NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

# *Developing standards and quality metrics for clinical phenotyping using EHR data in pragmatic clinical trials*

**Rachel Richesson, PhD, MPH. FACMI**
Professor of Informatics & Learning Health Sciences
School of Medicine, University of Michigan

@rrichesson

# Housekeeping

- All participants will be muted
- Enter **all questions** in the Zoom **Q&A/chat box** and send to Everyone
- Moderator will review questions from chat box and ask them at the end
- Want to continue the discussion? Associated podcast released about 2 weeks after Grand Rounds
- Visit impactcollaboratory.org
- Follow us on Twitter & LinkedIN:

@IMPACTcollab1 https://www.linkedin.com/company/65346172

# Learning Objectives

Upon completion of this presentation, you should be able to:

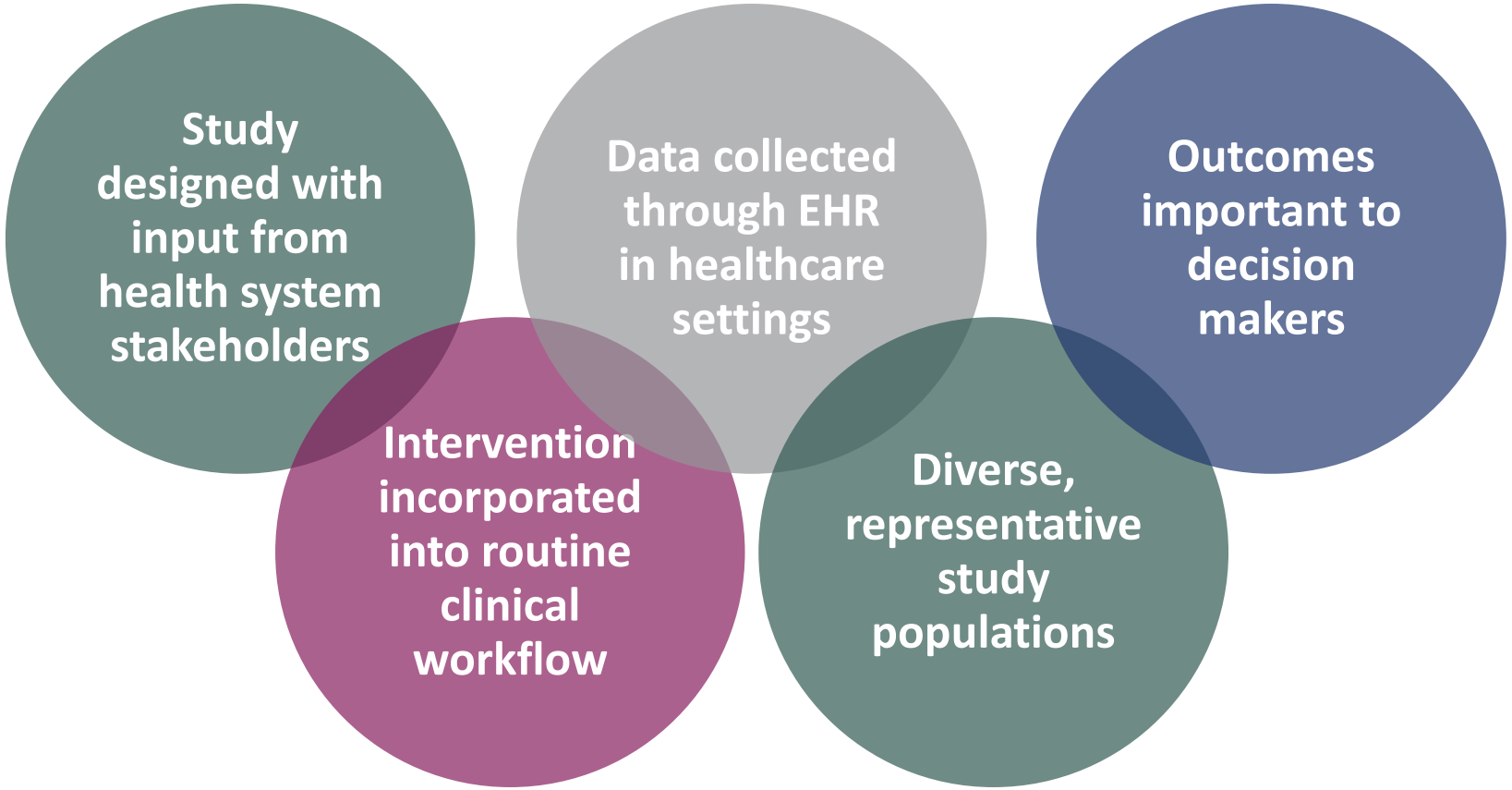
- Define computable phenotyping and discuss its relevance to pragmatic clinical trials
- Discuss approaches to find existing phenotypes and to assess their quality and appropriateness for certain uses
- Discuss the importance of reporting phenotype definition features and data quality assessment for pragmatic research

# OUTLINE

- NIH Pragmatics Trials Collaboratory and EHR experience
- Computable phenotypes and uses in pragmatic research
- Finding and assessing existing phenotypes for re-use
  - Challenges and Limitations

- The future and implications for the IMPACT Collaboratory

# Embedded PCTs Bridge Research & Clinical Care



# NIH Pragmatic Trials Collaboratory

**Initiated through the NIH Common Fund in 2012**

**Goal:** Strengthen the national capacity to implement cost-effective, large-scale research studies that engage healthcare delivery organizations as research partners


**Vision:** Support the design and execution of innovative pragmatic clinical trial Demonstration Projects to establish best practices and proof of concept

# Demonstration Projects

- Pragmatic trials embedded in healthcare
  systems to address questions of major public health importance
- Projects span multiple NIH Institutes, Centers, and Offices
- Projects have 1-year planning phase followed by implementation phase
- Coordinating Center supports methods-focused cores

# Cores

- Biostats and Study Design
- **Electronic Health Records**
- Ethics and Regulatory
- Health care Systems Interactions
- Patient Centered Outcomes
- Health Equity
- Implementation Science

# Use of EHR

| Trial | Eligibility Determination | Intervention Delivery | Outcome Assessment |
|---|---|---|---|
| ACP PEACE | ■ | | ■ |
| ABATE Infection | ■ | ■ | ■ |
| BackInAction | ■ | | ■ |
| BeatPain Utah | ■ | | ■ |
| EMBED | ■ | ■ | ■ |
| FM-TIPS | | | ■ |
| GGC4H | ■ | | ■ |
| GRACE | | | ■ |
| Hi-Lo | ■ | ■ | ■ |
| ICD-Pieces | ■ | ■ | ■ |
| IMPACt-LBP | | | ■ |
| LIRE | ■ | ■ | ■ |
| NOHARM | ■ | ■ | ■ |
| Nudge | ■ | | ■ |
| OPTIMUM | ■ | | ■ |
| PPACT | ■ | ■ | ■ |
| PRIM-ER | | ■ | |
| PROVEN | | | ■ |
| SPOT | ■ | ■ | ■ |
| STOP CRC | ■ | ■ | ■ |
| TiME | ■ | ■ | ■ |
| TSOS | ■ | ■ | ■ |

# No two EHRs are alike

- Even when sites are part of a single corporate entity, local coding varies
  - Cross-site data standardization is essential
  - Solution requires engagement of local data experts and time
- More sites = more work

Table 3. Examples of variations in original result units in the Mini-Sentinel laboratory results table source data

| | | | |
|---|---|---|---|
| *Glycosylated hemoglobin (HbA1c) original result units** | | | |
| % | %T.HGB | % TL HGB | % HGB |
| HEMOGLOBIN | %T.Hgb | % OF TOTAL | PERCENT |
| U | %T.Hgb | % of Hgb | Percent |
| %HB | % NGSP | % of total | HbA1c% |
| % OF T | %NGSP | %THb | %HbA1c |
| %AIC | % TOTAL HGB | %NGSP | % A1C |
| MG/DL | G/DL | mmol/mol† | Blank |
| % A1C | % A1c | %Hb | g/dL |
| NULL | %THb | | |
| *Platelet count original result units‡* | | | |
| Blank | FL | TH/UL | X10(3) |
| % | K/CMM | THOU/CMM | 1000/UL |
| /100 W | k/cmm | thou/cmm | X10(3)/MCL |
| /CMM | K/CU MM | thou/mm3 | X10(3)/UL |
| CMM | K/CUMM | THOU/UL | X10(6)/MCL |
| 10 3 L | K/MCL | THOUS/CU.MM | X10*9/L |
| 10X3UL | K/mcL | THOUS/MCL | X10E3/UL |
| 10^3/UL | K/UL | THOU/mcL | X1000 |
| 10*3/uL | k/uL | THOUS/UL | X10X3 |
| 10?3/uL | KU/L | Thou/uL | X10^3/UL |
| 10E3/uL | K/MM3 | THOUSA | x10 |
| 10e3/uL | K/mm3 | THOUSAND | X10?3/ul |
| 10e9/L | LB | THOUSAND/UL | X10E3/UL |
| E9/L | PLATELET CO | U | X10E3 |
| BIL/L | T/CMM | X 10-3/UL | K/A?L |
| bil/L | TH/MM3 | X 10(3)/UL | K/B5L |
| CU MM | th/mm3 | X10 3 | |

*Raebel et al, PDS 2014*

# No two EHRs are alike

- Tools do not transfer from one site to another (SPOT)
  - Local adaptation is necessary
  - Solution requires scarce IT resources and time


*Source: G Simon*

# The EHR is optimized for billing

- Integrating study-related data elements into the EHR has implications for clinical workflow and compliance
  - Pilot tests are critical
- Even minor modifications require allocation of scarce IT resources and leadership buy-in
- Engaging health care system decision-makers and EHR vendors is essential

PROFESSIONAL

# New ICD-10-CM Codes for Neurocognitive Disorders Effective October 1

*MICHAEL FIRST, M.D.*

Published Online: 24 Aug 2022 | https://doi.org/10.1176/appi.pn.2022.10.10.34

New ICD-10-CM Codes for Neurocognitive Disorders Effective October 1 | Psychiatric News

04:54 1x

*The coding changes for major and mild neurocognitive disorders represent the most consequential coding changes for DSM-5 disorders since the October 1, 2015, changeover from ICD-9-CM to ICD-10-CM.*

Every October 1, the ICD-10-CM codes for all of medicine are updated, resulting in the addition of new codes and the revision or deletion of existing codes. Only a small fraction of the 68,000 codes are actually affected; last year, 159 new codes were added, 25 codes were deleted, and 27 existing codes were revised. Given that all HIPAA-compliant health care entities are required to use the most up-to-date ICD-10-CM codes, clinicians and institutions need to keep on top of these coding changes, especially since the addition of new codes usually results in some existing codes becoming obsolete.

This year the coding changes are largely confined to major and mild neurocognitive disorders, but they represent the most consequential coding changes for *DSM-5* disorders since the October 1, 2015, changeover from ICD-9-CM to ICD-10-CM.

## Changes for Major Neurocognitive Disorder

The first three characters that make up the ICD-10-CM code for major neurocognitive disorder depend on the type of etiological medical condition and are unchanged:

https://psychnews.psychiatryonline.org/doi/10.1176/appi.pn.2022.10.10.34

## Table 2. ICD-10-CM codes for Major Neurocognitive Disorder

**F01 Major Vascular NCD**

*Codes sunsetted on September 30*

- **F01.50** Major vascular NCD, without behavioral disturbance
- **F01.51** Major vascular NCD, with behavioral disturbance

*Updated codes effective October 1*

- **F01.xy** Major vascular NCD

x=current severity, y=accompanying behavioral or psychological disturbance

- **F01.Ay** Major vascular NCD, mild...
- **F01.By** Major vascular NCD, moderate...
- **F01.Cy** Major vascular NCD, severe...

**.x11** ...with agitation
**.x2** ...with psychotic disturbance
**.x3** ... with mood symptoms
**.x4** ...with anxiety
**.x18** ...with other behavioral or psychological disturbance
**.x0** ...without accompanying behavioral or symptomatic disturbance

**F02 Major NCD due to another medical condition**

*Codes sunsetted on September 30*

- **F02.80** Major NCD due to AMC, without behavioral disturbance
- **F02.81** Major NCD due to AMC, with behavioral disturbance

*Updated codes effective October 1*

- **F02.xy** Major NCD due to [name of another medical condition]

x=current severity, y=accompanying behavioral or psychological disturbance

- **F02.Ay** Major NCD due to AMC , mild...
- **F02.By** Major NCD due to AMC, moderate...
- **F02.Cy** Major NCD due to AMC, severe...

