Big Data Phenomics in the VA

Mary Whooley MD
Director, VA Measurement Science QUERI
San Francisco VA Health Care System
University of California, San Francisco

Kelly Cho PhD MPH
Phenomics Lead, Million Veteran Program
VA Boston Health Care System
Harvard Medical School

Academy Health
Annual Research Meeting
June 27, 2017

DISCOVERY ★ INNOVATION ★ ADVANCEMENT

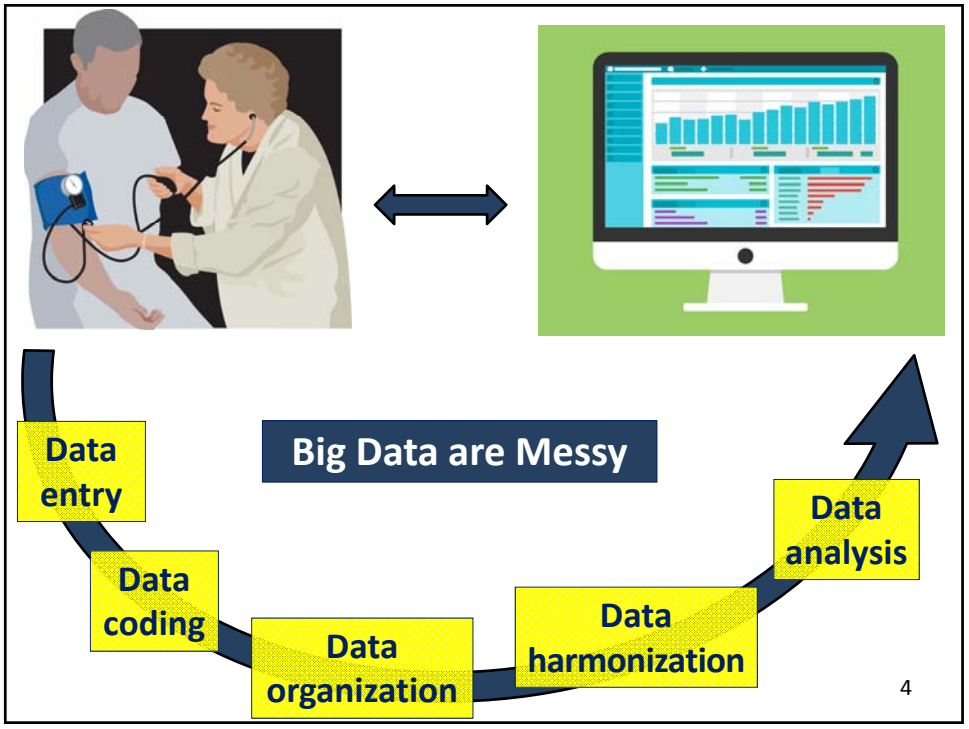
Outline

- Importance of data standardization and interoperability
- PCORnet and the Observational Medical Outcomes Partnership (OMOP) Common Data Model
- Million Veteran Program (use case)
- Coding algorithms for computable phenotypes

DISCOVERY ★ INNOVATION ★ ADVANCEMENT

2

3


Big Data are Messy
Data entry
Data coding
Data organization
Data harmonization
Data analysis

4

VA Information Systems Technology Architecture (VistA)

Example: How can we identify uncontrolled diabetics?

Value	LabChemTestName	units
7.5	A1C HGB	%
6.2	A1C GLYCOHEMOGLOBIN	% T HGB
5.9	GLYCOHEMOGLOBIN	% TOT HGB
8.3	GLYCOSYLATED HEMOGLOBIN	%TOTAL HBG
5.2	HBA1C	% gHbA1c
5.4	GLYCOSYL HEMOGLOBIN A1c	% HbA1C
5.7	HGB A1C	PERCENT
9.3	HEMOGLOBIN %A1C	%TOTAL
6.4	GLYCOHEMOGLOBIN (A1C)	%HGBA1C
7.3	GLYCOHEMOGLOBIN (GHB)	% HBG
11.2	GLYCOHGB	% OF TOTAL
6.8	HEMOGLOBIN A1C PANEL	% HBG
5.8	TOTAL A1C	% A1C

