
• Develop and refine hypotheses
• Provide external controls
• Observational research

CURE ID data sources

• Case reports from literature
• Clinician submitted infectious disease case reports
  1. What disease did your patient have?
  2. How did you make the diagnosis?
  3. **What made your patient’s infection difficult to treat?**
  4. What drug(s) did you use to address this difficult to treat infection?
  5. What was the patient’s outcome?
  6. Did the patient experience any adverse events?
A Platform to Capture Novel Uses of Existing Drugs

- Web-based tool
  - Computer, smartphone or mobile device
- Capture and share real-world experiences treating patients through a simple online case report form
- HIPAA compliant, contains no PII
- Newsfeed
- Link to www.clinicaltrials.gov

https://cure.ncats.io/explore
Real-World Data: Challenges

- Unstructured data
- Defining outcomes
- Privacy concerns
- Probabilistic linkage
- Data harmonization
Real-World Data: Data Harmonization

EXTRACT

TRANSFORM

LOAD

ETL
Real-World Data: Data Harmonization

• Centralized harmonization

• Sites use common data model at baseline

• Only for US data

• Limited to COVID-19

National COVID Cohort Collaborative (N3C)

The N3C is a partnership among the NCATS-supported Clinical and Translational Science Awards (CTSA) Program hubs, the National Center for Data to Health (CD2H), and NIGMS-supported Institutional Development Award Networks for Clinical and Translational Research (IDeA-CTR), with overall stewardship by NCATS. Collaborators will contribute and use COVID-19 clinical data to answer critical research questions to address the pandemic.

Scientists Use N3C Data to Identify Common Features of Long COVID

NIH-supported researchers used electronic health record data from the National COVID Cohort Collaborative (N3C) Data Enclave to identify people with long COVID and those likely to have it.

ncats.nih.gov/n3c
Observational Medical Outcomes Partnership (OMOP)

153 Controlled Vocabularies
9 Million concepts

ICD-10

NIH National Library of Medicine

SNOMED International

LOINC from Regenstrief
The Edge Tool

Proprietary Data Model

OMOP CDM V5.3.1
The Edge Tool

• Web-based decision support for concept mapping
• Base configuration settings for major EMR Vendors.
• Configuration management documentation tool
• Inspection Report of DevOps on ETL processes
• Data Quality Dashboard framework of 3,000+ data quality tests
• Collaborative cohort subset definition
• Perform de-identification and submission
• All open-source resources
The Edge Tool on Azure

ETL
- Perseus
- EMR base config
- Usagi
- White Rabbit
- Rabbit in the Hat

DevOps
- Data Quality Dashboard
- Documentation Engine
- Submission extraction
- Change control

Analysis
- Atlas
- WebAPI
- Hades
- R-Studio
- Methods Library

Azure SQL Server
OMOP Data Model
Vocabulary Management
Authentication and Authorization

Azure logo
The Edge Tool: De-identification

- Reassignment of Person IDs: Person IDs are regenerated sequentially from a sorted copy of the Person table. These new Person IDs are carried throughout the CDM to all tables that reference it.
- Date Shifting:
  - Each person is assigned a random date shift value between -186 and +186 days. All dates for that person are then shifted by that amount.
  - Birthdays: After date shifting a person’s birthday, the day is then set to the first of the new birth month. If the person would be > 89 years old then they are assigned a random birth year that would make them 90-99 years old.
- Date Truncation:
  - A user-defined Start and End date are used to exclude any date shifted data that falls outside of the target date range (E.G. Procedures, conditions occurrences, etc. Does not include Birthdates).
- Removal of other identifiers:
  - Other potentially identifying datapoints are removed from the dataset such as location_id, provider_id, and care_site_id
The Edge Tool Use Case: COVID-19

- Drug Repurposing
- Broad Impact
- Identifiable Cases
- Discreet Data
- Acute Disease
- Definitive Outcomes
Data Flow