**.x11** ...with agitation
**.x2** ...with psychotic disturbance
**.x3** ... with mood symptoms
**.x4** ...with anxiety
**.x18** ...with other behavioral or psychological disturbance
**.x0** ...without accompanying behavioral or symptomatic disturbance

**F03 Major NCD due to unknown etiology**

*Updated codes effective October 1*

*(Note: R41.9 will continue to apply to Unspecified Neurocognitive Disorder)*

- **F03.xy** Major NCD due to unknown etiology

x=current severity, y=accompanying behavioral or psychological disturbance

- **F03.Ay** Major NCD due unknown etiology, mild...
- **F03.By** Major NCD due to unknown etiology, moderate...
- **F03.Cy** Major NCD due to unknown etiology, severe...

**.x11** ...with agitation
**.x2** ...with psychotic disturbance
**.x3** ... with mood symptoms
**.x4** ...with anxiety
**.x18** ...with other behavioral or psychological disturbance
**.x0** ...without accompanying behavioral or symptomatic disturbance

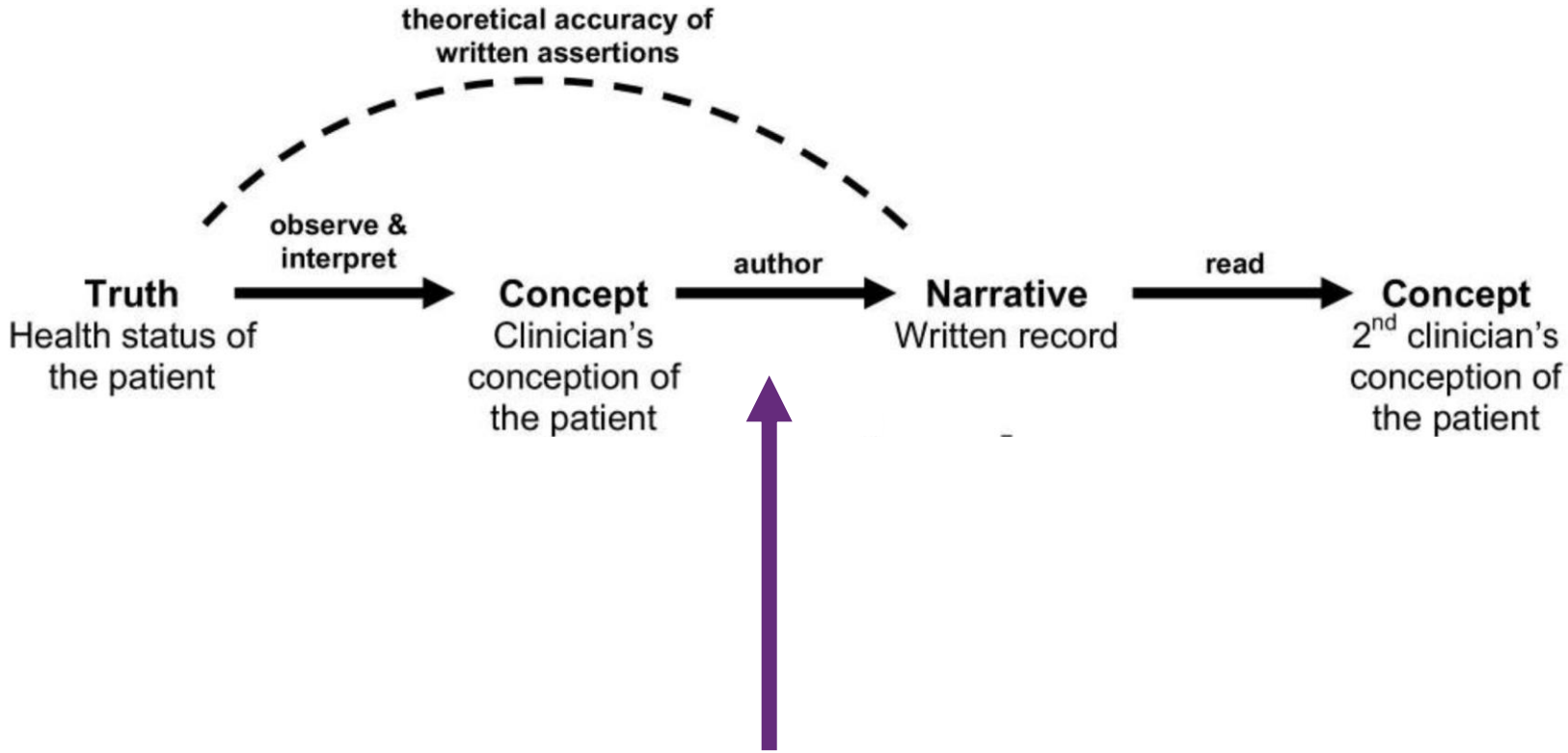



*Figure Adapted from: Hripcsak G, Elhadad N, Chen Y-H, Zhou L, Morrison FP. Using Empiric Semantic Correlation to Interpret Temporal Assertions in Clinical Texts. J Am Med Inform Assoc. 16:220–227.*

JAMIA
A SCHOLARLY JOURNAL OF INFORMATICS IN HEALTH AND BIOMEDICINE



Perspective

# Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory

Rachel L Richesson,[1] W Ed Hammond,[2,3,4] Meredith Nahm,[2] Douglas Wixted,[5] Gregory E Simon,[6] Jennifer G Robinson,[7] Alan E Bauck,[8] Denise Cifelli,[9] Michelle M Smerek,[5] John Dickerson,[8] Reesa L Laws,[8] Rosemary A Madigan,[9,10] Shelley A Rusincovitch,[3] Cynthia Kluchar,[11] Robert M Califf[5,12]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/amiajnl-2013-001926).

For numbered affiliations see end of article.

**Correspondence to**
Dr Rachel Richesson, Department of Informatics, Duke University School of Nursing, 2007 Pearson Bldg, 311 Trent Drive, Durham, NC 27710, USA; rachel.richesson@duke.edu

Received 15 April 2013
Revised 19 July 2013
Accepted 28 July 2013
Published Online First 16 August 2013

**ABSTRACT**

Widespread sharing of data from electronic health records and patient-reported outcomes can strengthen the national capacity for conducting cost-effective clinical trials and allow research to be embedded within routine care delivery. While pragmatic clinical trials (PCTs) have been performed for decades, they now can draw on rich sources of clinical and operational data that are continuously fed back to inform research and practice. The Health Care Systems Collaboratory program, initiated by the NIH Common Fund in 2012, engages healthcare systems as partners in discussing and promoting activities, tools, and strategies for supporting active participation in PCTs. The NIH Collaboratory consists of seven demonstration projects, and seven problem-specific working group 'Cores', aimed at leveraging the data captured in heterogeneous 'real-world' environments for research, thereby improving the efficiency, relevance, and generalizability of trials. Here, we introduce the Collaboratory, focusing on its Phenotype, Data Standards, and Data Quality Core, and present early observations from researchers implementing PCTs within large healthcare systems. We also identify gaps in knowledge and present an informatics research agenda that includes identifying methods for the definition and appropriate application of phenotypes in diverse healthcare settings, and methods for validating both the definition and execution of electronic health records based phenotypes.

## INTRODUCTION

The US healthcare system is poised to significantly enhance the relevance, number, speed, and cost-effectiveness of clinical trials by embedding them directly within the healthcare delivery system. This transformation[1] will be enabled by capabilities offered by electronic health records (EHRs) and patient-reported outcomes (PROs), changes in the organization and delivery of healthcare, and biological effects of new treatments, PCTs are designed to support clinical decision-making by evaluating interventions in 'real-world' practice conditions.[6] PCTs therefore recruit participants from heterogeneous practice settings, and pose challenges for reconciling the variation in healthcare operations, widely disparate information systems, and differences in data capture fidelity. The routine implementation of PCTs is a key element in achieving the vision of the learning health system,[7] but achieving this on a global scale will require innovations, including new ethical frameworks to assess consent and risk,[8 9] new methodologies to work with observational data, and more effective partnerships among healthcare systems.

Advancing our understanding and ability to conduct PCTs within healthcare systems using innovative approaches is a key focus of the NIH Collaboratory. The use of EHRs to support trial activities, including the identification of patient cohorts with precise clinical attributes, is an important component of this vision and the next generation of clinical trials. The Collaboratory is leveraging previous work in phenotype definition and execution, and adding new use cases and requirements to inform the practice of using EHRs for research, advancing the science for both informatics and evidence-based healthcare. In the following sections, we will describe the NIH Collaboratory and the Phenotype, Data Standards, and Data Quality (PSQ) 'Core' working group, including their early experiences with EHR data queries, standards considerations, and data quality activities. We will conclude with a proposed research agenda and suggested future directions.

## THE NIH COLLABORATORY

In 2012, The NIH Common Fund initiated the Health Care Systems Collaboratory (https://www.

https://doi.org/10.1136/amiajnl-2013-001926

*Journal of the American Medical Informatics Association*, 24(5), 2017, 996–1001
doi: 10.1093/jamia/ocx016
Advance Access Publication Date: 14 March 2017
Perspective

Perspective

# Pragmatic (trial) informatics: a perspective from the NIH Health Care Systems Research Collaboratory

**Rachel L Richesson,[1,2] Beverly B Green,[3] Reesa Laws,[4] Jon Puro,[5] Michael G Kahn,[6] Alan Bauck,[4] Michelle Smerek,[7] Erik G Van Eaton,[8] Meredith Zozus,[9] W Ed Hammond,[2] Kari A Stephens,[10] and Greg E Simon[3]**

[1]Division of Clinical Systems and Analytics, Duke University School of Nursing, Durham, NC, USA, [2]Duke Center for Health Informatics, Durham, NC, USA, [3]Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA, [4]Kaiser Permanente Center for Health Research, Portland, OR, USA, [5]OCHIN Inc, Portland, OR, USA, [6]Department of Pediatrics, University of Colorado, Denver, CO, USA, [7]Clinical Research Informatics, Duke Clinical Research Institute, Durham, NC, USA, [8]Department of Surgery, University of Washington, Seattle, WA, USA, [9]University of Arkansas for Medical Sciences, Little Rock, AR, USA and [10]Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA

Corresponding Author: Rachel Richesson, Duke University School of Nursing, 307 Trent Drive, Office 2007, Durham, NC 27710, USA. Phone: 919-681-0825. Fax: 919-681-8899. E-mail: rachel.richesson@dm.duke.edu

Received 20 November 2016; Revised 17 January 2017; Accepted 14 February 2017

**ABSTRACT**

Pragmatic clinical trials (PCTs) are research investigations embedded in health care settings designed to increase the efficiency of research and its relevance to clinical practice. The Health Care Systems Research Collaboratory, initiated by the National Institutes of Health Common Fund in 2010, is a pioneering cooperative aimed at identifying and overcoming operational challenges to pragmatic research. Drawing from our experience, we present 4 broad categories of informatics-related challenges: (1) using clinical data for research, (2) integrating data from heterogeneous systems, (3) using electronic health records to support intervention delivery or health system change, and (4) assessing and improving data capture to define study populations and outcomes. These challenges impact the validity, reliability, and integrity of PCTs. Achieving the full potential of PCTs and a learning health system will require meaningful partnerships between health system leadership and operations, and federally driven standards and policies to ensure that future electronic health record systems have the flexibility to support research.

