What is a Computable Phenotype and why do I care?

Robert M Califf MD
June 27th, 2014
Underlying Assumptions

• The chasm is growing between the need for evidence to support health/healthcare decisions and the availability of that evidence
  • Technology advancing rapidly
  • More awareness of the need for evidence to avoid hurting people through not knowing the best choice
• The issue is not intellectual, it is operational and financial
• The only way to close this chasm is through disruptive change in at least 3 spheres:
  • Capture data in the context of care delivery rather than creating an expensive, parallel universe of redundant data collected separately from patient care
  • Embed research in clinical care to reduce expensive redundant research operations
  • Streamline regulatory oversight and research operations while protecting research participants and adhering to their preferences
How I spent Monday and Tuesday

• We have a new group of people who didn’t exist before the invention of cardiopulmonary bypass—adults with congenital heart disease (ACHD)
• Before 1970 or so, they died in childhood because of defective hearts
• They now live into adulthood, but no one knows what to expect
• There are 1.5 to 2 million of these people and the numbers are growing every day (congenital heart defects occur in 0.8% of the population)
  • One of them is my 36 yo daughter
• There are 20 major different types of malformations, most of which would meet criteria for “orphan disease”
• NHLBI hosted a meeting to discuss research priorities for this population, given the fact that very little research funding has addressed the needs of these people
ACHD Priorities

• The problem is that almost nothing is known beyond old fashioned experience of experts and small studies—these people didn’t exist before.
• What can be expected in terms of longevity and freedom from stroke, heart failure and arrhythmia?
• What are the causes and consequences of attention deficit issues and cognitive difficulties associated with ACHD and cardiopulmonary bypass?
• Do the same medicines work to treat and prevent heart failure in patients with ACHD as in those without ACHD?
• When is reoperation, transplant or mechanical assist device indicated?
• How should pregnancy be handled?
• The answer to all these questions is essentially “We don’t know, but we have a lot of smart, well intentioned clinicians getting by as best they can”
  • My old mentor: “There are doctors who wander the wards and doctors who are armed with data”
• Almost all studies are single-center and biased by the specific referral base of the reporting institution.
The Obvious Solution

• A disease registry spanning the 100 or so specialty centers dealing with these patients
• This would enable delineation of clinical epidemiology and quality systems
• Problem: this was recommended to NIH by a working group 10 years ago; it hasn’t happened
  • NIH says it can’t fund a registry for every disease
  • Registries fare poorly in peer review compared with hypothesis driven research
• RCTs hard to design without knowledge of clinical epidemiology to estimate event rates
• Who you gonna call?
  • PCORnet?
USING TRADITIONAL CLINICAL RESEARCH METHODS WILL DOOM ADULTS WITH CONGENITAL HEART DISEASE TO A LIFETIME OF WELL-INTENTIONED BUT UNINFORMED HEALTH CARE
What if...

• The NHLBI, its investigators and relevant advocacy groups (patients) had access to data from up to 100 million EHRs in 11 CDRNs with consent from the patients to participate in studies
• With computable phenotypes and a parsimonious data set the community (patients, families, providers, administrators and policy makers) would have access to:
  • Prevalence data
  • Clinical outcomes (death, stroke, heart failure, arrhythmia, etc.)
  • Operations and procedures
  • Medications
• Precious dollars could be reserved for specific analyses, ancillary detailed data collection and interventional trials
General Form of Clinical Studies

• What are the operating characteristics of test/marker/finding X for disease/condition/outcome Y?
• How well does test/marker/finding X predict that outcome in people with disease/condition/outcome Y?
• What is the balance of risk and benefit compared with alternatives for treatment or delivery approach X for patients with disease/condition/outcome Y?
• Basically, the investigators need to characterize the population at the inception point for the study, characterize the intervention(s) and to measure the key outcomes
Specific Questions about Coarctation of the Aorta

• What is the true prevalence in the adult population?
• What is the expected trajectory of survival, stroke, atherosclerotic events, aortic valve replacement, arrhythmia
  • For the whole population
  • Stratified by likely risk factors and comorbidities
• Why do people with coarctation of the aorta have hypertension and accelerated atherosclerosis even when the coarctation is repaired?
• When is reoperation indicated, since recurrent coarctation is common over time?
Creating a Research Ready Data System for the Network

Common Data Model with demographics, procedures, meds, diagnoses and common outcomes

Computable Phenotypes for ACHD diagnostic groups

A research ready national infrastructure for patient-centered clinical research
Creating a Data System for Deep, Specialized Research in the Network

- Common Data Model and Computable Phenotypes
- Detailed disease specific data

= A national infrastructure for patient-centered clinical research
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
Coarctation of the Aorta: Simple Computable Phenotype?

