Standardized Representation for Electronic Health Record-Driven Phenotypes

April 8, 2014
AMIA Joint Summits for Translational Research

Rachel L. Richesson, PhD
Shelley A. Rusincovitch
Michelle M. Smerek
Jyotishman Pathak, PhD
Disclosure

• Rachel Richesson discloses that she has no relationships with commercial interests, ...
  .... nor do the other authors of this work
  Shelley Rusincovitch
  Michelle Smerek
  Jyotishman Pathak
Learning Objectives

After participating in this activity the learner should be better able to:

• Articulate the need for standardized representation of computable phenotypes
• Identify important metadata that should be included with computable phenotype definition
Outline

• Definitions
• Electronic Medical Records & Genomics Network (eMERGE)
• The NIH Healthcare Systems Collaboratory
  – Research Use Cases
• Criteria for designing computable phenotypes
  – User analysis
• Template requirements
• Discussion / Future
What is a Phenotype?

• Expression of genetic factors, influenced by environment

• Measurable biological (physiological, biochemical, and anatomical features), behavioral, or cognitive markers that are found more often in individuals with a disease than in the general population (MeSH definition)

• EHR Phenotyping – using data from EHRs to identify persons or populations with a condition or clinical profile. (“computable phenotype”)
  - ICD, CPT, labs, meds, vital signs, narrative notes
Different Definitions Yield Different Cohorts

A comparison of phenotype definitions for diabetes mellitus
The eMERGE Network

The mapping of the human genome has enabled new exploration of how genetic variations contribute to health and disease. To better realize this promise, researchers must now determine ways in which genetic make-up gives some individuals a greater chance of becoming sick with chronic conditions such as diabetes, Alzheimer’s, or heart disease. The goal of gaining this knowledge is to translate it to bedside practice and ultimately improve patient care.

The Electronic Medical Records and Genomics (eMERGE) Network is a national consortium organized by NHGRI to develop, disseminate, and apply approaches to research. It combines DNA biorepositories with electronic medical record (EMR) systems for large-scale, high-throughput genetic research with the ultimate goal of returning genomic testing results to patients in a clinical care setting. The Network is currently exploring more than a dozen phenotypes (with 13 additional electronic algorithms having already been published). Various models of returning clinical results have been implemented or planned for pilot at sites across the Network. Themes of bioinformatics, genomic medicine, privacy and community engagement are of particular relevance to eMERGE.

What makes eMERGE unique?

Each center participating in the Network is studying the relationship between genome-wide genetic variation and a common human trait. Such studies commonly involve testing hundreds of thousands of genetic variants called single nucleotide polymorphisms (SNPs) throughout the genome in people with and without the trait. A number of such studies are reporting an association between disease and a person’s genetic make-up, but those studies are typically costly and take a long time to complete.

The eMERGE model is exploring use of data from the EMR – clinical systems that represent an alternative methodology. Electronic medical records are one of the most exciting potential sources of data. We catalog and link EMR data to genetic samples obtained in the course of existing clinical research. The eMERGE model, there is no need to actively seek out cases and controls from residual tissue or blood samples. In the eMERGE model, there is no need to actively seek out cases and controls from residual tissue or blood samples. In the eMERGE model, there is no need to actively seek out cases and controls. Cases and controls are quickly and consistently identified from the EMR data, which is readily available. This approach is both cost-effective and time-efficient. More detailed information on phenotypes being explored in eMERGE can be found on the PheKB and other freely downloadable Resources page.