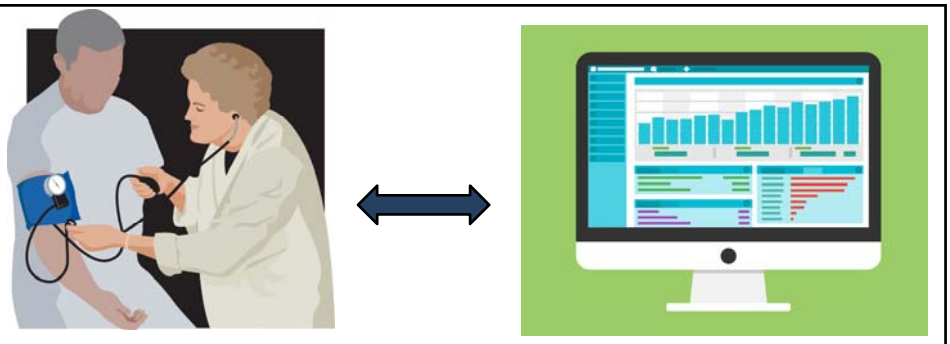
Logical Observation Identifiers Names and Codes

Example: How can we identify uncontrolled diabetics?

Value	LabChemTestName	units	LOINC
7.5	A1C HGB	%	4548-4
6.2	A1C GLYCOHEMOGLOBIN	% T HGB	4548-4
5.9	GLYCOHEMOGLOBIN	% TOT HGB	4548-4
8.3	GLYCOSYLATED HEMOGLOBIN	%TOTAL HBG	4548-4
5.2	HBA1C	% gHbA1c	4548-4
5.4	GLYCOSYL HEMOGLOBIN A1c	% HbA1C	4548-4
5.7	HGB A1C	PERCENT	4548-4
9.3	HEMOGLOBIN %A1C	%TOTAL	4548-4
6.4	GLYCOHEMOGLOBIN (A1C)	%HGBA1C	4548-4
7.3	GLYCOHEMOGLOBIN (GHB)	% HBG	4548-4
11.2	GLYCOHGB	% OF TOTAL	4548-4
6.8	HEMOGLOBIN A1C PANEL	% HBG	4548-4
5.8	TOTAL A1C	% A1C	4548-4

VA Corporate Data Warehouse Data Tables

- ☐ Allergy
- ☐ Appointment
- ☐ Consult
- ☐ CPRS Orders
- ☐ Dental
- ☐ Emergency Dept.
- ☐ Health Factors
- ☐ Health Benefits Request
- ☐ Immunization
- ☐ Inpatient
- ☐ Lab Microbiology
- ☐ Lab Chem
- ☐ Mental Health
- ☐ Non-VA Meds
- ☐ Outpatient
- ☐ Patient Enrollment
- ☐ Patient Insurance
- ☐ Patient Record Flag
- ☐ PCMM (Primary Care
- ☐ Pharmacy BCMA (Bar Code Medication
- ☐ Pharmacy Outpatient
- ☐ Pharmacy Patient
- ☐ Purchased Care (formerly fee)
- ☐ SStaff
- ☐ Surgery PRE, INTRA, and POST(
- ☐ VistA Waitlist
- ☐ VistA Compensation & Pension
- ☐ Vital Signs
- ☐ Women's Health



Big Data are Messy

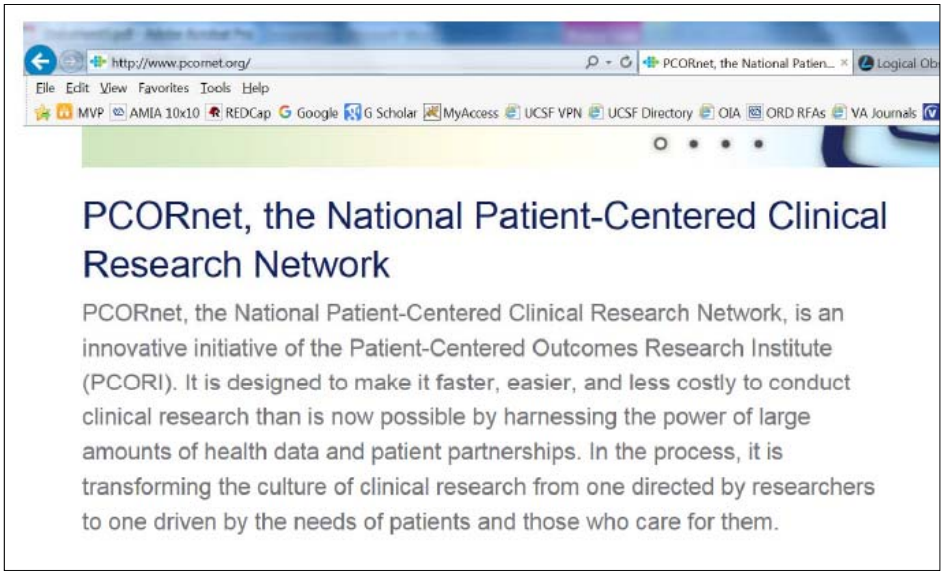
Data entry → Data coding → Data organization → Data harmonization → Data analysis

