Sentinel

Sentinel Innovation Day

Brigham and Women's Hospital
Founding Member, Mass General Brigham

Duke University
School of Medicine

VANDERBILT UNIVERSITY
MEDICAL CENTER

KAISER PERMANENTE®

Agenda

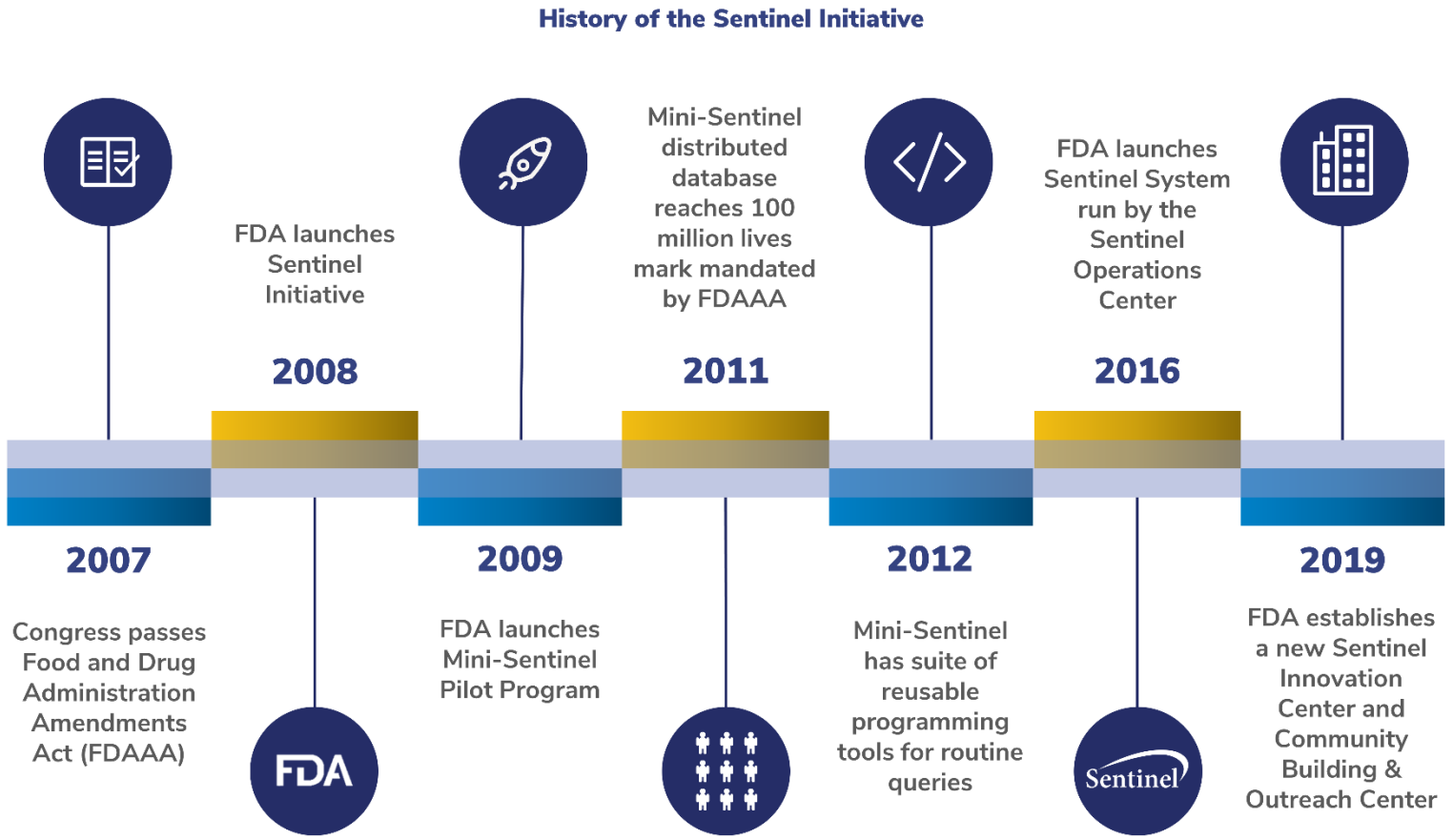
1. Welcome and Sentinel overview
2. FDA opening remarks
3. DI2: Representation of unstructured data across Common Data Models
4. DI3: Identification and mitigation of structured EHR source data mapping issues
5. FE1: Computable phenotyping framework
6. FE2: NLP tools for cohort identification, exposure assessment, covariate ascertainment
7. FE3: Improving probabilistic phenotyping of incident outcomes
8. CI1: Enhancing Causal Inference in the Sentinel System
9. CI2: A causal inference framework for Sentinel
10. Closing remarks

Overview

FDA's Sentinel system

2007 FDA Amendments Act mandates FDA to establish ***active surveillance system*** for monitoring drugs using electronic healthcare data

Through the Sentinel Initiative, FDA aims to assess the post-marketing safety of approved medical products



Sentinel Innovation Center (IC) Vision

Current Sentinel system limitations

Inability to identify certain study populations of interest from insurance claims

Inability to identify certain outcomes of interest from insurance claims

Other limitations (inadequate duration of follow-up, the need for additional signal identification tools)

Sentinel Innovation Center Initiatives

Data infrastructure (DI)

Feature engineering (FE)

- Emerging methods including machine learning and scalable automated natural language processing (NLP) approaches to enable computable phenotyping from unstructured EHR data

Causal inference (CI)

- Methodologic research to address specific challenges when using EHRs such as approaches to handle missing data, calibration methods for enhanced confounding adjustment

Detection analytics (DA)