Institutions sharing data via the Edge Tool

- VIRUS sites
- Emory
- JHU

Edge Tool

OMOP data format

CDISC/SDTM data format

Data access to VIRUS dataset

Data access to VIRUS COVID-19 Registry data

Data access to ISARIC + VIRUS dataset

Converting data to SDTM

CDISC/ADaM data format

Data Harmonization

CURE ID

SCCM

Merging Edge Tool Data + Data from VIRUS sites not using Edge Tool

EDGE tool data

CURE ID "publicly available" data set (40 variables)
Real-World Evidence: COVID-19

• Replicate findings of clinical trials
• Evaluate key agents: dexamethasone, baricitinib, tocilizumab
• Demonstrate utility of Edge Tool

Pilot site data: (March 2020- March 2022)

- Methodology: Data cleaning
  - Observational Medical Outcomes Partnership Common Data Model (OMOP CDM)
  - Originally, N = 12,129 patients
    - Inclusion criteria: 18+, no missing demographic data (ethnicity, race)
      - N = 11,297 patients
Limitations

- Unmeasured covariates
- Small sample size, control matching with replacement
- Single site
- Timing limitations
  - Shift-and-truncate
  - Data collection after dexamethasone was shown to be effective
- Covariates outside of data collected in trial not accounted for
- Context of treatment (timing and dosage) not accounted for
Future Directions: Beyond COVID

• Sepsis
  - Isolating organisms
  - Case definition

• Meningitis
  - Isolating organisms
  - Rarer cases of more interest

• Osteomyelitis
  - Isolating organisms
  - Linking encounters
  - Lost to follow up

Developing partnerships and infrastructure to provide sustainable resources to impact patient treatments globally

CDRC Stages

- **Stage 0** Define Diseases
- **Stage 1** Hypothesis from RWD
- **Stage 2** Confirming Hypothesis
- **Stage 3** Informing Clinical Practice
- **Stage 4** Regulatory Aims

Framework

- **Prioritization**
  - Repurposing
  - Development
  - Ranking
- **Pre-Clinical**
  - HTS
  - MOA, pathways
  - Omics
  - AI mining assoc
- **Real-World Data**
  - Literature
  - Clinician submitted
  - EHRs
  - Registry
- **Clinical Trials**
  - Platform
  - Adaptive
  - Embedded in clinical practice
- **Guidelines**
  - Publications
  - Professional societies
  - Advocacy
- **Policy**
  - Pathway for submission
  - Non-traditional paths
  - Level of evidence
- **Legislation**

Patient Impact

- **Disease Focus**
  - COVID-19 pilot, Rare Sarcomas (Angiosarcoma, PEComa, EHE)

Academia

Regulators

Foundations

Clinicians

Patients

Industry

Non-profits

Disease Societies
Thank You

<- scan for contact information


Edge Tool Screen Grabs
Concept Mapping Decision Support

https://github.com/SoftwareCountry/Perseus

https://www.youtube.com/watch?v=JBgliVKDOyc
Epic to OMOP ETL Project

Goals

- Publish a guide that can be computable (i.e., as a workflow) as well as easily searchable in html and pdf formats.
- Have the guide be compiled by the scripts so as new scripts are added, they are automatically added to the guide.
- Encode the scripts with markdown to enable navigation and readability.
- Have sufficient meta-data in the scripts to help users of the guide index the data for common questions.
- Whenever possible documentation should not be separated from the code.

