S29: PCORnet Implementation of PopMedNet Data Assessment Tools

Jessica Sturtevant, ScM
Harvard Pilgrim Health Care Institute & Harvard Medical School
PCORnet Distributed Research Network Operations Center (DRN OC)
Disclosure

• No disclosure or relationships with commercial interests.
Learning Objectives

• Evaluate extensibility of the PopMedNet Rapid Data Assessment Tool (RDAT) across different distributed research networks, diverse data sources and different common data models.
PCORnet Implementation of PopMedNet Data Assessment Tools

PCORNET
PCORnet Data Networks

11 (13) Clinical Data Research Networks (CDRNs) + 18 (20) Patient-Powered Research Networks (PPRNs) = PCORnet Phase 1 (Phase 2): A national infrastructure for patient-centered clinical research
PCORnet Goal

PCORnet seeks to improve the nation’s capacity to conduct clinical research by creating a large, highly representative, national patient-centered network that supports more efficient clinical trials and observational studies.
This map depicts the number of PCORI funded Patient-Powered or Clinical Data Research Networks that have coverage in each state.
PCORnet Distributed Research Network

• Must address confidentiality, privacy, regulatory and proprietary concerns of numerous data partners

• Distributed queries collect and report on aggregate information
Data Characterization

• Examine Common Data Model variables, formats, time trends
• Categorical variables examined using frequencies
• Continuous variables examined using distributions
• Output is summary level/aggregate
PCORnet Implementation of PopMedNet Data Assessment Tools

RAPID DATA ASSESSMENT TOOL
Distributed Research Network Operations Center (DRN OC)

• DRN OC performs data characterization (DC) for each DataMart (DM)
• DRN OC stores aggregate DC output locally
• Approved, aggregate DC output queried directly via the PCORnet Rapid Data Assessment Tool (RDAT)
• RDAT facilitates rapid, easy access to descriptive data by DataMart and Network
DRN OC Data Characterization Cycle

1. Data Characterization on Locked SAS Dataset
2. Data Characterization Review and Reporting
3. DataMart Approval
4. Approved Data Characterization Output to RDAT
5. RDAT Rapid High-level Descriptive Data by DataMart and Network
PCORnet DRN Query Tool: Rapid Data Assessment Tool (RDAT)

• PopMedNet (PMN)
  – Implemented for PCORnet Distributed Research Network (DRN) in 2014

• RDAT – PMN Query
  – Locally queries aggregate DC output
  – Supports preparatory-to-research (PTR)
  – Facilitates rapid data assessment to support clinical trials and observational research
# Standardized Data Characterization Output Model

RDAT Cross Network Extensibility

<table>
<thead>
<tr>
<th>Mini-Sentinel</th>
<th>NIH Collaboratory DRN</th>
<th>Health Data Collaboration</th>
<th>MDPHnet</th>
<th>PCORnet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical product safety surveillance</td>
<td>Research</td>
<td>Preparatory-to-research Feasibility</td>
<td>Public health surveillance</td>
<td>Research</td>
</tr>
</tbody>
</table>

18 Health Insurers/Integrated Delivery Systems

- Feasibility Project
  - 7 Health Insurers/Integrated Delivery Systems
- LIRE Project
  - 4 Health Insurers/Integrated Delivery Systems
- CRNnet
  - 10 Health Insurers/Integrated Delivery Systems
- IMEDS-Pfizer
  - 10 Health Insurers/Integrated Delivery Systems

- HCSRNnet
  - 14 Health Insurers/Integrated Delivery Systems

- 3 Multi-Site Medical Group Practices

11 Clinical Data Research Networks contributing 64 DataMarts
18 Patient-Powered Research Networks contributing 19 DataMarts

Current as of 10/27/2015
Rapid Data Assessment Tool (RDAT)

• Leveraged capabilities developed by the FDA Sentinel Project
  – Race
  – Ethnicity
  – Data Completeness

• PCORnet Enhancements
  – Age Distribution
  – Sex
  – Weight Distribution
  – Height Distribution
  – Diagnosis Codes
  – Procedure Codes
RDAT Development

PopMedNet Query Type: Rapid Data Assessment Tool (RDAT)

Leverage query terms developed by other PMN networks:
- Race
- Ethnicity
- Data

Develop query terms for PCORnet:
- Age
- Sex
- Height
- Weight
- DX
- PX
- DX, PX, etc.

