Update from the Phenotypes, Data Standards, Data Quality Core of the NIH HCS Research Collaboratory

NIH Collaboratory Grand Rounds
August 26, 2016

Rachel Richesson, PhD, MPH
Assoc. Professor, Informatics
Duke University School of Nursing
Outline

- PSQ Core and Charter
- Background and Landscape
- Phenotype-related activities
- Standards approach
- Data Quality Assessment
- Impact of PSQ core
- Future directions
# Members of the Phenotype Core of the NIH Collaboratory:

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alan Bauck</td>
<td>Kaiser Permanente Center for Health Research</td>
</tr>
<tr>
<td>Denise Cifelli</td>
<td>U. Penn.</td>
</tr>
<tr>
<td>John Dickerson</td>
<td>Kaiser Permanente Northwest</td>
</tr>
<tr>
<td>Pedro Gozalo</td>
<td>Brown Univ. School of Public Health &amp; Providence VA Health Services Research Service</td>
</tr>
<tr>
<td>Bev Green</td>
<td>Group Health</td>
</tr>
<tr>
<td>Chris Helker</td>
<td>U. Penn</td>
</tr>
<tr>
<td>Beverly Kahn</td>
<td>Suffolk Univ., Boston</td>
</tr>
<tr>
<td>Michael Kahn</td>
<td>Children’s Hospital of Colorado</td>
</tr>
<tr>
<td>Reesa Laws</td>
<td>Kaiser Permanente Center for Health Research</td>
</tr>
<tr>
<td>Melissa Leventhal</td>
<td>University of Colorado Denver</td>
</tr>
<tr>
<td>John Lynch</td>
<td>Connecticut Institute for Primary Care Innovation</td>
</tr>
<tr>
<td>Meghan Mayhew</td>
<td>Kaiser Permanente Center for Health Research</td>
</tr>
<tr>
<td>Rosemary Madigan</td>
<td>U. Penn</td>
</tr>
<tr>
<td>Vincent Mor</td>
<td>Brown Univ. School of Public Health &amp; Providence VA Health Services Research Service</td>
</tr>
<tr>
<td>George “Holt” Oliver</td>
<td>Parkland Health and Hospital System (UT Southwestern)</td>
</tr>
<tr>
<td>Jon Puro</td>
<td>OCHIN</td>
</tr>
<tr>
<td>Jerry Sheehan</td>
<td>National Library of Medicine</td>
</tr>
<tr>
<td>Greg Simon</td>
<td>Group Health</td>
</tr>
<tr>
<td>Kari Stephens</td>
<td>U. of Washington</td>
</tr>
<tr>
<td>Erik Van Eaton</td>
<td>U. of Washington</td>
</tr>
</tbody>
</table>

**Duke members:** Rachel Richesson, Michelle Smerek, Ed Hammond, Monique Anderson
Charter – Phenotype, Data Standards, and Data Quality Core (PSQ Core)

- Share experiences using EHR to support research in various disease domains and for various purposes.
- Identify generalizable approaches and best practices to promote the consistent use of practical methods to use clinical data to advance healthcare research.
- Suggest where tools are needed.
- Explore and advocate for cultural and policy changes related to the use of EHRs for identifying populations for research, including measures of quality and sufficiency.
The Landscape

• Little standardized data representation in EHRs
• What appears standard is not always so
  • Multiple sources of ICD-9-CM codes, lab values, and medication data
  • Use of codes varies by institution
• Coding systems change
• No standard representation or approach for phenotype definitions
• Reproducibility is a concern
• Data reflect patient and clinician/organizational factors
• Data quality is a concern
Imperfection of Clinical Data

Model by George Hripcsak, Columbia University, New York, USA
Additional Challenges with Clinical Data from Multiple Healthcare Systems

Questions for PCT:
Are data from different sites comparable? Valid? Reliable?

