NIH Collaboratory

Health Care Systems Research Collaboratory

pcornet

The National Patient-Centered Clinical Research Network

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

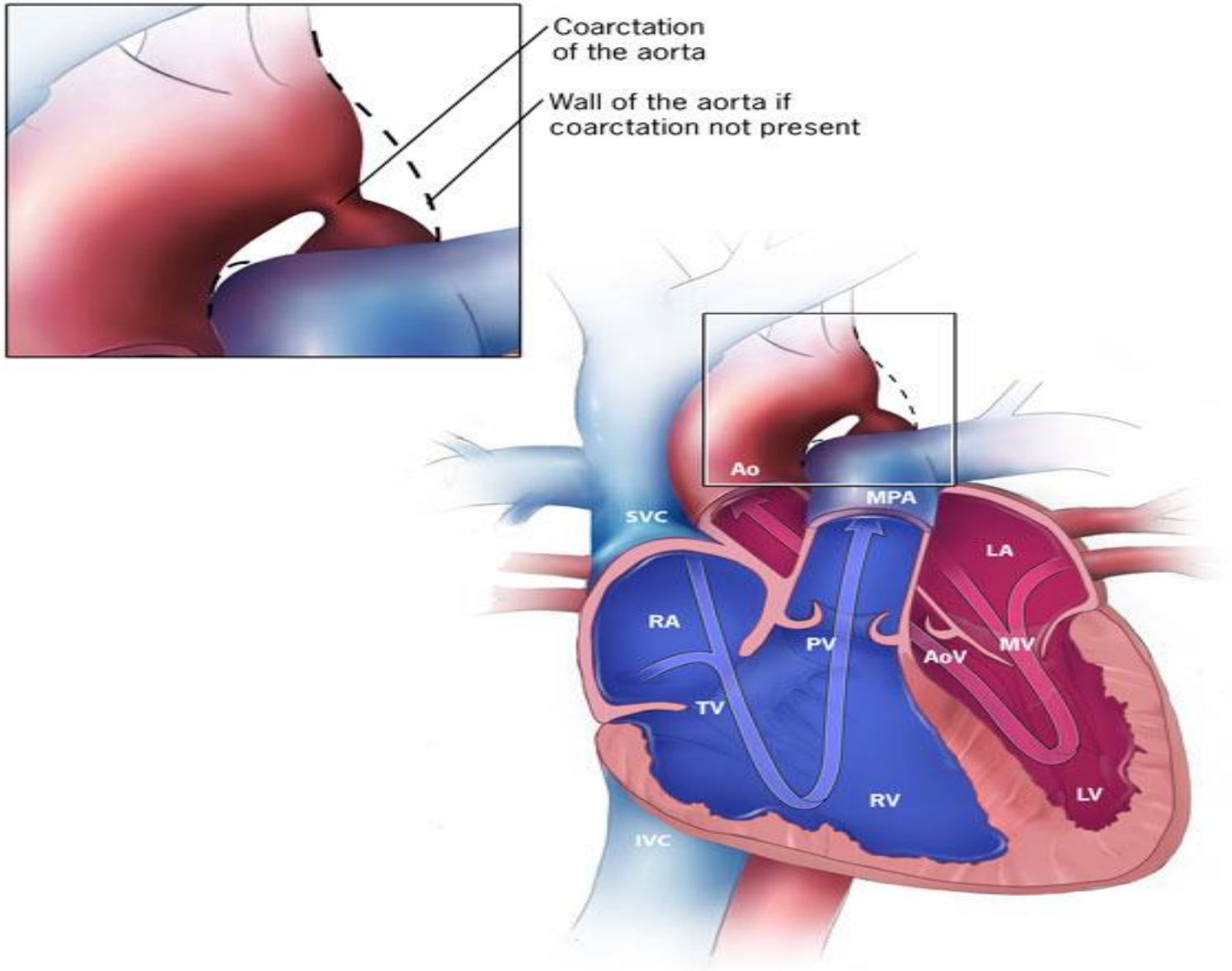
What if…