Future query term development:
- Smoking
- Date Range
- Dispensing
Typical RDAT Use

1. Question CDM Tables & Variables

2. RDAT Query of pertinent variables

3. High-level Descriptive Data by DataMart and Network

4. Descriptive Data and Charts used to Assess Data Usability

5. Inform DataMart Use in Studies and Trials
RDAT Query Response

PCORnet

DataMart 1
- Data Characterization Output
- Aggregate DC Output

DataMart 2
- Data Characterization Output
- Aggregate DC Output

DataMart 3
- Data Characterization Output
- Aggregate DC Output

Etc.
- Data Characterization Output
- Aggregate DC Output

DRN OC

RDAT

Query

www.amia.org
Rapid Data Assessment Tool

SINGLE DATAMART QUERY
The 15 PCORnet CDM Domains, v3.0

**CONDITION** v2.0  
A condition represents a patient’s diagnosed and self-reported health conditions and diseases. The patient’s medical history and current state may both be represented.

**DEATH** v3.0  
Reported mortality information for patients.

**DEATH_CAUSE** v3.0  
The individual causes associated with a reported death.

**DEMOGRAPHIC** v1.0  
Demographics record the direct attributes of individual patients.

**DIAGNOSIS** v1.0  
Diagnosis codes indicate the results of diagnostic processes and medical coding within healthcare delivery.

**DISPENSING** v2.0  
Outpatient pharmacy dispensing, such as prescriptions filled through a neighborhood pharmacy with a claim paid by an insurer. Outpatient dispensing is not commonly captured within healthcare systems.

**ENCOUNTER** v1.0  
Encounters are interactions between patients and providers within the context of healthcare delivery.

**HARVEST** v3.0  
Attributes associated with the specific PCORnet datamart implementation.

**LAB_RESULT_CM** v2.0  
Laboratory result Common Measures (CM) use specific types of quantitative and qualitative measurements from blood and other body specimens. These standardized measures are defined in the same way across all PCORnet networks.

**ENROLLMENT** v1.0  
Enrollment is a concept that defines a period of time during which all medically-attended events are expected to be observed. This concept is often insurance-based, but other methods of defining enrollment are possible.

**PCORNET_TRIAL** v3.0  
Patients who are enrolled in PCORnet clinical trials.

**PRESCRIBING** v3.0  
Provider orders for medication dispensing and/or administration.

**PRO_CM** v2.0  
Patient-Reported Outcome (PRO) Common Measures (CM) are standardized measures that are defined in the same way across all PCORnet networks. Each measure is recorded at the individual item level: an individual question/statement, paired with its standardized response options.

**PROCEDURES** v1.0  
Procedure codes indicate the discreet medical interventions and diagnostic testing, such as surgical procedures, administered within healthcare delivery.

**VITAL** v1.0  
Vital signs (such as height, weight, and blood pressure) directly measure an individual’s current state of attributes.
Age Group
Rapid Data Assessment Tool Query

Age Distribution among Selected Data Partners

- <0 yrs: 0.00%
- 0-1 yrs: 0.72%
- 2-4 yrs: 1.18%
- 5-9 yrs: 1.46%
- 10-14 yrs: 2.19%
- 15-18 yrs: 5.14%
- 19-21 yrs: 11.30%
- 22-44 yrs: 26.19%
- 45-64 yrs: 23.88%
- 65-74 yrs: 20.78%
- 75-110 yrs: 7.01%
- >110 yrs: 0.00%
- NULL or Missing: 0.00%
- Other: 0.14%

PCORnet Data Characterization (DC)
Standardized DC Output Model for Age Group
Race Frequency

<table>
<thead>
<tr>
<th>Race</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>88648115</td>
<td>62.78</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>247958</td>
<td>0.18</td>
</tr>
<tr>
<td>Asian</td>
<td>2937282</td>
<td>2.08</td>
</tr>
<tr>
<td>Black or African American</td>
<td>9249691</td>
<td>6.55</td>
</tr>
<tr>
<td>Native Hawaiian or Other Pacific Islander</td>
<td>592332</td>
<td>0.42</td>
</tr>
<tr>
<td>White</td>
<td>39538729</td>
<td>28</td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Race Distribution among Selected Data Partners