In addition, eMERGE focuses on ethical, legal, social, and policy issues such as privacy and...
<table>
<thead>
<tr>
<th>Title</th>
<th>Groups</th>
<th>Institutions</th>
<th>Data and Methods</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial Fibrillation - Demonstration Project</td>
<td>Vanderbilt - SD/RD Group</td>
<td>Vanderbilt University</td>
<td>CPT Codes, ICD 9 Codes, Natural Language Processing</td>
<td>Final</td>
</tr>
<tr>
<td>Cardiac Conduction (QRS)</td>
<td>eMERGE Phenotype WG</td>
<td>Vanderbilt University</td>
<td>CPT Codes, ICD 9 Codes, Laboratories, Medications, Natural Language Processing</td>
<td>Final</td>
</tr>
<tr>
<td>Cataracts</td>
<td>eMERGE Phenotype WG</td>
<td>Marshfield Clinic Research Foundation</td>
<td>CPT Codes, ICD 9 Codes, Medications, Natural Language Processing</td>
<td>Final</td>
</tr>
<tr>
<td>Clopidogrel Poor Metabolizers</td>
<td>Denny’s Group at Vandy, VESPA - Vanderbilt Electronic Systems for Pharmacogenomic Assessment</td>
<td></td>
<td>CPT Codes, ICD 9 Codes, Laboratories, Medications, Natural Language Processing</td>
<td>Final</td>
</tr>
<tr>
<td>Crohn’s Disease - Demonstration Project</td>
<td>Vanderbilt - SD/RD Group</td>
<td>Vanderbilt University</td>
<td>ICD 9 Codes, Medications, Natural Language Processing</td>
<td>Final</td>
</tr>
<tr>
<td>Dementia</td>
<td>eMERGE Phenotype WG</td>
<td>Group Health Cooperative</td>
<td>CPT Codes, ICD 9 Codes</td>
<td>Final</td>
</tr>
<tr>
<td>Diabetic Retinopathy</td>
<td>eMERGE Phenotype WG</td>
<td>Marshfield Clinic Research Foundation</td>
<td>CPT Codes, ICD 9 Codes, Medications, Natural Language Processing</td>
<td>Final</td>
</tr>
<tr>
<td>Drug Induced Liver Injury</td>
<td>eMERGE Phenotype WG</td>
<td>Columbia University</td>
<td>ICD 9 Codes, Laboratories, Medications, Natural Language Processing</td>
<td>Final</td>
</tr>
<tr>
<td>Rheumatoid Arthritis - Demonstration Project</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What is the Phenotype Portal?

Phenotyping is the process of identifying a cohort of patients based on certain diseases, symptoms or clinical findings. The Phenotype Portal is a tool funded by the SHARPn Project from the Office of the National Coordinator (ONC). It will enable clinicians and investigators to identify patient cohorts using electronic health record (EHR) data by leveraging informatics-based phenotyping processes. In turn, these cohorts will facilitate clinical trial enrollment, outcomes research, and inform clinical decision support. Currently, the field has various barriers in technological research and tool development, and Phenotype Portal is the first such platform for generating and executing Meaningful Use standards-based phenotyping algorithms that can be shared across multiple institutions and investigators.

Traditionally, a patient with diabetes is identified based on the symptom of abnormal blood glucose levels. However, understanding this disease involves more than just the presence of the symptom. Diabetes is a chronic disease that can be managed, and early detection can lead to better outcomes. The Phenotype Portal helps in identifying patients with poor control of diabetes, which is crucial for managing the disease effectively.

Diabetes: Hemoglobin A1c Poor Control

**Select an execution date range**

To: Dec 31 2012

**Diabetes: Hemoglobin A1c Poor Control**

- **Initial Patient Population**
  - AND: "Diagnosis: Active. Diabetes" starts before or during "Measurement Period"
  - AND: "Patient: Characteristic Birthdate: birth date" >= 16 year(s) starts before or during "Measurement Period"
  - AND: "Patient: Characteristic Birthdate: birth date" <= 75 year(s) starts before or during "Measurement Period"
  - OR:
    - "Encounter: Performed: Office Visit"
    - "Encounter: Performed: Face-to-Face Interaction"
    - "Encounter: Performed: Preventive Care Services - Established Office Visit, 18 and Up"
    - "Encounter: Performed: Preventive Care Services - Initial Office Visit, 18 and Up"
    - "Encounter: Performed: Preventive Care Services - Established Office Visit, 18 and Up"
    - "Encounter: Performed: Preventive Care Services - Initial Office Visit, 18 and Up"
    - "Encounter: Performed: Preventive Care Services - Established Office Visit, 18 and Up"
    - "Encounter: Performed: Preventive Care Services - Initial Office Visit, 18 and Up"
    - "Encounter: Performed: Annual Wellness Visit"
    - "During: "Measurement Period"