- 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

Coarctation of the Aorta



RA. Right Atrium
RV. Right Ventricle
LA. Left Atrium
LV. Left Ventricle

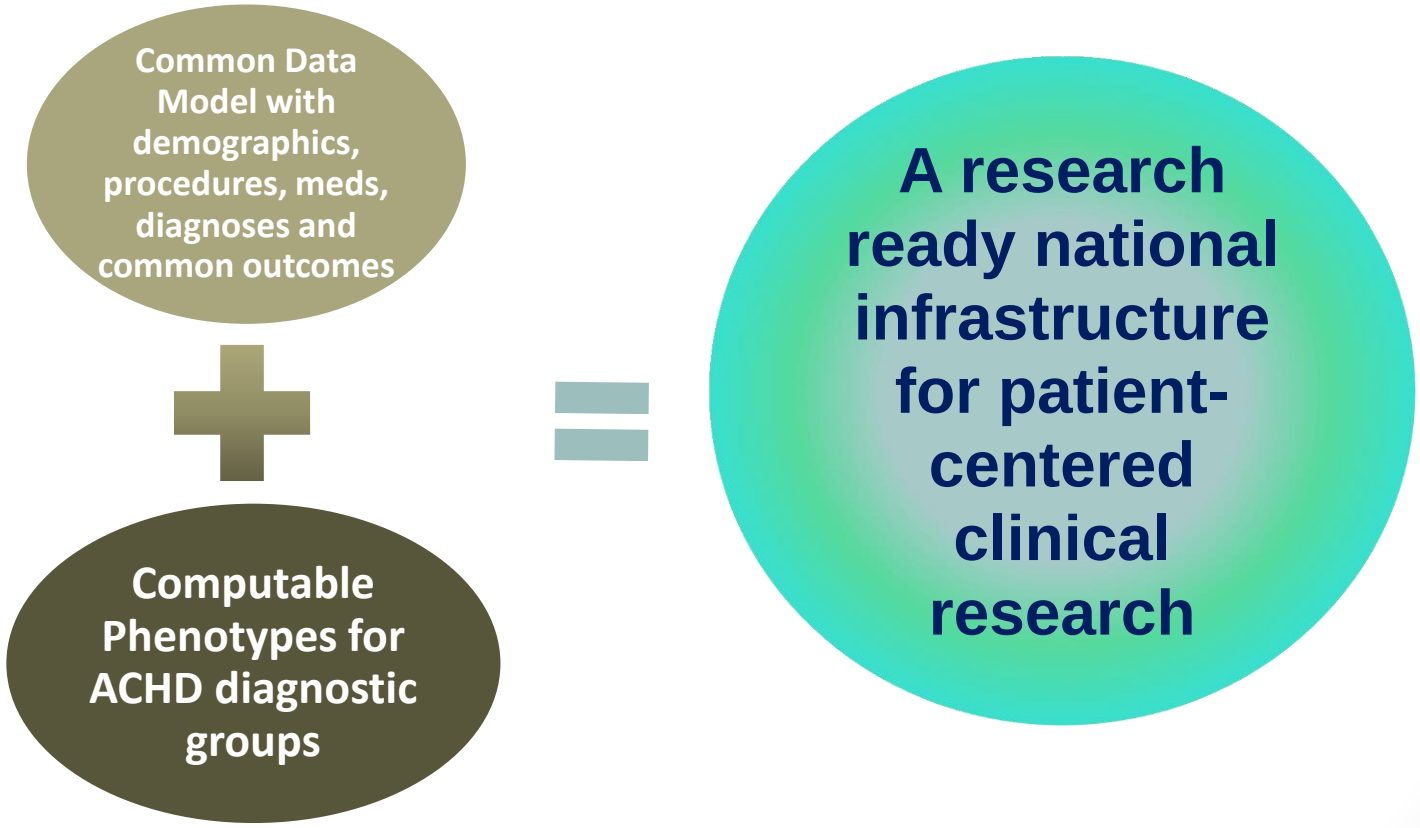
SVC. Superior Vena Cava
IVC. Inferior Vena Cava
MPA. Main Pulmonary Artery
Ao. Aorta