• ICD 9-- Q25.1
• ICD 10-- 747.10
• But....
  • Many of these people had repairs in childhood and now believe they are normal so they are not seeing specialists
  • Observation of ACHD specialists—many routine exams miss the scar on the chest or don’t ask why the scar is there
  • Coarctation associated with other congenital heart defects (bicuspid aortic valve for example) and other systemic risks
What Have we Learned about Computable Phenotypes from Common Diseases?
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 rigorous and 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. Each member site has EMR data linked to genetic samples obtained in the course of existing cohort studies from residual tissue or blood samples. In the eMERGE model, there is no need to actively recruit a study population. Cases and controls are quickly and consistently identified from the EMR database, 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>ICD 9 Codes, Medications</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>
</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.
Different Definitions Yield Different Cohorts

A comparison of phenotype definitions for diabetes mellitus
<table>
<thead>
<tr>
<th>Phenotype definitions:</th>
<th>Data domain criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD-9-CM 250.xx</td>
<td></td>
</tr>
<tr>
<td>CMS CCW</td>
<td></td>
</tr>
<tr>
<td>NYC A1c Registry</td>
<td></td>
</tr>
<tr>
<td>Diabetes-associated medications</td>
<td></td>
</tr>
<tr>
<td>DDC</td>
<td></td>
</tr>
<tr>
<td>SUPREME-DM</td>
<td></td>
</tr>
<tr>
<td>eMERGE†</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ICD-9-CM 250.xx and 250.x2 (excludes type 1 specific codes)</th>
<th>Expanded ICD-9-CM Codes (249.xx, 357.2, 362.0x, 366.41)</th>
<th>HbA1c</th>
<th>Fasting glucose</th>
<th>Random glucose</th>
<th>Abnormal OGTT</th>
<th>Diabetes-associated medications*</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD-9-CM 250.xx</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMS CCW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NYC A1c Registry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes-associated medications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DDC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUPREME-DM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eMERGE†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Medications vary by phenotype definition and are listed for each in the supplementary appendix (available online only).

†The eMERGE phenotype definition consists of five case scenarios with varying combinations of criteria. Any instance of type 1 specific codes (ie, 250.x1, 250.x3) results in the exclusion of the patient.

- = Sole criteria.
△ = Optional criteria, one of many.
☆ = Distinction made between inpatient and outpatient context.
\ = Distinction made for multiple instances and/or time points.


color=

A comparison of phenotype definitions for diabetes mellitus

Rachel L Richesson,1 Shelley A Rusincovitch,2 Douglas Wixted,3 Bryan C Batch,4 Mark N Feinglos,†4 Marie Lynn Miranda,3 W Ed Hammond,2,6 Robert M Califf,3,7 Susan E Spratt4
## Authoritative Sources of Phenotype Definitions (work in progress)

