

Practical Methods for Working with EHR Data

CIDA Big Data Seminar Series

Laura K. Wiley, PhD
Assistant Professor
Biomedical Informatics & Personalized Medicine

 @GenomeGal



University of Colorado
Anschutz Medical Campus

COLORADO CENTER FOR
PERS**NALIZED**
MEDICINE

Overview

- Introduction to EHRs
- Landscape of EHR-based Research
 - Types & Limitations of EHR Data
 - Phenome-Wide Association Studies
 - Computational Phenotyping
- CU Resources and How to Learn More

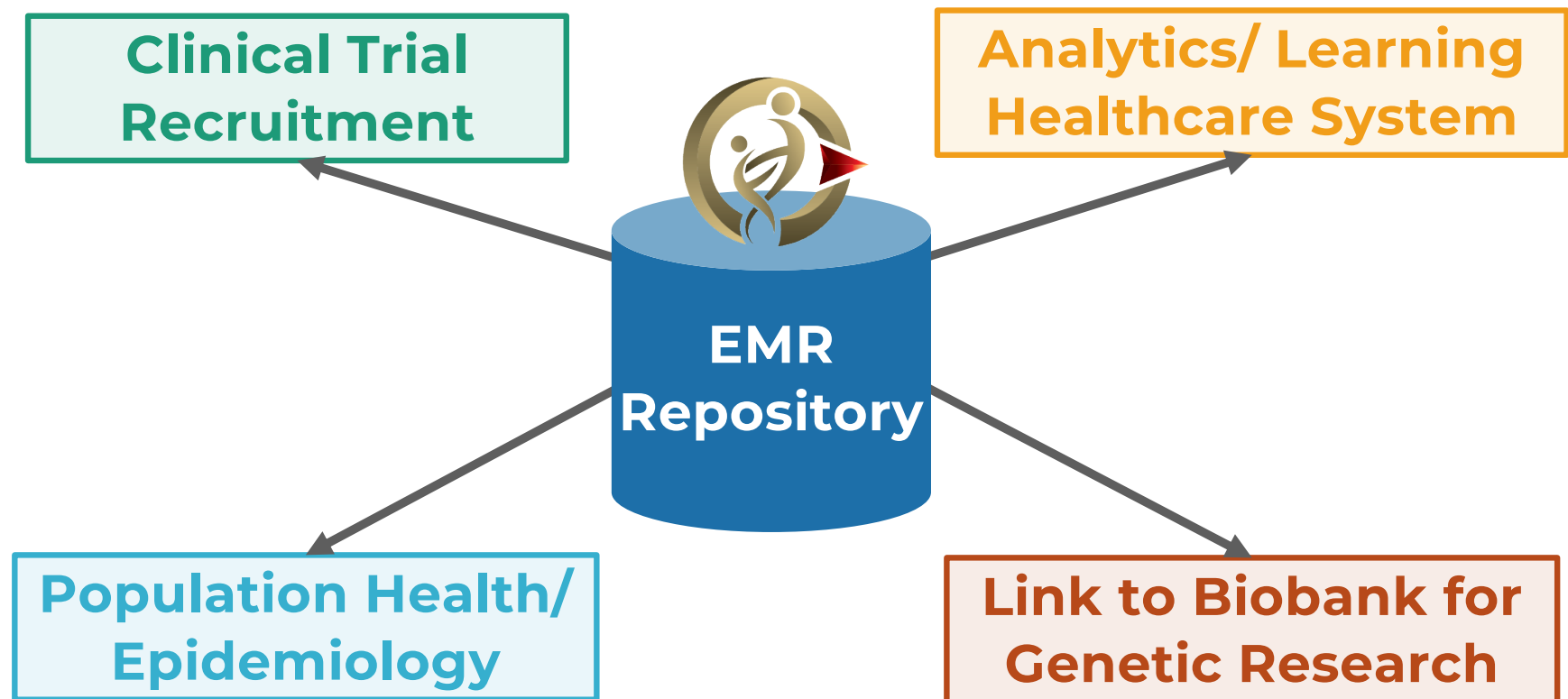
What is an EHR?

The diagram on the left illustrates the components of an EHR system, represented by yellow circles with arrows pointing to a central vertical bar. The components are:

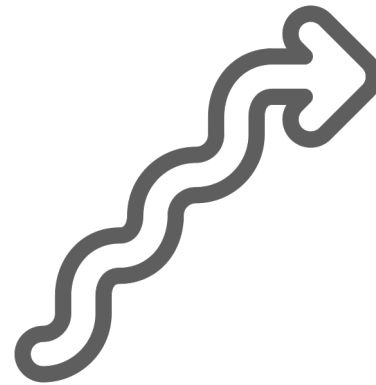
- clinic notes/forms
- meds
- labs
- demographics
- registries
- vitals
- billing codes
- outpatient encounters

The screenshot on the right shows the Epic EHR interface for a patient named Testmonday, Melissa A. The interface includes a top navigation bar with various tools and a main content area with tabs for Flowsheets, Summary, and Pastoral Services. The Flowsheets tab is active, showing a list of flowsheets and a detailed view of the 'Pastoral Services Follow-up' flowsheet. A 'Selection Form' dialog box is open, allowing the user to select options for the follow-up. The bottom of the screenshot shows the Windows taskbar with the time 4:01 PM on 5/20/2018.