- **Denominator**
  - AND: Initial Patient Population

- **Denominator Exclusions**
  - AND OR: "Occurrence A of Diagnosis. Active: Gestational Diabetes" ends before or during "Measurement Period"
  - OR: "Occurrence A of Diagnosis. Active: Gestational Diabetes" starts before or during "Measurement Period"

- **Numerator**
  - AND:
    - OR: NOT: "Occurrence A of Laboratory Test. Result: HbA1c Laboratory Test" during "Measurement Period"
    - OR: "Occurrence A of Laboratory Test. Result: HbA1c Laboratory Test" during "Measurement Period"
    - OR: "Occurrence A of Laboratory Test. Result: HbA1c Laboratory Test (result > 8%)"

- **Denominator Exceptions**

**Data Criteria (QDM Data Elements)**

- **Description**
  - Encounter: Performed: Preventive Care Services - Established Office Visit, 18 and Up using Preventive Care Services - Established Office Visit, 18 and Up
  - Encounter: Performed: Preventive Care Services - Initial Office Visit, 18 and Up
  - Encounter: Performed: Home Healthcare Services using Home Healthcare Services Grouping Value Set
Upcoming Events

Grand Rounds March 7: Bray Patrick-Lake (CTTI; PCORnet Executive Committee member), Sue Sheridan (PCORI), and Sean Tunis (CMTIP)
Patient Engagement in Infrastructure Development

Secretory’s Advisory Committee for Human Research Protections (SACHRP): March 12-13

Grand Rounds March 14: TBD 
TBD

Subscribe to our mailing list:
nih-collaboratory@dm.duke.edu.

Knowledge Repository

View Collaboratory products, resources, publication references, etc.

Collaboratory News

Stop CRC featured on NPR Health Blog
02/26/2014: Gloria Coranado, PhD, was recently featured on the NPR Health Blog discussing the Stop CRC study.

Joe Selby writes perspective piece for the New England Journal of Medicine on PCORI
02/13/14: Joe Selby, MD, MPH, Executive Director of PCORI, published a perspective piece in the latest issue of the New England Journal of Medicine on lessons learned in PCORI’s 3-year history.

First patient enrolled in Collaboratory trial
01/13/14: The TIME Demonstration Project, led by University of Pennsylvania’s Laura Cember, MD, has enrolled its first patient.

PubMed Related Articles

<table>
<thead>
<tr>
<th>URL</th>
<th>Publication Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rescuing clinical trials in the United States and beyond: A call for action.</td>
<td>2013/06</td>
<td>To promote consensus around the solutions needed to address the adverse trends in clinical research, the Duke Clinical Research Institute convened stakeholders from academia, industry, and government. This article summarizes the proceedings.</td>
</tr>
<tr>
<td>Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise</td>
<td>2013/05</td>
<td>To produce more rapid, responsive, and relevant research, we propose approaches that increase relevance via greater stakeholder involvement, speed research via innovative designs, streamline review processes.</td>
</tr>
<tr>
<td>Human subjects protections in community-engaged research: a research ethics framework</td>
<td>2010/03</td>
<td>This new framework for exploring the risks in community-engaged research can help academic researchers and community partners ensure the mutual respect that community-engaged research requires.</td>
</tr>
</tbody>
</table>

