Supporting Researchers with Electronic Patient Data

Thomas R. Campion, Jr., Ph.D.
Associate Professor of Research in Healthcare Policy & Research
Director, Research Informatics
Overview

• Problem
• Approach
  – Research Computing at Weill Cornell Medicine
  – Architecture for Research Computing in Health (ARCH)
• Discussion
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
Architecture for Research Computing in Health (ARCH)

Retrospective

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

Prospective

EDC integrated with EHR
EHR Interventions

Consent

WCM-NYP-CUMC Data Sharing

Biobank and Ancillary Omics
Data Core

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

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
Architecture for Research Computing in Health (ARCH)

Retrospective

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

Sas

Prospective

Data Integration
Architecture for Research Computing in Health (ARCH)

Retrospective

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

Prospective

Data Core

Data Integration

Scientific Computing
Architecture for Research Computing in Health (ARCH)

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

Prospective
- EHR Interventions

Data Core

Data Integration
Clinical Translation
Scientific Computing

Architecture for Research Computing in Health (ARCH)

Retrospective

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

Prospective

- EHR Interventions

Tools and Technologies:

- ACT
- OHDSI
- Epic
- Allscripts
- i2b2
- REDCap
- TrinetX
- SuperNLP
- NLP
- SQL Server
- Research Electronic Data Capture (REDCap)
Architecture for Research Computing in Health (ARCH)

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

Prospective
- Electronic Data Capture (EDC)
- EHR Interventions

Tools:
- Epic
- i2b2
- REDCap
- SQL Server
- SUPER
Architecture for Research Computing in Health (ARCH)

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

- Prospective
  - EHR Interventions

Epic
Step 1: obtain raw EHR data describing diagnoses assigned in patients encounters

<table>
<thead>
<tr>
<th>patient_num</th>
<th>enc_num</th>
<th>line</th>
<th>dx_id</th>
<th>primary_dx</th>
<th>comme...</th>
<th>dx_chronic</th>
<th>dx_unique</th>
<th>dx_ed_yn</th>
</tr>
</thead>
<tbody>
<tr>
<td>828</td>
<td>30147</td>
<td>3</td>
<td>935415</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>4</td>
<td>N</td>
</tr>
<tr>
<td>829</td>
<td>69154</td>
<td>1</td>
<td>12866</td>
<td>Y</td>
<td>NULL</td>
<td>N</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>830</td>
<td>42675</td>
<td>1</td>
<td>56131</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>2</td>
<td>N</td>
</tr>
<tr>
<td>831</td>
<td>42675</td>
<td>2</td>
<td>135365</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>3</td>
<td>N</td>
</tr>
<tr>
<td>832</td>
<td>63397</td>
<td>1</td>
<td>813626</td>
<td>Y</td>
<td>NULL</td>
<td>N</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>833</td>
<td>63397</td>
<td>2</td>
<td>807290</td>
<td>Y</td>
<td>NULL</td>
<td>N</td>
<td>2</td>
<td>N</td>
</tr>
<tr>
<td>834</td>
<td>42839</td>
<td>1</td>
<td>15598</td>
<td>Y</td>
<td>NULL</td>
<td>N</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>835</td>
<td>105950</td>
<td>1</td>
<td>8059885</td>
<td>Y</td>
<td>NULL</td>
<td>N</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>836</td>
<td>44243</td>
<td>1</td>
<td>24458</td>
<td>Y</td>
<td>NULL</td>
<td>N</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>837</td>
<td>13016</td>
<td>1</td>
<td>8028848</td>
<td>Y</td>
<td>NULL</td>
<td>N</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>838</td>
<td>32482</td>
<td>1</td>
<td>1005331</td>
<td>Y</td>
<td>NULL</td>
<td>N</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>839</td>
<td>79305</td>
<td>1</td>
<td>8078416</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>2</td>
<td>N</td>
</tr>
<tr>
<td>840</td>
<td>79305</td>
<td>2</td>
<td>6111</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>3</td>
<td>N</td>
</tr>
<tr>
<td>841</td>
<td>79305</td>
<td>3</td>
<td>8005450</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>4</td>
<td>N</td>
</tr>
<tr>
<td>842</td>
<td>79305</td>
<td>4</td>
<td>9834</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>5</td>
<td>N</td>
</tr>
<tr>
<td>843</td>
<td>79305</td>
<td>5</td>
<td>8059219</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>6</td>
<td>N</td>
</tr>
<tr>
<td>844</td>
<td>79305</td>
<td>6</td>
<td>8062109</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>7</td>
<td>N</td>
</tr>
<tr>
<td>845</td>
<td>79305</td>
<td>7</td>
<td>8091888</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>8</td>
<td>N</td>
</tr>
</tbody>
</table>
ARCH: EHR Reporting