Graphic courtesy of Alan Bauck, Kaiser Permanente Center for Health Research, 2011. (adapted)
Use of EHRs in Collaboratory PCTs

- PPACT needs to identify patients with chronic pain for the intervention. This is done in different EHR systems using a number of “phenotypes” for inclusion – e.g., neck pain, fibromyalgia, arthritis; long term opioid use.

- STOP CRC needs to continually identify screenings for colorectal cancer from each site, so must maintain master list of codes (CPT and local codes) related to fecal immunochemical test orders across multiple organizations.

- The TSOS trial needs to screen patients for PTSD on ED admission. How can different EHRs systems and patient data be leveraged to ensure consistency and efficiency of screening?
Use of EHRs in Collaboratory PCTs

- LIRE trail uses EHR data to identify cohorts (dynamically as radiology reports are produced), insertions based on rules in the EHR processing), and as primary source of outcome variables.

- The SPOT trial needs to identify possible suicide attempts (as study outcome measure) from different populations and information systems using a set of injury codes (in ICD-9-CM and ICD-10-CM).
Transparency and Reproducibility of PCTs

Multiple phenotype definitions:

Patient characteristics:

**Table 1. Patient Demographics and Baseline Characteristics**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Gentamicin-Colleague Sponge (n = 765)</th>
<th>Control (n = 749)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, median (IQR), y</td>
<td>64.2 (68.0-71.5)</td>
<td>64.9 (57.2-72.1)</td>
</tr>
<tr>
<td>White race</td>
<td>888 (91.2)</td>
<td>883 (91.2)</td>
</tr>
<tr>
<td>Weight, median (IQR), kg</td>
<td>98.0 (86.1-113.0)</td>
<td>98.8 (85.0-111.1)</td>
</tr>
<tr>
<td>Body mass index, median (IQR)</td>
<td>33.1 (30.2-37.2)</td>
<td>32.8 (30.0-36.2)</td>
</tr>
<tr>
<td>Body mass index &gt;30</td>
<td>574 (76.2)</td>
<td>563 (75.2)</td>
</tr>
<tr>
<td>Male sex</td>
<td>530 (70.4)</td>
<td>530 (70.5)</td>
</tr>
<tr>
<td>Medical history</td>
<td>659 (87.5)</td>
<td>659 (88.0)</td>
</tr>
<tr>
<td>History of hypertension</td>
<td>483 (65.5)</td>
<td>513 (68.5)</td>
</tr>
<tr>
<td>Current or history of smoking</td>
<td>406 (56.9)</td>
<td>450 (60.1)</td>
</tr>
<tr>
<td>Current smoking</td>
<td>136 (29.7)</td>
<td>123 (27.3)</td>
</tr>
<tr>
<td>History of chronic obstructive pulmonary disease</td>
<td>117 (15.5)</td>
<td>107 (14.3)</td>
</tr>
<tr>
<td>History of peripheral vascular disease</td>
<td>105 (13.9)</td>
<td>99 (11.9)</td>
</tr>
<tr>
<td>Previous median sternotomy</td>
<td>52 (6.5)</td>
<td>42 (5.6)</td>
</tr>
<tr>
<td>History of TIA or stroke</td>
<td>77 (10.2)</td>
<td>81 (10.8)</td>
</tr>
<tr>
<td>History of myocardial infarction</td>
<td>233 (31.9)</td>
<td>245 (32.7)</td>
</tr>
<tr>
<td>History of congestive heart failure</td>
<td>89 (11.8)</td>
<td>90 (12.0)</td>
</tr>
<tr>
<td>History of hyperlipidemia</td>
<td>619 (82.2)</td>
<td>607 (81.0)</td>
</tr>
<tr>
<td>Steroid use &lt;1 mo prior to surgery</td>
<td>28 (3.7)</td>
<td>33 (4.4)</td>
</tr>
<tr>
<td>Receiving diastole preoperatively</td>
<td>4 (0.5)</td>
<td>2 (0.3)</td>
</tr>
</tbody>
</table>

