Expansive Use of Data Repositories for Clinical Research

Thomas R. Campion, Jr., Ph.D.
Assistant Professor of Healthcare Policy & Research
Director, Research Informatics
Problem

- Obtaining electronic health record (EHR) data for clinical and translational research is difficult
  - Repurpose transactions for research
  - Use one or more tools
  - Understand strengths and limitations
  - Obtain approval
- Optimal approaches are unknown
- Research Informatics can help investigators
  - Obtain EHR data
  - Collect novel measures
  - Integrate data
ARCH: Architecture for Research Computing in Health

Retrospective

- Multi-institutional Data Sharing
- College-wide Cohort Discovery
- EHR Reporting
- Research Data Repositories
- Electronic Data Capture (EDC)
- EDC integrated with EHR

Prospective

- EHR Interventions

Consent

WCM-NYP-CUMC Data Sharing

Biobank and Ancillary Omics

Data Core

Data Integration  Clinical Translation  Compliance & Planning  Scientific Computing
ARCH: Architecture for Research Computing in Health

Retrospective

- Multi-institutional Data Sharing
- College-wide Cohort Discovery
- EHR Reporting
- Research Data Repositories
- Electronic Data Capture (EDC)
- EDC integrated with EHR

Prospective

Data Integration
ARCH: Architecture for Research Computing in Health

Retrospective

- Multi-institutional Data Sharing
- College-wide Cohort Discovery
- EHR Reporting
- Research Data Repositories
- Electronic Data Capture (EDC)

Prospective

EDC integrated with EHR

Data Integration

SAS
ARCH: Architecture for Research Computing in Health

Retrospective

- Multi-institutional Data Sharing
- College-wide Cohort Discovery
- EHR Reporting
- Research Data Repositories
- Electronic Data Capture (EDC)
- EDC integrated with EHR

Prospective

Model

ARCH: Architecture for Research Computing in Health

Retrospective

- Multi-institutional Data Sharing
- College-wide Cohort Discovery
- EHR Reporting
- Research Data Repositories
- Electronic Data Capture (EDC)
- EDC integrated with EHR

Prospective

- EHR Interventions

Model

Data Integration

Clinical Translation

ARCH: Architecture for Research Computing in Health

Retrospective

- Multi-institutional Data Sharing
- College-wide Cohort Discovery
- EHR Reporting
- Research Data Repositories
- Electronic Data Capture (EDC)
- EDC integrated with EHR

Prospective

- EHR Interventions

Data Integration

Clinical Translation

ARCH: Architecture for Research Computing in Health

Retrospective

Multi-institutional Data Sharing
College-wide Cohort Discovery
EHR Reporting
Research Data Repositories
Electronic Data Capture (EDC)
EDC integrated with EHR

Prospective

Epic
Allscripts
ARCH: Architecture for Research Computing in Health

Retrospective

Prospective

Multi-institutional Data Sharing
College-wide Cohort Discovery
EHR Reporting
Research Data Repositories
Electronic Data Capture (EDC)
EDC integrated with EHR

Epic
Allscripts
i2b2
ARCH: i2b2
ARCH: i2b2
ARCH: i2b2

~75 requests for MRNs
~90 IRB protocols
ARCH: i2b2

~75 requests for MRNs
~90 IRB protocols

**Original Article**

**Premedication Use in Preventing Acute Infliximab Infusion Reactions in Patients with Inflammatory Bowel Disease: A Single Center Cohort Study**

Stephanie L. Gold, MD,* Shirley Cohen-Mekelburg, MD,† Yecheskel Schneider, MD,† Nicole Shen, MD,‡ Alec Faggen, BA,§ Amanda Rupert, BA,‖ Ellen I. Schein, MD,*, Brian Bosworth, MD,‡ and Adam Steinlaut, MD‡

**Background:** Infliximab (IFX) is commonly used in patients with inflammatory bowel disease. One common side effect of IFX is an acute infusion reaction. Despite the lack of evidence supporting their use, clinicians use various premedications to prevent acute reactions. We evaluated the effectiveness of premedications in the prevention of acute IFX infusion reactions.

**Methods:** A retrospective cohort study was performed identifying patients with a diagnosis of inflammatory bowel disease who received IFX at our institution. Information about each IFX infusion was recorded, including the dose, infusion rate, use of premedications, and any reactions. Infusions were stratified into low and high risk. In the high- and low-risk groups, the relative risk was calculated for each premedication combination used in our institution.