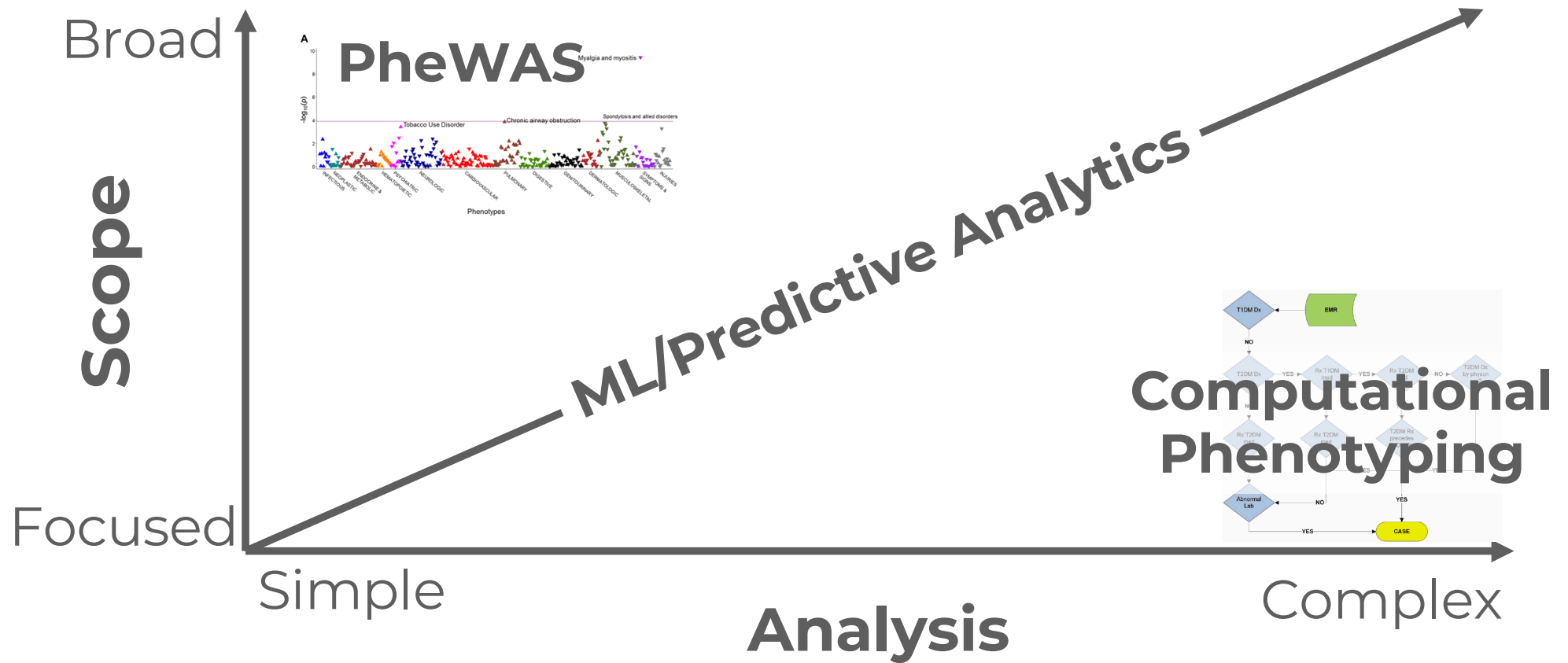
How to Leverage EHRs for Research?



EHRs Require Paradigm Shift



EHR – Based Research Landscape



What Types of Data are in an EHR?

Data
bases

Structured

Diagnosis codes		
Fake ID	ENTRY_DAT	CODE
34068	5/13/2001	41.85
37660	8/6/2002	79.99
140680	8/31/2003	79.99
23315	5/14/2003	112
75936	7/9/2004	117.9

Lab tests			
Fake ID	TEST	ENTRY_DAT	VALU
3536	pO2	1/23/1996	314
72921	LDL	2/5/1996	34
102460	pCO2	1/26/1996	45
135043	HDL	1/25/1996	35
135432	MonAb	1/24/1999	0.16

Semi-Structured

Problem lists:
---- Medications known to be prescribed:
Keppra 750 mg 1/2 tab q am and pm
Dexilant 60 mg by mouth daily
aspirin 325 mg 1 tablet by mouth daily
clopidogrel 75 mg tablet 1 tablet by mouth daily

---- Known adverse and allergic drug reactions:
Sulfa Drugs

---- known significant medical diagnoses:
Seizure disorder
Aneurysm
Heartburn

---- Known significant operative and invasive procedures:
2003 Appendectomy
2005 Stents put in **DATE [Aug 29 05]

Unstructured

Clinical notes
EXAM: BILATERAL DIGITAL SCREENING MAMMOGRAM WITH CAD, **DATE[Mar 16 01]:
COMPARISON: **DATE[Jul 01 01]
TECHNIQUE: Standard CC and MLO views of both breasts were obtained. FINDINGS: The breast parenchyma is heterogeneously dense. The pattern is extremely complex with postsurgical change seen in the right upper outer quadrant and scattered benign-appearing calcification seen bilaterally. A possible asymmetry is seen in the superior aspect of the left breast. The parenchymal pattern otherwise remains stable bilaterally, with no new distortion or suspicious calcifications. IMPRESSION: RIGHT: No interval change. No current evidence of malignancy.. LEFT: Possible developing asymmetry superior aspect left breast for which further evaluation by true lateral and spot compression views recommended. Ultrasound may also be needed.. RECOMMENDATION: Left diagnostic mammogram with additional imaging as outlined above.. A left breast ultrasound may also be needed. BI-RADS Category 0: Incomplete Assessment - Need additional imaging evaluation. IMPRESSION: RIGHT: No interval change. No current evidence of malignancy....