Preoperative diagnostic values

- Left ventricular ejection fraction, median (IQR), %: 55 (45-60) vs 55 (45-60)
- Serum glucose, median (IQR), mg/dL: 125 (101-130) vs 124 (103-126)
- Serum hemoglobin A1c, median (IQR), %: 6.6 (5.9-7.6) vs 6.6 (5.9-7.7)
- Hematocrit, median (IQR), %: 36 (34-42) vs 36 (34-42)
- Serum creatinine, median (IQR), mg/dL: 1.0 (0.9-1.3) vs 1.0 (0.9-1.2)
- Preoperative core temperature, median (IQR), °C: 97.6 (97.0-98.2) vs 97.7 (97.0-98.2)
- Preoperative hospital stay, median (IQR), d: 1.0 (0.0-3.0) vs 1.0 (0.0-3.0)
- Personnet risk score, median (IQR): 9.0 (6.0-14.5) vs 9.0 (6.0-16.0)

Abnormal Lab Results

**SUPREME-DM Phenotype**

- Definition: Adult Durham Population patients who meet ONE OR MORE of the following criteria during a DukeMed encounter between 2007-2011:
  - One or more instances of the specified ICD-9-CM diagnosis codes (see Table 7) on an inpatient encounter
  - OR 2 or more instances of the specified ICD-9-CM diagnosis codes (see Table 7) on outpatient encounter on separate days
  - OR 1 or more instances of active stand-alone medication (see Table 8) reported during outpatient medication reconciliation
  - OR 1 or more Oral Glucose Tolerance Test (OGTT) 2-hour 75g result >= 200 mg/dl where there is NO DIAGNOSIS CODE on the same encounter indicating pregnancy (V22, V23)
  - OR 2 or more hemoglobin A1c results >= 6.5% on 2 different days within 730 day span
  - OR 2 or more fasting glucose results >= 126 mg/dl on 2 different days within 730 day span
  - OR 2 or more random glucose results >= 200 mg on 2 different days within 730 day span
  - OR within a 730 days span on 2 different days:
    - Fasting glucose results >= 126 mg/dl
    - AND Random glucose results >= 200 mg
  - OR within a 750 day span can be same day
  - OR Hemoglobin A1c results >= 6.5%

**Abnormal HbA1c (NCV A1c Registry Definition)**

- Source: Hemoglobin laboratory results

- Definition: Adult Durham Population patients who meet ONE OR MORE of the following criteria during a DukeMed encounter between 2007-2011:
  - One or more instances of hemoglobin A1c results >= 6.5%
Reporting Pragmatic Clinical Trials

Introduction
Transparent reporting of clinical trials is essential for helping researchers, clinicians, patients, and other stakeholders understand the validity and reliability of the findings. Many have suggested that the quality of trial reporting is suboptimal and have sought consensus on the key elements of transparent reporting. To address this, a group of clinical trial methodologists and journal editors developed the CONSORT (Consolidated Standards of Reporting Trials) Statement. CONSORT is intended to improve transparency and dissemination of trial findings by providing a checklist and guidance for authors. The original CONSORT statement focused on the reporting of standard, two-group randomized controlled trials (RCTs) that compare an intervention with a control. Over the years, CONSORT has been expanded for clarity and revised, most recently in 2010, and now includes several official extensions to account for variations in trial design, interventions, and data (described in Appendix A).

Pragmatic Clinical Trials
The NIH Health Care Systems Research Collaboratory supports the design, execution, and dissemination of a set of Demonstration Projects, which are pragmatic clinical trials (PCTs) that address questions of major public health importance and are part of an effort to create a new infrastructure for collaborative research within healthcare systems. In contrast to RCTs, which elucidate a mechanical or biological process, PCTs are “designed for the primary purpose of informing decision makers regarding the comparative balance of benefits, burdens and risks of a biomedical or behavioral health intervention at the individual or population level.” To be clear, PCTs are on a continuum with traditional RCTs, and there are aspects of PCTs that make them either more exploratory or more pragmatic (described in Appendix B). Generally, a PCT is more pragmatic if the data are collected during routine clinical care (usually through the electronic health record); if there is some flexibility in the delivery of and adherence to the intervention; if a real-world population is included; and if the outcomes are relevant to patients and other decision makers.