- Development of signal detection approaches to account for and leverage differences in data content and structure of EHRs

Sentinel Innovation Center vision

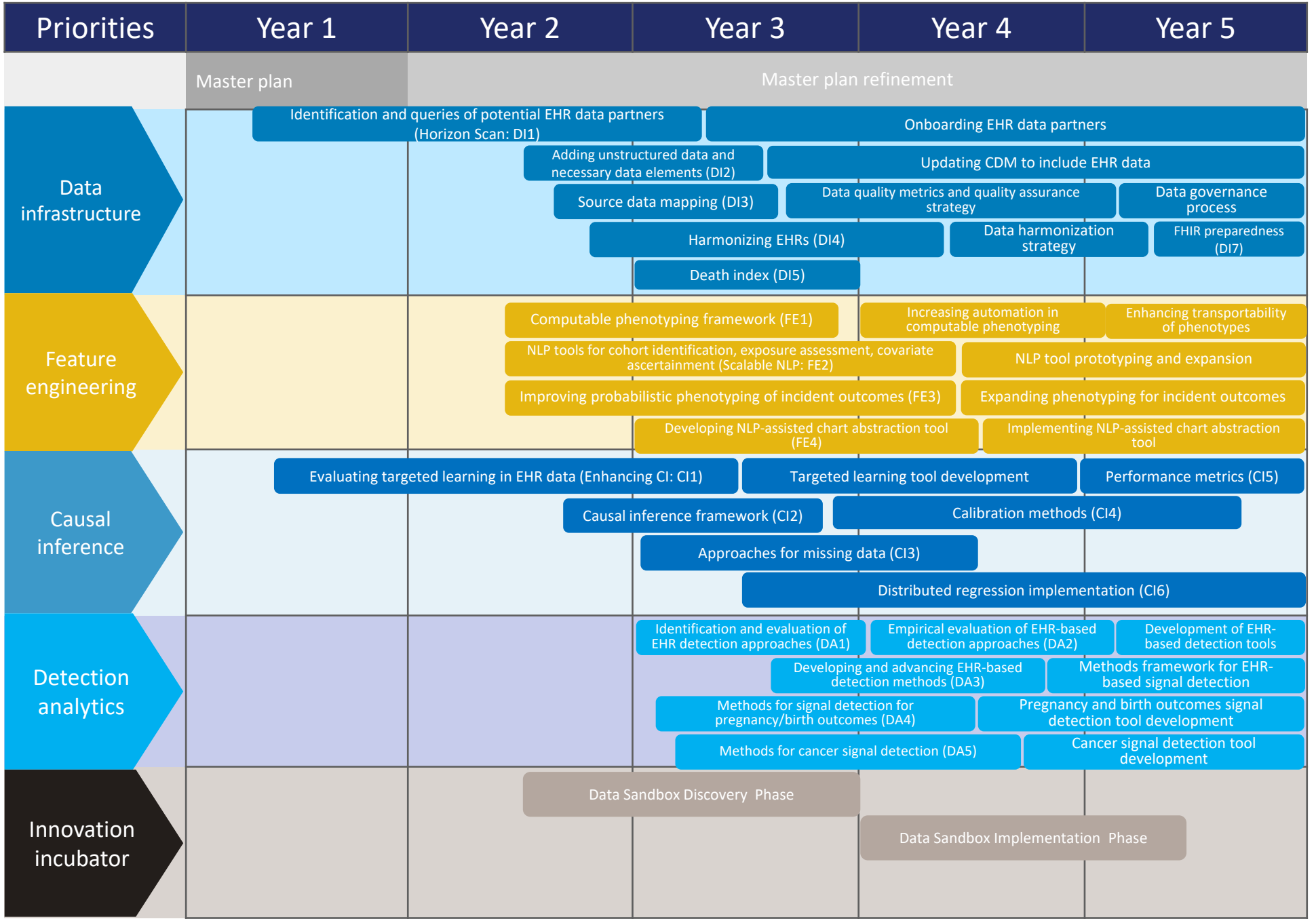
A query-ready, quality-checked distributed data network containing EHR for at least 10 million lives with reusable analysis tools