NLP
Tools

Wei, et. al.
Genome Med 2015
PMID: [25937834](https://pubmed.ncbi.nlm.nih.gov/25937834/)

Structured Data – ICD Codes

- International Classification of Disease (ICD)
- In US, we use ICD-10-CM (ICD-9-CM < October 2015)
- Diagnostic codes:
 - ICD-9-CM: ~13,500
 - ICD-10-CM: >70,000



W59.22XD
Struck by Turtle,
Subsequent Encounter

Structured Data – ICD Codes

False Positives



False Negatives



Structured Data – Laboratory Data

Positives

- Closer to Biology!
- Continuous Variable
 - Moar Power!
 - No controls!
- Look at Extremes

Challenges

- Multiple panels w/ same Labs
- Multiple lab values/ person
- Method & Reference range change over time
- UNITS!
- Junk
 - (PHONE, /, <4, Duplicate Order, Mislabeled Specimen, Result)

Structured(ish) Data - Medications

Formatting

- E-Prescribed: Name, size, frequency structured
- Dosing Instructions often semi or unstructured
- Also found in problem list and clinical notes -> unstructured

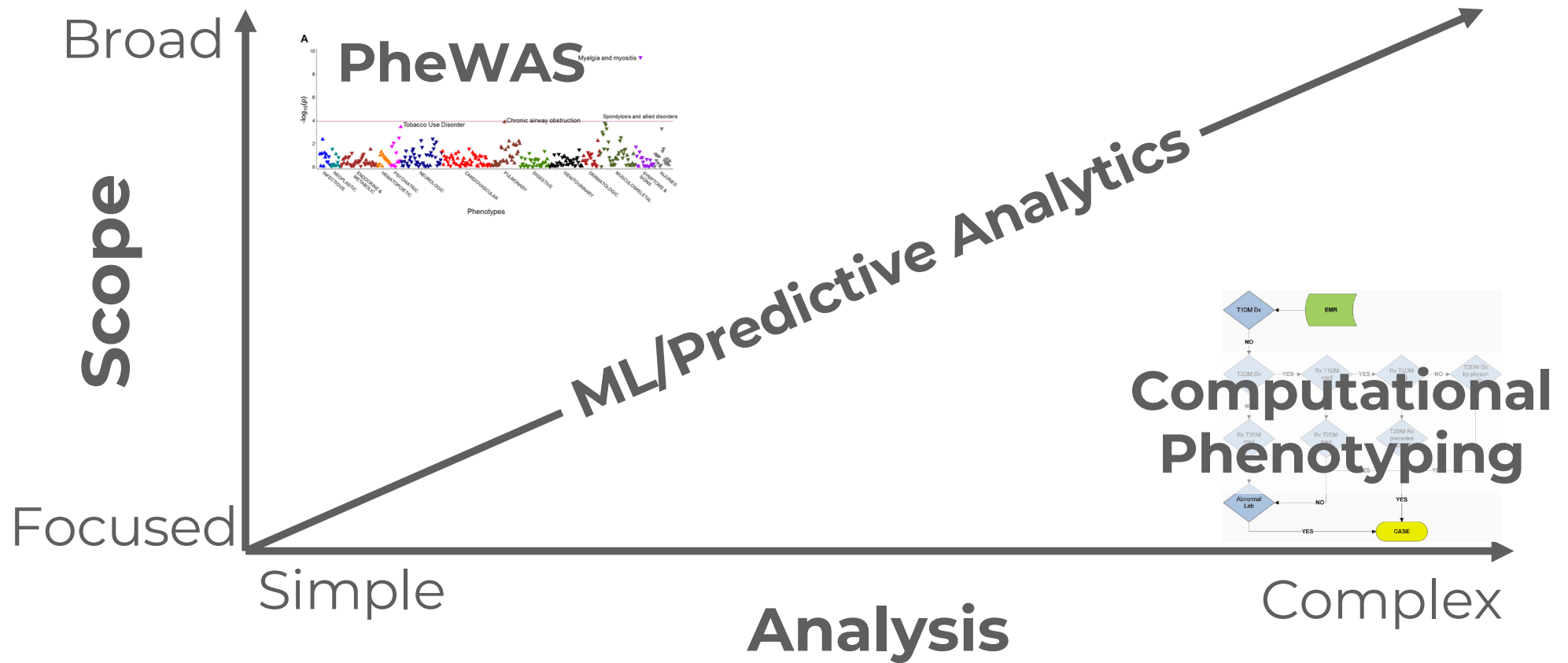
Challenges

- Atorvastatin = Lipitor
- Atorvastatin + Ezetimibe = Liptruzet
- Atorvastatin/Simvastatin both HMG-CoA Reductase Inhibitors
- Atorvastatin & Ezetimibe both cholesterol lowering drugs
- Order -> Fill -> Taking the Drug
- Identifying Discontinuation Difficult

Unstructured Data – Clinical Text

- Goal: Convert Unstructured Text -> Structured Data
- NLP Tools:
 - **Concept-indexing:**
 - “mad cow disease” → [C0120202](#)
 - “Bovine Spongiform Encephalopathy” → [C0120202](#)
 - **Negation/certainty tagging:**
 - “no evidence of diabetes”
 - **Identifying temporal expressions:**
 - “colonoscopy 5 years ago”
 - ... and much more

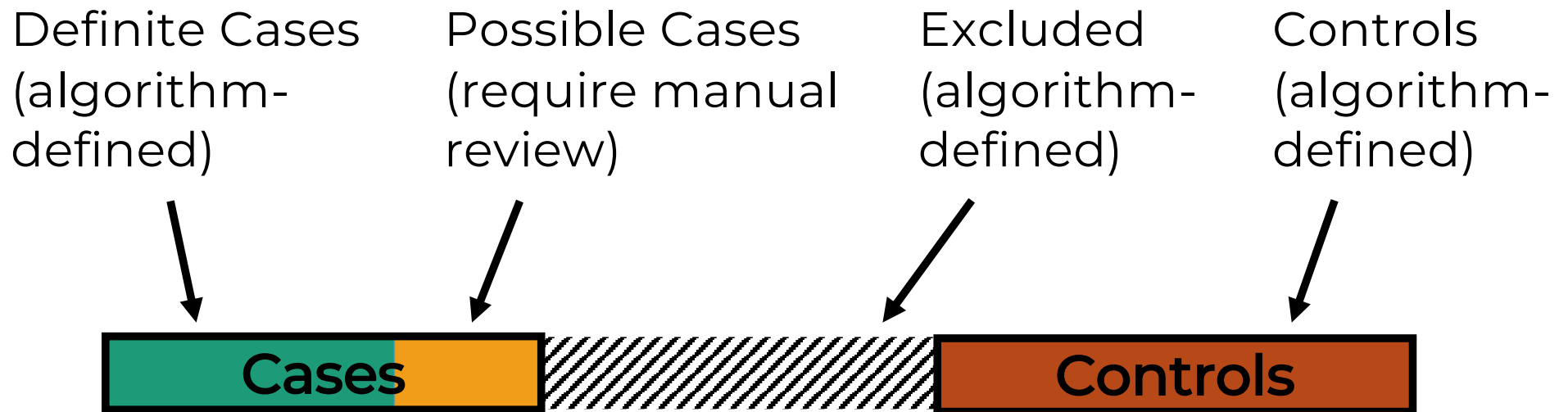
EHR – Based Research Landscape



What is (Computational) Phenotyping?