**Results:** Seven hundred seventy-three patients were identified; 578 patients (79%) met inclusion criteria and were included for analysis. Nineteen hundred eighty-nine high-risk infusions were initiated; 610 (36.8%) of these infusions were administered with premedications (diphenhydramine and/or hydrocortisone) and 313 (5.4%) reactions occurred. Six thousand one hundred four low-risk infusions were identified; 2,253 (36.9%) of these infusions had...
ARCH: Architecture for Research Computing in Health

Retrospective

- Multi-institutional Data Sharing
- College-wide Cohort Discovery
- EHR Reporting
- Research Data Repositories
- Electronic Data Capture (EDC)
- EDC integrated with EHR

Prospective

- NYC-CDRN
- Epic
- Allscripts
- i2b2
ARCH: Architecture for Research Computing in Health

Retrospective

Multi-institutional Data Sharing
College-wide Cohort Discovery
EHR Reporting
Research Data Repositories
Electronic Data Capture (EDC)
EDC integrated with EHR

Prospective

NYC-CDRN
New York City Clinical Data Research Network

Epic
Allscripts
i2b2

WCM
NYP
Mount Sinai
NYU Langone Medical Center
Montefiore
CDN
Clinical Directors Network
ARCH: Architecture for Research Computing in Health

Retrospective

Multi-institutional Data Sharing
College-wide Cohort Discovery
EHR Reporting
Research Data Repositories
Electronic Data Capture (EDC)

EDC integrated with EHR

Prospective

NYC-CDRN
New York City Clinical Data Research Network

TriNetX
i2b2

Epic
Allscripts
ARCH: Architecture for Research Computing in Health

Multi-institutional Data Sharing  
College-wide Cohort Discovery  
EHR Reporting  
Research Data Repositories  
Electronic Data Capture (EDC)  
EDC integrated with EHR

Retrospective  
Prospective

NYC-CDRN: New York City Clinical Data Research Network  
Epic  
Allscripts  
TriNetX  
i2b2  
ACT: The ACT Network  
Powered by NCATS CTSA Program
ARCH: Architecture for Research Computing in Health

Retrospective
- Multi-institutional Data Sharing
- College-wide Cohort Discovery
- Big Data Analytics
- EHR Reporting
- Research Data Repositories
- Electronic Data Capture (EDC)
- EDC integrated with EHR

Prospective

NYC-CDRN
New York City Clinical Data Research Network

OHDSI
Open Health Data Science Initiative

Epic

TriNetX

i2b2

ACT
The ACT Network
Powered by NCATS CTSA Program

Allscripts
ARCH: Architecture for Research Computing in Health

Retrospective

- Multi-institutional Data Sharing
- College-wide Cohort Discovery
- Big Data Analytics
- EHR Reporting
- Research Data Repositories
- Electronic Data Capture (EDC)

Prospective

- EDC integrated with EHR

NYC-CDRN
New York City Clinical Data Research Network

TriNetX

i2b2

OHDSI

NLP

Allscripts

Epic
ARCH: Architecture for Research Computing in Health
ARCH: Architecture for Research Computing in Health

Retrospective
- Multi-institutional Data Sharing
- College-wide Cohort Discovery
- Big Data Analytics
- EHR Reporting
- Research Data Repositories

Prospective
- Electronic Data Capture (EDC)
- EDC integrated with EHR

Tools and Providers:
- NYC-CDRN
- TriNetX
- ACT
- OHDSI
- Epic
- Allscripts
- REDCap
- SUPER
- NLP

Big Data Analytics
NLP
Allscripts
REDCap
SUPER

ARCH: SUPER REDCap

De-ID SUPER REDCap Project

Demographics And Enrollment

Adding new Record ID 29

Record ID

Demographics for ______

MRN

First name

Last name

Date of birth

Enrollment

Cohorts

Form Status
ARCH: SUPER REDCap
ARCH: SUPER REDCap
## Adjudicate data from ARCH WZ Data

Displayed below is the data fetched from the external source system. It will display all the mapped REDCap fields that have source data returned. To import the source data values into REDCap, select the source value by clicking the radio button for the desired value in the row. Some fields may have multiple values returned, so you must choose the best or most correct value. Once you made all your selections, press the Save button below to save the source values into REDCap.