Step 2: match raw EHR data describing diagnoses to ICD-9 codes

<table>
<thead>
<tr>
<th>patient_num</th>
<th>enc_num</th>
<th>line</th>
<th>dx_id</th>
<th>enc_icd_code</th>
<th>primary_dx_yn</th>
<th>comma_yn</th>
<th>dx_chronic_yn</th>
<th>dx_unique_yn</th>
<th>dx_ed_yn</th>
</tr>
</thead>
<tbody>
<tr>
<td>828</td>
<td>30147</td>
<td>3</td>
<td>935415</td>
<td>250.60</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>4</td>
<td>N</td>
</tr>
<tr>
<td>829</td>
<td>69154</td>
<td>1</td>
<td>12866</td>
<td>996.82</td>
<td>Y</td>
<td>NULL</td>
<td>N</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>830</td>
<td>42875</td>
<td>1</td>
<td>56131</td>
<td>389.18</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>2</td>
<td>N</td>
</tr>
<tr>
<td>831</td>
<td>42875</td>
<td>2</td>
<td>135365</td>
<td>388.30</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>3</td>
<td>N</td>
</tr>
<tr>
<td>832</td>
<td>63397</td>
<td>1</td>
<td>813625</td>
<td>555.1</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>833</td>
<td>63397</td>
<td>2</td>
<td>807290</td>
<td>NULL</td>
<td>Y</td>
<td>NULL</td>
<td>N</td>
<td>2</td>
<td>N</td>
</tr>
<tr>
<td>834</td>
<td>42839</td>
<td>1</td>
<td>15598</td>
<td>239.6</td>
<td>Y</td>
<td>NULL</td>
<td>N</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>835</td>
<td>105950</td>
<td>1</td>
<td>8059885</td>
<td>782.1</td>
<td>Y</td>
<td>NULL</td>
<td>N</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>836</td>
<td>44243</td>
<td>1</td>
<td>80450</td>
<td>281.1</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>837</td>
<td>13016</td>
<td>1</td>
<td>8028848</td>
<td>415.19</td>
<td>Y</td>
<td>NULL</td>
<td>N</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>838</td>
<td>32482</td>
<td>1</td>
<td>1005381</td>
<td>715.96</td>
<td>Y</td>
<td>NULL</td>
<td>N</td>
<td>1</td>
<td>N</td>
</tr>
<tr>
<td>839</td>
<td>79305</td>
<td>1</td>
<td>8078416</td>
<td>185</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>2</td>
<td>N</td>
</tr>
<tr>
<td>840</td>
<td>79305</td>
<td>2</td>
<td>6111</td>
<td>591.9</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>3</td>
<td>N</td>
</tr>
<tr>
<td>841</td>
<td>79305</td>
<td>3</td>
<td>8005450</td>
<td>198.5</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>4</td>
<td>N</td>
</tr>
<tr>
<td>842</td>
<td>79305</td>
<td>4</td>
<td>9834</td>
<td>787.91</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>5</td>
<td>N</td>
</tr>
<tr>
<td>843</td>
<td>79305</td>
<td>5</td>
<td>8059219</td>
<td>780.79</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>6</td>
<td>N</td>
</tr>
<tr>
<td>844</td>
<td>79305</td>
<td>6</td>
<td>8082109</td>
<td>279.00</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>7</td>
<td>N</td>
</tr>
<tr>
<td>845</td>
<td>79305</td>
<td>7</td>
<td>8091886</td>
<td>593.9</td>
<td>N</td>
<td>NULL</td>
<td>N</td>
<td>8</td>
<td>N</td>
</tr>
</tbody>
</table>
ARCH: EHR Reporting

Step 3: transform raw EHR data describing ICD-9 codes into dichotomous scientific variable