TV. Tricuspid Valve
MV. Mitral Valve
PV. Pulmonary Valve
AoV. Aortic Valve

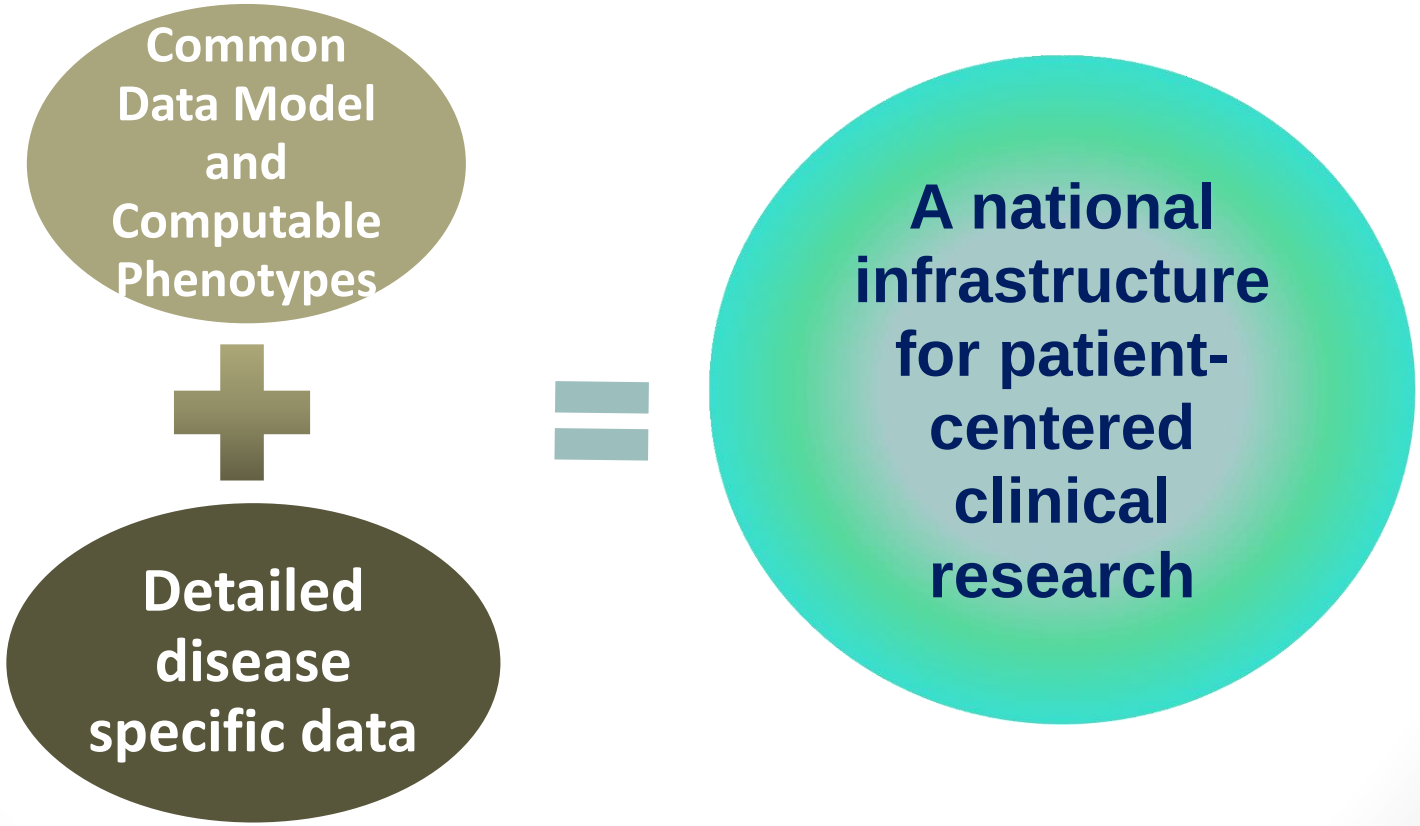
Specific Questions about Coarctation of the Aorta

- 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



Creating a Data System for Deep, Specialized Research in the Network



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 ro association between disease and a person's genetic make-up, but those studies are typically costly and take a long time to con

The eMERGE model is exploring use of data from the EMR – clinical systems that represer alternative methodology. Electronic medical records are one of the most exciting potentia member site has EMR data linked to genetic samples obtained in the course of existing co from residual tissue or blood samples. In the eMERGE model, there is no need to actively study population. Cases and controls are quickly and consistently identified from the EMR readily available. This approach is both cost-effective and time-efficient. More detailed inf phenotypes being explored in eMERGE can be found on on PheKB and other freely downl Resources page.