Assumptions

- Epic data model is proprietary, distribution should be through Epic customer portal (galaxy)
- Epic uses SQL Server so we will base our scripts on MS T-SQL
- There is setup necessary to configure an OMOP CDM database with tables, vocabulary, and constraints.
- There are 50+ SQL stored procedures to transform Epic Clarity data into the OMOP CDM 5.3.1
- There is a sequence of operations to orchestrate the transformation
- The SQL scripts use an Ansible YAML file with basic documentation capabilities (e.g., Comments *) for inline comments
- There are resources online available that are key to doing a transformation. In CDM github conventions and Book On OMOP
- Sites will need to change ETL scripts for site specific localizations where data is coming from flowsheets and smartdata forms.
- ETL scripts should be human readable

Users of the guide

- Database administrators implementing this ETL in their organization.
- Researchers wanting to see the clear provenance of data transformations from the clinical information system.

Questions the guide should facilitate

- Show the epic tables involved in the ETL process
- Show the source Epic clarity tables that feed into an OMOP Domain (condition, measurements, etc)
- Show the domains that are fed from a particular epic table
- Show the order of operations of the scripts

---

```sql
DECLARE @start_datetime DATETIME = getdate();
set @elapsed_seconds = DATEDIFF(second, @start_datetime, @end_datetime);
```
Data Quality Dashboard

DATA QUALITY ASSESSMENT

JOHNS HOPKINS MEDICINE ENTERPRISE

DataQualityDashboard Version: 1.0.0
Results generated at 2021-08-28 09:20:06 in 7 hours

Examples of DQ Checks from Kahn et al (2016)

- Atemporal Plausibility: 40% of labs outside of normal range
- Temporal Plausibility: Unexpected change in number of records from month to month
- Completeness: 62% of route_concept_id is missing
- Value Conformance: ICD9 codes in condition_concept_id
- Relational Conformance: visit_date and visit_datetime inconsistency in

<table>
<thead>
<tr>
<th></th>
<th>Verification</th>
<th>Validation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pass</td>
<td>Fail</td>
<td>Total</td>
</tr>
<tr>
<td>Plausibility</td>
<td>1938</td>
<td>75</td>
<td>2013</td>
</tr>
<tr>
<td>Conformance</td>
<td>643</td>
<td>3</td>
<td>646</td>
</tr>
<tr>
<td>Completeness</td>
<td>356</td>
<td>7</td>
<td>363</td>
</tr>
<tr>
<td>Total</td>
<td>2937</td>
<td>85</td>
<td>3022</td>
</tr>
</tbody>
</table>
Data Quality Dashboard

Data Quality Assessment

Johns Hopkins Medicine Enterprise

DataQualityDashboard Version: 1.0.0
Results generated at 2021-08-28 09:20:06 in 7 hours

<table>
<thead>
<tr>
<th>Verification</th>
<th>Validation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pass</td>
<td>Fail</td>
<td></td>
</tr>
<tr>
<td>Plausibility</td>
<td>1938</td>
<td>75</td>
</tr>
<tr>
<td>Conformance</td>
<td>643</td>
<td>3</td>
</tr>
<tr>
<td>Completeness</td>
<td>356</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>2937</td>
<td>85</td>
</tr>
</tbody>
</table>

Examples of DQ Checks from Kahn et al (2016)

- **Atemporal Plausibility**: 48% of labs outside of normal range
- **Temporal Plausibility**: Unexpected change in number of records from month to month
- **Completeness**: 62% of route_concept_id is missing
- **Value Conformance**: ICD9 codes in condition_concept_id
- **Relational Conformance**: visit_date and visit_datetime inconsistency in
Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study

Abel N. Khaw,1 M. Geoffrey Hayas,1 Laura Rasmussen-Tovell,1 Jennifer A. Pacheo,1 William K. Thompson,1 Loren L. Armstrong,1 Joshua C. Bemy,1 Peggy L. Pelot,2 Aaron W. Miller,2 Wei Q. Wei,1 Suzette J. Belinsky,1 Christopher G. Chute1
Cynthia L. Lardaro,4 Gail P. Jank1, David R. Craswell,1 Christopher S. Carlson,4 Katherine M. Newton,4 Wendy A. Watt,9 Rex I. Grisham,9 William L. Lowe9

PheKB >1,000 lines of code