<table>
<thead>
<tr>
<th>Race</th>
<th>Height</th>
<th>Weight</th>
<th>BMI</th>
<th>Diabetes</th>
<th>Hypertension</th>
<th>Obesity</th>
<th>NSAID</th>
<th>Antidiabetic</th>
<th>Current_Cigarette_Smoker</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 WHITE</td>
<td>5' 4&quot;</td>
<td>2096</td>
<td>22.84</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>2 OTHER COMBINATIONS NOT DESCRIBED</td>
<td>6' 0&quot;</td>
<td>4704</td>
<td>39.47</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>3 WHITE</td>
<td>5' 5&quot;</td>
<td>3168</td>
<td>34.72</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>4 WHITE</td>
<td>5' 4.5&quot;</td>
<td>2352</td>
<td>24.85</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>5 ASIAN</td>
<td>5' 9&quot;</td>
<td>2656</td>
<td>25.60</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>6 DECLINED</td>
<td>6' 0&quot;</td>
<td>3648</td>
<td>30.92</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>7 DECLINED</td>
<td>5' 6&quot;</td>
<td>1827.2</td>
<td>18.44</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>8 OTHER COMBINATIONS NOT DESCRIBED</td>
<td>4' 9.5&quot;</td>
<td>1872</td>
<td>25.11</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>9 OTHER COMBINATIONS NOT DESCRIBED</td>
<td>5' 2&quot;</td>
<td>2448</td>
<td>28.92</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>10 ASIAN</td>
<td>5' 9&quot;</td>
<td>3072</td>
<td>28.34</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>11 NULL</td>
<td>5' 0&quot;</td>
<td>3609.6</td>
<td>44.06</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>12 DECLINED</td>
<td>5' 1.024&quot;</td>
<td>2704</td>
<td>31.95</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>13 WHITE</td>
<td>6' 1&quot;</td>
<td>3056</td>
<td>26.08</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>14 NULL</td>
<td>6' 0&quot;</td>
<td>3408</td>
<td>3827.82</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>15 NULL</td>
<td>6' 2&quot;</td>
<td>3284.8</td>
<td>25.99</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>16 DECLINED</td>
<td>5' 8&quot;</td>
<td>2560</td>
<td>26.63</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>17 DECLINED</td>
<td>5' 8&quot;</td>
<td>2640</td>
<td>25.09</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>18 NULL</td>
<td>5' 1&quot;</td>
<td>2769.8</td>
<td>32.72</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>
**ARCH: EHR Reporting**

Step 4: repeat process to define additional diagnosis dichotomous variables

<table>
<thead>
<tr>
<th>Race</th>
<th>Height</th>
<th>Weight</th>
<th>BMI</th>
<th>Diabetes</th>
<th>Hypertension</th>
<th>Obesity</th>
<th>NSAID</th>
<th>Antidiabetic</th>
<th>Current_Cigarette_Smoker</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WHITE</td>
<td>5' 4&quot;</td>
<td>2096</td>
<td>22.64</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>2</td>
<td>OTHER COMBINATIONS NOT DESCRIBED</td>
<td>6' 0&quot;</td>
<td>4704</td>
<td>39.47</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>WHITE</td>
<td>5' 5&quot;</td>
<td>3168</td>
<td>34.72</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>4</td>
<td>WHITE</td>
<td>5' 4.5&quot;</td>
<td>2352</td>
<td>24.85</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>5</td>
<td>ASIAN</td>
<td>5' 9&quot;</td>
<td>2656</td>
<td>25.60</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>6</td>
<td>DECLINED</td>
<td>6' 0&quot;</td>
<td>3648</td>
<td>30.92</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>7</td>
<td>DECLINED</td>
<td>5' 6&quot;</td>
<td>1827.2</td>
<td>18.44</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>8</td>
<td>OTHER COMBINATIONS NOT DESCRIBED</td>
<td>4' 9.5&quot;</td>
<td>1872</td>
<td>25.11</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>9</td>
<td>OTHER COMBINATIONS NOT DESCRIBED</td>
<td>5' 2&quot;</td>
<td>2448</td>
<td>28.92</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>10</td>
<td>ASIAN</td>
<td>5' 9&quot;</td>
<td>3072</td>
<td>28.34</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>11</td>
<td>NULL</td>
<td>5' 0&quot;</td>
<td>3609.6</td>
<td>44.06</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>12</td>
<td>DECLINED</td>
<td>5' 1.024&quot;</td>
<td>2704</td>
<td>31.95</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>13</td>
<td>WHITE</td>
<td>6' 1&quot;</td>
<td>3056</td>
<td>26.98</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>14</td>
<td>NULL</td>
<td>6' 0&quot;</td>
<td>3408</td>
<td>3827.82</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>15</td>
<td>NULL</td>
<td>6' 2&quot;</td>
<td>3284.8</td>
<td>26.99</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>16</td>
<td>DECLINED</td>
<td>5' 8&quot;</td>
<td>2560</td>
<td>26.63</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>17</td>
<td>DECLINED</td>
<td>5' 8&quot;</td>
<td>2640</td>
<td>28.09</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>18</td>
<td>NULL</td>
<td>5' 1.1&quot;</td>
<td>2769.6</td>
<td>32.72</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>
Step 4: repeat process to transform raw EHR data for medications and other domains (raw data for medications and other domains not pictured)