- Unknown: 62.78%
- American Indian or Alaska Native: 0.18%
- Asian: 2.08%
- Black or African American: 6.55%
- Native Hawaiian or Other Pacific Islander: 0.42%
- White: 28.00%
- Missing: 0.00%
### Sex Frequency

<table>
<thead>
<tr>
<th>Sex</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguous</td>
<td>312</td>
<td>0.04</td>
</tr>
<tr>
<td>Female</td>
<td>386649</td>
<td>44.62</td>
</tr>
<tr>
<td>Male</td>
<td>360638</td>
<td>43.80</td>
</tr>
<tr>
<td>No Information</td>
<td>35286</td>
<td>4.29</td>
</tr>
<tr>
<td>NULL or Missing</td>
<td>28379</td>
<td>3.20</td>
</tr>
<tr>
<td>Other</td>
<td>194</td>
<td>0.02</td>
</tr>
<tr>
<td>Unknown</td>
<td>6699</td>
<td>0.81</td>
</tr>
<tr>
<td>Values outside of CDM specifications</td>
<td>27304</td>
<td>3.32</td>
</tr>
<tr>
<td>Unselected Categories</td>
<td>0</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Sex Distribution among Selected Data Partners**

- **Ambiguous**: 0.04%
- **Female**: 44.52%
- **Male**: 43.80%
- **No Information**: 4.29%
- **NULL or Missing**: 3.20%
- **Other**: 0.02%
- **Unknown**: 0.81%
- **Values outside of CDM specifications**: 3.32%
- **Unselected Categories**: 0.00%
Weight
Rapid Data Assessment Tool Query

Weight Distribution among Selected Data Partners*

*Data represents distribution of weight among selected data partners. The chart shows the percentage of data partners falling within different weight ranges.
Height Frequency

<table>
<thead>
<tr>
<th>Height</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-10 in</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>11-20 in</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>21-45 in</td>
<td>11062</td>
<td>1.34</td>
</tr>
<tr>
<td>46-52 in</td>
<td>24449</td>
<td>2.97</td>
</tr>
<tr>
<td>53-58 in</td>
<td>78345</td>
<td>9.51</td>
</tr>
<tr>
<td>59-64 in</td>
<td>211150</td>
<td>25.64</td>
</tr>
<tr>
<td>65-70 in</td>
<td>245806</td>
<td>29.85</td>
</tr>
<tr>
<td>71-76 in</td>
<td>210697</td>
<td>25.56</td>
</tr>
<tr>
<td>77-82 in</td>
<td>13916</td>
<td>1.69</td>
</tr>
<tr>
<td>83-88 in</td>
<td>558</td>
<td>0.07</td>
</tr>
<tr>
<td>89-94 in</td>
<td>316</td>
<td>0.04</td>
</tr>
<tr>
<td>&lt;0 in</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>NULL or Missing</td>
<td>20737</td>
<td>2.52</td>
</tr>
<tr>
<td>95+ in</td>
<td>4</td>
<td>0.00</td>
</tr>
<tr>
<td>Other</td>
<td>6519</td>
<td>0.79</td>
</tr>
</tbody>
</table>
Rapid Data Assessment Tool

MULTIPLE DATAMART QUERY
## Sex Frequency Across DataMarts

<table>
<thead>
<tr>
<th>Sex</th>
<th>N</th>
<th>AAAA</th>
<th>N</th>
<th>BBBB</th>
<th>N</th>
<th>CCCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguous</td>
<td>40</td>
<td>0.03</td>
<td>78</td>
<td>0.06</td>
<td>11</td>
<td>0.01</td>
</tr>
<tr>
<td>Female</td>
<td>54561</td>
<td>43.11</td>
<td>60421</td>
<td>47.43</td>
<td>61224</td>
<td>46.29</td>
</tr>
<tr>
<td>Male</td>
<td>52861</td>
<td>41.77</td>
<td>58233</td>
<td>45.71</td>
<td>61412</td>
<td>46.43</td>
</tr>
<tr>
<td>No Information</td>
<td>260</td>
<td>0.21</td>
<td>1562</td>
<td>1.23</td>
<td>712</td>
<td>0.54</td>
</tr>
<tr>
<td>NULL or Missing</td>
<td>12145</td>
<td>9.6</td>
<td>1123</td>
<td>0.88</td>
<td>53</td>
<td>0.04</td>
</tr>
<tr>
<td>Other</td>
<td>25</td>
<td>0.02</td>
<td>12</td>
<td>0.01</td>
<td>124</td>
<td>0.09</td>
</tr>
<tr>
<td>Unknown</td>
<td>160</td>
<td>0.13</td>
<td>5609</td>
<td>4.4</td>
<td>841</td>
<td>0.64</td>
</tr>
<tr>
<td>Values outside of CDM specifications</td>
<td>6502</td>
<td>5.14</td>
<td>351</td>
<td>0.28</td>
<td>7891</td>
<td>5.97</td>
</tr>
<tr>
<td>Unselected Categories</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Sex Frequency Across DataMarts