### Table 1: Primary Phenotype Sources

<table>
<thead>
<tr>
<th>Source</th>
<th>Comments</th>
<th>http</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Classifications Software (CCS), also known as AllHq Bundles</td>
<td>Only based upon diagnosis codes, but very large listing of conditions; this is the basis for most SEDI variables.</td>
<td><a href="http://www.hcup-us.ahrq.gov/toolssoftware/ccs.jsp">http://www.hcup-us.ahrq.gov/toolssoftware/ccs.jsp</a></td>
</tr>
<tr>
<td>CMS Chronic Conditions Warehouse (CCW)</td>
<td>Only based upon diagnosis codes and procedure codes; clinical review to date has felt that inclusion logic can be overly broad.</td>
<td><a href="http://www.ccwdata.org/web/guest/condition-categories">http://www.ccwdata.org/web/guest/condition-categories</a></td>
</tr>
<tr>
<td>Mini-Sentinal</td>
<td>Exhaustively researched definitions, but limited number of phenotypes represented.</td>
<td><a href="http://www.mini-sentinal.org/assessments/diagnoses_and_medical_procedures/default.aspx">http://www.mini-sentinal.org/assessments/diagnoses_and_medical_procedures/default.aspx</a></td>
</tr>
<tr>
<td>eMERGE Network and PhenxS phenotypes library</td>
<td>Probably the most well-recognized phenotyping source at present, but limited number of phenotypes represented; should be carefully evaluated because core mission of genomic studies can result in exclusionary logic inappropriate for the SEDI population health focus.</td>
<td><a href="http://www.phenx.org/phenotypes">http://www.phenx.org/phenotypes</a> <a href="http://www.ncbi.nlm.nih.gov/pubmed/21494599">http://www.ncbi.nlm.nih.gov/pubmed/21494599</a></td>
</tr>
</tbody>
</table>

### Table 2: Secondary Phenotype Sources

<table>
<thead>
<tr>
<th>Source</th>
<th>Comments</th>
<th>http</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint Commission</td>
<td>The CMS/Joint Summit QualityNet is generally the better source, not just using the Joint Commission directly.</td>
<td><a href="http://www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures.aspx">http://www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures.aspx</a></td>
</tr>
<tr>
<td>World Health Organization (WHO) Global Burden of Disease</td>
<td>In general, this may be useful for mental health, but probably not helpful for most clinical condition phenotypes.</td>
<td><a href="http://www.who.int/healthinfo/global_burden_disease/data_sources_methods/en/index.html">http://www.who.int/healthinfo/global_burden_disease/data_sources_methods/en/index.html</a></td>
</tr>
</tbody>
</table>

### Meaningful Use

This area needs further research. Does MU publish specific phenotypes for disease conditions? Most documentation appears related to attestation of technical capacities, especially in stage 1, not clinical definitions. | [http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/MeaningfulUse.html](http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/MeaningfulUse.html) |

---

**Presented by Shelley Rusincovitch at Collaboratory Grand Rounds, Nov. 2013.**
Challenges in Applying Computable Phenotypes in Practice

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

• Timing of observations/measurements vs. inception of study

• Fragmentation of care and incomplete data

• Data quality concerns

• This is not “push button research”—methods expertise and “sleeves rolled up” data curation is required
Important Metadata

• 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.
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 should 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
Presenting Baseline Characteristics for Clinical Study Reporting ("Table 1")

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:
- 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 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 different days within 730 day span
- OR 2 or more fasting glucose results >= 126 mg/dL on different days within 730 day span
- OR 2 or more random glucose results >= 200 mg on different days within 730 day span
- OR within a 750 day 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):
  - Hemoglobin A1c results >= 6.5%
  - OR within 200 mg/dL
- OR within 126 mg/dL

**Abnormal Lab 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%
- OR one or more fasting glucose results >= 126 mg/dL within 365 day span
- OR one or more random glucose results >= 200 mg/dL within 365 day span.

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

**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%
# 1. Table of Contents

1. Table of Contents .......................................................... 1
2. PCORnet Common Data Model (CDM): Overview and Development.................................................................................. 2
   2.1. License and Use ................................................................ 2
   2.2. Overview ...................................................................... 2
   2.3. Design and Modeling for the PCORnet CDM ...................... 3
   2.4. Development Notes ....................................................... 4
   2.5. Future State and Revision Process .................................... 5
   2.6. Comments on Protected Health Information (PHI) .......... 5
3. Individual Table Specifications .............................................. 6
   3.1. Table: DEMOGRAPHIC .................................................. 7
   3.2. Table: ENROLLMENT .................................................... 10
   3.3. Table: ENCOUNTER ...................................................... 12
   3.4. Table: DIAGNOSIS ...................................................... 17
   3.5. Table: PROCEDURE ..................................................... 19
   3.6. Table: VITAL .............................................................. 21
4. Glossary of Terms ................................................................ 23
5. History of Releases and Modifications ................................. 25
## CDRNs: disease Cohorts