2020 → **2024**

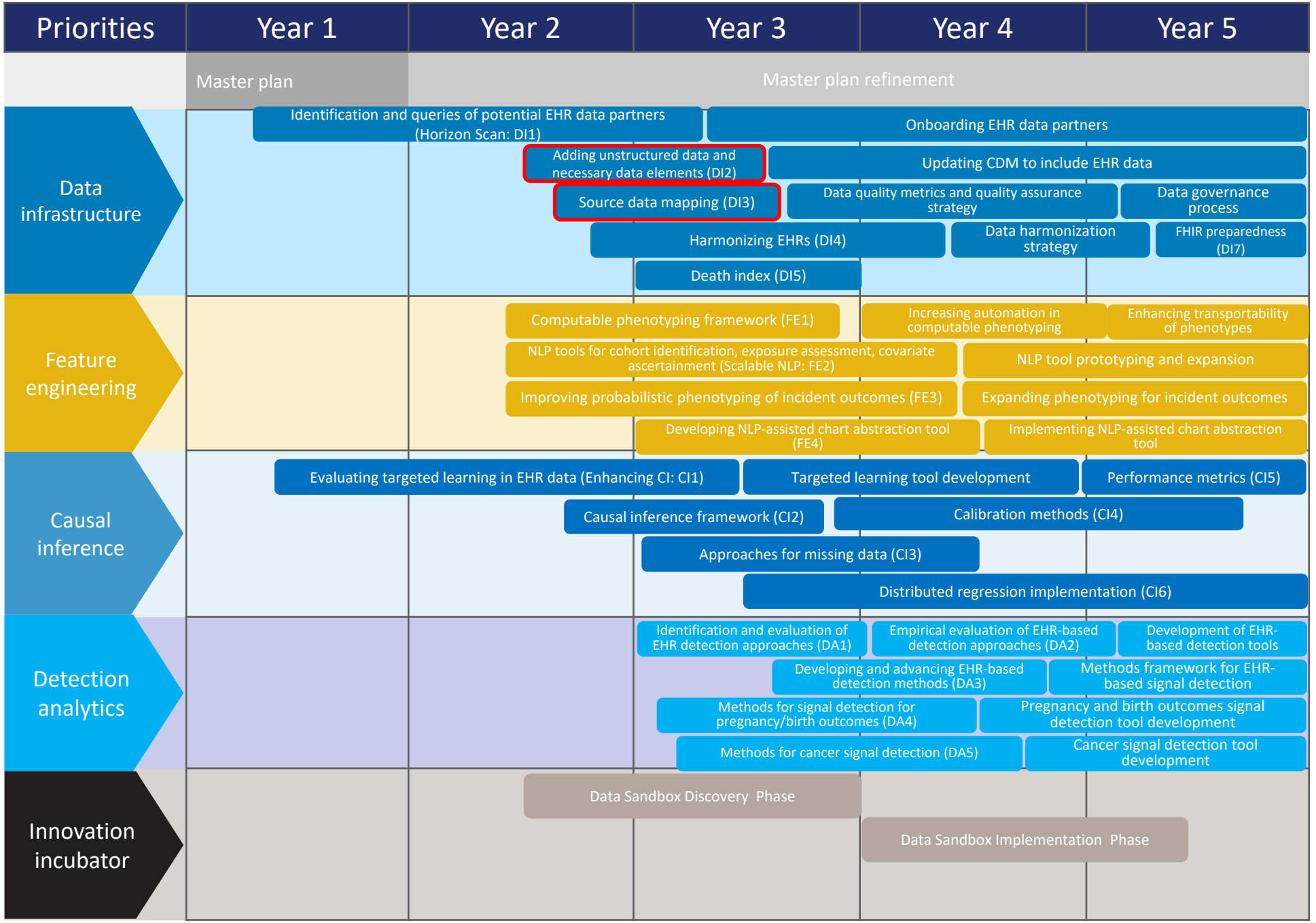
Desai et al. npj Digital Medicine (2021) 4:170

IC Master Plan:

A snapshot of ongoing and future activities



- **DI2**: Representation of unstructured data across Common Data Models
- **DI3**: Identification and mitigation of structured EHR source data mapping issues



Challenges and Opportunities in Integrating Electronic Health Record (EHR) Data in Sentinel

Keith Marsolo, PhD
Associate Professor
Department of Population Health Sciences
Duke Clinical Research Institute
Duke University School of Medicine

4/28/22

Purpose

IC Projects -- Highlight Challenges and Opportunities

As the Sentinel Innovation Center works to establish an infrastructure of administrative claims linked with electronic health record (EHR) data on 10 million+ lives:

- Focus = two projects that develop aspects of the infrastructure needed to bring EHR data into the Sentinel framework

Each highlights potential challenges and opportunities presented by EHR

DI2: Representation of unstructured data across Common Data Models

DI3: Identification and mitigation of structured EHR source data mapping issues

DI2: Representation of Unstructured Data Across Common Data Models

Incorporating Unstructured Data into a Common Data Model

Goal: To guide the Sentinel Network on **how best to incorporate information derived from unstructured data into a Common Data Model (CDM) framework.**

Objectives:

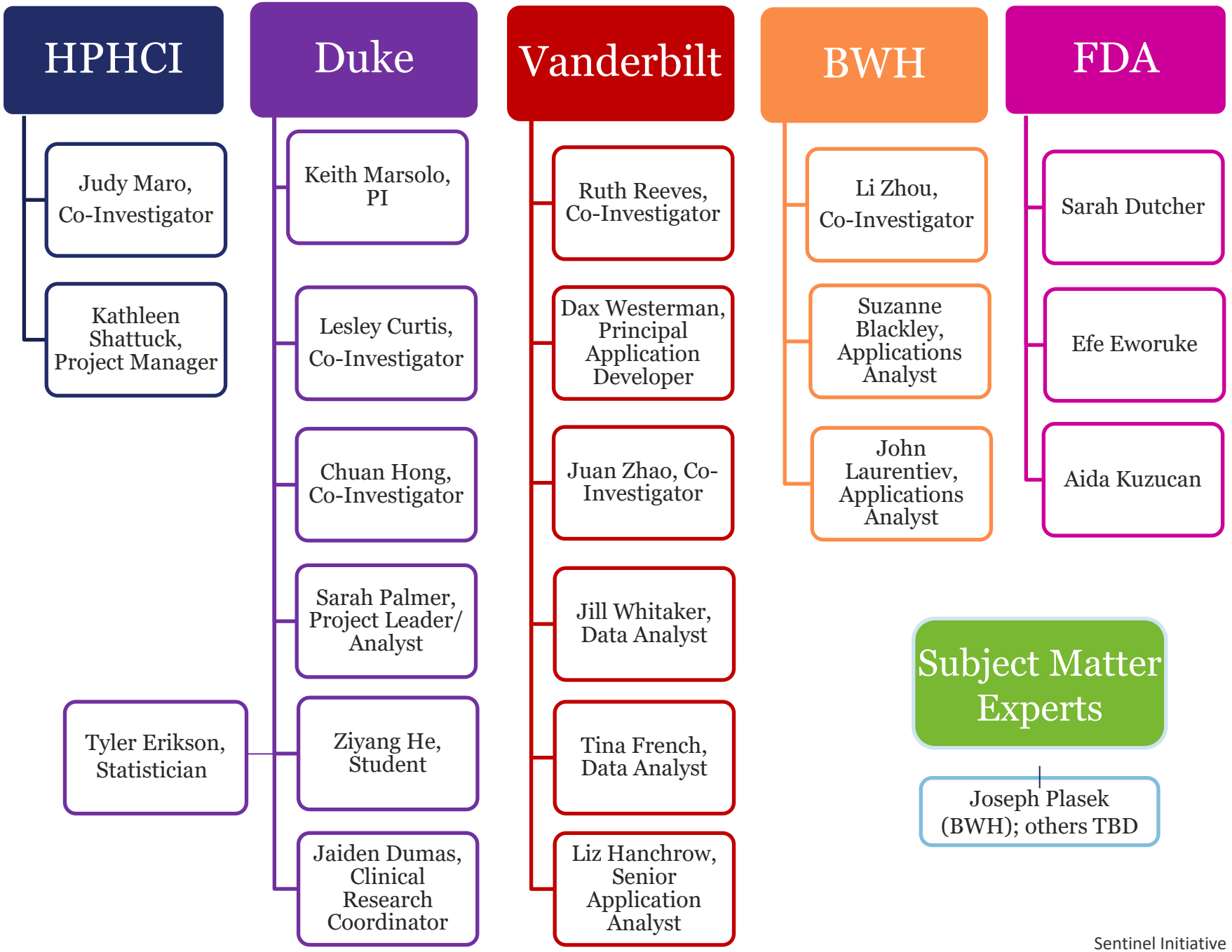
1) *What information is important?* – Identify the priority elements that should be derived from unstructured data