Sex Distribution within AAAA

Sex Distribution within BBBB

Sex Distribution within CCCC
# Age Group Across DataMarts

<table>
<thead>
<tr>
<th>Age</th>
<th>N</th>
<th>AAAA</th>
<th>N</th>
<th>BBBB</th>
<th>N</th>
<th>CCCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0 yrs</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>0-1 yrs</td>
<td>1203</td>
<td>0.95</td>
<td>964</td>
<td>0.76</td>
<td>1123</td>
<td>0.86</td>
</tr>
<tr>
<td>2-4 yrs</td>
<td>2064</td>
<td>1.63</td>
<td>2531</td>
<td>1.99</td>
<td>1456</td>
<td>1.11</td>
</tr>
<tr>
<td>5-9 yrs</td>
<td>2237</td>
<td>1.77</td>
<td>2658</td>
<td>2.09</td>
<td>2389</td>
<td>1.82</td>
</tr>
<tr>
<td>10-14 yrs</td>
<td>3072</td>
<td>2.43</td>
<td>2987</td>
<td>2.35</td>
<td>2578</td>
<td>1.97</td>
</tr>
<tr>
<td>15-18 yrs</td>
<td>7748</td>
<td>6.13</td>
<td>7849</td>
<td>6.16</td>
<td>6998</td>
<td>5.35</td>
</tr>
<tr>
<td>19-21 yrs</td>
<td>12816</td>
<td>10.13</td>
<td>14523</td>
<td>11.4</td>
<td>13677</td>
<td>10.45</td>
</tr>
<tr>
<td>22-44 yrs</td>
<td>37892</td>
<td>29.96</td>
<td>36176</td>
<td>28.4</td>
<td>35236</td>
<td>26.91</td>
</tr>
<tr>
<td>45-64 yrs</td>
<td>26783</td>
<td>21.18</td>
<td>29535</td>
<td>23.19</td>
<td>31442</td>
<td>24.02</td>
</tr>
<tr>
<td>65-74 yrs</td>
<td>23564</td>
<td>18.63</td>
<td>22527</td>
<td>17.69</td>
<td>30941</td>
<td>23.63</td>
</tr>
<tr>
<td>75-110 yrs</td>
<td>8921</td>
<td>7.05</td>
<td>7411</td>
<td>5.82</td>
<td>5023</td>
<td>3.84</td>
</tr>
<tr>
<td>&gt;110+ yrs</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>NULL or Missing</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>Other</td>
<td>157</td>
<td>0.12</td>
<td>206</td>
<td>0.16</td>
<td>56</td>
<td>0.04</td>
</tr>
</tbody>
</table>
Age Group Across DataMarts