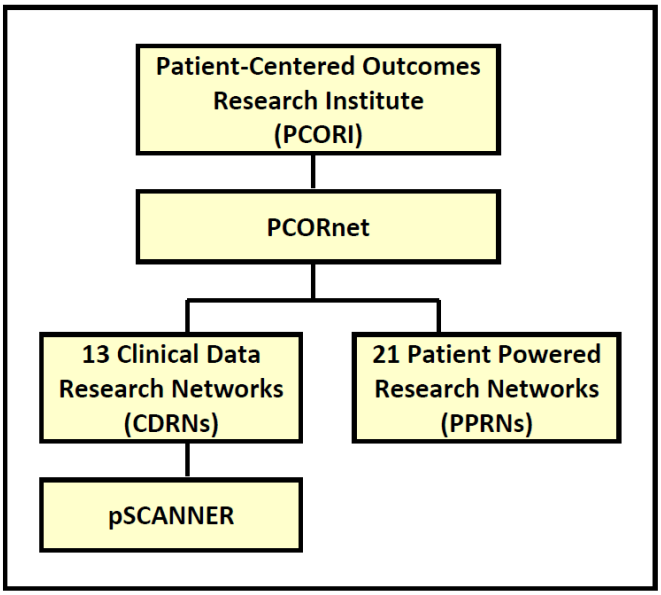
10

Outline

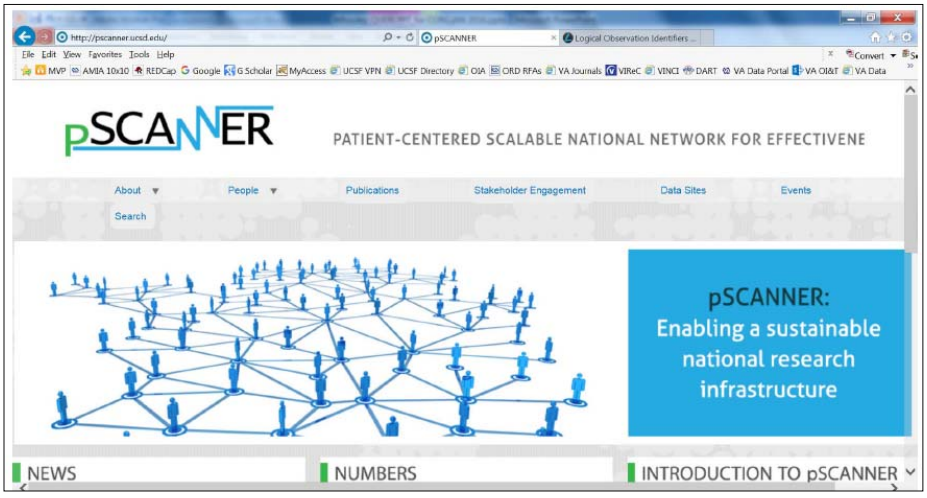
- Importance of data standardization and interoperability
- **PCORnet and the Observational Medical Outcomes Partnership (OMOP) Common Data Model**
- **Million Veteran Program (use case)**
- **Coding algorithms for computable phenotypes**

http://www.pcornet.org/



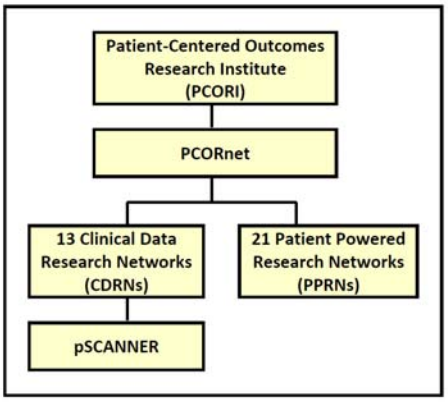


13



http://pscanner.ucsd.edu/

14



Organizations participating in pSCANNER

(1) UC San Diego
(2) UC San Francisco
(3) UC Irvine
(4) UC Davis
(5) UC Los Angeles
(6) VA (>150 medical centers)
(7) University of Southern California
(8) University of Colorado
(9) University of Washington
(10) San Mateo Medical Center
(11) Cedars-Sinai Medical Center

http://pscanner.ucsd.edu/

Transforming the National Department of Veterans Affairs Data Warehouse to the OMOP Common Data Model