<table>
<thead>
<tr>
<th>Race</th>
<th>Height</th>
<th>Weight</th>
<th>BMI</th>
<th>Diabetes</th>
<th>Hypertension</th>
<th>Obesity</th>
<th>NSAID</th>
<th>Antidiabetic</th>
<th>Current_Cigarette_Smoker</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHITE</td>
<td>5' 4&quot;</td>
<td>2096</td>
<td>22.84</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>OTHER COMBINATIONS NOT DESCRIBED</td>
<td>6' 0&quot;</td>
<td>4704</td>
<td>39.47</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>WHITE</td>
<td>5' 5&quot;</td>
<td>3169</td>
<td>34.72</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>WHITE</td>
<td>5' 4.5&quot;</td>
<td>2352</td>
<td>24.85</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>ASIAN</td>
<td>5' 9&quot;</td>
<td>2656</td>
<td>25.60</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>DECLINED</td>
<td>6' 0&quot;</td>
<td>3648</td>
<td>30.92</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>DECLINED</td>
<td>5' 6&quot;</td>
<td>1827.2</td>
<td>18.44</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>OTHER COMBINATIONS NOT DESCRIBED</td>
<td>4' 9.5&quot;</td>
<td>1872</td>
<td>25.11</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>OTHER COMBINATIONS NOT DESCRIBED</td>
<td>5' 2&quot;</td>
<td>2448</td>
<td>28.92</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>ASIAN</td>
<td>5' 9&quot;</td>
<td>3072</td>
<td>28.34</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>NULL</td>
<td>5' 0&quot;</td>
<td>3609.6</td>
<td>44.06</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>DECLINED</td>
<td>5' 1.024&quot;</td>
<td>2704</td>
<td>31.95</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>WHITE</td>
<td>6' 1&quot;</td>
<td>3056</td>
<td>26.08</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>NULL</td>
<td>6' 0&quot;</td>
<td>3408</td>
<td>3827.82</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>NULL</td>
<td>6' 2&quot;</td>
<td>3284.8</td>
<td>25.99</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>DECLINED</td>
<td>5' 8&quot;</td>
<td>2560</td>
<td>26.63</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>DECLINED</td>
<td>5' 8&quot;</td>
<td>2640</td>
<td>25.09</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>NULL</td>
<td>5' 1&quot;</td>
<td>2769.6</td>
<td>32.72</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
</tbody>
</table>
Architecture for Research Computing in Health (ARCH)

Retrospective

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

Prospective

EHR Interventions

i2b2
ARCH: i2b2
ARCH: i2b2
ARCH: i2b2
ARCH: i2b2
ARCH: i2b2
ARCH: i2b2
ARCH: i2b2

i2b2 Query & Analysis Tool

Query Tool

Query Name: 250: Diabetes m@18:43:05

Temporal Constraint: Treat all groups independently

Groups:

- **Group 1**
  - 250: Diabetes mellitus
  - [Outpatient] Encounter Diagnosis
  - [Outpatient] Medical History
  - [Outpatient] Problem List
  - 240-246: Disorders of thyroid gland
  - 249-259: Diseases of other endocrine glands
  - 250: Secondary diabetes mellitus
  - 250: Diabetes mellitus without mention of complication
  - 250:1: Diabetes with ketoacidosis
  - 250:2: Hyperosmolar (nonketotic) coma
  - 250:3: Diabetes with other coma
  - 250:4: Diabetes with renal manifestations
  - 250:5: Diabetes with ophthalmic manifestations

- **Group 2**
  - [Outpatient] Encounter Diagnosis
  - [Outpatient] Medical History
  - [Outpatient] Problem List
  - 250:0: Diabetes mellitus without mention of complication

- **Group 3**
  - [Outpatient] Encounter Diagnosis
  - [Outpatient] Medical History
  - [Outpatient] Problem List
  - 250:1: Diabetes with ketoacidosis
  - 250:2: Hyperosmolar (nonketotic) coma
  - 250:3: Diabetes with other coma
  - 250:4: Diabetes with renal manifestations
  - 250:5: Diabetes with ophthalmic manifestations