2) *What NLP tools are in use & how are they used?; What information is available within a note?* – Assess the overall availability of the priority elements within the Sentinel ecosystem

3) *How to best represent information derived from unstructured text?* – Recommend how those priority elements should be represented in the Sentinel Common Data Model

Project completion date: **May 31, 2022** (to be extended)

Project team



Objective 1 – What information is important?

Process:

Generated list of concepts from commonly-used NLP pipelines (commercial & open-source)

- Focused mainly on broad categories, not specific items, unless called out in documentation (e.g., medications, not aspirin)
- Looked at the basic functionality provided by each tool, not every research project
- Generated "good enough" list – stopped when we reached saturation

FDA reviewed list, identified any missing elements & assigned priority rankings (high / medium / low) - highest priority given to those concepts not easily obtained from claims that are also important for drug safety studies

End Product:

Set of priority elements to be derived from unstructured text.

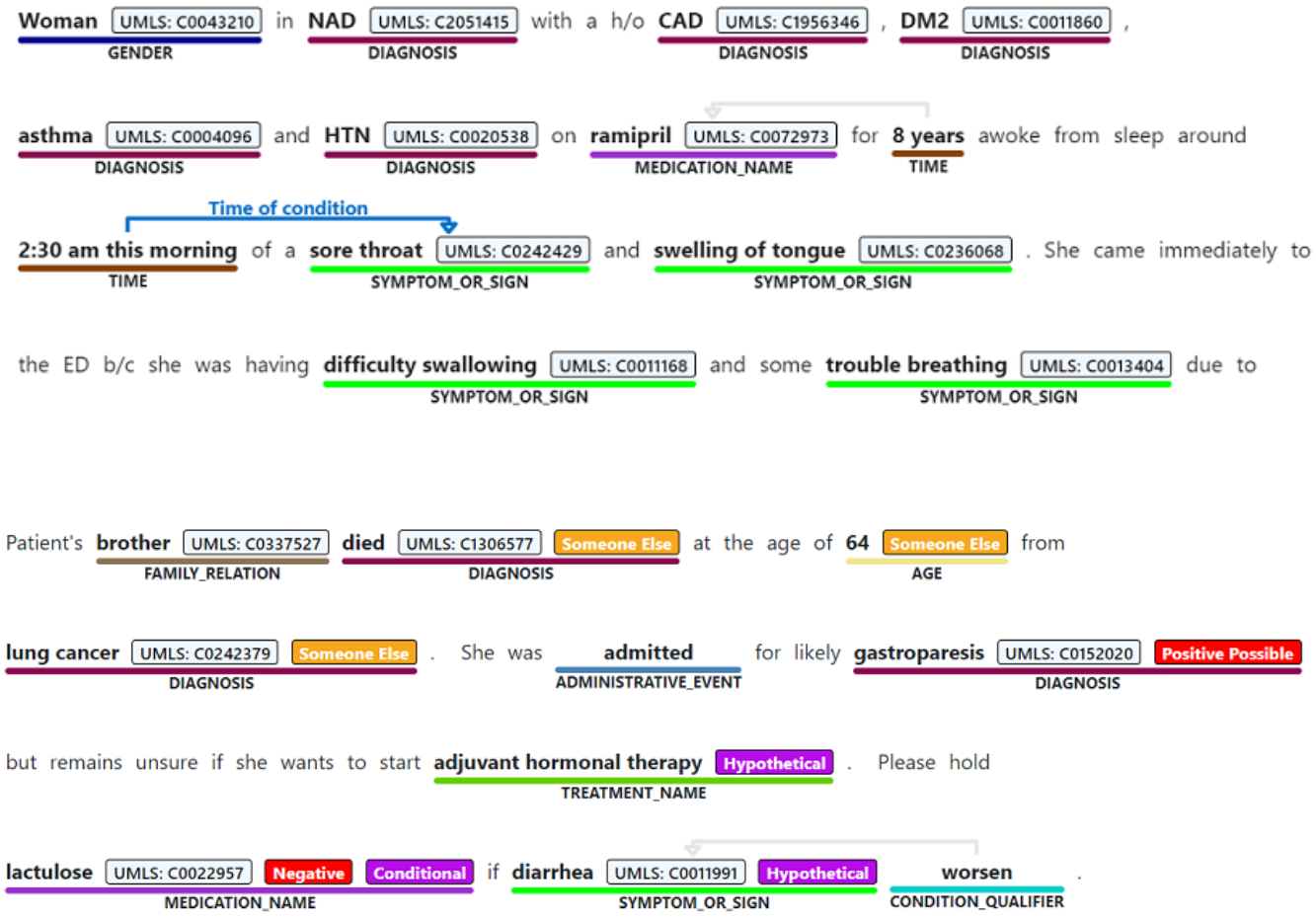


Image source: https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/how-tos/text-analytics-for-health

Example priority rankings (subset)

Concepts from existing tools

Domain	Concept(s)	Priority	Notes
Cancer	Site	High	Several ARIA insufficiency rankings due to lack of data on cancer (e.g., staging)
	Histology	High	
	Procedure	High	
Condition	Diagnoses	Medium	Often captured in claims
	Signs / Symptoms	High	Less available in claims, useful in different aspects of studies
	Family History (Type)	Medium	Useful in some studies, but not all
	Medical History (Type)	High	Often gaps in EHR data, medical history important to capture
Medication	Class	Low	Can be inferred from drug name

Missing concepts

Concept(s)	Priority	Notes
Timing & duration of medication	High	Particularly important for inpatient medications
Physical findings (e.g., vital signs)	High	Key covariate for FDA studies, under-captured in claims
Indication for a drug	High	Rationale for why a drug is given
Oxygen support	High	Relevant for many COVID-19 studies
Death (date) & cause	Low*	Capture of death data varies by Sentinel Data Partner

Objective 2 – What NLP tools are in use and how are they used? What information is available within a note?

Process:

Distributed survey to partners within the Sentinel ecosystem to assess their NLP capabilities (e.g., tool(s) used, notes processed, concepts extracted, etc.) – understand how well the current state of NLP use aligns with the priority concepts identified by FDA

Perform chart annotations at 2 sites (Vanderbilt, Brigham & Women's Hospital) to assess availability of priority elements within 2 different use cases *(in progress)*

End Product:

Survey responses from Partners on their ability to extract priority data elements from unstructured text, and statistics on the overall availability of priority data elements within the unstructured data as determined by chart annotation.

NLP capabilities survey (initial results)

Distributed to 14 Sentinel Data Partners & 8 partners affiliated with the Innovation Center

A total of 17 responses received (13 from Sentinel Data Partners)

- 12 use NLP in some capacity
- 50% for project-specific research; 50% for research & "operational" purposes

Wide variety of tools used / notes processed (type, number of years)

Scope of concepts extracted also varies widely

- 9 of 12 report being able to extract Diagnoses (highest percentage)
- Handful of other concepts extracted by >50% of respondents (e.g., cancer site & histology, smoking status, signs & symptoms)