In addition, eMERGE focuses on ethical, legal, social, and policy issues such as privacy an

Phenotypes

Group - Any -

Include Methods ICD 10 Codes

Exclude Methods

Mine Only - Any -

Apply

Title	Groups	Institutions	Data and Methods	Status
Atrial Fibrillation - Demonstration Project	Vanderbilt - SD/RD Group	Vanderbilt University	CPT Codes, ICD 9 Codes, Natural Language Processing	Final
Cardiac Conduction (QRS)	eMERGE Phenotype WG	Vanderbilt University	CPT Codes, ICD 9 Codes, Laboratories, Medications, Natural Language Processing	Final
Cataracts	eMERGE Phenotype WG	Marshfield Clinic Research Foundation	CPT Codes, ICD 9 Codes, Medications, Natural Language Processing	Final
Clopidogrel Poor Metabolizers	Denny's Group at Vandy, VESPA - Vanderbilt Electronic Systems for Pharmacogenomic Assessment		CPT Codes, ICD 9 Codes, Laboratories, Medications, Natural Language Processing	Final
Crohn's Disease - Demonstration Project	Vanderbilt - SD/RD Group	Vanderbilt University	ICD 9 Codes, Medications, Natural Language Processing	Final
Dementia	eMERGE Phenotype WG	Group Health Cooperative	ICD 9 Codes, Medications	Final
Diabetic Retinopathy	eMERGE Phenotype WG	Marshfield Clinic Research Foundation	CPT Codes, ICD 9 Codes, Medications, Natural Language Processing	Final
Drug Induced Liver Injury	eMERGE Phenotype WG	Columbia University	ICD 9 Codes, Laboratories, Medications, Natural	Final

Most Recent Phenotypes

- Severe Early Childhood Obesity
- Warfarin dose/response
- Drug Induced Liver Injury
- Clopidogrel Poor Metabolizers
- Rheumatoid Arthritis - Demonstration Project

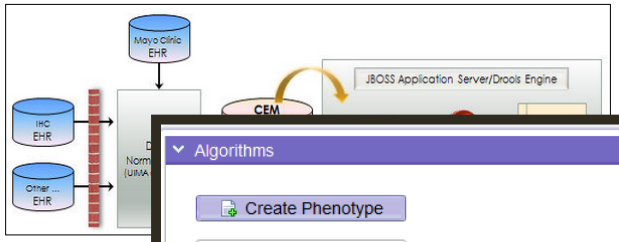
SHARPn.org Phenotype Portal

Log In | Register

Home Phenotypes Help

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 patien
towards creating a ur
government agencies

News and Updates

Date	News
July 10, 2013	NQF 2014 eMeasures have been uploaded.
September 24, 2013	QDM Phenotyping Translator in now integrated with the portal.
September 22, 2013	All of the Eligible Provider Clinical Quality Measures (CQMs)
December 10, 2012	Phenotype Portal now uses CTS2 value set service.
August 12, 2013	NQF 2014 beta translator now integrated with the portal.
June 17, 2012	Updates with new algorithms.
June 07, 2012	Release version 1.0 of Phenotype Portal.
June 01, 2012	Part of the Office of the National Coordinator for Health
June 01, 2012	We propose research that will generate a framework of

Recently Uploaded

Algorithms

Create Phenotype

Upload Phenotype

Phenotypes

- Disease of the skin and subcutaneous tissue
- Diseases of the blood and blood forming organs
- **Diseases of the circulatory system (8)**
- Diseases of the digestive system
- Diseases of the genitourinary system
- **Diseases of the musculoskeletal system (1)**
- **Diseases of the nervous system (5)**
- **Diseases of the respiratory system (5)**
- **Endocrine, nutritional and metabolic disease**
  - **Diseases of other endocrine glands (6)**
    - **Diabetes mellitus (6)**
      - *Diabetes: Eye Exam*
      - *Diabetes: Foot Exam*
      - *Diabetes: Hemoglobin A1c Poor Control*
      - *Diabetes: Low Density Lipoprotein (LDL)*
      - *Diabetes: Urine Protein Screening*
      - *Hemoglobin A1c Test for Pediatric Patier*
    - Diseases of thymus gland
    - Disorders of adrenal glands
    - Disorders of parathyroid gland
    - Disorders of the pituitary gland and its hypo
    - Other disorders of pancreatic internal secre
    - Other endocrine disorders
    - Ovarian dysfunction
    - Polyglandular dysfunction
    - Secondary diabetes mellitus
    - Testicular dysfunction
  - **Disorders of lipoid metabolism (2)**
  - Disorders of thyroid gland