### Fetching data for Record ID "29"

**New items:** 3  
**Time of last data fetch:** less than a minute ago

<table>
<thead>
<tr>
<th>REDCap Field</th>
<th>REDCap Date/Time</th>
<th>ARCH WZ Data Date/Time</th>
<th>REDCap Current Value</th>
<th>ARCH WZ Data Source Value</th>
<th>Import?</th>
</tr>
</thead>
<tbody>
<tr>
<td>first_name &quot;first name&quot;</td>
<td>-</td>
<td>-</td>
<td>A</td>
<td>Ztest</td>
<td>false</td>
</tr>
<tr>
<td>last_name &quot;Last name&quot;</td>
<td>-</td>
<td>-</td>
<td>Z</td>
<td>Ztest</td>
<td>true</td>
</tr>
<tr>
<td>date_of_birth &quot;Date of birth&quot;</td>
<td>-</td>
<td>-</td>
<td>1955-12-16</td>
<td>1955-12-16</td>
<td>true</td>
</tr>
</tbody>
</table>

Save  | Cancel
ARCH: SUPER REDCap

De-ID SUPER REDCap Project

Laboratory Results

Editing existing Record ID 29

No new items from source system (View)

Record ID 29

Visit date

1 K/μL = 1000 cells per micro liter
1 M/μL = 1000000 cells per micro liter
% = Percentage

Platelets (K/μL)

Neutrophils (K/μL)

Monocytes (K/μL)

Lymphocytes (K/μL)

Leukocytes (K/μL)

Erythrocytes (M/μL)
ARCH: SUPER REDCap

De-ID SUPER REDCap Project

Laboratory Results

Record ID 29

Record ID 29

Visit date: 02-02-2015

1 K/uL = 1000 cells per micro liter
1 M/uL = 1000000 cells per micro liter

% = Percentage

Platelets (K/uL)

Neutrophils (K/uL)

Monocytes (K/uL)

Lymphocytes (K/uL)

Leukocytes (K/uL)
**ARCH: SUPER REDCap**

Image of a computer screen displaying a REDCap data adjudication form. The form is titled "Adjudicate data from ARCH WZ Data." It shows a table with columns for REDCap fields, REDCap dates/times, ARCH WZ data source dates/times, REDCap current values, ARCH WZ data source values, and import/export options. The table contains data for platelets values over time, with dates ranging from 2009-05-29 to 2009-12-11.

---

**Description:**

- The top section of the form provides instructions on how to adjudicate data from an external source into REDCap.
- It explains that the screen displays mapped REDCap fields with data that has been fetched.
- Users can select the source value by clicking the radio button for the desired value in the row.
- Some fields may have multiple values, so users must choose the best or most correct value.

**Table Content:**

<table>
<thead>
<tr>
<th>REDCap Field</th>
<th>REDCap Date/Time</th>
<th>ARCH WZ Data Source Date/Time</th>
<th>REDCap Current Value</th>
<th>ARCH WZ Data Source Value</th>
<th>Import?</th>
</tr>
</thead>
<tbody>
<tr>
<td>platelets &quot;Platelets (K/uL)&quot;</td>
<td>2010-02-02 (00:00)</td>
<td>2009-05-29 00:00</td>
<td>265</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-10-07 00:00</td>
<td>2009-05-29 00:00</td>
<td>287</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-11-02 00:00</td>
<td>2009-05-29 00:00</td>
<td>188</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-11-03 00:00</td>
<td>2009-05-29 00:00</td>
<td>219</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-11-03 00:00</td>
<td>2009-05-29 00:00</td>
<td>317</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-11-09 00:00</td>
<td>2009-05-29 00:00</td>
<td>245</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-11-17 00:00</td>
<td>2009-05-29 00:00</td>
<td>184</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-11-20 00:00</td>
<td>2009-05-29 00:00</td>
<td>230</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-11-29 00:00</td>
<td>2009-05-29 00:00</td>
<td>172</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-12-11 00:00</td>
<td>2009-05-29 00:00</td>
<td>214</td>
<td></td>
<td>reset</td>
</tr>
</tbody>
</table>
ARCH: SUPER REDCap

Displayed below is the data fetched from the external source system. It will display all the mapped REDCap fields that have source data returned. To import the source data values into REDCap, select the source value by clicking the radio button for the desired value in the row. Some fields may have multiple values returned, so you must choose the best or most correct value. Once you made all your selections, press the Save button below to save the source values into REDCap.