<table>
<thead>
<tr>
<th>Organization</th>
<th>Common Disease Cohort</th>
<th>Rare Disease Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADVANCE</td>
<td>Diabetes</td>
<td>HIV &amp; hepatitis C virus co-infection</td>
</tr>
<tr>
<td>CAPriCORN</td>
<td>Anemia; asthma</td>
<td>Sickle cell disease; recurrent <em>C. difficile</em> colitis</td>
</tr>
<tr>
<td>Greater Plains Collaborative</td>
<td>Breast cancer</td>
<td>Amyotrophic lateral sclerosis</td>
</tr>
<tr>
<td>Louisiana Clinical Data Research Network</td>
<td>Diabetes</td>
<td>Sickle cell disease; rare cancers</td>
</tr>
<tr>
<td>NYC-CDRN</td>
<td>Diabetes</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>Mid-South CDRN</td>
<td>Coronary heart disease</td>
<td>Sickle cell disease</td>
</tr>
<tr>
<td>PEDSnet</td>
<td>Inflammatory bowel disease</td>
<td>Hypoplastic left heart syndrome</td>
</tr>
<tr>
<td>PORTAL</td>
<td>Colorectal cancer</td>
<td>Severe congenital heart disease</td>
</tr>
<tr>
<td>pSCANNER</td>
<td>Congestive heart failure</td>
<td>Kawasaki disease</td>
</tr>
<tr>
<td>PaTH</td>
<td>Atrial fibrillation</td>
<td>Idiopathic pulmonary fibrosis</td>
</tr>
<tr>
<td>SCIHLS</td>
<td>Osteoarthritis</td>
<td>Pulmonary arterial hypertension</td>
</tr>
</tbody>
</table>
PPRNs represent a number of conditions...

<table>
<thead>
<tr>
<th>Organization</th>
<th>Principal Investigator</th>
<th>Condition</th>
<th>Population Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accelerated Cure Project for Multiple Sclerosis</td>
<td>Robert McBurney</td>
<td>Multiple sclerosis</td>
<td>20,000</td>
</tr>
<tr>
<td>American Sleep Apnea Association</td>
<td>Susan Redline</td>
<td>Sleep apnea</td>
<td>50,000</td>
</tr>
<tr>
<td>Cincinnati Children's Hospital Medical Center</td>
<td>Peter Margolis</td>
<td>Pediatric Crohn's disease and ulcerative colitis</td>
<td>15,000</td>
</tr>
<tr>
<td>COPD Foundation</td>
<td>Richard Mularski</td>
<td>Chronic obstructive pulmonary disease</td>
<td>50,000</td>
</tr>
<tr>
<td>Crohn’s and Colitis Foundation of America</td>
<td>R. Balfour Sartor</td>
<td>Inflammatory bowel disease (Crohn’s disease and ulcerative colitis)</td>
<td>30,000</td>
</tr>
<tr>
<td>Global Healthy Living Foundation</td>
<td>Seth Ginsberg</td>
<td>Arthritis (rheumatoid arthritis; spondyloarthritis), musculoskeletal disorders (osteoporosis), and inflammatory conditions (psoriasis)</td>
<td>50,000</td>
</tr>
<tr>
<td>Massachusetts General Hospital</td>
<td>Andrew Nierenberg</td>
<td>Major depressive disorder and bipolar disorder</td>
<td>50,000</td>
</tr>
<tr>
<td>University of California, San Francisco</td>
<td>Mark Pletcher</td>
<td>Cardiovascular health</td>
<td>100,000</td>
</tr>
<tr>
<td>University of South Florida</td>
<td>Rebecca Sutphen</td>
<td>Hereditary breast &amp; ovarian cancer</td>
<td>17,000</td>
</tr>
</tbody>
</table>
....including rare diseases