Fern FitzHenry[1,2], Jesse Brannen[1], Jason Denton[1,2], Jonathan R. Nebeker[3,4], Scott L. DuVall[3,4], Freneka Minter[1,2], Jeffrey Scehnet[3], Brian Sauer[3,4], Lucila Ohno-Machado[5], Michael E. Matheny[1,2]

[1]Tennessee Valley Healthcare System, Veterans Affairs Medical Center, Nashville, TN; [2]Vanderbilt University, Nashville, TN; [3]VA Salt Lake City Health Care System, Salt Lake City, UT; [4]University of Utah, Salt Lake City, UT; [5]Bioinformatics and Systems Biology, University of California, San Diego, CA;

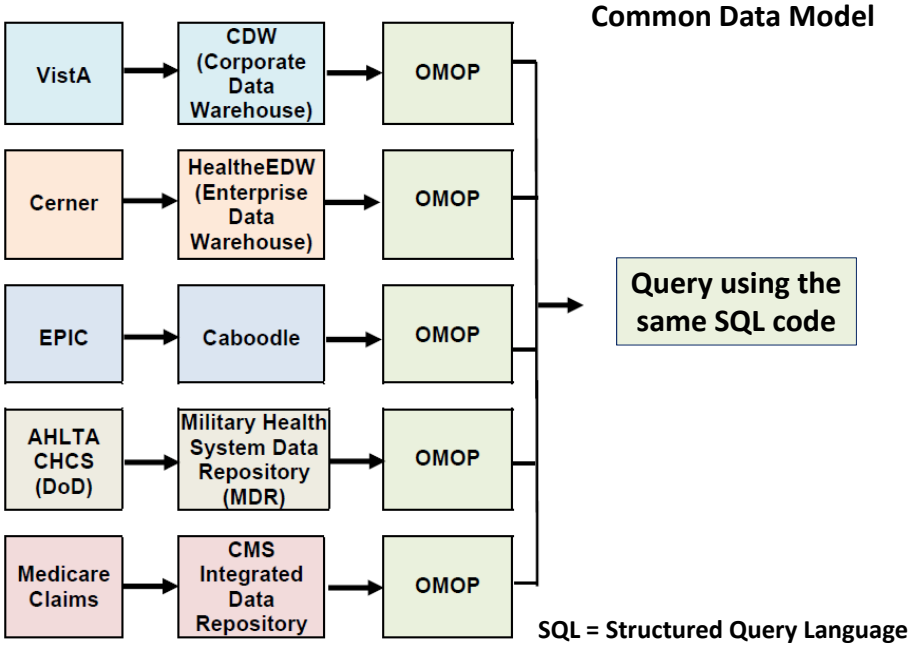
Abstract: To describe the conversion of the national Department of Veterans Affairs (VA) healthcare network's corporate data warehouse to the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) suitable for distributed observational research. Observational outcomes from electronic medical record systems are becoming more important in comparative effectiveness research, particularly as post marketing surveillance research.[1]

2000 to present

- **16 million unique patients**
- **11 million w/ at least one encounter**
- **5 million deaths**
- **3 billion procedures**
- **2.5 billion conditions**
- **973,000 providers**