Fetching data for Record ID "29" using ±150 days

New items: 8
Time of last data fetch: just now
Display all items (all forms)
Display only this form's items

<table>
<thead>
<tr>
<th>REDCap Field</th>
<th>REDCap Date/Time</th>
<th>ARCH WZ Data Source Date/Time</th>
<th>REDCap Current Value</th>
<th>ARCH WZ Data Source Value</th>
<th>Import?</th>
</tr>
</thead>
<tbody>
<tr>
<td>platelets &quot;Platelets (K/UL)&quot;</td>
<td>2010-02-02 (00:00)</td>
<td>2009-09-29 00:00</td>
<td>265</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-10-07 00:00</td>
<td>2009-10-07 00:00</td>
<td>287</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-11-02 00:00</td>
<td>2009-11-02 00:00</td>
<td>188</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-11-03 00:00</td>
<td>2009-11-03 00:00</td>
<td>219</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-11-03 00:00</td>
<td>2009-11-03 00:00</td>
<td>317</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-11-09 00:00</td>
<td>2009-11-09 00:00</td>
<td>245</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-11-17 00:00</td>
<td>2009-11-17 00:00</td>
<td>184</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-11-20 00:00</td>
<td>2009-11-20 00:00</td>
<td>230</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-11-29 00:00</td>
<td>2009-11-29 00:00</td>
<td>172</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td>2009-12-11 00:00</td>
<td>2009-12-11 00:00</td>
<td>714</td>
<td></td>
<td>reset</td>
</tr>
</tbody>
</table>
ARCH: SUPER REDCap

De-ID SUPER REDCap Project

Laboratory Results

<table>
<thead>
<tr>
<th>Record ID</th>
<th>29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visit date</td>
<td>02-02-2010</td>
</tr>
<tr>
<td>1 K/uL = 1000 cells per micro liter</td>
<td></td>
</tr>
<tr>
<td>1 M/uL = 100 000 cells per micro liter</td>
<td></td>
</tr>
<tr>
<td>% = Percentage</td>
<td></td>
</tr>
<tr>
<td>Platelets (K/uL)</td>
<td>184</td>
</tr>
<tr>
<td>Neutrophils (K/uL)</td>
<td>2.4</td>
</tr>
<tr>
<td>Monocytes (K/uL)</td>
<td>0.5</td>
</tr>
<tr>
<td>Lymphocytes (K/uL)</td>
<td>2.2</td>
</tr>
<tr>
<td>Leukocytes (K/uL)</td>
<td>5.2</td>
</tr>
</tbody>
</table>
ARCH: Architecture for Research Computing in Health

Retrospective

- Multi-institutional Data Sharing
- College-wide Cohort Discovery
- EHR Reporting
- Research Data Repositories
- Electronic Data Capture (EDC)
- EDC integrated with EHR

Prospective
ARCH: Architecture for Research Computing in Health

- Retrospective
  - Multi-institutional Data Sharing
  - College-wide Cohort Discovery
  - EHR Reporting
  - Research Data Repositories
  - Electronic Data Capture (EDC) integrated with EHR

- Prospective

Epic

REDCap
ARCH: Architecture for Research Computing in Health

Retrospective

Multi-institutional Data Sharing
College-wide Cohort Discovery
EHR Reporting
Research Data Repositories
Electronic Data Capture (EDC)
EDC integrated with EHR

Prospective

Epic
Allscripts™
REDCap

Weill Cornell Medicine
New York-Presbyterian
ARCH: Architecture for Research Computing in Health

- Retrospective
  - Multi-institutional Data Sharing
  - College-wide Cohort Discovery
  - EHR Reporting
  - Research Data Repositories

- Prospective
  - Electronic Data Capture (EDC) integrated with EHR

Epic

Allscripts™

CompuRecord
ARCH: Architecture for Research Computing in Health

Retrospective

- Multi-institutional Data Sharing
- College-wide Cohort Discovery
- EHR Reporting
- Research Data Repositories
- Electronic Data Capture (EDC)
- EDC integrated with EHR

Epic

Allscripts™

CompuRecord

REDCap

Biobank

Prospective
ARCH: Architecture for Research Computing in Health

Retrospective

Multi-institutional Data Sharing
College-wide Cohort Discovery
EHR Reporting
Research Data Repositories

Prospective
Electronic Data Capture (EDC)
EDC integrated with EHR

Multi-institutional Data Sharing
College-wide Cohort Discovery
EHR Reporting
Research Data Repositories
Electronic Data Capture (EDC)
EDC integrated with EHR

ARCH: Architecture for Research Computing in Health

Multi-institutional Data Sharing
College-wide Cohort Discovery
EHR Reporting
Research Data Repositories
Electronic Data Capture (EDC)
EDC integrated with EHR

ARCH: Architecture for Research Computing in Health
ARCH: Research Data Repository

Disparate Sources

Research Data Repositories (RDRs)

Centralized Integration

Aggregate → Transform

Epic
Allscripts
CompuRecord
REDCap
IProfiler

Other
ARCH: Research Data Repository

Disparate Sources  Research Data Repositories (RDRs)