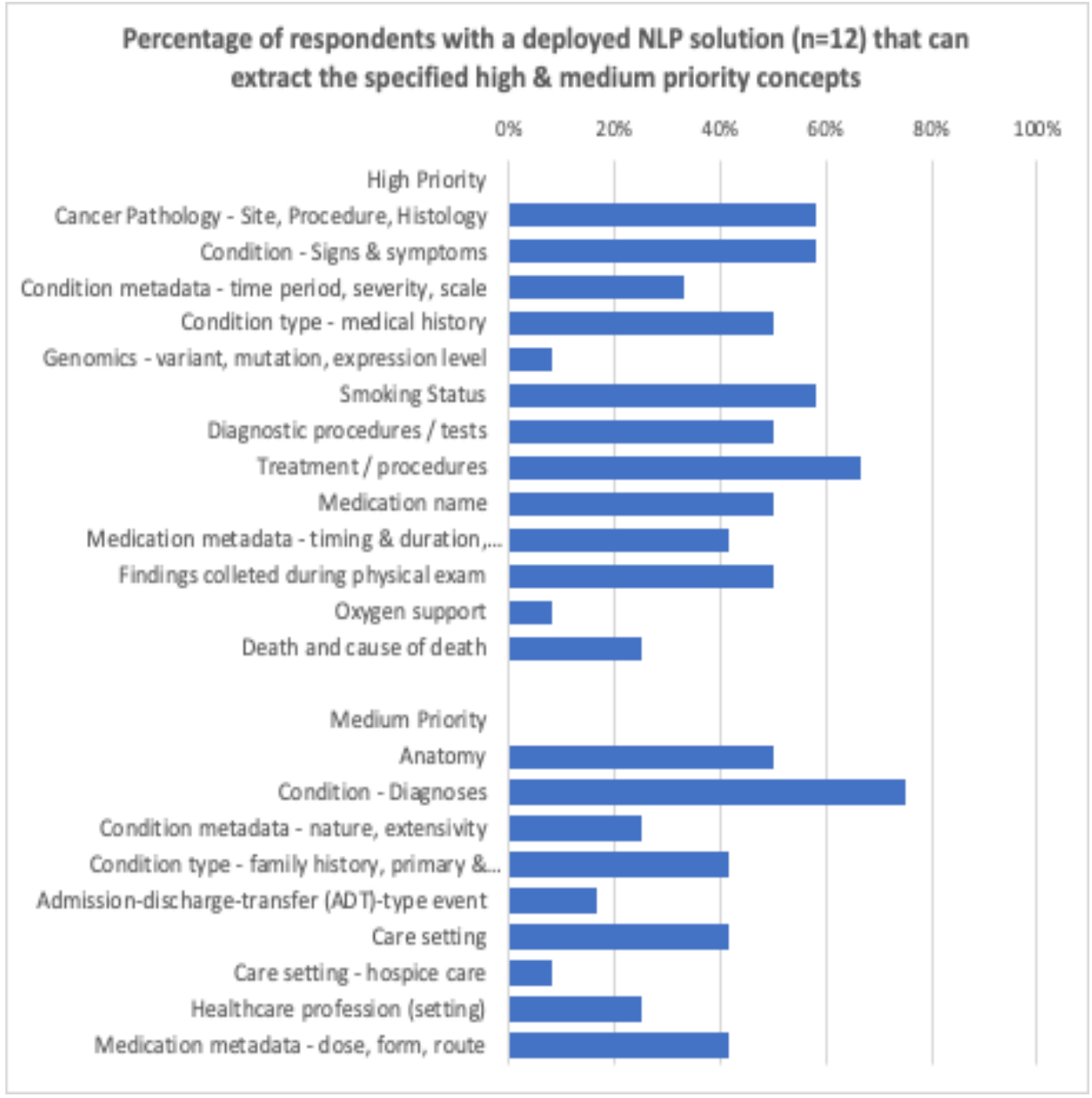


Chart annotation - Motivation

Vision

A future state where Sentinel partners with access to EHR data have processed all / some of their clinical notes through an NLP pipeline (or pipelines).

- Some projects may require the development of new pipelines/classifiers,
- Others will rely on the "stock" NLP outputs.

We want to use those derived data elements in a Sentinel analysis.

Issues to consider:

- What note type(s) need to have been processed?
- What time frame had to have been covered?

Example

Looking for history of MI:

- patient had MI 10 years ago

Can we assume it is mentioned in the note at every visit, or just a subset (i.e., first visit with a new provider; every visit for the 2 years after the event, etc.)?

Chart annotation (in progress)

Focus on two use cases

- Hospitalized patients with COVID-19
- Cancer

For both, we propose to look at a subset of notes, since we will not necessarily be able to assume that (future) partners will have run NLP on everything (e.g., all hospital discharge summaries are included, but not respiratory therapist notes)

Purpose is not to develop a classifier or a pipeline, but to describe the information contained in the notes of the patients in each cohort

Hospitalized patients with COVID-19

Population:

- Index event - inpatient encounter with an admitting diagnosis of COVID-19 between April 1, 2020 and December 31, 2021
- Limit to patients who are age >= 18 at the time of admission.

Sampling strategy:

- Cohort 1 – patients without a billing code for supplemental oxygen. Select 35 patients at random.
- Cohort 2 – patients with a billing code for supplemental oxygen. Select 35 patients at random.

Analysis:

- Primary – Pull the discharge summary associated with the hospitalization and annotate priority concepts (e.g., oxygen use, conditions, medication exposure & metadata, smoking status)
- Secondary – For a subset of patients in each cohort (5-10, randomly selected), run a query to identify all notes that include keywords related to oxygen use. Review note / paragraph / sentences around the keyword and determine whether it indicates oxygen use.

Rationale for design choices:

- The secondary analysis will allow us to characterize the degree of "missingness" related to oxygen use, as discharge summaries are not expected to contain the full detail related to oxygen use
- Discharge summaries were chosen because if we are planning to use pre-computed NLP concepts in an analysis, discharge summaries are more likely to be processed across a network than specialty notes (e.g., respiratory therapy)
- Stratifying by billing codes for supplemental oxygen should ensure there is a mix of patients who did and did not receive oxygen compared with a purely random sample of hospitalized patients

Cancer

Population:

Index event

- Patients with a prescription/order for darzalex (daratumumab) and with no prescription/order for darzalex in the prior 3 years
- Index event should be between January 1, 2016 and November 30, 2021.

Sampling strategy:

- Select 30 patients at random from the cohort
- Annotate the physician note(s) associated with the visit where the patient was prescribed the medication (assume new prescription occurs in the outpatient setting)

Analysis:

- Annotate selected concepts (e.g., conditions, medications, smoking status, those specific to label);
- Determine if available concepts are sufficient to determine indication behind prescription

DARZALEX example

Medication-related concepts
Diagnosis-related concepts
Concepts that are expected to be primarily NLP-based

1. in combination with bortezomib, melphalan and prednisone for the treatment of patients with newly diagnosed multiple myeloma who are **ineligible for autologous stem cell transplant**

2. in combination with lenalidomide and dexamethasone, or bortezomib and dexamethasone, for the treatment of patients with multiple myeloma who have received at least one prior therapy <*list of candidate therapies required to define this part*>

3. in combination with pomalidomide and dexamethasone for the treatment of patients with multiple myeloma who have received at least two prior therapies including lenalidomide and a proteasome inhibitor

4. as monotherapy, <*exclude patients with concurrent candidate therapies*> for the treatment of patients with multiple myeloma who have received at least three prior lines of therapy including a proteasome inhibitor (PI) and an immunomodulatory agent or who **are double-refractory** to a PI and an immunomodulatory agent.

Objective 3 – How to best represent information derived from unstructured text? (in progress)

Process:

Assess current approaches for representing data derived from unstructured text (from other Common Data Models, NLP tools, etc.)

Describe tradeoffs between approaches (e.g., ease of querying, burden on partners, strengthens and weaknesses of different terminologies)

End Product:

Develop set of recommendations for the Sentinel Operations Center as they make decisions on extending the Sentinel Common Data Model

DI3: Identification and mitigation of structured EHR source data mapping issues

Mapping of EHR Data and developing quality metrics

Goal:
To assess the mapping of structured electronic health record (EHR) data to reference terminologies and to develop quality metrics to allow for comparisons across domains within a data source to further identify issues.