- Translates observed characteristics or manifestations of a human condition/disease
- Utilizes clinical criteria from the EMR to identify subjects meeting definition

What is Computational Phenotyping?



Computational Phenotyping Overview

Population

uhealth



Colorado Center for Personalized Medicine
Biobank

Inclusion

Must have

Exclusion

Must *NOT* Have

Case/Control Definitions

Steroid-Induced Osteonecrosis

Population:

Vanderbilt BioVU 

Inclusion:

Any **Corticosteroid**, No Minimum Dose
Intravenous, Intramuscular, or Oral
(Not Inhaled)
At least 14 days

Exclusion:

Alcohol Abuse, Sickle Cell, Gaucher, Legge-
Calve-Perthes, HIV, Organ Transplant

Cases:

ICD9: **733.4X**

OR

Keywords: osteonecrosis, avascular
necrosis, osteochondritis dessicans

Controls:

Not a Case

AND

No record of bisphosphonates

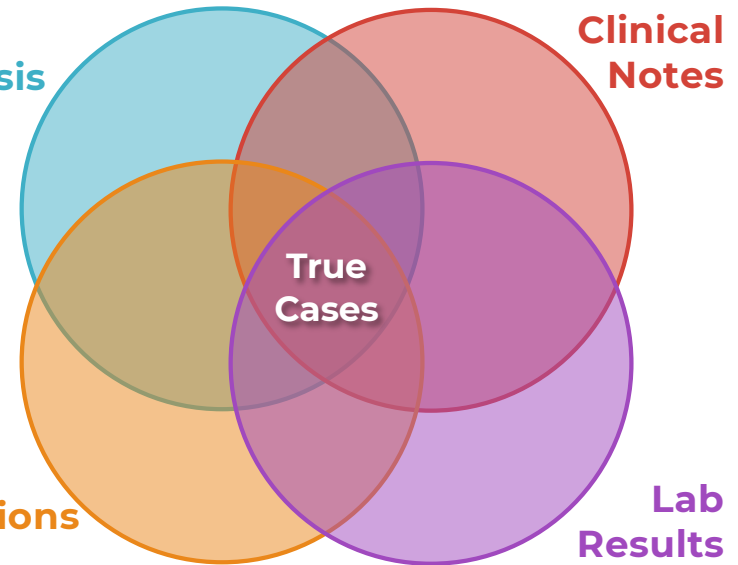
Billing/
Diagnosis
Codes

Clinical
Notes

True
Cases

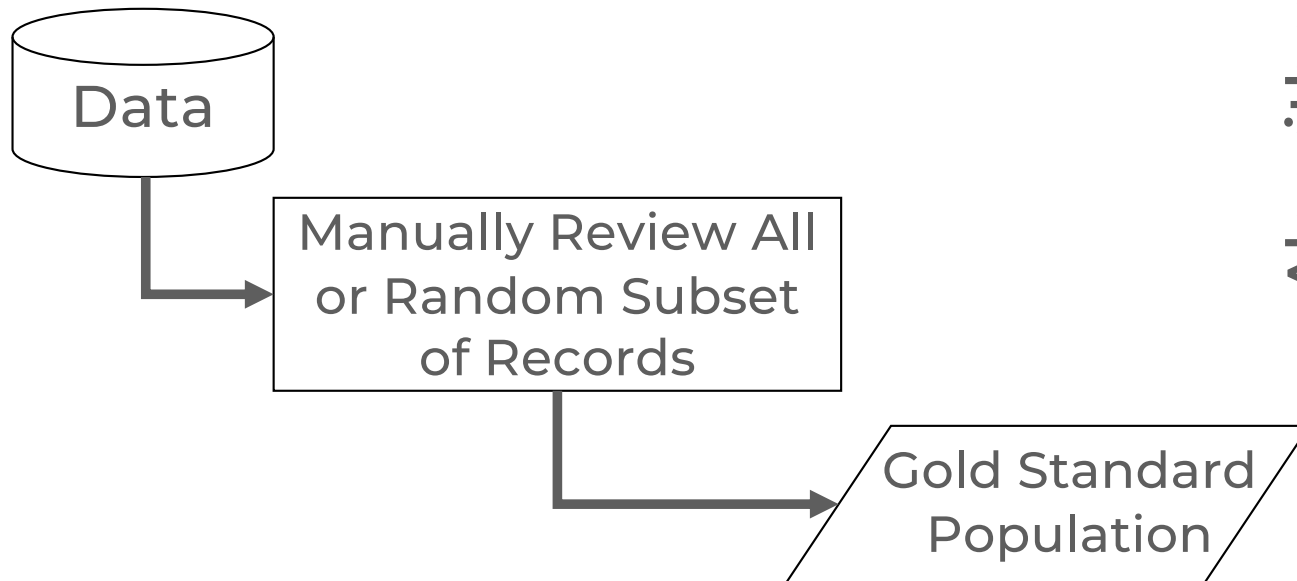
Medications

Lab
Results



How Do You Know It Worked?