Abstract presented Nov 2015
Am Medical Informatics Assoc

Mapping to Observational Medical Outcomes Partnership (OMOP) Common Data Model



SQL = Structured Query Language

Observational Outcomes Partnership (OMOP) Common Data Model Implementations

> 600 million patients worldwide

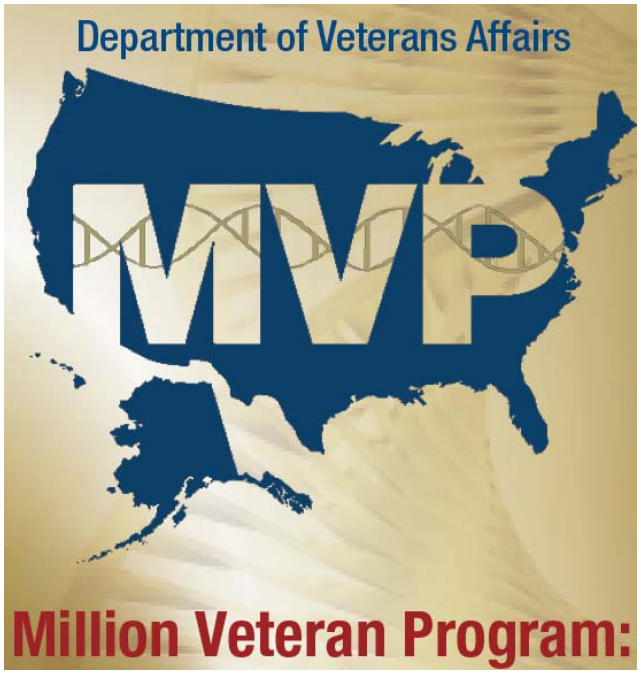
DISCOVERY ★ INNOVATION ★ ADVANCEMENT

Research & Development

18

Outline

- Importance of data standardization and interoperability
- PCORnet and the Observational Medical Outcomes Partnership (OMOP) Common Data Model
- **Million Veteran Program (use case)**
- **Coding algorithms for computable phenotypes**



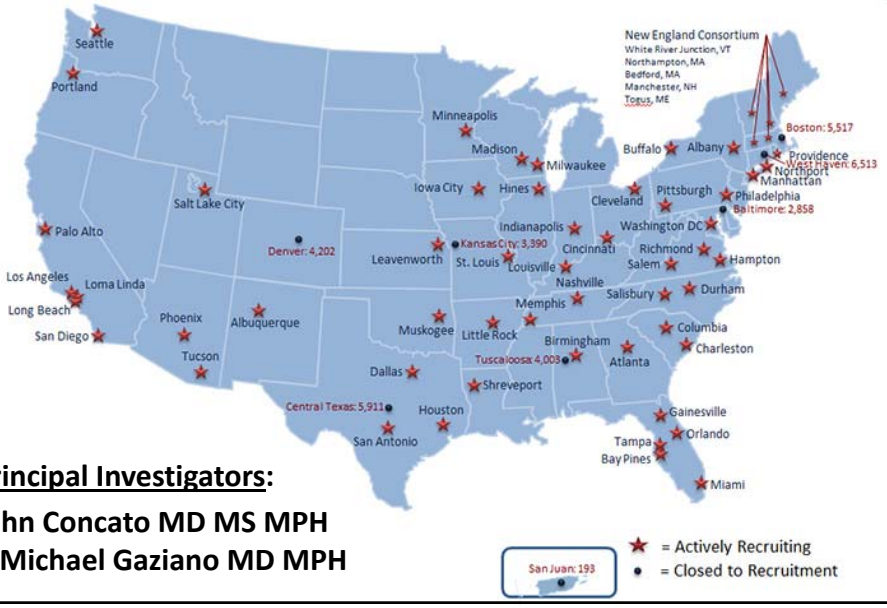
Million Veteran Program (MVP)

- National VA research initiative aiming to enroll one million users of the VHA in an observational cohort
- Over 500,000 patients already enrolled
- Blood collection for genotyping and storage
- Access to electronic medical record
- Goal is to create database of genomic, military exposure, lifestyle and electronic health information

DISCOVERY ★ INNOVATION ★ ADVANCEMENT

Veterans Health Administration Research & Development Improving Veterans' Lives www.research.va.gov