Other Research Updates in the News ... >
<table>
<thead>
<tr>
<th>Institution/PI</th>
<th>Demo Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaiser Foundation Hospitals/ Gloria D. Coronado</td>
<td>Strategies and Opportunities to Stop Colon Cancer in Priority Populations</td>
</tr>
<tr>
<td>Kaiser Foundation Hospitals/ Lynn Larson DeBar</td>
<td>Collaborative Care for Chronic Pain in Primary Care</td>
</tr>
<tr>
<td>University of Pennsylvania/ Laura M. Dember</td>
<td>Pragmatic Trials in Maintenance Hemodialysis</td>
</tr>
<tr>
<td>University of California, Irvine/ Susan Huang</td>
<td>Decreasing Bioburden to Reduce Healthcare-Associated Infections and Readmissions</td>
</tr>
<tr>
<td>University of Washington/ Jeffrey Jarvik</td>
<td>A Pragmatic Trial of Lumbar Image Reporting with Epidemiology (LIRE)</td>
</tr>
<tr>
<td>University of Iowa/ Gary Rosenthal</td>
<td>Nighttime Dosing of Anti-Hypertensive Medications: A Pragmatic Clinical Trial</td>
</tr>
<tr>
<td>Group Health Cooperative/ Gregory E. Simon</td>
<td>Pragmatic trial of population-based programs to prevent suicide attempt</td>
</tr>
</tbody>
</table>
## Research Use Cases for Phenotypes

<table>
<thead>
<tr>
<th>Use case</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic association studies</td>
<td><em>PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations (Denny et al, 2010)</em></td>
</tr>
<tr>
<td>Identifying patients for prospective trials.</td>
<td><em>Identifying eligible study patients for the Nighttime Dosing of Anti-Hypertensive Medications Project.</em></td>
</tr>
<tr>
<td>Describing patient cohorts for analysis of existing data. (e.g., CER, health svvs research)</td>
<td><em>What are the effects of different anti-diabetic drugs on chronic kidney disease (CKD) outcomes in patients newly diagnosed with type 2 diabetes.</em></td>
</tr>
<tr>
<td>Presenting baseline characteristics or conditions to describe research populations for clinical trials.</td>
<td><em>What are the differences in intervention and control groups ... in a study on....?</em></td>
</tr>
</tbody>
</table>
## Research Use Cases – *cont’d.*

<table>
<thead>
<tr>
<th>Use case</th>
<th>Example research question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presenting primary outcomes to test the trial hypothesis.</td>
<td><em>Can interventions for persons at risk for suicidal behavior reduce the incidence of suicidal behavior?</em></td>
</tr>
<tr>
<td></td>
<td><em>Can a team-based program effectively help patients manage chronic pain?</em></td>
</tr>
<tr>
<td></td>
<td><em>Does the timing of administration of high blood pressure medications improve the effectiveness (i.e., reduce future events.)</em></td>
</tr>
<tr>
<td>The evaluation of supportive tools for providers that are embedded within EHR systems and clinical workflows.</td>
<td><em>Can customized provider-directed information... support the routine assessment for suicide risk?</em></td>
</tr>
<tr>
<td></td>
<td>.. support the handling of incidental findings from lumbar imaging studies?</td>
</tr>
</tbody>
</table>
Presenting Baseline Characteristics for Clinical Trial Reporting

**Multiple phenotype definitions:**

**SUPREME-DM Phenotype**

- Definition:
  - Adult Durham Population patients who meet **ONE OR MORE** of the following criteria during a DukeMed encounter between 2007-2011:
    - OR 1 or more instances of the specified ICD-9-CM diagnosis codes (see table 7) on an inpatient encounter
    - OR 1 or more instances of the specified ICD-9-CM diagnosis codes (see table 7) on outpatient encounters 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 days span
    - OR 2 or more fasting glucose results >= 126 mg/dl on 2 different days within 730 days span
    - OR 2 or more random glucose results >= 200 mg on 2 different days within 730 days 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 730 days span (can be same day):
      - Hemoglobin A1c results >= 6.5%
      - AND Fasting glucose results >= 126 mg/dl

**Abnormal Lab Results**

Source: Laboratory results

**Definition:**

- Adult Durham Population patients who meet **ONE OR MORE** of the following criteria during a DukeMed encounter between 2007-2011:
  - OR 1 or more instances of hemoglobin A1c results >= 6.5% OR 1 or more fasting glucose results >= 126 mg/dl within 365 days span OR 1 or more random glucose results >= 200 mg/dl within 365 days span