Gold-Standard Population

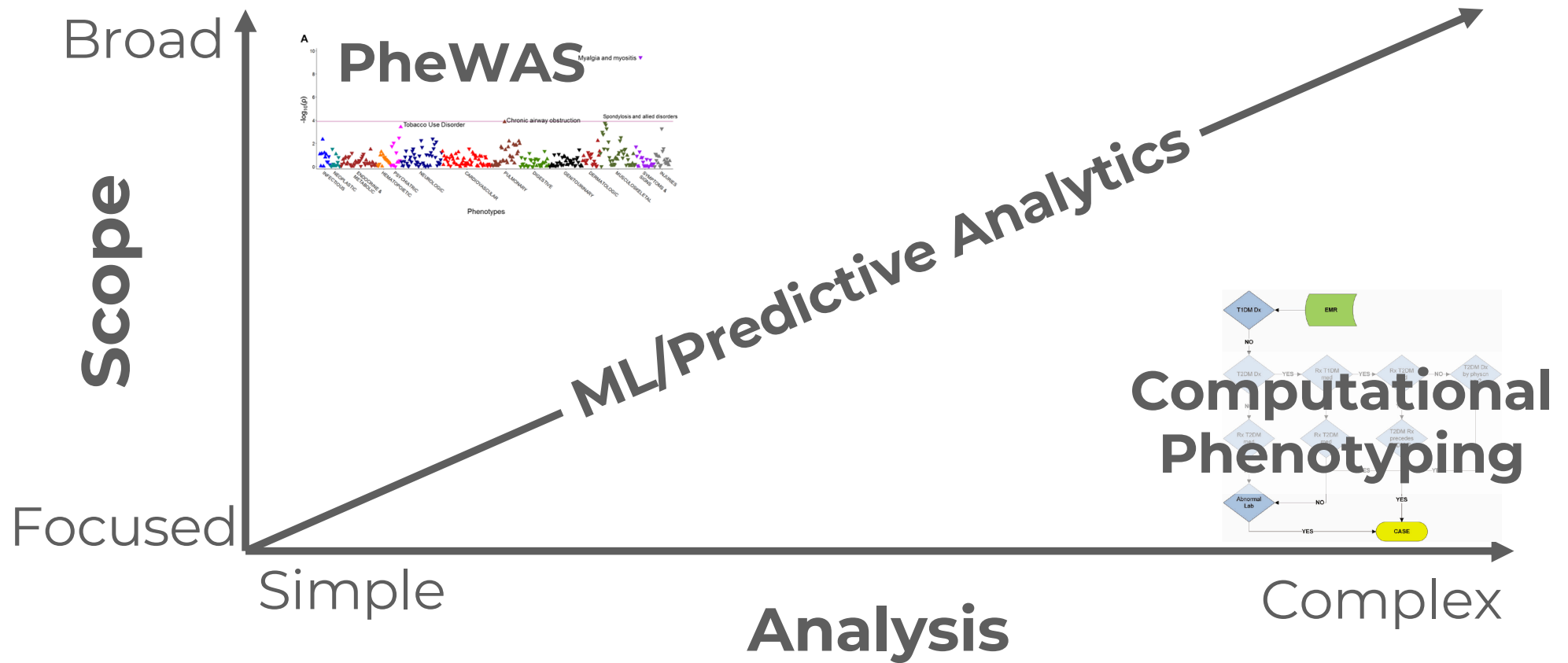


		Gold Standard Truth	
		+	-
Algorithm	+	True Positive	False Positive
	-	False Negative	True Negative

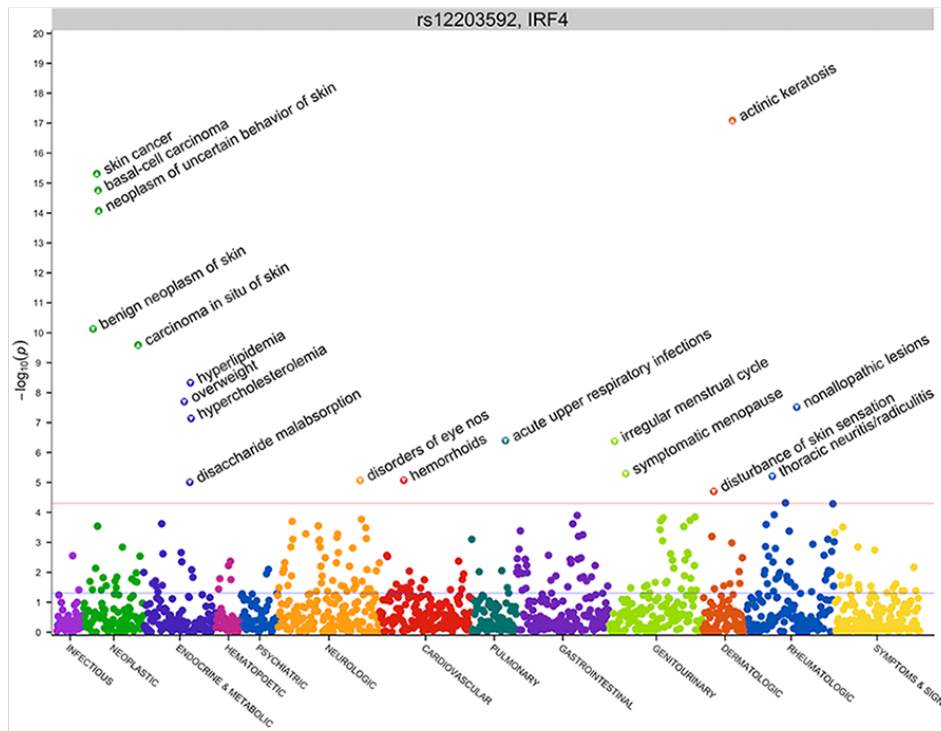
Caution – Danger Ahead

- Defining Controls is *HARD*
 - Clinical?
 - Biological?