Diabetes: Hemoglobin A1c Poor Control

Select an execution date range

From : Jan 1 2012

To : Dec 31 2012

Execute...

File Info | Criteria | Summary | Demographics

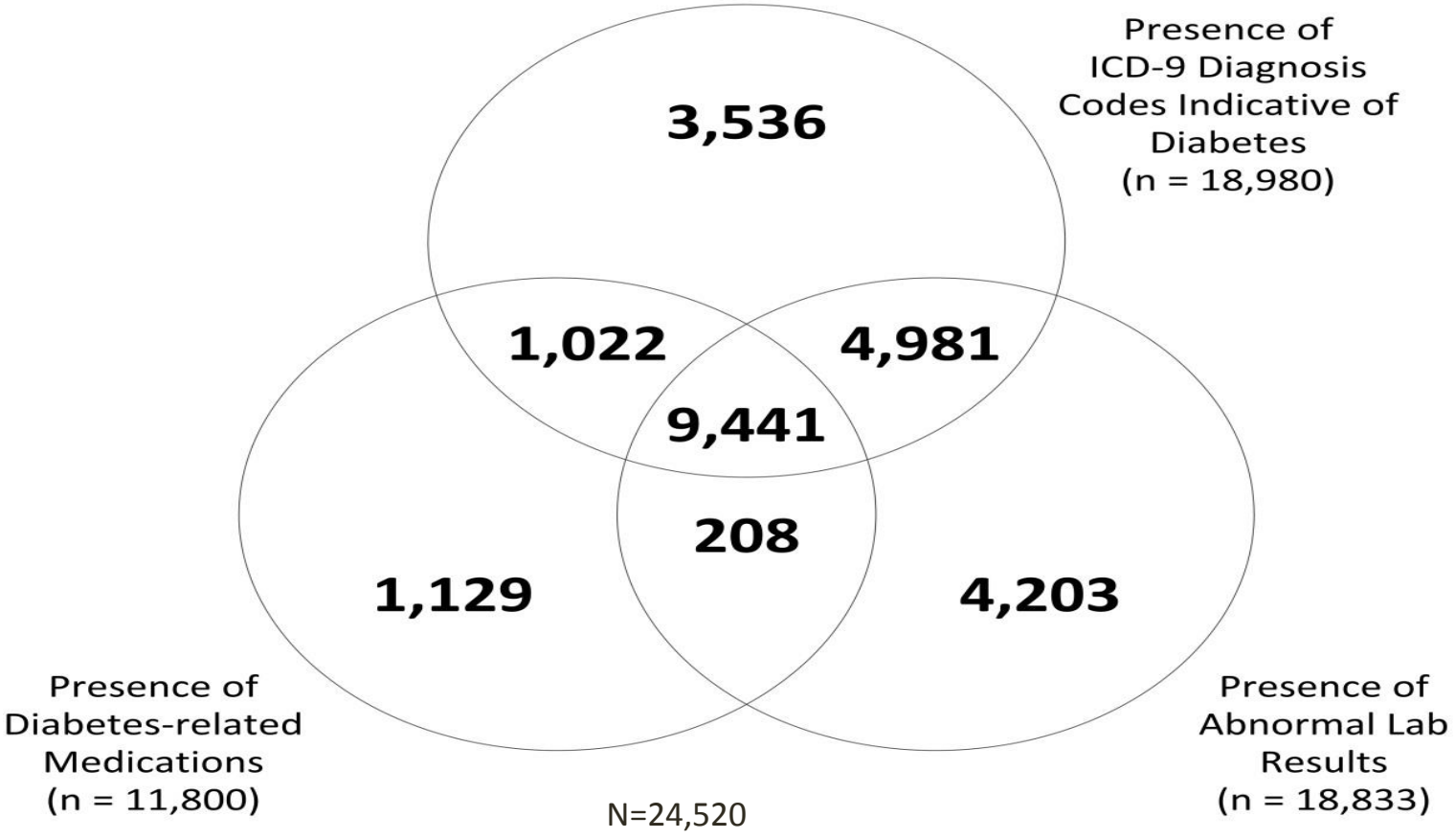
Diabetes: Hemoglobin A1c Poor Control

- **Initial Patient Population =**
  - AND: "Diagnosis, Active: Diabetes" starts before or during "Measurement Period"
  - AND: "Patient Characteristic Birthdate: birth date" >= 18 year(s) starts before start of "Measurement Period"
  - AND: "Patient Characteristic Birthdate: birth date" <= 75 year(s) starts before start of "Measurement Period"
  - AND:
    - OR: "Encounter, Performed: Office Visit"
    - OR: "Encounter, Performed: Face-to-Face Interaction"
    - OR: "Encounter, Performed: Preventive Care Services - Established Office Visit, 18 and Up"
    - OR: "Encounter, Performed: Preventive Care Services-Initial Office Visit, 18 and Up"
    - OR: "Encounter, Performed: Home Healthcare Services"
    - OR: "Encounter, Performed: Annual Wellness Visit"
    - during "Measurement Period"
- **Denominator =**
  - AND: "Initial Patient Population"
- **Denominator Exclusions =**
  - AND NOT: "Occurrence A of Diagnosis, Active: Gestational Diabetes" ends before start of "Measurement Period"
  - AND: "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:
      - AND: MOST RECENT:"Occurrence A of Laboratory Test, Result: HbA1c Laboratory Test" during "Measurement Period"
      - AND: "Occurrence A of Laboratory Test, Result: HbA1c Laboratory Test (result > 9 %)"
- **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 an
Encounter, Performed: Home Healthcare Services using Home Healthcare Services Grouping Value Set

Different Definitions Yield Different Cohorts

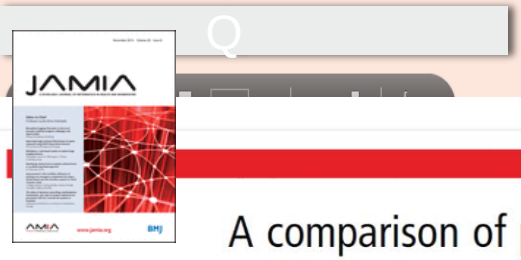


Research and applications

A comparison of phenotype definitions for diabetes mellitus

Phenotype definitions:	Data domain criteria							
	ICD-9-CM 250.xx	ICD-9-CM 250.x0 and 250.x2 (excludes type 1 specific codes)	Expanded ICD-9-CM Codes (249.xx, 357.2, 362.0x, 366.41)	HbA1c	Fasting glucose	Random glucose	Abnormal OGTT	Diabetes-associated medications*
ICD-9-CM 250.xx	●							
CMS CCW	▲ ✳ \\		▲ ✳ \\					
NYC A1c Registry				●				
Diabetes-associated medications								●
DDC		▲	▲	▲ \\	▲ \\	▲ \\	▲ \\	▲
SUPREME-DM	▲ ✳ \\		▲ ✳ \\	▲ \\	▲ \\	▲ \\	▲	▲
eMERGE†		● ✳ \\		▲	▲	▲		▲

*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.



Research and applications

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,[5] W Ed Hammond,[2,6] Robert M Califf,[3,7] Susan E Spratt[4]

Authoritative Sources of Phenotype Definitions (work in progress)