<table>
<thead>
<tr>
<th>Organization</th>
<th>Principal Investigator</th>
<th>Condition</th>
<th>Population Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALD Connect, Inc.</td>
<td>Florian Eichler</td>
<td>Adrenoleukodystrophy</td>
<td>3,000</td>
</tr>
<tr>
<td>Arbor Research Collaborative for Health</td>
<td>Bruce Robinson</td>
<td>Primary nephrotic syndrome; focal segmental glomerulosclerosis; minimal change disease; and membranous nephropathy multiple sclerosis</td>
<td>1,250</td>
</tr>
<tr>
<td>Duke University</td>
<td>Laura Schanberg</td>
<td>Juvenile rheumatic disease</td>
<td>9,000</td>
</tr>
<tr>
<td>Epilepsy Foundation</td>
<td>Janice Beulow</td>
<td>Aicardi syndrome; Lennox-Gastaut syndrome; Phelan-McDermid syndrome; hypothalamic hamartoma; Dravet syndrome, tuberous sclerosis</td>
<td>1,500</td>
</tr>
<tr>
<td>Genetic Alliance, Inc.</td>
<td>Sharon Terry</td>
<td>Alström syndrome; dyskeratosis congenital; Gaucher disease; hepatitis; inflammatory breast cancer; Joubert syndrome; Klinefelter syndrome &amp; associated conditions; psoriasis; metachromatic leukodystrophy; pseudoxanthoma elasticum</td>
<td>50- 50,000</td>
</tr>
<tr>
<td>Immune Deficiency Foundation</td>
<td>Kathleen Sullivan</td>
<td>Primary immunodeficiency diseases</td>
<td>1,250</td>
</tr>
<tr>
<td>Parent Project Muscular Dystrophy</td>
<td>Holly Peay</td>
<td>Duchenne and Becker muscular dystrophy</td>
<td>4,000</td>
</tr>
<tr>
<td>Phelan-McDermid Syndrome Foundation</td>
<td>Megan O’Boyle</td>
<td>Phelan-McDermid syndrome</td>
<td>737</td>
</tr>
<tr>
<td>University of Pennsylvania</td>
<td>Peter Merkel</td>
<td>Vasculitis</td>
<td>500</td>
</tr>
<tr>
<td>Rare Diseases in PCORnet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>(n=45)</em></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adrenoleukodystrophy</th>
<th>Gaucher disease</th>
<th>Pediatric Ulcerative Colitis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aicardi Syndrome</td>
<td>Granulomatosis with Polyangiitis</td>
<td>Phelan-McDermid Syndrome</td>
</tr>
<tr>
<td>alpha-1 antitrypsin deficiency</td>
<td>Hypoplastic left heart syndrome</td>
<td>Primary Immunodeficiency Diseases</td>
</tr>
<tr>
<td>Alström syndrome</td>
<td>Hypothalamic Hamartoma</td>
<td>Primary Nephrotic Syndrome (Focal Segmental Glomerulosclerosis)</td>
</tr>
<tr>
<td>Amyotrophic Lateral Sclerosis</td>
<td>Inflammatory breast cancer (rare form of common disease)</td>
<td>Pseudoxanthoma elasticum</td>
</tr>
<tr>
<td>Becker muscular dystrophy</td>
<td>Joubert syndrome</td>
<td>Pulmonary artery hypertension</td>
</tr>
<tr>
<td>Chronic Granulomatous Disease</td>
<td>Juvenile Rheumatic Disease</td>
<td>Idiopathic pulmonary fibrosis</td>
</tr>
<tr>
<td>Churg-Strauss Syndrome</td>
<td>Kawasaki Disease</td>
<td>Rare Cancers</td>
</tr>
<tr>
<td>Co-infection with HIV and hepatitis C virus</td>
<td>Klinefelter syndrome and associated conditions</td>
<td>Selective IgA Deficiency</td>
</tr>
<tr>
<td>Common Variable Immunodeficiency</td>
<td>Lennox-Gastaut Syndrome</td>
<td>Severe Combined Immunodeficiency</td>
</tr>
<tr>
<td>Cystic fibrosis</td>
<td>Membranous Nephropathy [MN]</td>
<td>Severe Congenital Heart Disease</td>
</tr>
<tr>
<td>DiGeorge Syndrome</td>
<td>Metachromatic leukodystrophy</td>
<td>Sickle Cell Disease</td>
</tr>
<tr>
<td>Dravet Syndrome</td>
<td>Microscopic Polyangiitis</td>
<td>Recurrent C. Difficile</td>
</tr>
<tr>
<td>Duchenne muscular dystrophy</td>
<td>Minimal Change Disease</td>
<td>Tuberous Sclerosis</td>
</tr>
<tr>
<td>Dyskeratosis congenital</td>
<td>Pediatric Crohn's disease</td>
<td>X-Linked Agammaglobulinemia</td>
</tr>
</tbody>
</table>
Resources now on Collaboratory Website