**Key words**: pragmatic clinical trial, demonstration project, National Institutes of Health, clinical informatics, electronic health records

## INTRODUCTION

The growing use of electronic health records (EHRs) has increased the potential of pragmatic clinical trials (PCTs), randomized controlled trials designed for generalizability, often involving multiple clinical sites and broad eligibility criteria.[1,2] In contrast to traditional clinical trials, in which the goal is to evaluate new treatments under highly controlled conditions, PCTs are comparative effectiveness

to determine whether health interventions actually work in the "real world." Hence rapid, efficient implementation of PCTs will be key to a successful learning health system.[4] PCTs are also a source of "real-world evidence" that can inform therapeutic development, outcomes research, patient care, research on health care systems, quality improvement, safety surveillance, and well-controlled effectiveness studies.[5]

The Health Care Systems Research Collaboratory is funded by

https://doi.org/10.1093/jamia/ocx016

# General Recommendations

- **Engage systems** as research partners to access local IT staff
- **Frequent communication** among staff and research teams
- **Systematic data quality tests** throughout; require planning, time & staff
- **Use & develop standards** to augment EHR systems with additional data collection (equiv. across sites)
  - Use elements from a standard library
  - Promote standard research data elements in EHRs
  - Post phenotype definitions to a public repository

# Computable Phenotype Definition

- Specifications for identifying patients or populations with a given characteristic or condition of interest using data that are routinely collected in EHRs or ancillary data sources.
- EHR-based condition definition

# Example

Diabetes defined as[1]:

ICD-9 codes

- one inpatient discharge diagnosis (ICD-9-CM 250.x, 357.2, 366.41, 362.01-362.07)

or any combination of two of the following events occurring within 24 months of each other:

Lab codes

- A1C ≥ 6.5% (48 mmol/mol)
- fasting plasma glucose ≥ 126 mg/dl (7.0 mmol/L)
- random plasma glucose ≥ 200 mg/dl (11.1 mmol/L)
- 2-h 75-g OGTT ≥ 200 mg/dl
- outpatient diagnosis code (same codes as inpatient)
- anti-hyperglycemic medication dispense (see details below)

Medication codes

- NDC in associated list
- **…etc., etc…**

# Multiple phenotype definitions exist

*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%

**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:

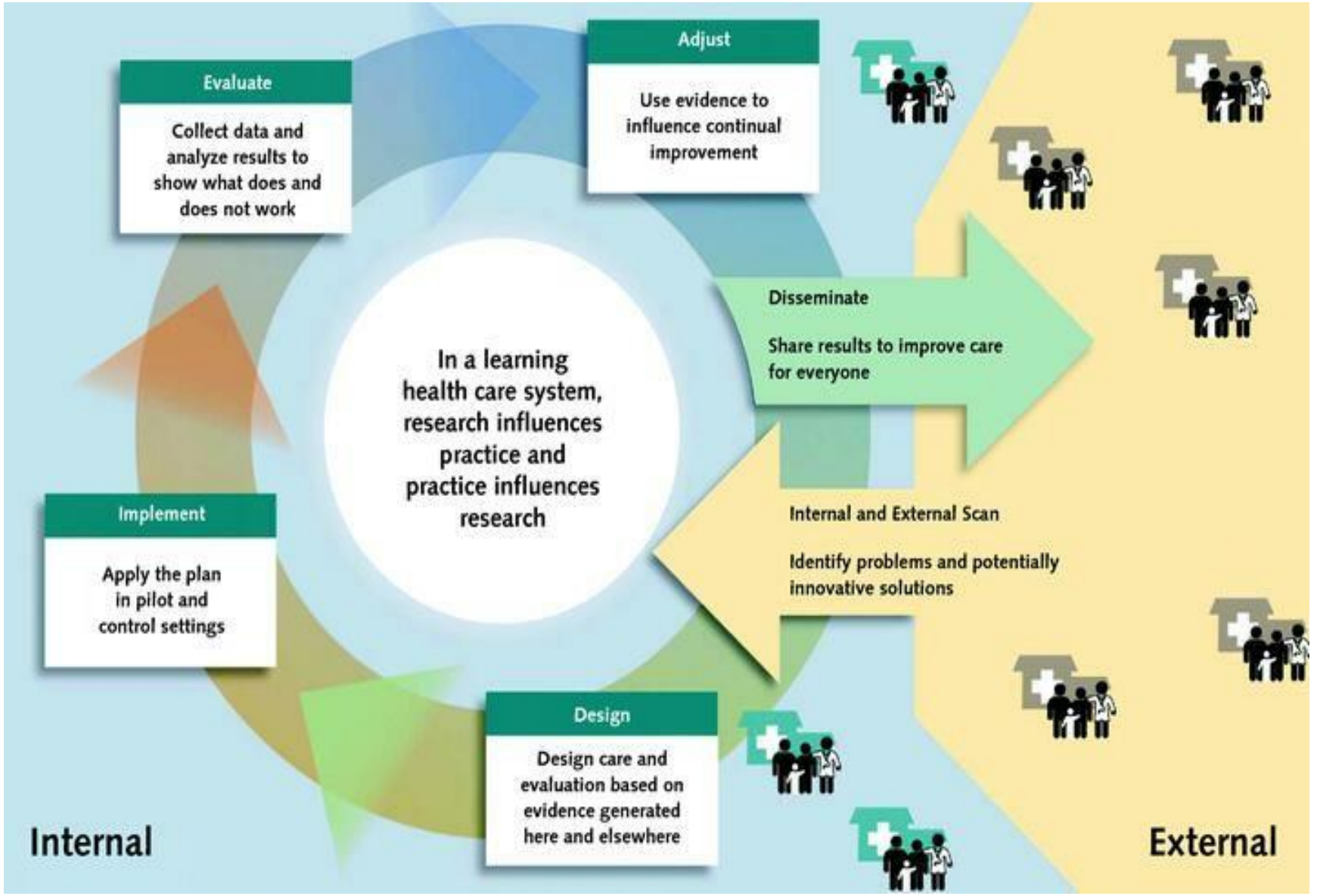
- One or more instances of hemoglobin A1c results >= 6.5%

# Benefits of Sharing & Re-Using Phenotypes

- Development and conduct of new multi-site studies
- Efficiencies of re-using executable phenotype code
- Comparability of EHR-derived data sets
- Comparison of study results and aggregation of evidence
- Reporting of data sets or results (e.g., ClinicalTrials.gov, NIH)
- Description of research populations in medical journals



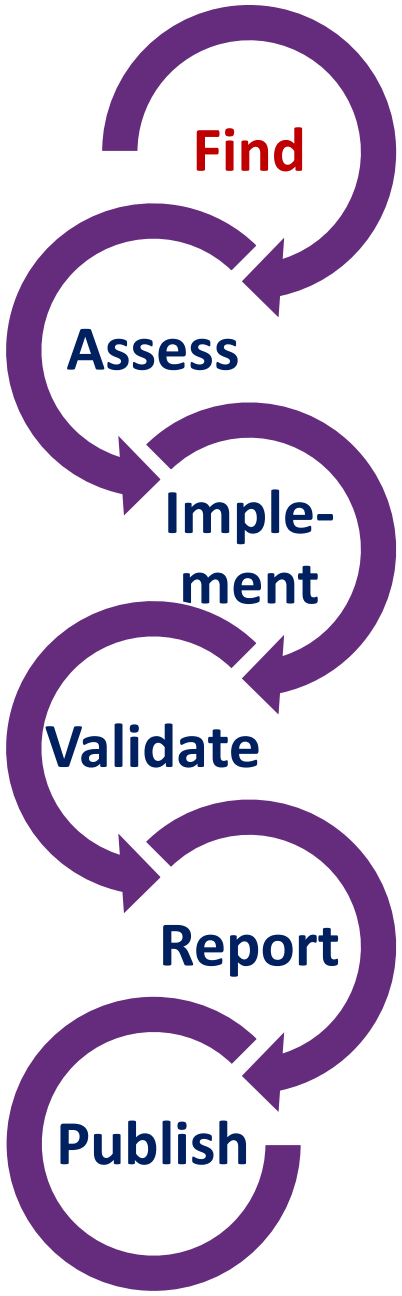
Richesson RL, Smerek MM, Blake Cameron C. A framework to support the sharing and reuse of computable phenotype definitions across health care delivery and clinical research applications. EGEMS (Wash DC) 2016;4(3):1232. - PMC - PubMed






Ann Intern Med. 2012; 157(3):207-210. doi:10.7326/0003-4819-157-3-201208070-00012
https://www.acponline.org

# RE-USE →



# Where can I find phenotype definitions?

# Chronic Conditions

The CCW contains two versions of the Chronic Conditions: 30 CCW Chronic Conditions (2017 forward) and 27 CCW Chronic Conditions (1999–2020). CMS developed the 27 CCW Chronic Condition variables from algorithms validated from the research literature and criteria used by other federal sources. In 2020, CMS contracted an expert panel to refine and enhance these algorithms, resulting in the 30 CCW Chronic Condition algorithms.

The Chronic Conditions File Enhancement White Paper document provides more detail on the differences between the two versions and recommendations for researchers.

## 30 CCW Chronic Conditions (2017 forward)

There are **30** CCW Chronic Condition categories, available for file years 2017 forward. These reference only ICD-10 diagnosis codes and have modified look-back periods, qualifying claims, and codes.

All variables listed here are currently available in the **Master Beneficiary Summary File (MBSF)** in the **MBSF_CHRONIC_*YYYY*** file.

30 CCW Chronic Conditions Algorithms and Change History

- Acute Myocardial Infarction
- Alzheimer's Disease
- Anemia
- Asthma
- Atrial Fibrillation and Flutter
- Benign Prostatic Hyperplasia
- Cancer, Breast
- Cancer, Colorectal
- Cancer, Endometrial
- Cancer, Lung
- Cancer, Prostate
- Cancer, Urologic (Kidney, Renal Pelvis, and Ureter) **NEW!**
- Cataract
- Chronic Kidney Disease
- Chronic Obstructive Pulmonary Disease
- Depression, Bipolar, or Other Depressive Mood Disorders
- Diabetes
- Glaucoma
- Heart Failure and Non-Ischemic Heart Disease
- Hip/Pelvic Fracture
- Hyperlipidemia
- Hypertension
- Hypothyroidism*
- Ischemic Heart Disease
- Non-Alzheimer's Dementia†
- Osteoporosis With or Without Pathological Fracture
- Parkinson's Disease and Secondary Parkinsonism **NEW!**
- Pneumonia, All-cause **NEW!**
- Rheumatoid Arthritis/Osteoarthritis
- Stroke/Transient Ischemic Attack

* Within the 27 CCW Chronic Conditions, this condition is "Acquired Hypothyroidism."
† Within the 27 CCW Chronic Conditions, this condition is "Alzheimer's Disease, Related Disorders, or Senile Dementia."

# 30 CCW Chronic Conditions Algorithms

MBSF_CHRONIC_{*YYYY*} FILE | REVISED 02/2022

## Alzheimer's Disease

**Reference Period:**

2 years

**Number/Type of Claims to Qualify[1]:**

At least 1 inpatient/SNF/HHA claim **OR** 2 HOP/carrier claims with DX codes

**Valid ICD-10 Codes[2]:**

G30.0, G30.1, G30.8, G30.9 (any DX on the claim)

PheKB

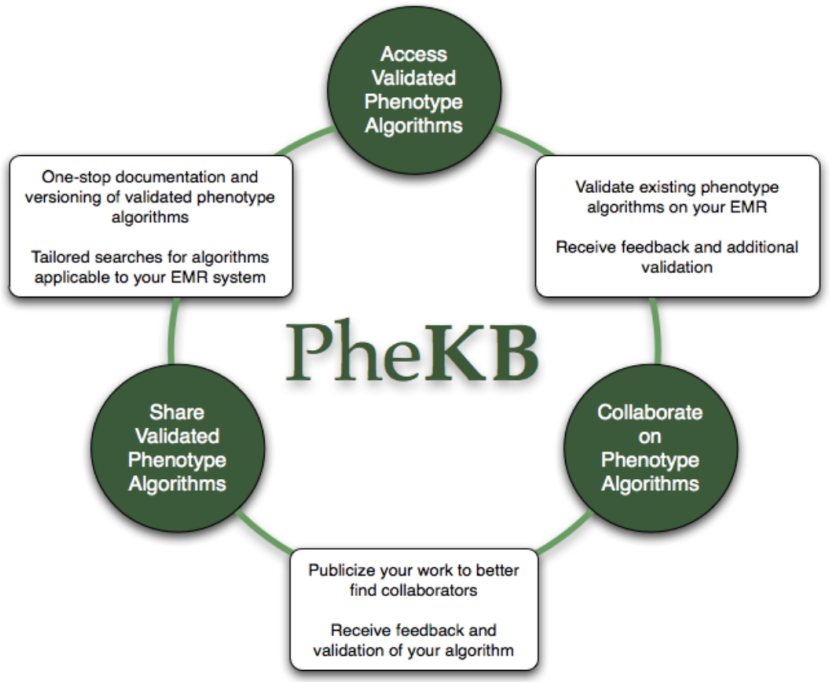
a knowledgebase for discovering phenotypes from electronic medical records

Login | Request Account

Search

Home | Phenotypes | Resources | Contact Us

# What is the Phenotype KnowledgeBase?




Health Data is becoming an increasing important source for clinical and genomic research. Researchers create and iteratively refine algorithms using structured and unstructured data to better identify cohorts of subjects within the health data.

The Phenotype Knowledgebase website, PheKB, is a collaborative environment to building and validating electronic algorithms to identify characteristics of patients within health data. PheKB was functionally designed to enable such a workflow and has purposefully integrated tools and standards that guide the user in efficiently navigating each of these stages from early stage development to public sharing and reuse. PheKB has tools to enable cross-site collaboration for algorithm development, validation, and sharing for reuse with confidence.