Currently enrolling at >50 VHA Facilities

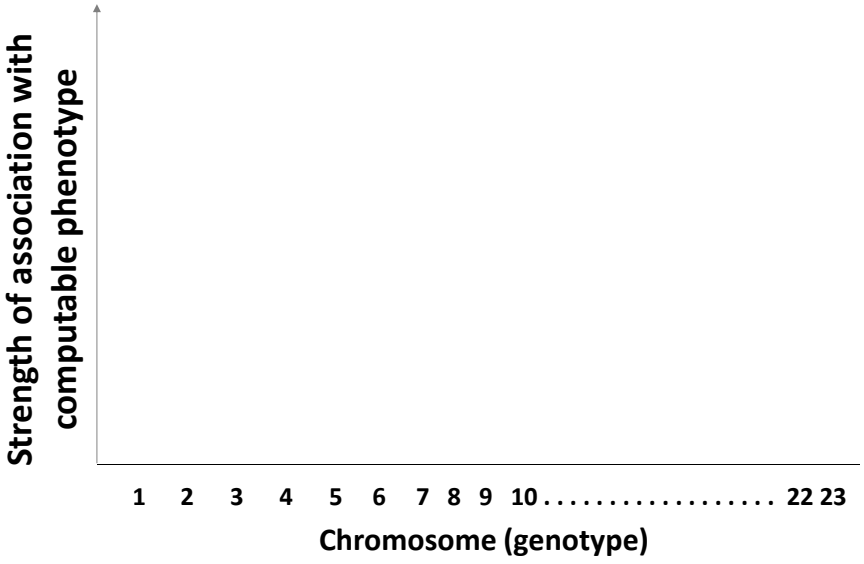


Principal Investigators:

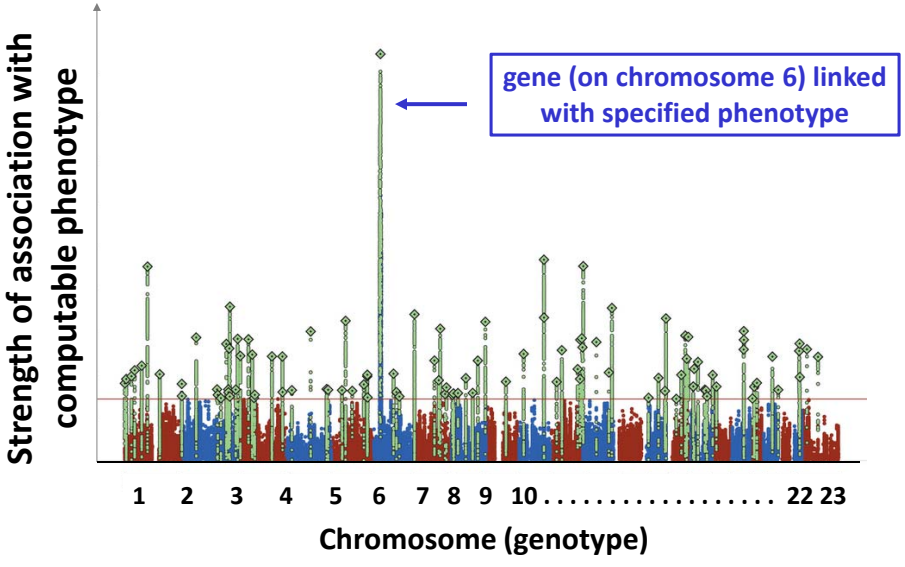
John Concato MD MS MPH

J. Michael Gaziano MD MPH

Genome-wide association study (GWAS): identify genotype(s) associated with specified phenotype



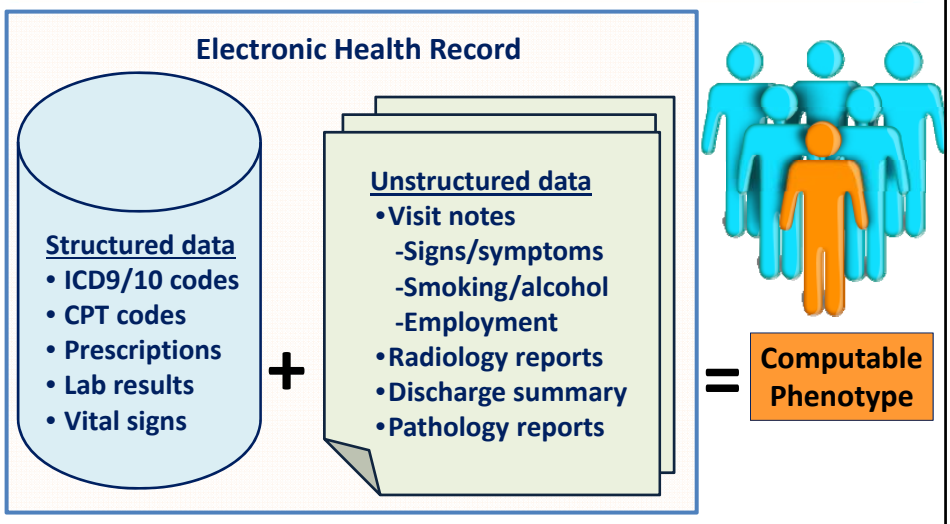
Genome-wide association study (GWAS): identify genotype(s) associated with specified phenotype