Knowledge Repository

https://www.nihcollaboratory.org/Products/Forms/AllItems.aspx

Three phenotype definition recommendations (sex, race/ethnicity, and type 2 diabetes mellitus)
Phenotype literature search suggestions document

Living Textbook

“Electronic Health Records-Based Phenotyping” Topic Chapter:
http://sites.duke.edu/rethinkingclinicaltrials/ehr-phenotyping/

Phenotype recommendations from the Knowledge Repository are featured on the new “Tools for Research” page:
http://sites.duke.edu/rethinkingclinicaltrials/tools-for-research/

Page describing the Table 1 Project:
http://sites.duke.edu/rethinkingclinicaltrials/ehr-phenotyping/table-1-project/
Upcoming Events
Grand Rounds March 7: Bray Patrick-Lake (CTTI; PCORnet
Executive Committee member), Sue Sheridan (PCORI), and
Sean Tunis (CMTP)
Patient Engagement in Infrastructure Development
Secretary’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,
Type 2 Diabetes Mellitus Phenotype Definitions

From the NIH Collaboratory Phenotypes, Data Standards, and Data Quality Core

Available at: https://www.nihcollaboratory.org/Pages/Knowledge-Repository.aspx

Background: The Phenotypes, Data Standards, and Data Quality Core of the NIH Health Care Systems Research Collaboratory is developing a series of recommendations for the collection/query of data from electronic health records (EHRs) and/or ancillary systems for person characteristics and clinical features to support standardized reporting of baseline characteristics of research populations in interventional and observational studies.

Purpose of this document: This document represents our synthesis of existing phenotype definitions that have been used in diabetes research and population health activities. Using guidelines for the evaluation of existing phenotypes, our informatics and EHR phenotyping experience, and specialized clinical/research expertise, we suggest a suite of phenotype definitions, each appropriate for a particular purpose. The following is our recommendation, complete with a justification and supporting information and resources, for explicit EHR-derived phenotype definitions for diabetes. However, neither the Collaboratory nor the NIH has formally endorsed these definitions or their use in the data collection or reporting of this condition at this time (see disclaimer).

Audience: This document and supporting information is directed to clinical researchers and research sponsors who are making decisions about the data to use for studies. These documents should provide specifications and guidance that will assist researchers in making informed and deliberate choices about EHR data to use in research studies. The supporting information is intended to empower them to have conversations with operational data specialists at their institutions regarding the local implementation and use of these standard specifications. In addition, research sponsors can use these recommendations to proactively define data collection requirements for researchers.

Comments: We encourage comment, including updated information on formal validation or institutional experience with any of the referenced definitions, or suggestion/correction/clarification of our supporting information or interpretation. Please direct comments to: nih-collaboratory@dm.duke.edu.
Resources now on Collaboratory Website

Knowledge Repository

https://www.nihcollaboratory.org/Products/Forms/AllItems.aspx

Three phenotype definition recommendations (sex, race/ethnicity, and type 2 diabetes mellitus)
Phenotype literature search suggestions document

Living Textbook

“Electronic Health Records-Based Phenotyping” Topic Chapter:
http://sites.duke.edu/rethinkingclinicaltrials/ehr-phenotyping/

Phenotype recommendations from the Knowledge Repository are featured on the new “Tools for Research” page:
http://sites.duke.edu/rethinkingclinicaltrials/tools-for-research/

Page describing the Table 1 Project:
http://sites.duke.edu/rethinkingclinicaltrials/ehr-phenotyping/table-1-project/
Special Thanks

• Collaboratory Working Group(s) and PCORnet Task Force(s)
• Rachel Richesson (slides), Ed Hammond, Michelle Smerek, Meredith Zozus, Darcy Louzao, Jerry Sheehan, Leslie Curtis, Monique Anderson, Cindy Kluchar, Shelley Rusincovitch, Beverly Green, Reesa Laws, Alan Bauk, Greg Simon, Jennifer Robinson, Rosemary Madigan, Denise Cifelli, Chris Heckler, John Dickerson, Michael Kahn