## Most Recent Phenotypes

- HIV
- Functional seizures
- RxNorm RxCUI codes for Cancer Therapies
- Type 1 Diabetes
- Body Mass Index (BMI)

https://phekb.org

# Phenotypes

| Title | Institution | Phenotype Attributes | Owner Phenotyping Groups | Status | Type |
|---|---|---|---|---|---|
| Abdominal Aortic Aneurysm ( AAA ) | Geisinger | CPT Codes, ICD 9 Codes, Vital Signs | eMERGE Geisinger Group | Final | Disease or Syndrome |
| ACE Inhibitor (ACE-I) induced cough | Vanderbilt University | CPT Codes, ICD 9 Codes, Medications, Natural Language Processing | eMERGE Vanderbilt Group | Final | Drug Response - adverse effect or efficacy |
| ADHD phenotype algorithm | CHOP | ICD 9 Codes, Medications, Natural Language Processing | eMERGE CHOP Group | Final | Disease or Syndrome |
| Anxiety algorithm | CHOP | CPT Codes, ICD 10 Codes, ICD 9 Codes, Medications | eMERGE CHOP Group | Final | Disease or Syndrome |
| Appendicitis | Cincinnati Children's Hospital Medical Center | CPT Codes, ICD 9 Codes, Medications, Natural Language Processing | eMERGE CCHMC/BCH Group | Final | Disease or Syndrome |
| Asthma | CHOP | ICD 9 Codes, Laboratories, Medications, Natural Language Processing | eMERGE CHOP Group | Final | Disease or Syndrome |

# Implementations and Datasets For This Phenotype

Phenotype | Data Dictionaries | Implementations/Datasets

Upload a New Implementation

| Implementation Details | Case PPV | Control PPV | Dataset/Dictionary |
|---|---|---|---|
| **ADHD Validation (CCHMC)**<br>*Cincinnati Children's Hospital Medical Center*<br>*Cases: 0 Controls: 0 (Case, Control)*<br>*Uploaded: 09/11/2014* | 0.891304 | 0.95 | cchmc_adhd_cases_demo.csv<br>ADHD Data Dictionary_demographics (12).csv |
| | | | cchmc_adhd_cases_hx.csv<br>ADHD Data Dictionary_hx of ADHD (12).csv |
| | | | cchmc_adhd_cases_med.csv<br>ADHD Data Dictionary_hx of meds (12).csv |
| | | | cchmc_adhd_cases_psych.csv<br>ADHD Data Dictionary_ hx of psych cond(12).csv |
| | | | cchmc_adhd_cases_encounters.csv<br>ADHD Data Dictionary_hx of visits (12).csv |
| **CHOP implementation**<br>*CHOP*<br>*Cases: 0 Controls: 0 (Case, Control)*<br>*Uploaded: 01/21/2015* | 0.96 | 0.96 | No datasets uploaded |
| **Harvard ADHD**<br>*Cases: 80 Controls: 1581 (Case, Control)*<br>*Uploaded: 12/23/2015* | | | adhd_CasesControls.csv<br>ADHD Data Dictionary_demographics (12).csv |
| **ADHD Implementation - Columbia**<br>*Columbia University*<br>*Cases: 5 Controls: 1294 (Case, Control)*<br>*Uploaded: 11/30/2017* | | | Columbia_Adhd_Demographics_2016May.csv<br>ADHD Data Dictionary_demographics (12).csv |
| | | | Columbia_Adhd_hx_of_ADHD_2016May.csv<br>ADHD Data Dictionary_hx of ADHD (12).csv |
| | | | Columbia_Adhd_hx_of_other_psychiatric_conditions_2016May.csv<br>ADHD Data Dictionary_ hx of psych cond(12).csv |
| | | | Columbia_Adhd_visits_2016May.csv<br>ADHD Data Dictionary_hx of visits (12).csv |

HDR UK Phenotype Library

Home | Phenotypes | Concepts | API | About | Log in

The HDR UK Phenotype Library is a comprehensive, open access resource providing the research community with information, tools and phenotyping algorithms for UK electronic health records.

Search our Phenotype library

| 791 | 1618 | 106627 | 25 | 16 |
|---|---|---|---|---|
| Phenotypes | Concepts | Clinical Codes | Data Sources | Coding Systems |

## A Reference Catalogue of Human Diseases

**Connected.** The Phenotype Library is accessible via an API to support interoperability, is integrated with health dataset information in HDR-UK's Innovation Gateway, and hosts content from numerous contributing organisations.

**Patient-focused.** The Library is enabling important research to improve patient health and well-being. Content spans major disease areas, including heart disease, cancer, COVID-19 and other common and rare human health conditions. Curated collections from contributors such as the HDR UK BREATHE Hub for respiratory health share clinical expertise to tackle critical research questions.

**Cutting-edge.** Built with a focus on computability, this resource aims to drive the next generation of research methods. Integration with Phenoflow enables executable implementations of the phenotypes in our collection, while the API and R package client facilitate integration of the Library content directly into other analysis workflows.

https://phenotypes.healthdatagateway.org/

# Phenotypes

dementia

Applied Filters: *Search:* dementia ×

**Filters** *68 Record(s)*

Type

Collection

Coding System

Data Source

Date

Authorship

Refresh Clear

Order By: Relevance Results Per Page: 20

« 1 2 3 4 »

## PH859 - Dementia Alzheimer Vascular Mixed Nonspecific

Richard Hoile, Naji Tabet, Helen Smith, Stephen Bremner, Jackie Cassell, Elizabeth Ford

Read codes v2 *Disease or Syndrome* *2022-04-04*

ClinicalCodes Repository Phenotype Library

## PH862 - Specific Dementias

Richard Hoile, Naji Tabet, Helen Smith, Stephen Bremner, Jackie Cassell, Elizabeth Ford

Read codes v2 *Disease or Syndrome* *2022-04-04*

ClinicalCodes Repository Phenotype Library

## PH473 - Dementia

Robert L Grant, Vari M Drennan, Greta Rait, Irene Petersen, Steve Iliffe

Read codes v2 *Disease or Syndrome* *2021-10-26*

ClinicalCodes Repository Phenotype Library

Export Phenotype Print

Home

Definition

Implementation

Publications

Clinical Code Lists

API

Version History

# Dementia

Kuan V, Denaxas S, Gonzalez-Izquierdo A, Direk K, Bhatti O, Husain S, Sutaria S, Hingorani M, Nitsch D, Parisinos C, Lumbers T, Mathur R, Sofat R, Casas JP, Wong I, Hemingway H, Hingorani A

| | |
|---|---|
| Type | Disease or Syndrome |
| ID | PH148 |
| Version ID | 296 |
| Data Sources | CPRD GOLD , HES Admitted Patient Care data for CPRD GOLD |
| Valid event data range | 01/01/1999 - 01/07/2016 |
| Sex | Female, Male |
| Agreement Date | 2019-05-20 |
| Coding system | Read codes v2 ICD10 codes Med codes |
| Tags /Collections | CALIBER Phenotype Library |

## Definition

At the specified date, a patient is defined as having had 'Dementia' IF they meet the criteria for any of the following on or before the specified date. The earliest date on which the individual meets any of the following criteria on or before the specified date is defined as the first event date:

Primary care

1. 'Dementia' diagnosis or history of diagnosis during a consultation

OR Secondary care (ICD10)

1. ALL diagnoses of 'Dementia' or history of diagnosis during a hospitalization

## Implementation

**PhenoFlow Implementation:** https://kclhi.org/phenoflow/phenotype/download/433

## Publications

- Kuan V., Denaxas S., Gonzalez-Izquierdo A. et al. A chronological map of 308 physical and mental health conditions from 4 million individuals in the National Health Service. The Lancet

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Who We Are | OHDSI Updates & News | Standards | Software Tools | OHDSI Studies | Book of OHDSI | Resources | New To OHDSI?

EHDEN Academy | This Week In OHDSI/Community Calls | Events/Collaborations | Workgroups | Forms For Workgroups, MS Teams

NEW: Our Journey – Where The OHDSI Community Has Been, And Where We Are Going | 2022 Europe Symposium | Follow OHDSI/Newsletters

Home › Resources › Libraries › Phenotype library

# Phenotype library

A common challenge we all face is developing standard definitions for identifying patients with a particular medical condition or exposed to a specific intervention. Our phenotype workgroup is researching and developing strategies for establishing a standardized, evidence-based approach to constructing algorithms to define disease phenotypes that can be used in observational analytics (as cohort criteria, covariates, and outcomes). The group is exploring the entire continuum of possibilities, from the expert-derived consensus-building approach (e.g. eMERGE) to vocabulary-driven approaches to machine learning techniques applied to clinical sources.

As phenotypes are developed and released, we will post details on this page, so check back regularly…

https://www.ohdsi.org/resources/libraries/phenotype-library/

# Phenotype Phebruary • Daily Threads & What We Learned

"Phenotype Phebruary" was a community-wide initiative to both develop and evaluate phenotypes for health outcomes that could be investigated by the community. Patrick Ryan introduced this initiative in both a video presentation and a forum post, and each of the conversations around the "28 phenotypes for 28 days" are being held within the OHDSI forums.

This page will provide direct links to each forum post, which is where conversations around each specific phenotype should be held. The video on the right includes "phun phacts" shared about each phenotype during our weekly community calls.


## Daily Phenotype Phebruary Links

*(future dates are subject to change)*

**Feb. 1** • Type 2 Diabetes Mellitus
**Feb. 2** • Type 1 Diabetes Mellitus
**Feb. 3** • Atrial Fibrillation
**Feb. 4** • Multiple Myeloma
**Feb. 5** • Alzheimer's Disease
**Feb. 6** • Hemorrhagic Events
**Feb. 7** • Neutropenia
**Feb. 8** • Kidney Stones

https://www.ohdsi.org/phenotype-phebruary/

# Phenotype Phebruary Day 5- Alzheimer’s Disease

General

**Patrick_Ryan** Feb 5

Team:

Day 5 of Phenotype Phebruary. Still lots of methodological topics to discuss and disease areas to investigate. Today, I’ll try to start a conversation of the phenotype that was most highly voted on across our community: Alzheimer’s disease.

**Clinical description:**

Alzheimer’s disease is a progressive neurodegenerative disorder and the most common cause of dementia (loss of cognitive functions interfering with daily activities), representing 60-80% of cases (according to Alzheimer’s Association). Intitial symptoms of Alzheimer’s disease may be short-term memory loss and other difficulties associated with mild cognitive impairment, such as word-finding, visual/spatial issues, and general confusion. Diagnosis of Alzheimer’s disease may involve neurological exam, including brain MRI or CT scans, to identify other potential causes of dementia other than Alzheimer’s, and mental cognitive status tests. Drugs approved for use in Alzheimer’s disease include cholinesterase inhibitors (such as donepezil, galantamine, or rivastigmine) and memantine, which are primarily aimed at treating cognitive symptoms. In 2021, aducanumab was approved by US FDA on the basis of clinical trial data suggesting reduction of amyloid beta plaque. Alzheimer’s disease risk increases with age, with most cases detected after 65 years old. Prevalence of AD is higher in females



https://forums.ohdsi.org/t/phenotype-phebruary-day-5-alzheimers-disease/15806

(“Alzheimer disease”[MeSH Terms] OR “Alzheimer”[Title/Abstract])

AND

((“retrospective cohort”) OR (Epidemiology[MeSH Terms]) OR (Epidemiologic Methods[MeSH Terms]) OR (phenotype[Title/Abstract]) OR (insurance) OR (claims) OR (database) OR (Diseases Category/epidemiology[MeSH Terms]) OR (Validation Study[Publication Type]) OR (Validation Studies as Topic[MeSH Terms]) OR (Sensitivity and Specificity[MeSH Terms]) OR (Predictive Value of Tests[MeSH Terms]) OR (Reproducibility of Results[MeSH Terms]) )

AND

((Medicaid) OR (Medicare) OR (Truven) OR (Optum) OR (Medstat) OR (“Nationwide Inpatient Sample”) OR (“National Inpatient Sample”) OR (PharMetrics) OR (PHARMO) OR (ICD-9[Title/Abstract]) OR (ICD-10[Title/Abstract]) OR (IMS[Title/Abstract]) OR (“electronic medical records”[Text Word]) OR (Denmark/epidemiology[MeSH Terms]) OR (Veterans Affairs[Title/Abstract]) OR (“Premier database”[Title/Abstract]) OR (“National Health Insurance Research Database”[Title/Abstract]) OR (Outcome Assessment[Title/Abstract]) OR (“insurance database”[Title/Abstract]) OR (Database Management System[MeSH Terms]) OR (Medical Records Systems, Computerized[MeSH Terms]) OR (“Positive predictive value”[Title/Abstract]) )