EHR – Based Research Landscape

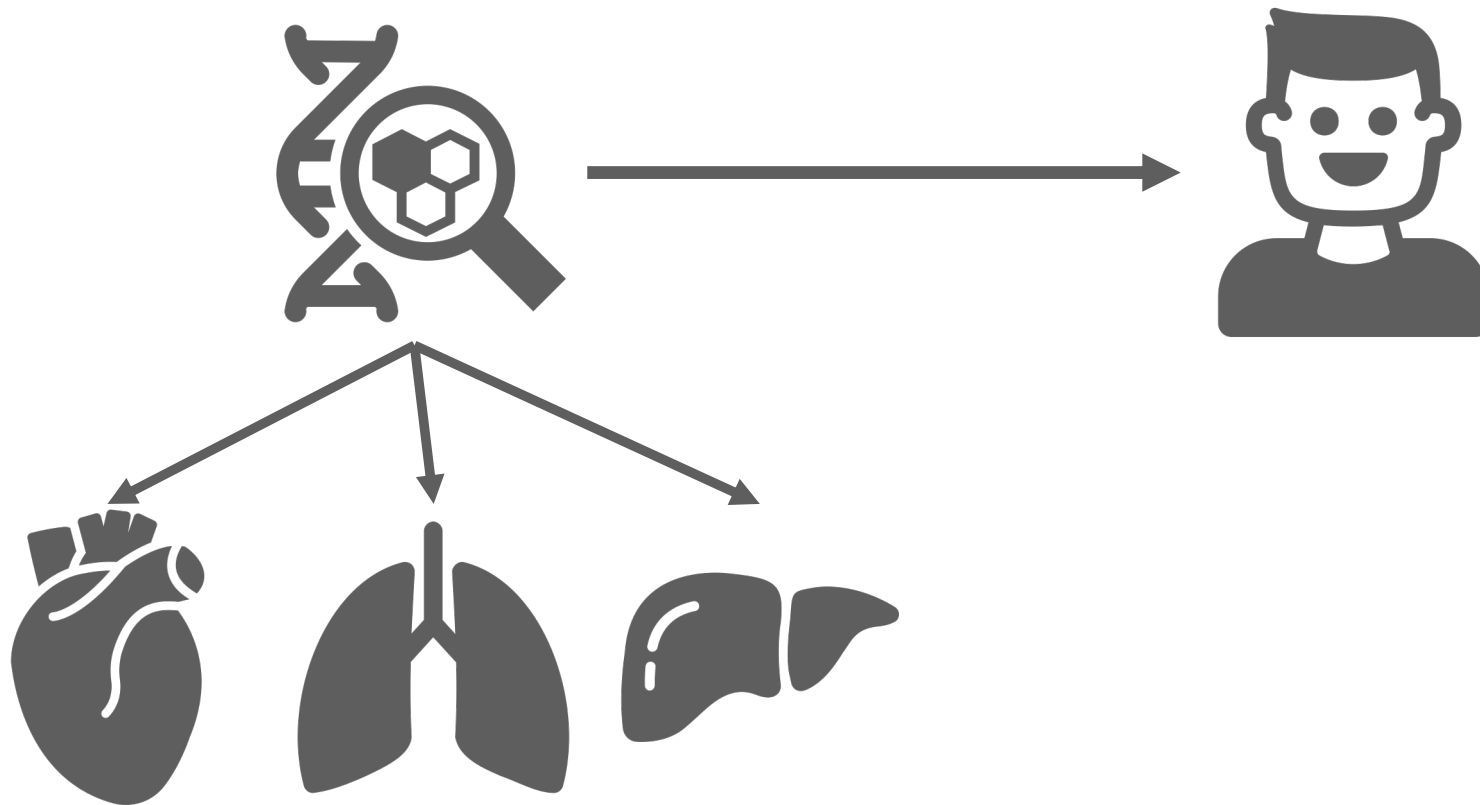


Phenome-Wide Association Study (PheWAS)



Denny JC et al. [PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations](#). Bioinformatics. 2010 May 1;26(9):1205-10.

When to Use PheWAS?

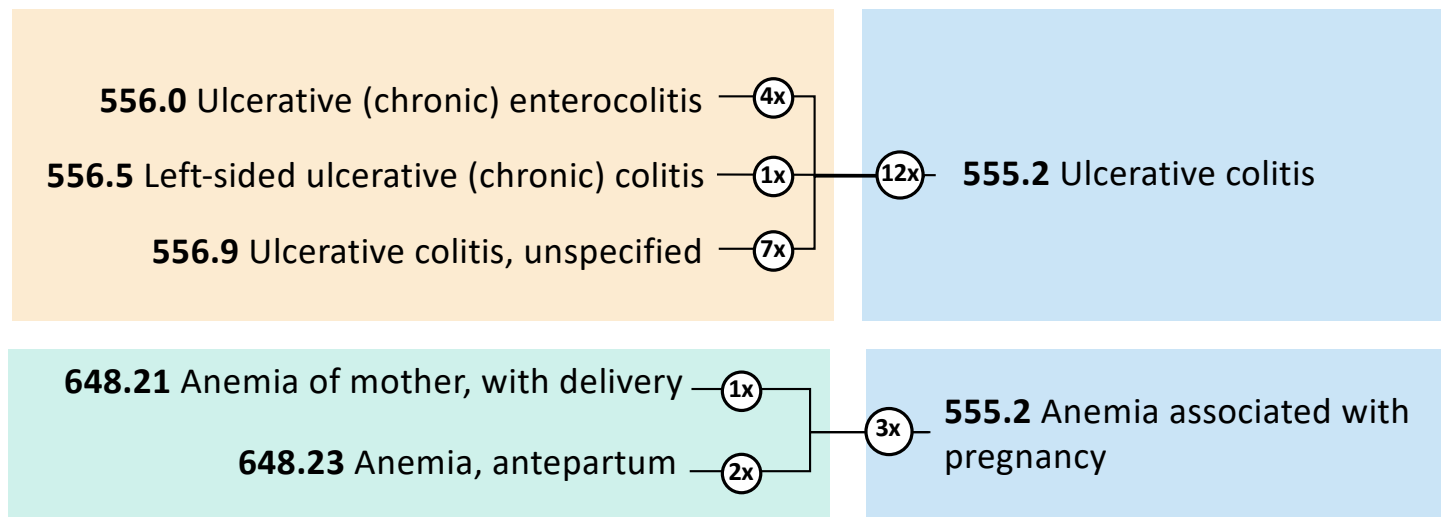


How Does PheWAS Work?