Query Status

Finished Query: "250: Diabetes m@18:43:05"
Compute Time: 2.2 secs

Number of patients for "250: Diabetes m@18:43:05"
patient_count: 70434
ARCH: i2b2
ARCH: i2b2
ARCH: i2b2
ARCH: i2b2
Architecture for Research Computing in Health (ARCH)

Retrospective

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

Prospective

- EHR Interventions
### De-ID SUPER REDCap Project

#### Demographics And Enrollment

<table>
<thead>
<tr>
<th>Record ID</th>
<th>29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics for _____</td>
<td></td>
</tr>
<tr>
<td>MRN</td>
<td></td>
</tr>
<tr>
<td>First name</td>
<td></td>
</tr>
<tr>
<td>Last name</td>
<td></td>
</tr>
<tr>
<td>Date of birth</td>
<td></td>
</tr>
</tbody>
</table>

**Enrollment**

- Breast
- Leukemia
- Lymphoma
ARCH: SUPER REDCap

Weill Cornell Medical College
Clinical and Translational Science Center

De-ID SUPER REDCap Project

Demographics And Enrollment

Adding new Record ID 29

Record ID 29

Demographics for ____

MRN

601987

First name

Last name

Date of birth

Enrollment

Cohorts

Breast

Leukemia

Lymphoma

Form Status
Save record and fetch data

Now that you have entered a value for the Source Identifier Field, click the Save button below to save the value and automatically begin fetching data from the source system.
**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

- **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>first_name</td>
<td>-</td>
<td>-</td>
<td>A</td>
<td>Ztest</td>
<td>reset</td>
</tr>
<tr>
<td>last_name</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>date_of_birth</td>
<td>-</td>
<td>-</td>
<td>1955-12-16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Save**
- **Cancel**
## De-ID SUPER REDCap Project

### Demographics And Enrollment

#### Editing existing Record ID 29

<table>
<thead>
<tr>
<th>Record ID</th>
<th>29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics for _____</td>
<td></td>
</tr>
<tr>
<td>MRN</td>
<td>601987</td>
</tr>
<tr>
<td>First name</td>
<td>A</td>
</tr>
<tr>
<td>Last name</td>
<td>Ztest</td>
</tr>
<tr>
<td>Date of birth</td>
<td>1955-12-16</td>
</tr>
<tr>
<td>Enrollment</td>
<td></td>
</tr>
<tr>
<td>Cohorts</td>
<td></td>
</tr>
<tr>
<td>Breast</td>
<td></td>
</tr>
</tbody>
</table>
## Laboratory Results

### Editing existing Record ID 29

<table>
<thead>
<tr>
<th>Record ID</th>
<th>29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visit date</td>
<td></td>
</tr>
</tbody>
</table>
| 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) |  |
### Laboratory Results

**Editing existing Record ID 29**

<table>
<thead>
<tr>
<th>Record ID</th>
<th>29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visit date</td>
<td>02-02-2010</td>
</tr>
</tbody>
</table>

1 K/μL = 1000 cells per micro liter  
1 M/μL = 100000 cells per micro liter  
% = Percentage

<table>
<thead>
<tr>
<th>Platelets (K/μL)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutrophils (K/μL)</td>
<td></td>
</tr>
<tr>
<td>Monocytes (K/μL)</td>
<td></td>
</tr>
<tr>
<td>Lymphocytes (K/μL)</td>
<td></td>
</tr>
<tr>
<td>Leukocytes (K/μL)</td>
<td></td>
</tr>
</tbody>
</table>
### 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" using ±150 days

**New items:** 8  
**Time of last data fetch:** just now

<table>
<thead>
<tr>
<th>REDCap Field</th>
<th>REDCap Date/Time</th>
<th>REDCap 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>2009-09-29 00:00</td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-10-07 00:00</td>
<td>287</td>
<td>2009-10-07 00:00</td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-11-02 00:00</td>
<td>188</td>
<td>2009-11-02 00:00</td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-11-03 00:00</td>
<td>219</td>
<td>2009-11-03 00:00</td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-11-03 00:00</td>
<td>317</td>
<td>2009-11-03 00:00</td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-11-09 00:00</td>
<td>245</td>
<td>2009-11-09 00:00</td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-11-17 00:00</td>
<td>184</td>
<td>2009-11-17 00:00</td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-11-20 00:00</td>
<td>230</td>
<td>2009-11-20 00:00</td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-11-20 00:00</td>
<td>172</td>
<td>2009-11-20 00:00</td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-12-11 00:00</td>
<td>714</td>
<td>2009-12-11 00:00</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**