NOT

(“Clinical Trial”[pt] OR “Editorial”[pt] OR “Letter”[pt] OR “Randomized Controlled Trial”[pt] OR “Clinical Trial, Phase I”[pt] OR “Clinical Trial, Phase II”[pt] OR “Clinical Trial, Phase III”[pt] OR “Clinical Trial, Phase IV”[pt] OR “Comment”[pt] OR “Controlled Clinical Trial”[pt] OR “Letter”[pt] OR “Case Reports”[pt] OR “Clinical Trials as Topic”[Mesh] OR “double-blind”[All] OR “placebo-controlled”[All] OR “pilot study”[All] OR “pilot projects”[Mesh] OR “Prospective Studies”[Mesh] OR “Genetics”[Mesh] OR (“Genotype”[Mesh]) OR (biomarker[Title/Abstract]))

condition

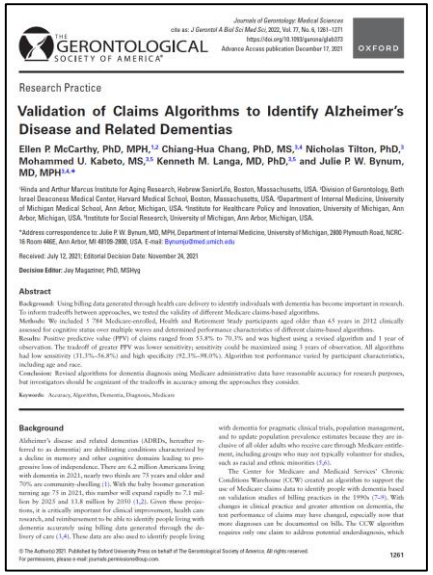
Type of study

Database study

Non-observational research

**Table 1.** ADRD Algorithm Specifications

| | CCW | Bynum-EM | Bynum-Standard |
|---|---|---|---|
| Observation Period | 3 Years | 1 Year and 3 Years | 1 Year and 3 Years |
| ICD-9-CM Diagnosis Codes | 331.0, 331.11, 331.19, 331.2, 331.7, 290.0, 290.10, 290.11, 290.12, 290.13, 290.20, 290.21, 290.3, 290.40, 290.41, 290.42, 290.43, 294.0, 294.10, 294.11, 294.20, 294.21, 294.8, 797 | 331.0, 331.11, 331.19, 331.2, 331.7, **331.82, 331.89,** 290.0, 290.10, 290.11, 290.12, 290.13, 290.20, 290.21, 290.3, 290.40, 290.41, 290.42, 290.43, **290.8**, 294.0, 294.10, 294.11, 294.20, 294.21, 797 | 331.0, 331.11, 331.19, 331.2, 331.7, **331.82, 331.89,** 290.0, 290.10, 290.11, 290.12, 290.13, 290.20, 290.21, 290.3, 290.40, 290.41, 290.42, 290.43, **290.8**, 294.0, 294.10, 294.11, 294.20, 294.21, 797 |
| Claims Files and Qualifying Claims | | | |
| MEDPAR | Any inpatient or SNF claim | Any inpatient or SNF claim | Any inpatient or SNF claim |
| Home Health Agency | Any claim* | Any claim | Any claim |
| Hospice | | Any claim | Any claim |
| HOF for outpatient medical services | Any claim* | Includes only claims from Rural Health Clinics, Federally Qualified Health Centers, and Critical Access Hospitals—Payment Option II | Includes only claims from Rural Health Clinics, Federally Qualified Health Centers, and Critical Access Hospitals—Payment Option II |
| Carrier (Provider) File for services from physicians and other health care providers | Any claim*, *excluding claims* with BETOS codes of D1A, D1B, D1C, D1D, D1E, D1F, D1G (for durable medical equipment), or O1A (for ambulance services) | Any claim for evaluation and management (E&M) by a provider *Includes only claims* BETOS "M" codes: M1A, M1B, M2A, M2B, M2C, M3, M4A, M4B, M5A, M5B, M5C, or M6 | Any claim<br>**This algorithm requires two or more qualifying Carrier or HOF claims at least 7 days apart.* |



Journals of Gerontology: Medical Sciences
cite as: J Gerontol A Biol Sci Med Sci, 2022, Vol. 77, No. 6, 1261–1271
https://doi.org/10.1093/gerona/glab373
Advance Access publication December 17, 2021

THE GERONTOLOGICAL SOCIETY OF AMERICA® | OXFORD

Research Practice

**Validation of Claims Algorithms to Identify Alzheimer's Disease and Related Dementias**

**Ellen P. McCarthy, PhD, MPH,[1,2] Chiang-Hua Chang, PhD, MS,[3,4] Nicholas Tilton, PhD,[3] Mohammed U. Kabeto, MS,[3,5] Kenneth M. Langa, MD, PhD,[3,6] and Julie P. W. Bynum, MD, MPH[3,4,*]**

[1]Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, Massachusetts, USA. [2]Division of Gerontology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA. [3]Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan, USA. [4]Institute for Healthcare Policy and Innovation, University of Michigan, Ann Arbor, Michigan, USA. [5]Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA.

*Address correspondence to: Julie P. W. Bynum, MD, MPH, Department of Internal Medicine, University of Michigan, 2800 Plymouth Road, NCRC-16 Room 446E, Ann Arbor, MI 48109-2800, USA. E-mail: Bynumju@med.umich.edu

Received: July 12, 2021; Editorial Decision Date: November 24, 2021

Decision Editor: Jay Magaziner, PhD, MSHyg

**Abstract**

Background: Using billing data generated through health care delivery to identify individuals with dementia has become important in research. To inform tradeoffs between approaches, we tested the validity of different Medicare claims-based algorithms.

Methods: We included 1 784 Medicare-enrolled, Health and Retirement Study participants aged older than 65 years in 2012 clinically assessed for cognitive status over multiple waves and determined performance characteristics of different claims-based algorithms.

Results: Positive predictive value (PPV) of claims ranged from 53.8% to 70.3% and was highest using a revised algorithm and 1 year of observation. The tradeoff of greater PPV was lower sensitivity; sensitivity could be maximized using 3 years of observation. All algorithms had low sensitivity (31.1%–56.8%) and high specificity (92.1%–98.0%). Algorithm test performance varied by participant characteristics, including age and race.

Conclusion: Revised algorithms for dementia diagnosis using Medicare administrative data have reasonable accuracy for research purposes, but investigators should be cognizant of the tradeoffs in accuracy among the approaches they consider.

Keywords: Accuracy, Algorithm, Dementia, Diagnosis, Medicare

**Background**

Alzheimer's disease and related dementias (ADRDs, hereafter referred to as dementia) are debilitating conditions characterized by a decline in memory and other cognitive domains leading to progressive loss of independence. There are 6.2 million Americans living with dementia in 2021, nearly two thirds are 75 years and older and 70% are community-dwelling (1). With the baby boomer generation turning age 75 in 2021, this number will expand rapidly to 7.1 million by 2025 and 13.8 million by 2050 (1,2). Given these projections, it is critically important for clinical improvement, health care research, and reimbursement to be able to identify people living with dementia accurately using billing data generated through the delivery of care (3,4). These data are also used to identify people living with dementia for pragmatic clinical trials, population management, and to update population prevalence estimates because they are inclusive of all older adults who receive care through Medicare entitlement, including groups who may not typically volunteer for studies, such as racial and ethnic minorities (5,6).

The Center for Medicare and Medicaid Services' Chronic Conditions Warehouse (CCW) created an algorithm to support the use of Medicare claims data to identify people with dementia based on validation studies of billing practices in the 1990s (7–9). With changes in clinical practice and greater attention on dementia, the test performance of claims may have changed, especially now that more diagnoses can be documented on bills. The CCW algorithm requires only one claim to address potential underdiagnosis, which



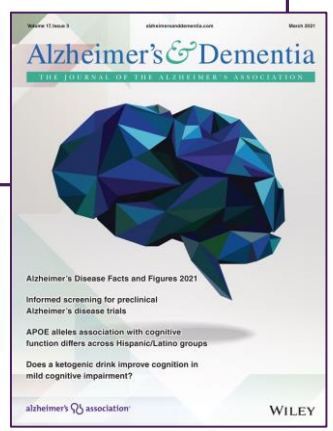


McCarthy EP, Chang CH, Tilton N, Kabeto MU, Langa KM, Bynum JPW. **Validation of Claims Algorithms to Identify Alzheimer's Disease and Related Dementias**. J Gerontol A Biol Sci Med Sci. 2022 Jun 1;77(6):1261-1271. doi: 10.1093/gerona/glab373. PMID: 34919686; PMCID: PMC9159657.


https://pubmed.ncbi.nlm.nih.gov/34919686/

**TABLE 1.** Alzheimer's disease and related disorders

| ICD-9 code | |
|---|---|
| 3310 | Alzheimer's disease |
| 2900 | Senile dementia, uncomplicated |
| 29010 | Presenile dementia, uncomplicated |
| 29011 | Presenile dementia with delirium |
| 29012 | Presenile dementia with delusional features |
| 29013 | Presenile dementia with depressive features |
| 29020 | Senile dementia with delusional features |
| 29021 | Senile dementia with depressive features |
| 2903 | Senile dementia with delirium |
| 29040 | Vascular dementia, uncomplicated |
| 29041 | Vascular dementia, with delirium |
| 29042 | Vascular dementia, with delusions |
| 29043 | Vascular dementia, with depressed mood |
| 2940 | Amnestic disorders in conditions classified elsewhere |
| 29410 | Dementia in conditions classified elsewhere without behavioral disturbance |
| 29411 | Dementia in conditions classified elsewhere with behavioral disturbance |
| 29420 | Dementia, unspecified, without behavioral disturbance |
| 29421 | Dementia, unspecified, with behavioral disturbance |
| 2948 | Other persistent mental disorders due to conditions classified elsewhere |
| 797 | Senility without mention of psychosis |
| ICD-10 codes | |
| G300 | Alzheimer's disease with early onset |
| G301 | Alzheimer's disease with late onset |
| G308 | Other Alzheimer's disease |
| G309 | Alzheimer's disease, unspecified |
| F0150 | Vascular dementia without behavioral disturbance |
| F0151 | Vascular dementia with behavioral disturbance |
| F0280 | Dementia in other diseases classified elsewhere without behavioral disturbance |
| F0281 | Dementia in other diseases classified elsewhere with behavioral disturbance |
| F0390 | Unspecified dementia without behavioral disturbance |
| F0391 | Unspecified dementia with behavioral disturbance |
| F04 | Amnestic disorder due to known physiological condition |
| R4181 | Age-related cognitive decline |




Jain S, Rosenbaum PR, Reiter JG, et al. Using Medicare claims in identifying Alzheimer's disease and related dementias. Alzheimer Dementia. 2021;17:515–524.

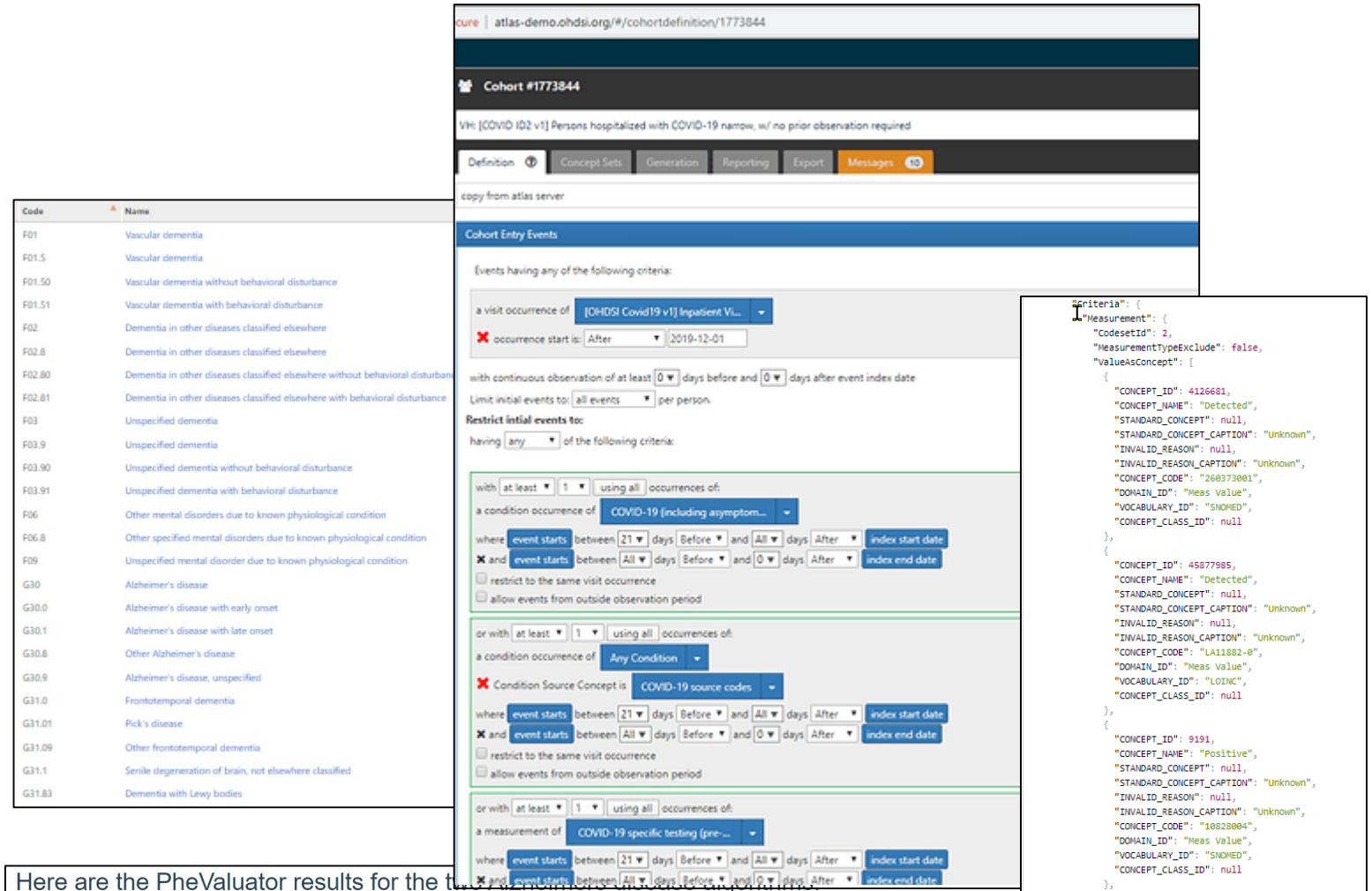
https://www-ncbi-nlm-nih-gov.proxy.lib.umich.edu/pmc/articles/PMC8296851/