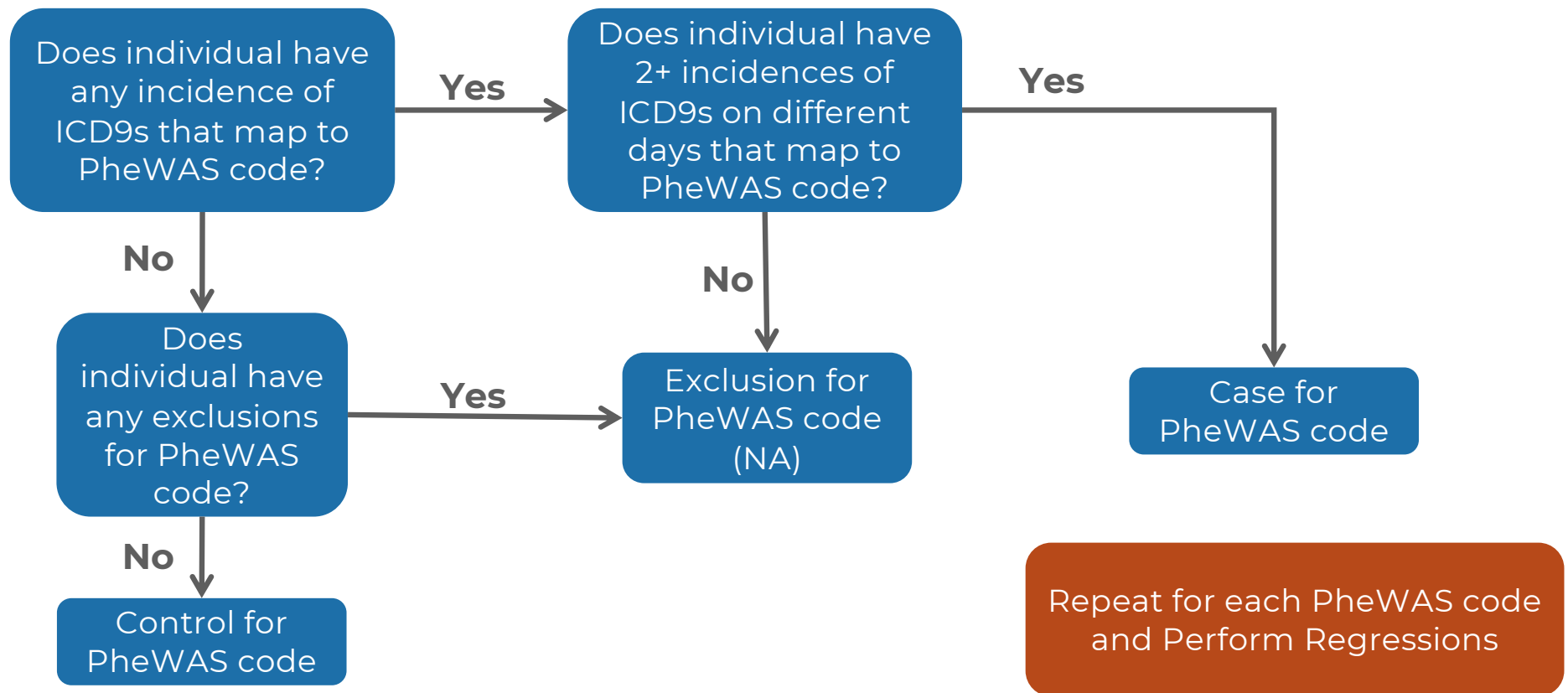
**ICD billing
code in EMR**



**PheWAS
code**



How Does PheWAS Work?



PheWAS R Package

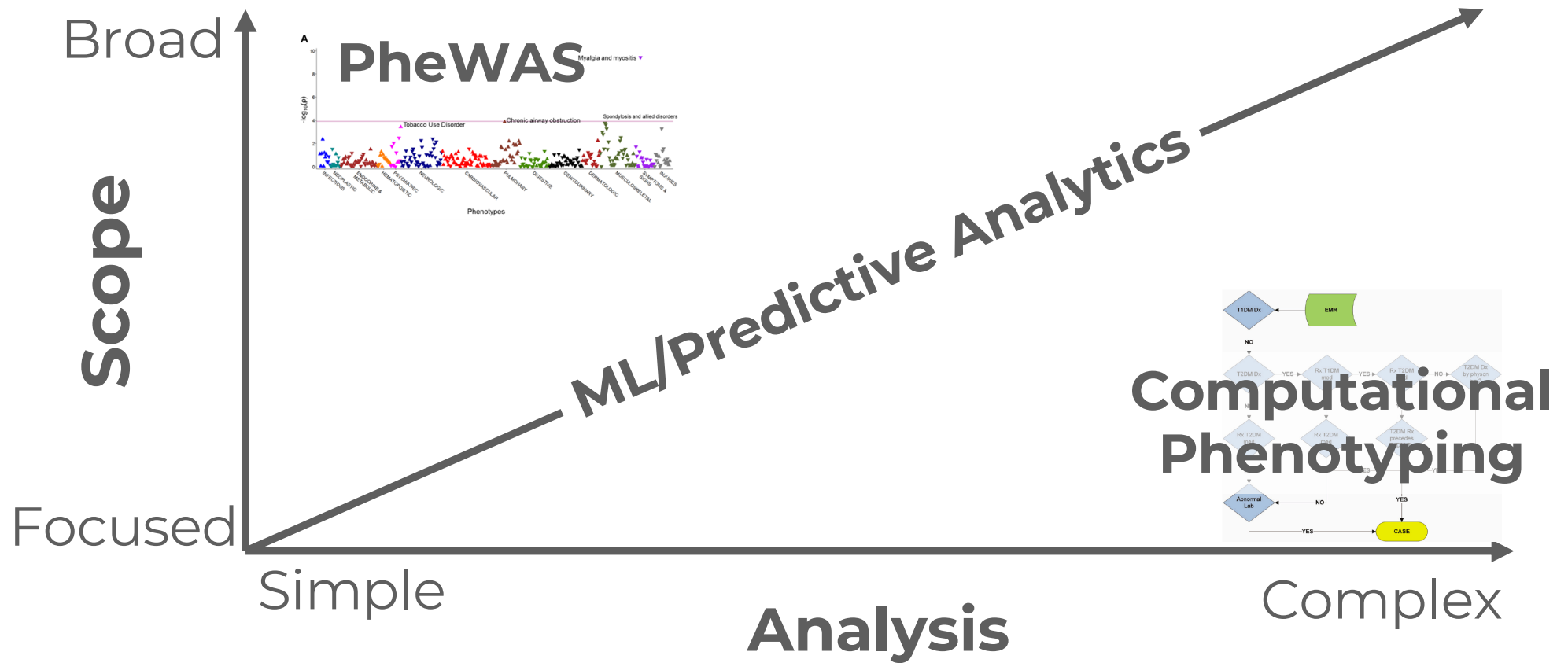
- <https://github.com/PheWAS/PheWAS>
- Contains:
 - Translation Tables (ICD9-CM, ICD10-CM, ICD10)
 - Analysis Scripts
 - Multiple Testing Correction (Bonferroni & FDR)