Purpose of this Template
This template is intended to help authors with the transparent reporting of their PCT. Though we have looked to the CONSORT guidance and extensions wherever possible, new areas are emerging related to PCTs that the CONSORT checklist and guidance do not address. These include, for example, reporting around the secondary use of EHR data, wider stakeholder and health system involvement in the conduct of PCTs, and special ethical and regulatory considerations for PCTs.

(Will be posted to Living Text site soon...)

Prepared by: Coordinating Center Staff Science Writers
Reviewed by: Kevin Weinfurt, PhD
Specifications regarding data from EHRs or administrative systems

• “How the population of interest was identified. Researchers should explicitly reference any specific standards, data elements, or controlled vocabularies used, and provide details of strategies for translating across coding systems where applicable.”

• “Each clinical phenotype (EHR-based condition definition) used should be clearly defined and study reports should reference a location for readers to obtain the detailed definitional logic....The use of national repository for phenotype definitions, such as PheKB or NLM VSAC is preferred. GitHub or other repository for code...”

• “Process and results from assessment of the quality of the data (should be informed by Collaboratory PSQ Core recommendations for Data Quality)”

• “Data management activities during the study, including description of different data sources or processes used at different sites. (Note that the data quality assessment recommendations are particularly relevant to monitor data quality across sites that have different information systems and data management plans for the study.)”

• “The plan for archiving or sharing the data after the study, including specific definitions for clinical phenotypes and specifications for coding system (name and version) for any coded data.....”
Collaboratory Approach to Phenotype Definitions

**Definitions on Collaboratory website**

**Selection and planning**

Justification and guidance for use in Pragmatic Trials

**Implementation**

Human readable phenotype, collaboration, versioning, public dissemination

**Phenotype Definitions Used**

Type 2 Diabetes Mellitus Phenotype Definitions

**Populations**: Patients w/ chronic pain
Patients w/ imaging studies for lower back pain
Patients who are candidates for CRC screening

**Confounders or Risks**: Diabetes
Hypertension

**Outcomes**: Mortality
Suicide attempt
• Ideally, *research and clinical definitions should be semantically equivalent*.  
  i.e., they should identify equivalent populations.
Path to Re-Usable Phenotype Definitions

• Access
• Evaluate and compare
• Facilitate use and reporting
• Explore incentives

• Engage:
  • Research sponsors
  • SDOs
  • Policy makers
The CRAP License

The Community Research and Academic Programming License (CRAPL) is an academic-strength open source license by the well-known professor Matt Might. Its purpose is to encourage academic collaboration.

Terms:

• Any evidence of having been properly tested or verified is coincidental.

• You agree to hold the Author free from shame, embarrassment or ridicule for any hacks, kludges or leaps of faith found within the Program.

• You recognize that any request for support for the Program will be discarded with extreme prejudice.
Data Quality White Paper

- The use of population-level data is essential to explore, measure, and report “data quality” so that the results can be appropriately interpreted.

- Need adequate data and methods to detect the likely and genuine variation between populations at different trial sites and/or intervention groups.

- Recommend formal assessment of accuracy, completeness, and consistency for key data elements.

- Should be described, reported, and informed by workflows.

https://www.nihcollaboratory.org/Products/Assessing-data-quality_V1%200.pdf
Data Quality Recommendations: Use

- Have you read DQ recommendations and considered using?
  - 50% had read
  - 25% read upon contact for survey
  - 25% had not read/unknown
- Did you have DQ plans in place before you knew about the DQ recommendations?
  - 100% had DQA plans in place with application
- Have implemented or are in the process of implementing DQ recommendations?
  - 25% Yes
  - 75% NA or Have own plan
- Are you using a CDM?
  - 62.5% no
  - 25% yes Mini Sentinel, HMORN
  - 12.5% Project specific CDM
Data Quality Challenges