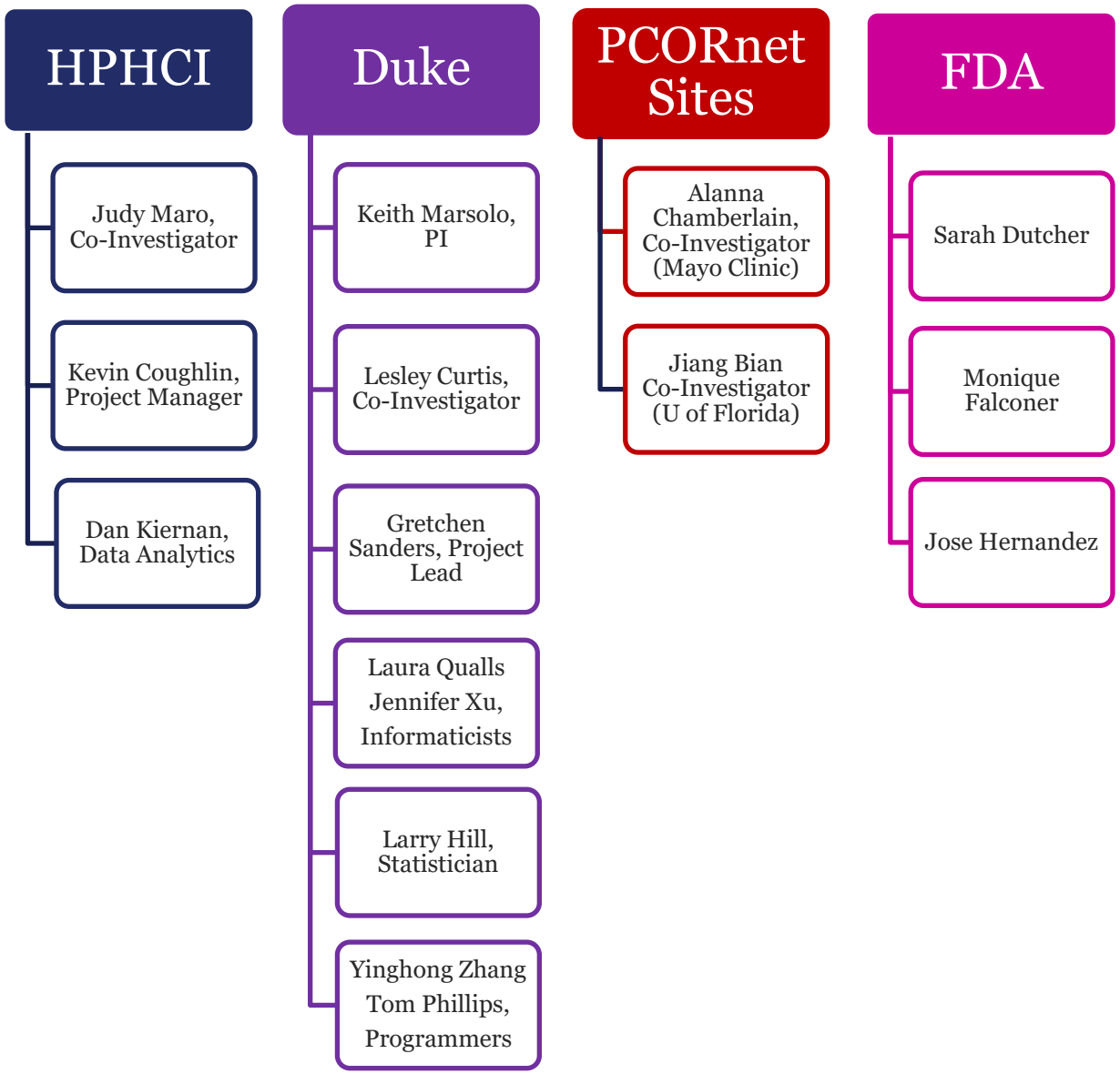
Objectives:

1) Develop procedures to assess the mapping of structured EHR data to reference terminologies for laboratory results, medication orders and administrations (inpatient and outpatient) & characterize the severity of issues that are uncovered

2) Develop standardized metrics related to medications & laboratory results that allow for comparisons across domains within a data source using profiles of records across time, care setting, population, etc. This work will supplement the Sentinel Operations Center's Data Quality Measures (DQM) in EHRs project by defining *new* metrics for assessments that *are not* routinely conducted in EHR datasets.

Project completion date: **September 30, 2022**

Project team



Motivation – Harmonization of EHR data sources

Many EHR data domains (e.g., medication orders, laboratory results) are not captured in standard formats

To use these data for research or for data exchange, must harmonize to a reference standard

Examples shown within these slides are taken from the National Patient-Centered Clinical Research Network (PCORnet®), but the same challenges exist regardless of the source

For analyses that leverage linked claims-EHR data, findings from this project can provide guidance on the types of EHR data to be included in a CDM and how to ensure and verify accurate transformation