# How assimilate this information?

- Different code list formats
- Different lists of codes

- *What are the differences?*
- *Are they impactful?*

- Concept set
- Iteration
- Testing & review



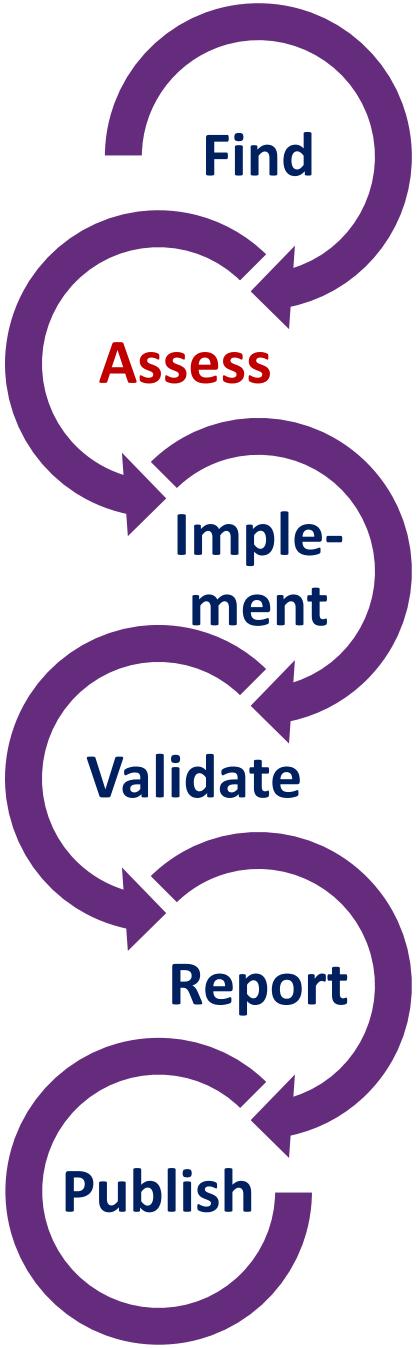
# Summary – Sources of Phenotypes

- Published literature
- Research networks
- CMS resources for code lists and value sets
  - AHRQ CCC, eCQMs and NLM VSAC

- Code repositories: GitHub
- Enhanced code repositories (w/ tools & data): OHDSI

- Phenotype repositories
  - PheKB, HDR-UK

# RE-USE →



C. Blake Cameron, MD, MBI
Nephrologist
Duke University

# A USER'S GUIDE TO COMPUTABLE PHENOTYPES

By

C. Blake Cameron, M.D., M.B.I.

## TABLE OF CONTENTS



NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

https://dcricollab.dcri.duke.edu/sites/NIHKR/KR/Blake_Users_Guide_to_Computable_Phenotypes.pdf

# Phases of Review and Reviewer Roles

| | Admin | MD or Clinical | Clinical Research | Informatics / Data Analyst |
|---|---|---|---|---|
| <u>Overall</u>:<br>Who, What, Where, When, Why? | X | X | X | X |
| <u>Clinical</u>:<br>Is algorithm valid in my patient population for my intended purpose? | | X | X | |
| <u>Technical</u>:<br>Implementation feasibility: documentation quality, concordance with local data models | | | X | X |

Laura K. Wiley

## Characterizing Variability of EHR-Driven Phenotype Definitions

[PDF] from medrxiv.org

Authors: Pascal S Brandt, Abel Kho, Yuan Luo, Jennifer A Pacheco, Theresa L Walunas, Hakon Hakonarson, George Hripcsak, Cong Liu, Ning Shang, Chunhua Weng, Nephi Walton, David S Carrell, Paul K Crane, Eric Larson, Christopher G Chute, Iftikhar Kullo, Robert Carroll, Josh Denny, Andrea Ramirez, Wei-Qi Wei, Jyoti Pathak, Laura K Wiley, Rachel Richesson, Justin B Starren, Luke V Rasmussen

Publication date: 2022/1/1

Journal: medRxiv

Publisher: Cold Spring Harbor Laboratory Press

Related work in-progress:

- Literature review to assess phenotype reporting
- Health equity implications for clinical phenotypes

***2 questions to assess a phenotype algorithm***:

1.) Does it capture the right patients?

2.) Does it have performance metrics that meet my intended needs?

## Identifying Heart Failure from Electronic Health Records: A Systematic Evidence Review

Authors: Rebecca T Levinson, Jennifer R Malinowski, Suzette J Bielinski, Luke V Rasmussen, Quinn S Wells, Veronique L Roger, Laura K Wiley

Publication date: 2021/1/1

Source: medRxiv

Publisher: Cold Spring Harbor Laboratory Press

Description: Background

Heart failure (HF) is a complex syndrome associated with significant morbidity and healthcare costs. Electronic health records (EHRs) are widely used to identify patients with HF and other phenotypes. Despite widespread use of EHRs for phenotype algorithm development, it is unclear if the characteristics of identified populations mirror those of clinically observed patients and reflect the known spectrum of HF phenotypes.

Methods

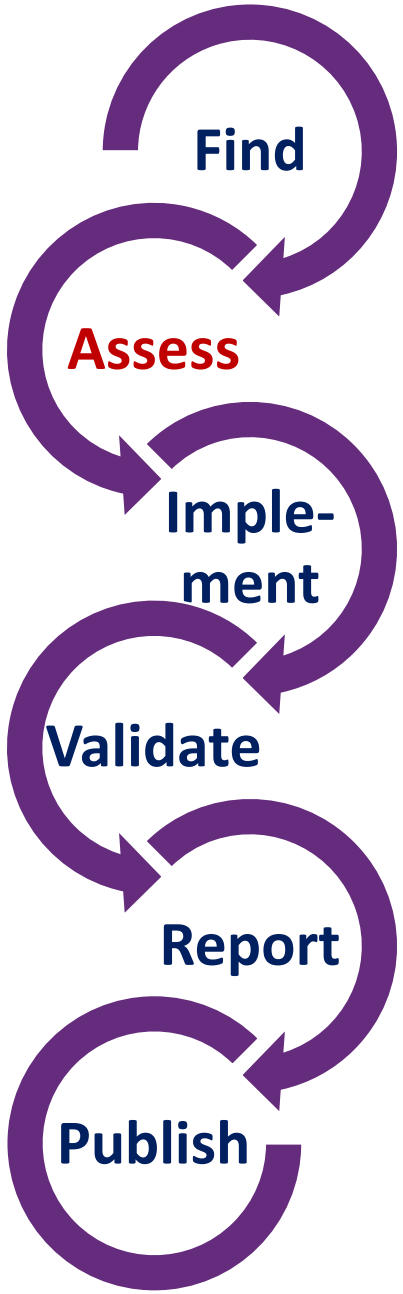
We performed a subanalysis within a larger systematic evidence review to assess the different methods used for HF algorithm development and their application to research and clinical care. We queried PubMed for articles published up to November 2020. Out of 318 studies screened, 25 articles were included for primary analysis and 15 studies using only International Classification of Diseases (ICD) codes were evaluated for secondary analysis. Results are reported descriptively.

# Targeted Patients

- Type of dementia
  - Alzheimer's Disease
  - Vascular Dementia
  - Frontotemporal Dementia
  - Lewy Body Disease
  - Mixed forms

- Severity/stage
- Presence of behavioral symptoms
- Cognitive impairment due to dementia

# RE-USE →

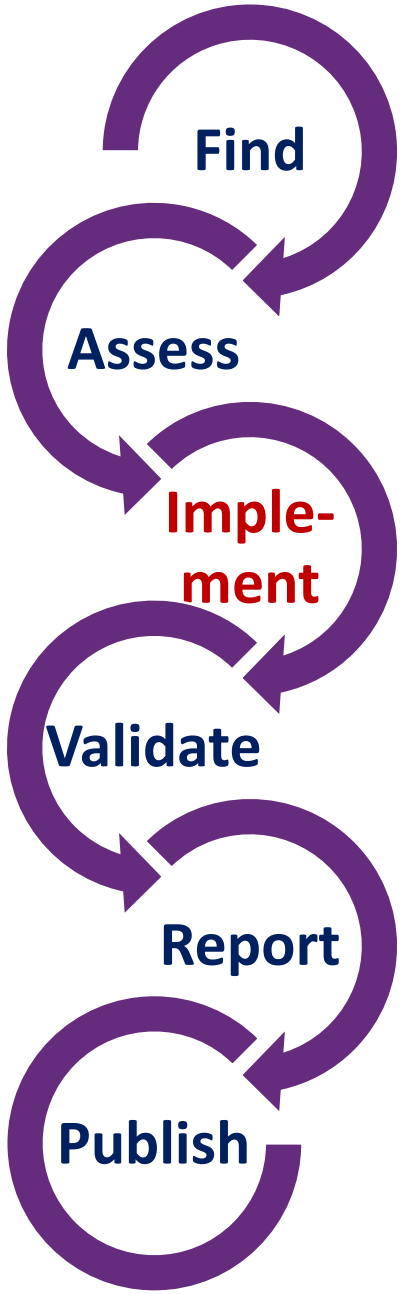



Purpose of Phenotype?

Performance metrics?

- Sensitivity?
- Specificity?
- PPV?

# RE-USE →



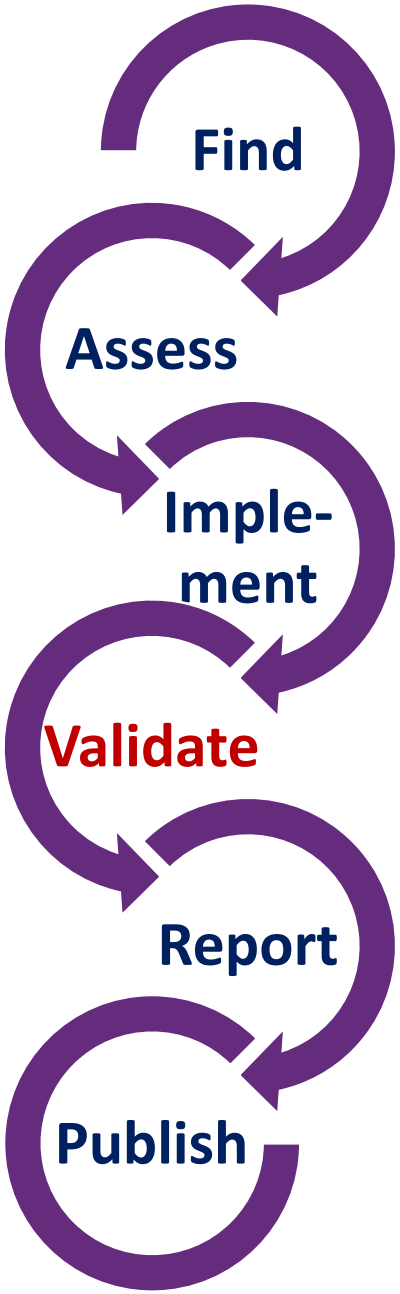

Documentation clear?

Do you have required data?