<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></td>
<td>2009-10-07 00:00</td>
<td>287</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-11-02 00:00</td>
<td>188</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-11-03 00:00</td>
<td>219</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-11-09 00:00</td>
<td>317</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-11-10 00:00</td>
<td>245</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-11-17 00:00</td>
<td>184</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-11-20 00:00</td>
<td>230</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-11-29 00:00</td>
<td>172</td>
<td></td>
<td>reset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2009-12-11 00:00</td>
<td>714</td>
<td></td>
<td>reset</td>
</tr>
</tbody>
</table>

Save | Cancel
## Laboratory Results

### Editing existing Record ID 29

<table>
<thead>
<tr>
<th>Record ID</th>
<th>29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visit date</td>
<td>02-02-2010</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<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: SUPER REDCap on FHIR
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
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

REDCap
**ARCH:** Architecture for Research Computing in Health

<table>
<thead>
<tr>
<th>Retrospective</th>
<th>Prospective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-institutional Data Sharing</td>
<td>EDC integrated with EHR</td>
</tr>
<tr>
<td>College-wide Cohort Discovery</td>
<td>Research Data Repositories</td>
</tr>
<tr>
<td>EHR Reporting</td>
<td>Electronic Data Capture (EDC)</td>
</tr>
</tbody>
</table>

**Epic**

**Allscripts**

**REDCap**
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

Tools:
- Epic
- REDCap
- 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

Prospective

- Epic
- REDCap
- Allscripts™
- CompuRecord
- Biobank
ARCH: Architecture for Research Computing in Health

Retrospective

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

Prospective

- Research Data Repositories
- Electronic Data Capture (EDC)
- EDC integrated with EHR

Weill Cornell Medicine

Epic
Biobank REDCap
CompuRecord
Allscripts

NYP
ARCH: Research Data Repository
ARCH: Research Data Repository

Data

RDR

SUPER

Tools

i2b2

REDCap

Other

Microsoft SQL Server
ARCH: Research Data Repository

**Tools**
- i2b2

**Data**
- RDR
- SUPER
- REDCap
- Epic
- Allscripts
- Other
- SQL Server

**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
- Epic
- Allscripts
- Other

**Workflows**
- Study 1
- Study 2
- Study n: future

**Workflows**
- Microsoft SQL Server
ARCH: Research Data Repositories

- All of Us Research Program (Kaushal)
- Anesthesiology (Turnbull)
- Center for Advanced Digestive Care (Shah)
- Clinical and Translational Neuroscience Unit (Kamel)
- Dalio Institute (Min)
- Health Informatics (Pathak/Zhang)
- Leukemia Program (Roboz)
- Myeloproliferative Neoplasms (Scandura)
- Neurological Surgery (Hoffman)
- Ophthalmology (D’Amico)
- Pediatric Epilepsy (Grinspan)
- Pediatric Pulmonology (Ono)
- Pulmonary and Critical Care (Schenck)
- Urology (Hu)
- Trauma and Surgical Critical Care (Winchell)
ARCH RDRs: Data Catalog

Dynamic Epilepsy Database (DED)

Data Provider: Architecture for Research Computing in Health (ARCH)

description
A data set aggregating both patient-level and procedure-level data for WCM patients with epilepsy (n = ~28,000). The patient-level DED view includes both basic demographic data and a quick view of each patient’s most recent imaging and neuropsychological procedures, including excerpts from free text radiology reports. The procedure-level view contains detailed information about each imaging and neuropsychological procedure, including performing provider, CPT code, and other elements.

Record Period Start: Jan. 1, 2007
Record Period End: None
Keywords: ARCH RDR

Mitochondrial Biomarkers for the Onset and Severity of Sepsis (MBOSS)

Data Provider: Architecture for Research Computing in Health (ARCH)

description
A data mart designed to study early predictors of sepsis and the associated outcomes of different treatment approaches in a group of WCM patients presenting in the emergency department and consented to a specific protocol (n = ~300). Contains data from inpatient flowsheets, diagnoses, medications, microbiology lab results, blood bank systems, and others, including calculated critical care variables, such as Sequential Organ Failure (SOFA) score and others. Data can also be connected to banked specimens.

http://datacatalog.weill.cornell.edu
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
Acknowledgments

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

ARCH Leadership
- Curtis Cole, MD
- John Leonard, MD
- Mark Weiner, MD

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
  – TRAC: https://webapps.nyp.org/trac
  – Data Catalog: http://datacatalog.weill.cornell.edu

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