Centralized Integration

Epic  Allscripts

Aggregate  Transform  Anesthesiology
ARCH: Research Data Repository

Disparate Sources

Research Data Repositories (RDRs)

Centralized Integration

Aggregate → Transform → Center for Advanced Digestive Care

Disparate Sources: Epic, Allscripts, CompuRecord, REDCap, iProfiler

Research Data Repositories (RDRs): Center for Advanced Digestive Care
ARCH: Research Data Repository

Disparate Sources

Research Data Repositories (RDRs)

Centralized Integration

Aggregate → Transform

Leukemia Program
ARCH: Research Data Repository

Disparate Sources

Research Data Repositories (RDRs)

Epic
Allscripts
REDCap
Other

Centralized Integration

Aggregate ➔ Transform ➔ Neurological Surgery
ARCH: Research Data Repository
ARCH: Research Data Repository
ARCH: Research Data Repository

Tools

- i2b2

Data

- RDR
- REDCap
- Epic
- Allscripts
- Other

Workflows

- Discovery
  - Hypothesis generation
  - Patient counts
- Collection
  - Capture of novel measures
  - Annotation of existing data
- Analysis
  - Scientific variable definition
  - Hypothesis testing
ARCH: Research Data Repository

**Tools**
- i2b2

**Workflows**
- **Discovery**
  - Hypothesis generation
  - Patient counts
- **Collection**
  - Capture of novel measures
  - Annotation of existing data
- **Analysis**
  - Scientific variable definition
  - Hypothesis testing

**Data**
- RDR
- SUPER
- REDCap
- Microsoft SQL Server

**Workflows**:
- Study 1
- Study 2
- Study n: future
ARCH: RDR Startup and Maintenance

- **Startup**
  - Approval from ARCH Steering Committee
  - Creation of project team
  - Prioritization of workflows, tools, and projects
  - One-time implementation cost

- **Maintenance**
  - Data refresh, hosting, and security
  - 40 hours of custom development
  - Process for expanded projects
  - Annual fee
ARCH: Research Data Repositories

1. Anesthesiology
2. Center for Advanced Digestive Care
3. Englander Institute for Precision Medicine
4. Myeloproliferative Neoplasms
5. Neurology/Clinical and Translational Neuroscience Unit
6. Ophthalmology
7. Pathak Lab
8. Pediatric Epilepsy
9. Pediatric Pulmonary
10. Pulmonary & Critical Care
11. Neurological Surgery
12. Urology
13. Trauma
ARCH: Architecture for Research Computing in Health

- Multi-institutional Data Sharing
- College-wide Cohort Discovery
- Big Data Analytics
- EHR Reporting
- Research Data Repositories
- Electronic Data Capture (EDC)
- EDC integrated with EHR

- NYC-CDRN
- TriNetX
- ACT
- i2b2
- NLP
- Epic
- NHDSI
- Allscripts
- REDCap
- SUPER
- REDCap
- SQL Server
Discussion

- Research Informatics can help investigators
  - Obtain EHR data
  - Collect novel research data
  - Integrate data
- ARCH matches investigators with right tools and services with respect to
  - Study design
  - Data sources
  - Cost
- One approach does not meet all needs
  - Efficient evidence generation, not tool usage, is the goal
  - Premium offerings can partially offset cost of free services
Acknowledgments

**Funding**
- Clinical and Translational Science Center (UL1 TR000457)
- Joint Clinical Trials Office

**ARCH Leadership**
- Curtis Cole, MD
- Stephen Johnson, PhD
- John Leonard, MD
- Jyotishman Pathak, PhD
- Vinay Varughese

**ITS Research Informatics**
- Sajjad Abedian
- Prakash Adekkanattu
- Marcos Davila
- Xiaobo Fuld
- David Kraemer
- Steven Flores
- Joseph Kabariti
- Ryan McGregor
- Sean Pompea
- Julian Schwartz
- Evan Sholle
- Scott Turner
- Jacob Weiser
Questions

• Access and inquiries
  – ARCH inquiries: arch-support@med.cornell.edu
  – i2b2 access: i2b2-support@med.cornell.edu

• Web resources
  – ARCH: http://arch.weill.cornell.edu
  – NYC-CDRN: http://www.nyccdrn.org
  – TRAC: https://webapps.nyp.org/trac

• Tom Campion
  – thc2015@med.cornell.edu
  – 646-962-2345
  – http://vivo.med.cornell.edu/display/cwid-thc2015