• Time-consuming
• Require population data (in addition to trial-specific data)
• Data retention requirements and related storage issues
• The cost of storage can be substantial
  • There are many storage options that impact cost, availability and completeness of data.
  • Medical record retention regulations are governed by state law and very widely in terms of retention time requirements and the amount of information.
Areas of Impact

• Technical Challenges
  • Methods, tools, best practices
  • Measuring quality
  • Quantification of differences across populations

• Culture changes
  • Can we identify and endorse “good enough”? 
  • Create culture of sharing and tools to support this
Dissemination

• “Living Textbook”

• Posters/presentations on Phenotype Template, and Methods for Development and Evaluation

• Manuscript (informatics journal) on EHR Phenotyping experience and strategies of Demonstration Projects
Future Plans

• Strategy for data standards
• ICD-9/10 (guidance for researchers)
• Cultural change/education/creativity regarding data quality
  • Getting specific about which quality dimensions are critical
  • Expecting data quality assessment
    • Comparison-based, i.e., data verification or reproducibility-based, i.e., multiple analyses on data from different sources
  • Using assessment results to answer how good is good enough?
    • Practicality versus perfection - how can we help draw some lines on the balance
• Integrate efforts and work products with other computable phenotyping initiatives (e.g., Big Data to Knowledge [BD2K], biosharing.org, CEDAR, Precision Medicine Initiative).
## Members of the Phenotype Core of the NIH Collaboratory:

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alan Bauck</td>
<td>Kaiser Permanente Center for Health Research</td>
</tr>
<tr>
<td>Denise Cifelli</td>
<td>U. Penn.</td>
</tr>
<tr>
<td>John Dickerson</td>
<td>Kaiser Permanente Center for Health Research, Kaiser Permanente Northwest</td>
</tr>
<tr>
<td>Pedro Gozalo</td>
<td>Brown Univ. School of Public Health &amp; Providence VA Health Services Research Service</td>
</tr>
<tr>
<td>Bev Green</td>
<td>Group Health</td>
</tr>
<tr>
<td>Chris Helker</td>
<td>U. Penn</td>
</tr>
<tr>
<td>Beverly Kahn</td>
<td>Suffolk Univ., Boston</td>
</tr>
<tr>
<td>Michael Kahn</td>
<td>Children’s Hospital of Colorado</td>
</tr>
<tr>
<td>Reesa Laws</td>
<td>Kaiser Permanente Center for Health Research</td>
</tr>
<tr>
<td>Melissa Leventhal</td>
<td>University of Colorado Denver</td>
</tr>
<tr>
<td>John Lynch</td>
<td>Connecticut Institute for Primary Care Innovation</td>
</tr>
<tr>
<td>Meghan Mayhew</td>
<td>Kaiser Permanente Center for Health Research</td>
</tr>
<tr>
<td>Rosemary Madigan</td>
<td>U. Penn</td>
</tr>
<tr>
<td>Vincent Mor</td>
<td>Brown Univ. School of Public Health &amp; Providence VA Health Services Research Service</td>
</tr>
<tr>
<td>George “Holt” Oliver</td>
<td>Parkland Health and Hospital System (UT Southwestern)</td>
</tr>
<tr>
<td>Jon Puro</td>
<td>OCHIN</td>
</tr>
<tr>
<td>Jerry Sheehan</td>
<td>National Library of Medicine</td>
</tr>
<tr>
<td>Greg Simon</td>
<td>Group Health</td>
</tr>
<tr>
<td>Kari Stephens</td>
<td>U. of Washington</td>
</tr>
<tr>
<td>Erik Van Eaton</td>
<td>U. of Washington</td>
</tr>
</tbody>
</table>

**Duke members:** Rachel Richesson, Michelle Smerek, Ed Hammond, Monique Anderson
Acknowledgements

The work presented here was funded by the National Institutes of Health (NIH) Common Fund, through a cooperative agreement (U54 AT007748) from the Office of Strategic Coordination within the Office of the NIH Director.