Table 1: Primary Phenotype Sources

Source	*Comments*
Clinical Classifications Software (CCS), also known as AHRQ Bundles	Only based upon diagnosis codes, but very large listing of conditions; this is the basis for most early SEDI variables. http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp
CMS Chronic Conditions Warehouse (CCW)	Only based upon diagnosis codes and procedure codes; clinical review to date has felt that inclusion logic can be overly broad. https://www.ccwdata.org/web/guest/condition-categories http://www.ncbi.nlm.nih.gov/pubmed/21649659
Mini-Sentinel	Exhaustively researched definitions, but limited number of phenotypes represented. http://www.mini-sentinel.org/assessments/diagnoses_and_medical_procedures/default.aspx
eMERGE Network and PheKB phenotypes library	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. http://www.phekb.org/phenotypes http://www.ncbi.nlm.nih.gov/pubmed/21269473
Quality Net (joint effort of CMS and Joint Commission)	Separates measures between inpatient basis and outpatient basis. Go to the "specifications manual" option; the appendixes contain specific listings of ICD-9 code tables, medication tables, and CPT codes. This is one of the only CPT code groupings that we've seen so far (CPT licensure is very restrictive), but QualityNet only includes for outpatient context. https://www.qualitynet.org/
National Drug File Reference Terminology (NDF-RT)	Search on a term (eg, diabetes) using "contains" and "name/code" specifiers. The results tab for "view all" contains the "may_treat" relationship of conditions to drugs. http://nciterms.nci.nih.gov/ncitbrowser/pages/vocabulary.jsf?dictionary=National%20Drug%20File%20-%20Reference%20Terminology
Professional society guidelines	These are an important source for definitions of abnormal laboratory results and specific ranges, which are often not represented in other definitions. Examples: American Diabetes Association, National Kidney Foundation, American College of Cardiology
Major and well-recognized clinical trials and registries using EHR data to identify cohorts	Clinical and expert guidance can be important for identification of these pivotal trials; another potential technique might be to limit results to high-impact journals via a PubMed search.

Table 2: Secondary Phenotype Sources

Source	*Comments*
Joint Commission	**The CMS/Joint Summit QualityNet is generally the better source, not using the Joint Commission directly.** This organization evaluates hospital adherence with federal regulations, and publishes a specifications manual for inpatient quality measures. Appendix A.1 lists the definitions for specific conditions, mostly based upon ICD-9. A limitation is that these definitions are centered on inpatient admissions, and may not be applicable in an outpatient setting. http://www.jointcommission.org/specifications_manual_for_national_hospital_inpatient_quality_measures.aspx
World Health Organization (WHO) Global Burden of Disease	**In general, this may be useful for mental health, but probably not helpful for most clinical condition phenotypes.** The Global Burden of Disease classifications include both ICD-9 and ICD-10 diagnosis code groupings. See "cause-specific documentation" for individual conditions (eg, cerebrovascular disease, diabetes mellitus, etc). The diagnosis codes are not granular (eg, it just lists 250 for diabetes mellitus), due to global application, and the clinical conditions are very broad. May be somewhat out of date; it appears that the classifications date back to 2000; the last formal GBD update appears to have been 2004, although this is difficult to ascertain from their website. However, there are a lot of mental health classifications, which may be useful. http://www.who.int/healthinfo/global_burden_disease/data_sources_methods/en/index.html
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/Meaningful_Use.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")

Patient characteristics:

Table 1. Patient Demographics and Baseline Characteristics

Characteristic	No. (%) of Patients[a] Gentamicin-Collagen Sponge (n = 753)	Control (n = 749)
Patient demographics		
Age, median (IQR), y	64.2 (58.0-71.5)	64.9 (57.2-72.1)
White race	688 (91.4)	683 (91.2)
Weight, median (IQR), kg	98.0 (86.1-113.0)	98.8 (85.0-111.1)
Body mass index, median (IQR)	33.1 (30.2-37.2)	32.8 (30.0-36.2)
Body mass index >30	574 (76.2)	563 (75.2)
Male sex	530 (70.4)	530 (70.8)
Medical history		
History of hypertension	659 (87.5)	659 (88.0)
History of diabetes	493 (65.5)	513 (68.5)
Current or history of smoking	458 (60.8)	450 (60.1)
Current smoking	136 (29.7)	123 (27.3)
History of chronic obstructive pulmonary disease	117 (15.5)	107 (14.3)
History of peripheral vascular disease	105 (13.9)	89 (11.9)
Previous median sternotomy	52 (6.9)	42 (5.6)
History of TIA or stroke	77 (10.2)	81 (10.8)
History of myocardial infarction	233 (31.0)	245 (32.7)
History of congestive heart failure	89 (11.8)	90 (12.0)
History of hyperlipidemia	619 (82.2)	607 (81.0)
Steroid use ≤1 mo prior to surgery	28 (3.7)	33 (4.4)
Receiving dialysis preoperatively	4 (0.5)	2 (0.3)
Preoperative diagnostic values		
Left ventricular ejection fraction, median (IQR), %	55 (45-60)	55 (45-60)
Serum glucose, median (IQR), mg/dL	125 (101-160)	124 (103-167)
Serum hemoglobin A_{1c}, median (IQR), %	6.5 (5.9-7.6)	6.6 (5.9-7.7)
Hematocrit, median (IQR), %	39 (36-42)	39 (36-42)
Serum creatinine, median (IQR), mg/dL	1.0 (0.9-1.3)	1.0 (0.9-1.2)
Preoperative core temperature, median (IQR), °C	97.6 (97.0-98.2)	97.7 (97.0-98.2)
Preoperative hospital stay, median (IQR), d	1.0 (0-3.0)	1.0 (0-3.0)
Parsonnet risk score, median (IQR)[b]	9.0 (6.0-14.5)	9.0 (6.0-16.0)