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

Source: Glucose hemoglobin laboratory results

**Definition:**

- Adult Durham Population patients who meet **ONE OR MORE** of the following criteria during a DukeMed encounter between 2007-2011:
  - OR 1 or more instances of hemoglobin A1c results >= 6.5%
“Standardized” Phenotype Definitions

• Explicit, consistent, and computable definitions can support:
  – development and conduct of new multi-site studies (interventional and observational)
  – comparability of EHR-derived data sets
  – comparison of study results and aggregation of evidence
  – reporting of data sets or results (e.g., ClinicalTrials.gov)
  – better practices for describing research populations in publication submissions to medical journals
Challenges for Research Applications

• Computable phenotype requirements are:
  – Condition-specific
  – Design-specific
  – Protocol-specific

• Timing of observations/measurements vs. inception of trial
• Fragmentation of care and incomplete data
• Data quality concerns
Desirable Features– URU*

• **Understandable**
  - Clearly defined data constructs
  - Clearly defined data source
  - Clearly defined purpose
  - Human readable (researchers and operations)

• **Reproducible**
  - Clearly defines the data elements and coding systems
  - Explicit specifications (~high quality documentation)
  - Computability and machine interpretation

• **Usable**
  - Accessibility and updates
  - Intellectual Property considerations
  - Specifications and implementation guidance

*URU coined by Keith Campbell, MD, PhD*
Desirable Features— “URU + U”

- Understandable
- Reproducible
- Usable

- Useful
  - Validation (results and methods)
  - Uses data elements and coding systems that are widely implemented
  - Community acceptance -- “Standardized” across sites or research communities

*URU coined by Keith Campbell, MD, PhD*
Important Metadata

(aka - things consumers look for)

• Feasibility
  – Encounter basis (inpatient, outpatient)
  – Data domains (e.g., diagnosis, medications) and sources (orders, claims)
  – Coding systems (e.g., ICD-9-CM, ICD-10-CM)
  – Multiple time points
  – Phenotyping modalities (structured database queries, NLP, optical character recognition, etc. )
  – Combination of structured and unstructured EMR data

• Appropriateness of phenotype definition
  – Intent of phenotype → taxonomy of research purposes
  – Discriminatory intent
  – Representational adequacy
Important Metadata (cont’d)

• Quality of phenotype definition
  – Developer
  – Reviewers (public vetting)
  – Performance metrics and validation
  – Applied in published studies, registries, etc.

• Disease characteristics
  – chronic, acute, transient

• State of diagnostics
  – Do quantitative measures and indicators of disease exist?

• Special considerations
  – Impact of incomplete data
  – Aggregate data to identify quality issues or differential coding practices at different institutions.
Related activity (approach)

• Desiderata (~good common sense criteria) for designing computable phenotype definitions
  – R01: NATIONAL INFRASTRUCTURE FOR STANDARDIZED AND PORTABLE EHR PHENOTYPING ALGORITHMS
  – PI’s: Jyoti Pathak (Mayo), Josh Denny (Vanderbilt) and Will Thompson (Northwestern)
  – Online surveys and user analysis on algorithm desiderata (i.e., minimum information model)
  – Focus groups in progress for gathering user requirements for algorithm creation and authoring

• More details: [http://informatics.mayo.edu/PheMA](http://informatics.mayo.edu/PheMA)
Things to think about…. For standardization

• Need to link to information models and emerging standards from EHRs and research networks
  – NQF Quality Data Model (QDM)
  – PCORnet
  – HMORN
  – OMAP
  – HL7 Domain Analysis Models / EHR Profiles

• Need to engage potential users and stakeholders to ensure uptake / future endorsement or adoption
Acknowledgments

• Thanks to the many members of the eMERGE and Collaboratory projects that made this possible.

• This presentation was made possible by NIH grants:
  – U54 AT007748-02 (The Collaboratory)
  – U01HG006379 (eMERGE)
  – R01GM105688 (PheMA)