You Have A Significant Hit! Now What?

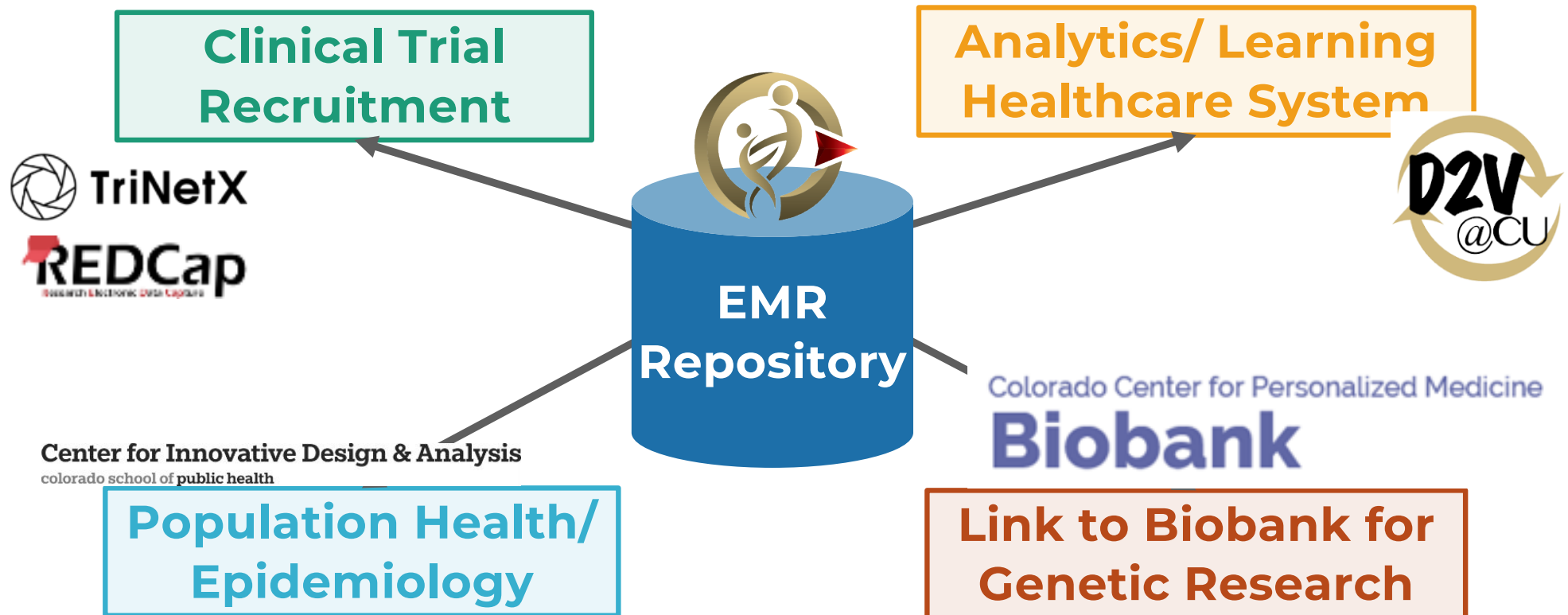
- PheWAS Code Correlation – Which is the driver?
- Check Genotype & Phenotype Frequencies
- Is this plausible/meaningful?
- *Look* at individual patients w/ Code

PheWAS codes are NOT
phenotypes

EHR – Based Research Landscape



CU Resources



Interested in Learning More?



Clinical Data Science Specialization

Created By:



University of Colorado
Anschutz Medical Campus

LearnClinicalDataScience.org

Industry Partner:

Google Cloud

Hosted By:

Coursera

FREE For CU Faculty,
Staff, & Students!

Instructions

1. Log in to my.cu.edu
2. Open "Training" (if student) or "CU Resources Home" > "Training" (if staff/faculty)
3. Select "CU on Coursera"
4. Click "Join for Free"
5. Click "Log in with University of Colorado"



How To Get Help

- Health Data Compass: www.healthdatacompass.org
- Colorado Center for Personalized Medicine Biobank: www.cobiobank.org
- CIDA: <http://www.ucdenver.edu/academics/colleges/PublicHealth/research/centers/CBC/Pages/welcome.aspx>
- D2V: <http://www.ucdenver.edu/academics/colleges/medicalschool/programs/D2V/Pages/D2V.aspx>