Abbreviations: IQR, interquartile range; TIA, transient ischemic attack.
SI conversion factors: To convert creatinine to µmol/L, multiply by 88.4; glucose to mmol/L, multiply by 0.0555.
[a] Unless otherwise indicated.
[b] Theoretical range is 0 to 148; 50% in Parsonnet et al[11] had a score between 0 and 9.

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[3]
- 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)[4]
- 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 day span on 2 different days:
  - Fasting glucose results >= 126 mg/dl
  - AND Random glucose results >= 200 mg
- OR within a 730 day 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:

- 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)

Source:
Glycated 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%

pcornet

Common Data Model (CDM) Specification, Version 1.0

Released by the Data Standards, Security and Network Infrastructure (DSSNI) Task Force on May 30, 2014

1. Table of Contents

CDRNs: disease Cohorts

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

PPRNs represent a number of conditions…

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

….including rare diseases

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

Rare Diseases in PCORnet

(n=45)

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

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/

NIH Collaboratory

Rethinking Clinical Trials™

Health Care Systems Research Collaboratory

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

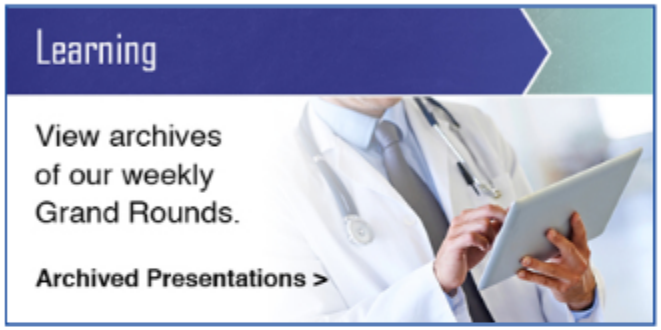
Secretary's Advisory Committee for Human Research Protections (SACHRP): March 12-13

Grand Rounds March 14: TBD
TBD

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

Knowledge Repository

View Collaboratory products, resources, publication references, etc.



Rethinking Clinical Trials™

Featured Topics

Articles, presentations, and other products related to specific topics of interest.

- Regulatory Update related to SUPPORT Trial
- Demonstration Projects - Regulatory and Ethics Discussions
- NIH Collaboratory Communication Channels Chart

NEW

Educational Presentations: Archives

02-21-14: Sharon Terry
Participant Engagement: Tools to Meet People Where They Are

02-14-14: Eric Larson
Engaging Health Systems in Research Partnerships

Collaboratory News

Stop CRC featured on NPR Health Blog
02/26/2014: Gloria Coronado, 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 the University of Pennsylvania's Laura Dember, MD, has enrolled its first patient.

Other Research Updates in the News ... >

PubMed Related Articles

URL	Publication Date	Description
Rescuing clinical trials in the United States and beyond: A call for action.	2013/06	To promote consensus around the solutions needed to address the adverse trends in clinical research, the Duke Clinical Research Institute convenedstakeholders from academia, industry, and government. This article summarizes the proceedings.
Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise	2013/05	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.
Human subjects protections in community-engaged research: a research ethics framework	2010/03	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.

NIH Collaboratory *Rethinking Clinical Trials™*
Health Care Systems Research Collaboratory

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