Age Distribution within AAAAA

Age Distribution within BBBB

Age Distribution within CCCC

www.amia.org
## Weight Group Across DataMarts

<table>
<thead>
<tr>
<th>Weight Range</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0 lbs</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0-1 lbs</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2-6 lbs</td>
<td>34</td>
<td>0.03</td>
<td>2</td>
<td>0.03</td>
<td>2</td>
<td>0.03</td>
</tr>
<tr>
<td>7-12 lbs</td>
<td>121</td>
<td>0.09</td>
<td>294</td>
<td>0.23</td>
<td>12</td>
<td>0.01</td>
</tr>
<tr>
<td>13-20 lbs</td>
<td>285</td>
<td>0.22</td>
<td>487</td>
<td>0.38</td>
<td>256</td>
<td>0.19</td>
</tr>
<tr>
<td>21-35 lbs</td>
<td>512</td>
<td>0.39</td>
<td>632</td>
<td>0.49</td>
<td>333</td>
<td>0.25</td>
</tr>
<tr>
<td>36-50 lbs</td>
<td>1254</td>
<td>0.95</td>
<td>768</td>
<td>0.6</td>
<td>1256</td>
<td>0.95</td>
</tr>
<tr>
<td>51-75 lbs</td>
<td>2771</td>
<td>2.11</td>
<td>1063</td>
<td>0.83</td>
<td>2893</td>
<td>2.19</td>
</tr>
<tr>
<td>76-100 lbs</td>
<td>3699</td>
<td>2.81</td>
<td>2741</td>
<td>2.14</td>
<td>11714</td>
<td>8.86</td>
</tr>
<tr>
<td>101-125 lbs</td>
<td>12186</td>
<td>9.27</td>
<td>14602</td>
<td>11.41</td>
<td>14502</td>
<td>10.96</td>
</tr>
<tr>
<td>126-150 lbs</td>
<td>13239</td>
<td>10.07</td>
<td>19369</td>
<td>15.14</td>
<td>14337</td>
<td>10.84</td>
</tr>
<tr>
<td>151-175 lbs</td>
<td>14997</td>
<td>11.41</td>
<td>27172</td>
<td>21.24</td>
<td>21836</td>
<td>16.51</td>
</tr>
<tr>
<td>176-200 lbs</td>
<td>21658</td>
<td>16.48</td>
<td>25132</td>
<td>19.64</td>
<td>21184</td>
<td>16.01</td>
</tr>
<tr>
<td>201-225 lbs</td>
<td>20762</td>
<td>16.81</td>
<td>25461</td>
<td>19.9</td>
<td>20069</td>
<td>15.17</td>
</tr>
<tr>
<td>226-250 lbs</td>
<td>12543</td>
<td>9.54</td>
<td>5623</td>
<td>4.39</td>
<td>14404</td>
<td>10.89</td>
</tr>
<tr>
<td>251-275 lbs</td>
<td>10356</td>
<td>7.88</td>
<td>2004</td>
<td>1.57</td>
<td>8506</td>
<td>6.43</td>
</tr>
<tr>
<td>276-300 lbs</td>
<td>9804</td>
<td>7.46</td>
<td>1134</td>
<td>0.89</td>
<td>30</td>
<td>0.02</td>
</tr>
<tr>
<td>301-350 lbs</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NULL or Missing</td>
<td>3252</td>
<td>2.47</td>
<td>999</td>
<td>0.78</td>
<td>441</td>
<td>0.33</td>
</tr>
<tr>
<td>350+ lbs</td>
<td>386</td>
<td>0.29</td>
<td>22</td>
<td>0.02</td>
<td>471</td>
<td>0.36</td>
</tr>
<tr>
<td>Other</td>
<td>3577</td>
<td>2.72</td>
<td>451</td>
<td>0.35</td>
<td>39</td>
<td>0.03</td>
</tr>
</tbody>
</table>
Weight Group Across DataMarts

Weight Distribution within AAAA

Weight Distribution within BBBB

Weight Distribution within CCCC
RDAT Development Next Steps

- Develop access controls for viewing DM, CDRN or aggregate network results
- Add terms such as smoking, dispensing, etc.
- Enhance visualizations
PCORnet Next Steps

• Inform PCORnet DataMart participation in observational research studies and clinical trials
• Rapid access to data characterization results for PTR
• Inform study design (e.g. missingness, ranges, variable population)
Questions

Thank you!

Jessica Sturtevant
jessica.lynn@post.harvard.edu
Acknowledgements

The PCORnet Distributed Research Network is a product of many data partners and collaborators. We gratefully acknowledge their work and individual contributions.

- Harvard Pilgrim Health Care Institute (HPHCI)
  - Jeffrey Brown
  - Kyle Erickson
  - Chayim Herzig-Marx
  - Casie Horgan
  - Lindsey Petro
  - Gabrielle Purcell
  - Jose Rodriguez

- Patient-Centered Outcomes Research Institute (PCORI)
  - Maryan Zirkle

- Duke Medicine
  - Lesley Curtis
  - Ed Hammond
  - Deb Harris
  - Jenny Ibarra
  - Michael Park
  - Laura Qualls
  - Shelley Rusincovitch
  - James Topping
Further Reading

• Data Quality/DRN

• PCORnet