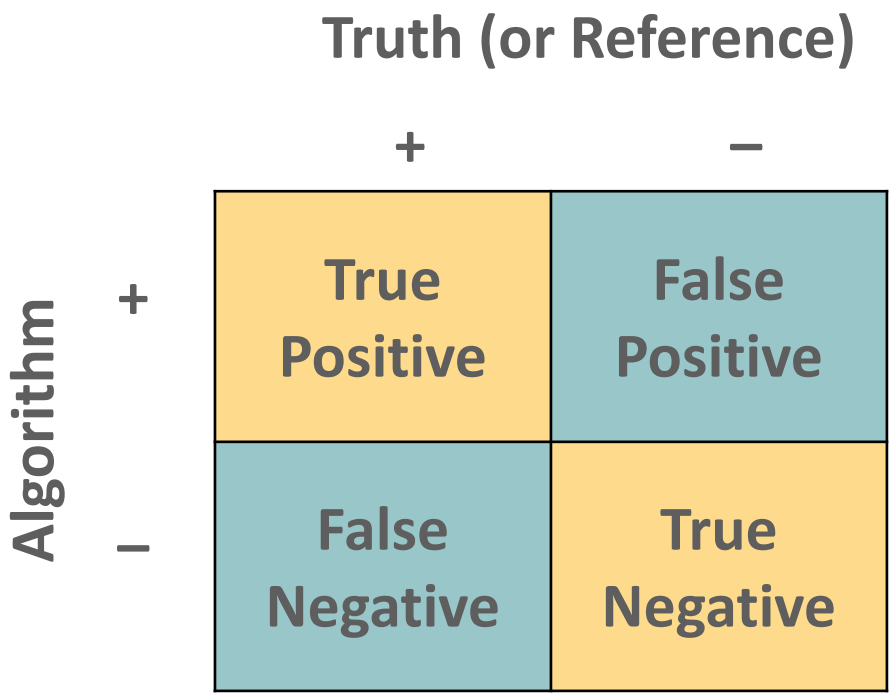
Is data formatted correctly?

Do you have capability?

# RE-USE →

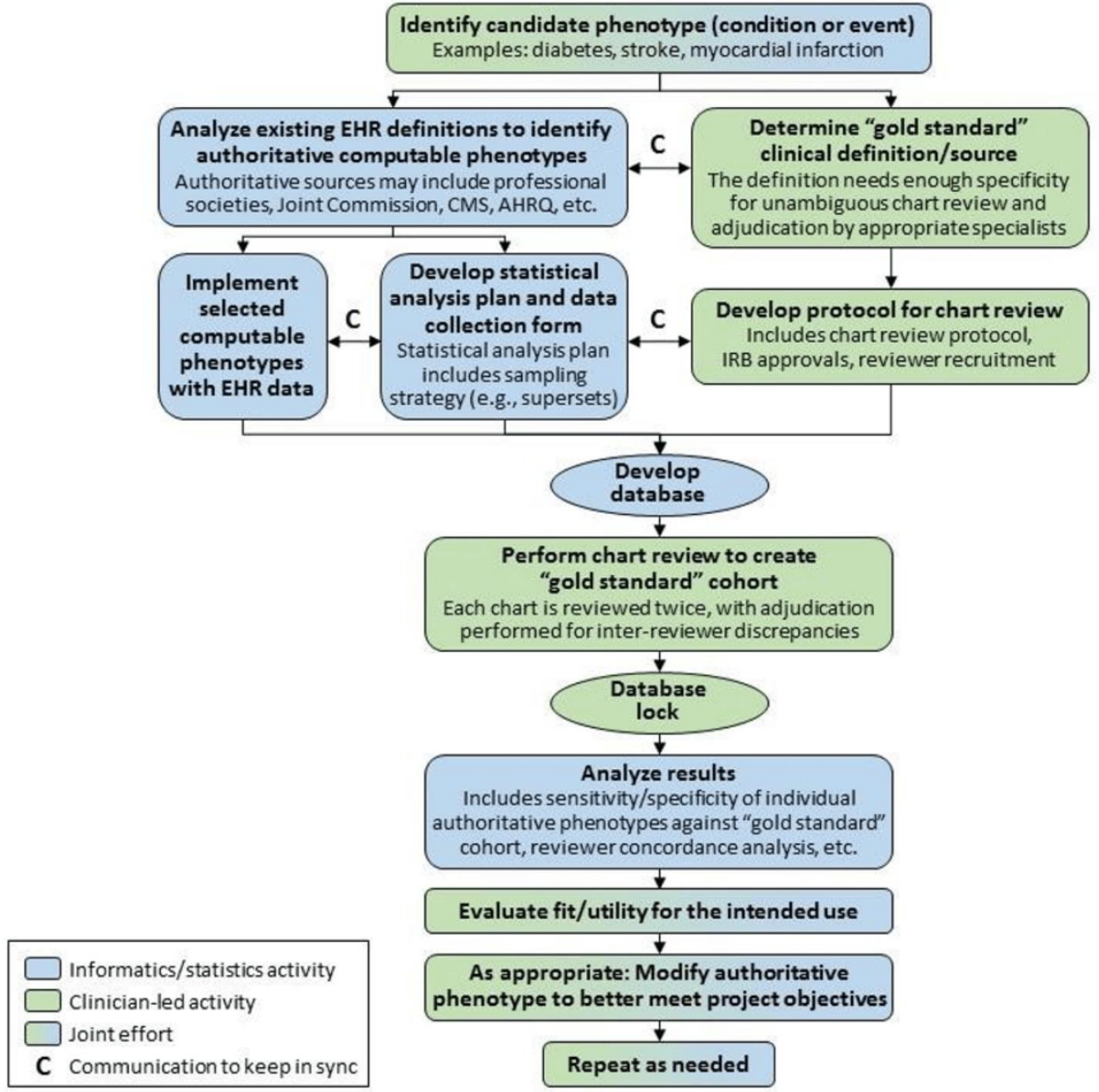


# Validation metrics




- **Sensitivity: TP/(TP+FN)**
- **Specificity: TN/(TN+FP)**
- **PPV: TP/(TP+FP)**
- **NPV: TN/(TN+FN)**

# Phenotype Evaluation Process




S. Rusincovitch.
https://rethinkingclinicaltrials.org/chapters/conduct/electronic-health-records-based-phenotyping/using-phenotypes-in-pcts-how-do-i-get-started/

# Types of Validation

- Gold Standard
  - Manually review patient records to find truth
- Comparative Gold Standard
  - Derive reference labels from another source – e.g. enrolled population, registry data, patient reported outcomes, etc.
- “Silver Standard”
  - Use “fuzzy” labels, probabilistic models, etc.

# Data Quality

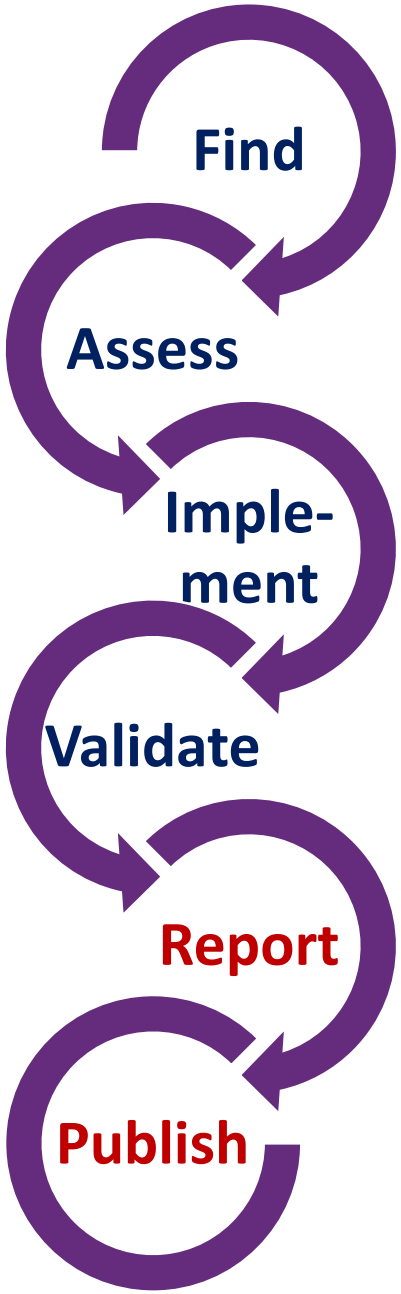
- Quality of data can affect results of phenotype-based queries
- Recognize that EHR and other healthcare data are not collected for research
- Data quality assessment should accompany phenotype validation
- Workflow assessment at each site should be included

**Table 1.** Data Quality Dimensions Determining Fitness for Use of Research Data

| Dimension | Conceptual definition | Operational examples |
|---|---|---|
| Completeness | Presence of the necessary data | Presence of necessary data elements, percent of missing values for a data element, percent of records with sufficient data to calculate a required variable (e.g., an outcome) |
| Accuracy | Closeness of agreement between a data value and the true value* | Percent of data values found to be in error based on a gold standard, percent of physically implausible values, percent of data values that do not conform to range expectations |
| Consistency | Relevant uniformity in data across clinical investigation sites, facilities, departments, units within a facility, providers, or other assessors | Comparable proportions of relevant diagnoses across sites, comparable proportions of documented order fulfillment (e.g., returned procedure report for ordered diagnostic tests) |

https://dcricollab.dcri.duke.edu/sites/NIHKR/KR/Assessing-data-quality_V1%200.pdf

# RE-USE →




Purpose and context

Specifications

Validation Methods Results

Experience

# What is needed for phenotype re-use at scale?

- Platform to search and browse existing phenotype definitions
- Standard review information & metadata
- Incentives to report information & metadata

OXFORD

(GIGA)n SCIENCE

*GigaScience*, 10, 2021, 1–13

https://doi.org/10.1093/gigascience/giab059
Review

REVIEW

# Desiderata for the development of next-generation electronic health record phenotype libraries

Martin Chapman [1,*], Shahzad Mumtaz [2], Luke V. Rasmussen [3], Andreas Karwath [4], Georgios V. Gkoutos [4], Chuang Gao [2], Dan Thayer [5], Jennifer A. Pacheco [3], Helen Parkinson [6], Rachel L. Richesson [7], Emily Jefferson [2], Spiros Denaxas [8] and Vasa Curcin [1]

[1]Department of Population Health Sciences, King's College London, London, SE1 1UL, UK; [2]Health Informatics Centre (HIC), University of Dundee, Dundee, DD1 9SY, UK; [3]Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA; [4]Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, B15 2TT, UK; [5]SAIL Databank, Swansea University, Swansea, SA2 8PP, UK; [6]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, CB10 1SD, UK; [7]Department of Learning Health Sciences, University of Michigan Medical School, MI 48109, USA and [8]Institute of Health Informatics, University College London, London, NW1 2DA, UK

[*]**Correspondence address**. Martin Chapman, 3.07 Addison House, Guy's Campus, King's College London, London, SE1 1UL, UK. E-mail: martin.chapman@kcl.ac.uk http://orcid.org/0000-0002-5242-9701

https://pubmed.ncbi.nlm.nih.gov/34508578/

# Desiderata for Phenotype Libraries

- Support modelling languages
- Support NLP & ML–based definitions
- Support multi-dimensional descriptions

*modelling*

- Support versioning and data provenance
- Support modular relationships between phenotypes

*logging*

- Communicate implementation information
- Support tooling for multiple programming language implementations
- Support tooling that provides connectivity with multiple data standards

*implementation*

- Support a defined validation process
- Automate multiple validation techniques
- Enable feedback

*validation*

- Expose a standard API
- Offer advanced search capabilities
- Include comprehensive metadata

*sharing & warehousing*

NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

# Opportunities for the Future

HealthIT.gov
Interoperability Standards Advisory (ISA)

Official Website of The Office of the National Coordinator for Health Information Technology (ONC)

Connect with us:

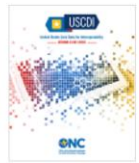
Home | About the ISA | ISA Content | ISA Publications | Recent ISA Updates | USCDI | SVAP

Home > USCDI

For data class description and applicable standards supporting data elements, click to view the USCDI Version 1 (July 2020 errata) in PDF format below.

Click to View USCDI V1

Click to View USCDI V2



Click to View USCDI V3

## United States Core Data for Interoperability (USCDI)

The United States Core Data for Interoperability (USCDI) is a standardized set of health data classes and constituent data elements for nationwide, interoperable health information exchange. Review the USCDI Fact Sheet to

**A USCDI "Data Class" is an aggregation of various Data Elements by a co**

**A USCDI "Data Element" is the most granular level at which a piece of d**

**For example, Date of Birth is a Data Element rather than its component**

**USCDI ONC New Data Element & Class (ONDEC) Submission System**

HL7 FHIR® Patient_Reported_Outcomes Implementation Guide For Comment Ballot

Home | Implementation Guidance | Profiles and Extensions | Terminology | Capability Statements | Downloads

This page is part of the Patient Reported Outcomes (PRO) FHIR IG (v0.1.0: STU 1 Ballot 1) based on FHIR v3.5.0. . For a full list of available versions, see the Directory of published versions

TOC Home

# Patient Reported Outcomes FHIR Implementation Guide

## 0.1 Introduction

Table of Contents:
- Introduction
- Guidance to the readers

The Patient Reported Outcomes (PRO) FHIR Implementation Guide (IG) will focus on capturing and exchanging patient reported outcome data electronically using the FHIR standard. The data that is captured will be made available to both providers and authorized researchers. While the PRO FHIR IG can be applied to multiple use cases, the current requirements have been drawn from [PCORnet] use cases and implementations. The capabilities described as part of the IG are intended to be leveraged to build US data infrastructure for a Learning Health System (LHS).