Outline

- Importance of data standardization and interoperability
- PCORnet and the Observational Medical Outcomes Partnership (OMOP) Common Data Model
- Million Veteran Program (use case)
- **Coding algorithms for computable phenotypes**

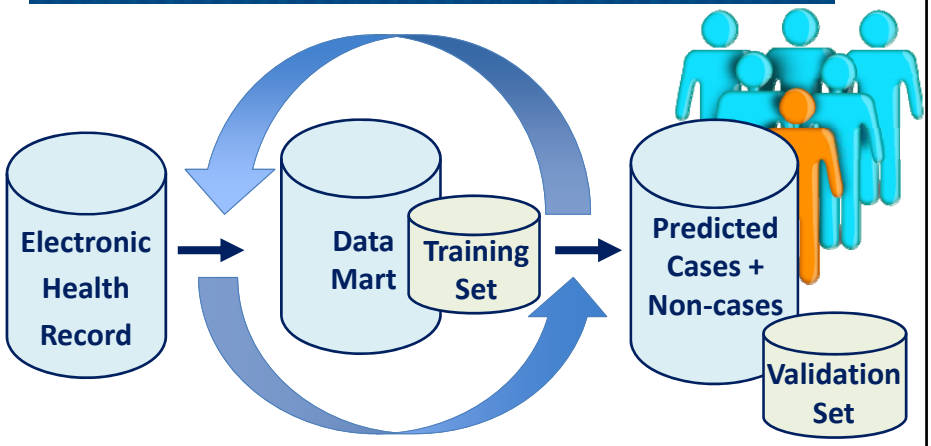
What is a computable phenotype?



Phenotype Algorithms – https://phekb.org/phenotypes

Phenotype	Methods	Owner
Atrial Fibrillation	CPT Codes, ICD 9 Codes, Natural Language Processing	Vanderbilt
Dementia	ICD 9 Codes, Medications	eMERGE Univ Washington
Heart Failure	CPT, ICD 9 Codes, Labs, Meds, Natural Language Processing	eMERGE Mayo
Coronary Disease	CPT Codes, ICD 9 Codes	PCORI MidSouth CDRN
Sleep Apnea	CPT Codes, ICD 9 Codes	Beth Israel Deaconess
Type 2 Diabetes	ICD 9 Codes, Labs, Medications	eMERGE Northwestern
Venous Thromboembolism	CPT, ICD 9 Codes, Vital Signs Natural Language Processing	eMERGE Mayo

What is Natural Language Processing?



1. Identify cases and non-cases (often requires chart review)

2. Iteratively refine & test classification algorithm

3. Validate final algorithm (probabilistic approach)

28

Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network

Katherine M Newton,[1] Peggy L Peissig,[2] Abel Ngo Kho,[3] Suzette J Bielinski,[4] Richard L Berg,[2] Vidhu Choudhary,[2] Melissa Basford,[5] Christopher G Chute,[6] Iftikhar J Kullo,[7] Rongling Li,[8] Jennifer A Pacheco,[3] Luke V Rasmussen,[3] Leslie Spangler,[1] Joshua C Denny[9]

J Am Med Inform Assoc 2013

Extracting research-quality phenotypes from electronic health records to support precision medicine

Wei-Qi Wei[1] and Joshua C Denny[1,2*]

Genome Medicine 2015

29

MVP Phenomics Group

Mission:

1) to provide a phenotyping framework for MVP Phenomics Science
2) to manage and coordinate resources for MVP phenotyping projects
3) to play a leading role towards "*Mapping the Human Phenome*"

Organization:

Kelly Cho PhD MPH	Lead, MVP Phenotyping
Scott DuVall PhD	Lead, MVP-VINCI Collaboration
Jackie Honerlaw RN MPH	Manager, Phenomics Core
Kevin Malohi BS	Manager, VINCI Data Services
Mai Nguyen PhD	Manager, MVP Data Analytics
Anne Ho MPH	Lead, MVP Data Management
David Gagnon MD PhD	Lead, Biostatistics and Data Science

DISCOVERY ★ INNOVATION ★ ADVANCEMENT

Research & Development

30

Summary – Big Data Phenomics in the VA

- Big data are messy
- VA EHR data have been mapped to national VA Corporate Data Warehouse (CDW)
- CDW data have been transformed to OMOP Common Data Model
- Million Veteran Program actively using these data
- Phenotype algorithms can be shared at PheKB.org

32