PRO FHIR IG will leverage the US-Core IG and profiles for the resources that overlap with US-Core. PRO FHIR IG will also leverage the Structured Data Capture (SDC) FHIR IG. In addition the IG will create profiles and extensions necessary for PRO purposes which do not exist in US-Core and SDC FHIR IG.

The next section provides a road map for the reader to walk through the implementation guide.

## 0.1 Guidance to the readers

The following table will provide a road map to the reader to follow and absorb the content of the implementation guide.

| Topic to Read | What it Contains and its relationship to PRO IG | Where can I find the content ? |
|---|---|---|
| Basic Definitions | The set of definitions applicable to the PRO FHIR IG. (Definition of"Supported or MUST Support", Usage of Code Bindings in US Core Profiles.). | US-Core Definitions |
| PRO Overview | The artifact provides background on Patient Reported Outcomes, Patient Reported Outcome Measures and other PRO related topics. | PRO Overview |
| Profiles | The artifact defines the various profiles, extensions and resources that make up the PRO FHIR IG. | Profiles |
| Capability Statements | The artifact defines the various capability statements (implementation requirements) for each PRO actor that make up the PRO FHIR IG. | Capability Statements |
| Implementation Guidance | The artifact contains guidance and examples that will help implementers of PRO FHIR IG. | Implementation Guidance |

This is the Continuous Integration Build of FHIR (will be incorrect/inconsistent at times).
See the Directory of published versions

Content | Examples | Detailed Descriptions | Mappings | Profiles & Extensions | R

## 6.3 Resource Provenance - Content

| Security Work Group | Maturity Level: 3 | Trial Use | Security Category: Not Classified |
|---|---|---|---|

## 2.5.9 Resource Questionnaire - Detailed Descriptions

| FHIR Infrastructure Work Group | Maturity Level: 3 | Trial Use | Security Category: Business |
|---|---|---|---|

Detailed Descriptions for the elements in the Questionnaire resource.

JAMA Network Open

Original Investigation | Geriatrics

# Information Sharing Practices Between US Hospitals and Skilled Nursing Facilities to Support Care Transitions

Julia Adler-Milstein, PhD; Katherine Raphael, BA; Terrence A. O'Malley, MD; Dori A. Cross, PhD

## Abstract

**IMPORTANCE** Patient transitions from hospitals to skilled nursing facilities (SNFs) require robust information sharing. After a decade of investment in health information technology infrastructure and new incentives to promote hospital-SNF coordination in the US, the current state of information sharing at this critical transition is unknown.

**OBJECTIVE** To measure the completeness, timeliness, and usability of information shared by hospitals when discharging patients to SNFs, and to identify relational and structural characteristics associated with better hospital-SNF information sharing.

**DESIGN, SETTING, AND PARTICIPANTS** Survey of 500 SNFs from a US nationally representative sample (265 respondents representing 471 hospital-SNF pairs; response rate of 53.0%) that collected detailed data on information sharing that supports care transitions from each of the 2 hospitals from which they receive the largest volume of patient referrals. Survey administration occurred between January 2019 and March 2020.

**MAIN OUTCOMES AND MEASURES** Overall assessment of information completeness, timeliness, and usability using 5-point Likert scales. Detailed measures, including (1) completeness—routine sharing of 23 specific information types; (2) timeliness—how often information arrived after the patient; and (3) usability—whether information was duplicative, extraneous, or not tailored to SNF needs. In addition, 8 relational characteristics (eg, shared staffing, collaborative meetings, and referral volume) and 10 structural characteristics (eg, size, ownership, and staffing) were assessed as potential factors associated with better information sharing.

**RESULTS** Of 471 hospital-SNF pairs, 64 (13.5%) reported excellent performance on all 3 dimensions of information sharing, whereas 141 (30.0%) were at or below the mean performance on all dimensions. Social status (missing in 309 pairs [65.7%]) and behavioral status (missing in 319 pairs [67.7%]) were the most common types of missing information. Receipt of hospital information was delayed, sometimes (159 pairs [33.8%]) or often (77 pairs [16.4%]) arriving after the patient. In total, 358 pairs [76.0%] reported at least 1 usability shortcoming. Having a hospital clinician on site at the SNF was associated in multivariate analysis with more complete (odds ratio, 1.72; 95% CI, 1.07-2.78; $P = .03$), timely (odds ratio, 1.76; 95% CI, 1.08-2.88; $P = .02$), and usable (odds ratio, 1.64; 95% CI, 1.02-2.63; $P = .04$) information sharing. Hospital accountable care organization participation was associated with more timely information sharing (odds ratio, 1.88; 95% CI, 1.13-3.14; $P = .02$).

## Key Points

**Question** What is the current state of information sharing to support care transitions between hospitals and skilled nursing facilities (SNFs) in the US, and what characteristics are associated with better sharing?

**Findings** In a US nationally representative survey that included responses from 471 hospital-SNF pairs about information sharing, SNFs reported that key information was often missing (functional, mental, and behavioral status as well as whom to contact at the hospital with follow-up questions), delayed (often arriving after the patient), and difficult to use (discharge documents with duplicative and extraneous information). Having a hospital clinician on site at the SNF was associated with more complete, timely, and usable information sharing.

**Meaning** This study finds shortcomings across numerous dimensions of information sharing, raising concerns about patients' transitional care experience from hospitals to SNFs.

+ Invited Commentary

+ Supplemental content

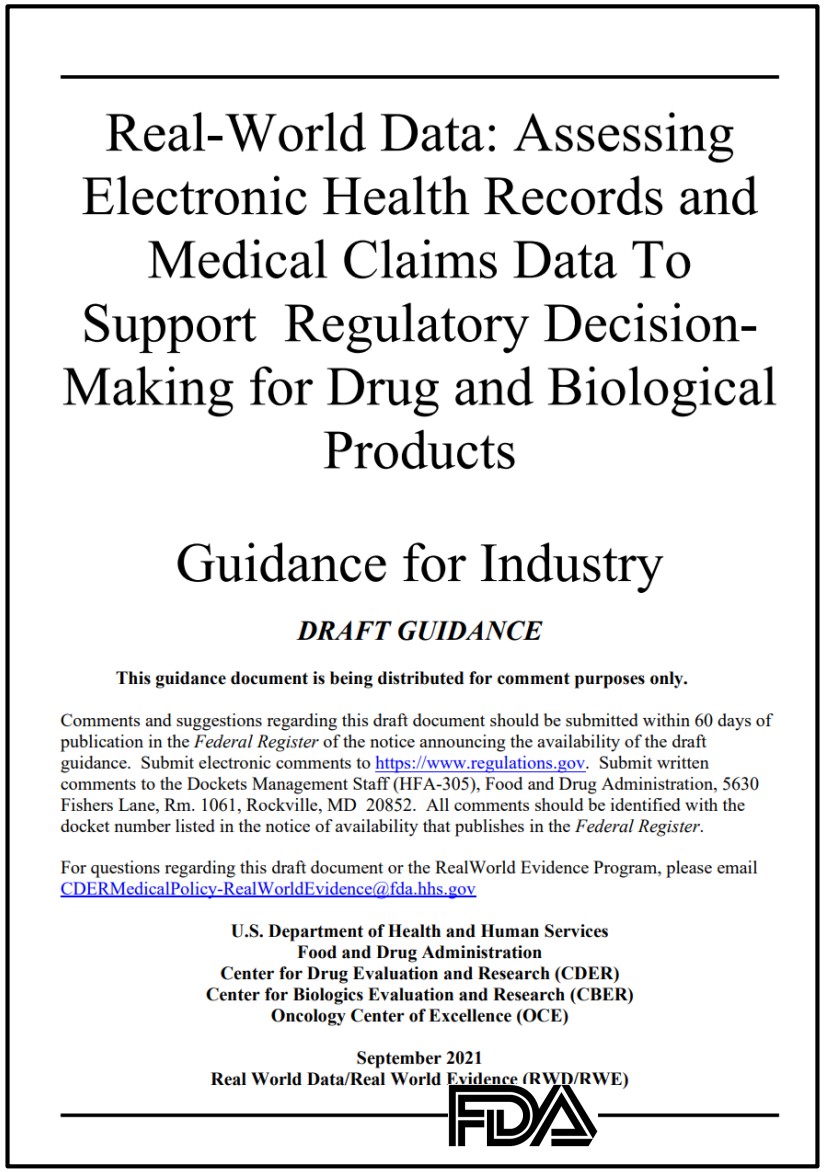
Author affiliations and article information are listed at the end of this article.

NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

https://pubmed-ncbi-nlm-nih-gov.proxy.lib.umich.edu/33443582/

"A computable phenotype definition should include **metadata** and **supporting information about the definition, its intended use, the clinical rationale or research justification for the definition, and data assessing validation** in various health care settings."

The computable phenotype definition, composed of data elements and phenotype algorithm, should be described in the protocol and study report and should also be available in a computer-processable format. Clinical validation of the computable phenotype definition should be described in the protocol and study report."



Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products

Guidance for Industry

***DRAFT GUIDANCE***

**This guidance document is being distributed for comment purposes only.**

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document or the RealWorld Evidence Program, please email CDERMedicalPolicy-RealWorldEvidence@fda.hhs.gov

**U.S. Department of Health and Human Services**
**Food and Drug Administration**
**Center for Drug Evaluation and Research (CDER)**
**Center for Biologics Evaluation and Research (CBER)**
**Oncology Center of Excellence (OCE)**

**September 2021**
**Real World Data/Real World Evidence (RWD/RWE)**

https://www.fda.gov/media/152503/download

*"I'm afraid you've had a paradigm shift."*


AMIA 25x5
Reducing Documentation Burden

NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

# The Living Textbook

*of Pragmatic Clinical Trials*

www.rethinkingclinicaltrials.org


NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

## ELECTRONIC HEALTH RECORDS–BASED PHENOTYPING

### SECTION 1

# Introduction

+ Contributors

In the context of electronic health records (EHRs), a "computable phenotype," or simply "phenotype," is a clinical condition or characteristic that can be ascertained by means of a computerized query to an EHR system or clinical data repository using a defined set of data elements and logical expressions. These queries can identify patients with particular conditions and can be used to support a variety of purposes, including population management, quality measurement, and observational and interventional research. Standardized computable phenotypes can facilitate large-scale pragmatic clinical trials across multiple healthcare systems while ensuring reliability and reproducibility (Richesson et al 2013).

In this chapter, we offer an overview of considerations for identifying, defining, and evaluating computable phenotypes, focusing in particular on standardization efforts within the NIH Pragmatic Trials Collaboratory.

Next Section

## SECTIONS



## RESOURCES

**Advances at the Intersection of Digital Health, Electronic Health Records and Pragmatic Clinical Trials**: An NIH Collaboratory Grand Rounds EHR Workshop Series

Keynote: Can the COVID-19 Crisis Lead to Evolution of the Evidence Generation Ecosystem?; NIH Collaboratory Grand Rounds; May 1, 2020

# USING ELECTRONIC HEALTH RECORD DATA IN PRAGMATIC CLINICAL TRIALS

## SECTION 1

# Introduction

— Contributors

Rachel Richesson, MS, PhD, MPH Richard Platt, MD, MSc Gregory Simon, MD, MPH Lesley Curtis, PhD Reesa Laws, BS

Adrian Hernandez, MD, MSH Jon Puro, MPA-HA Doug Zatzick, MD Erik van Eaton, MD, FACS Vincent Mor, PhD

**Contributing Editor**
Karen Staman, MS

Some material in this chapter is based on the Acquiring and Using Electronic Health Record Data white paper originally written by Zozus et al.

Using electronic health record (EHR) data for research is fundamentally different than collecting the research data prospectively, as is traditional for controlled clinical trials. Several features of EHR systems create these important differences, most importantly being the lack of investigator control over data collection and recording processes in health care facilities. Other factors include the lack of standard definitions for identifying patient cohorts and study-specific outcomes, the challenges associated with completeness of longitudinal data, and potential errors in linkage of records across systems. All of these challenge investigators to assure and demonstrate that data are of adequate quality to support research conclusions. While many of the issues addressed in this chapter apply to a broad range of study designs that might use data from the EHR, this chapter describes the use

## SECTIONS



## RESOURCES

Acquiring and Using Electronic Health Record Data

NIA IMPACT COLLABORATORY
TRANSFORMING DEMENTIA CARE

# Questions?

richessr@med.umich.edu

@IMPACTcollab1

@rrichesson

IMPACTcollaboratory.org