Representing a medication in RxNorm

	RxNorm Term Type	Information encoded				Example medication representation
	Description	Ingredient(s)	Strength	Dose Form	Brand Name	Original string - Augmentin XR 12 HR 1000 MG Extended Oral Release Tablet
Most Granular	Semantic Branded Drug	X	X	X	X	Augmentin XR 12 HR 1000 MG Extended Release Oral Tablet
	Semantic Clinical Drug	X	X	X		12 HR Amoxicillin 1000 MG / Clavulanate 62.5 MG Extended Release Oral Tablet
	Brand Name Pack	X	X	X	X	N/A
	Generic Pack	X	X	X		N/A
	Semantic Branded Drug Form	X		X	X	Amoxicillin / Clavulanate Extended Release Oral Tablet [Augmentin]
	Semantic Clinical Drug Form	X		X		Amoxicillin / Clavulanate Extended Release Oral Tablet
↓	Semantic Branded Dose Form Group*			X	X	Augmentin Oral Product; Augmentin Pill (Requires two records)
	Semantic Clinical Dose Form Group*	X		X		Amoxicillin / Clavulanate Oral Product; Amoxicillin / Clavulanate Pill (Requires two records)
	Semantic Branded Drug Component	X	X		X	Amoxicillin 1000 MG / Clavulanate 62.5 MG [Augmentin]
	Brand Name				X	Augmentin
	Multiple Ingredients	X				Amoxicillin / Clavulanate
	Semantic Clinical Drug Component*	X	X			Amoxicillin 1000 MG; Clavulanate 62.5 MG (Requires two records)
	Precise Ingredient	X				N/A
Least Granular	Ingredient*	X				Amoxicillin; Clavulanate (Requires two records)
Non-specific	Dose Form			X		Extended Release Oral Tablet
	Dose Form Group*			X		Oral Product; Pill (Requires two records)
	Prescribable Name					
	Synonym					
	Tall Man Lettering Synonym					

* Denotes term types that require multiple records to represent multi-ingredient medications

Within the PCORnet Common Data Model, medication orders and administrations (at most sites) are coded using RxNorm

RxNorm is an interoperability standard maintained by the National Library of Medicine that represents medication orders and administrations at various levels of granularity

Even if Sentinel leverages a different standard to represent EHR-based medications, data partners may still need to transform data to/from RxNorm

PCORnet has defined a set of preferred "tiers" for the different RxNorm Term Types

		RxNorm Term Type	Information encoded			
	Term Type	Description	Ingredient(s)	Strength	Dose Form	Brand Name
Tier 1	SBD	Semantic Branded Drug	X	X	X	X
	SCD	Semantic Clinical Drug	X	X	X	
	BPCK	Brand Name Pack	X	X	X	X
	GPCK	Generic Pack	X	X	X	
Tier 2	SBDF	Semantic Branded Drug Form	X		X	X
	SCDF	Semantic Clinical Drug Form	X		X	
	SBDG	Semantic Branded Dose Form Group*			X	X
	SCDG	Semantic Clinical Dose Form Group*	X		X	
	SBDC	Semantic Branded Drug Component	X	X		X
	BN	Brand Name				X
	MIN	Multiple Ingredients	X			
Tier 3	SCDC	Semantic Clinical Drug Component*	X	X		
	PIN	Precise Ingredient	X			
	IN	Ingredient*	X			
Tier 4 (Do not use)	DF	Dose Form			X	
	DFG	Dose Form Group*			X	
	PSN	Prescribable Name				
	SY	Synonym				
	TMSY	Tall Man Lettering Synonym				

* Denotes term types that require multiple records to represent multi-ingredient medications

Example quality issue – medication mapping

Highest-volume medication records by RxNorm code				Highest-volume medication records by name (within the EHR)			Percent Agreement
Rank based on Code	RxNorm Code	Medication name (derived from RxNorm code)	Record Count by Code	Rank based on Name	Medication name (from EHR)	Record Count by Name	
1	Null or missing		1257171	1	Null or missing	1257171	100%
2	313002	Sodium Chloride 9 MG/ML Injectable Solution	801348	2	Sodium Chloride	1007029	79.6%
3	307668	Acetaminophen 32 MG/ML Oral Suspension	321510	3	Acetaminophen 300MG / Codeine Phosphate 15 MG Oral Tablet	511779	
4	197803	Ibuprofen 20 MG/ML Oral Suspension	293209	4	Ibuprofen 20 MG/ML / Pseudoephedrine Hydrochloride 3 MG/ML Oral Suspension	293218	
5	540930	Water 1000 MG/ML Injectable Solution	286133	5	Water 1000 MG/ML Injectable Solution	287011	99.6%
6	309778	Glucose 50 MG/ML Injectable Solution	285557	6	Glucose 50 MG/ML / Potassium Chloride 0.01 MEQ/ML / Sodium Chloride 0.0342 MEQ/ML Injectable Solution	286108	99.8%

Shading indicates a discordance in medications (e.g., RxNorm code represents a single ingredient in RxNorm vs. multi-ingredient order within the EHR)

Objective 1: Methods to assess mapping of structured EHR data to reference terminologies

General approach:

- Develop queries to assess mapping of medication orders, medication administrations and laboratory tests – limit analysis to the top 200 by volume
- For each medication / lab, generate statistics on all the different combinations within the structured fields and “raw” source fields
- For example, for a given medication name, summarize the number of records/patients for associated RxNorm codes, dose units, dose forms, as well as the corresponding ”raw” fields

RAW Medication Name	RxNorm Code	CDM Dose Unit	RAW Dose Unit	Number of Records	Number of Patients
CALCIUM CARBONATE 300 MG (750 MG) CHEWABLE TABLET	1044532	Other		2	2
	1044532	Other	mg of elemental	13	11
	1044532	Other	mg of salt	50564	14817
	1044532	Other	tablet	1	1
	1484737	Other		3	2
	1484737	Other	mg of elemental	4	3
	1484737	Other	mg of salt	51092	14887
	1484737	Other	tablet	2	2

Example statistics for Dose Unit for a single medication

Objective 1: Evaluation

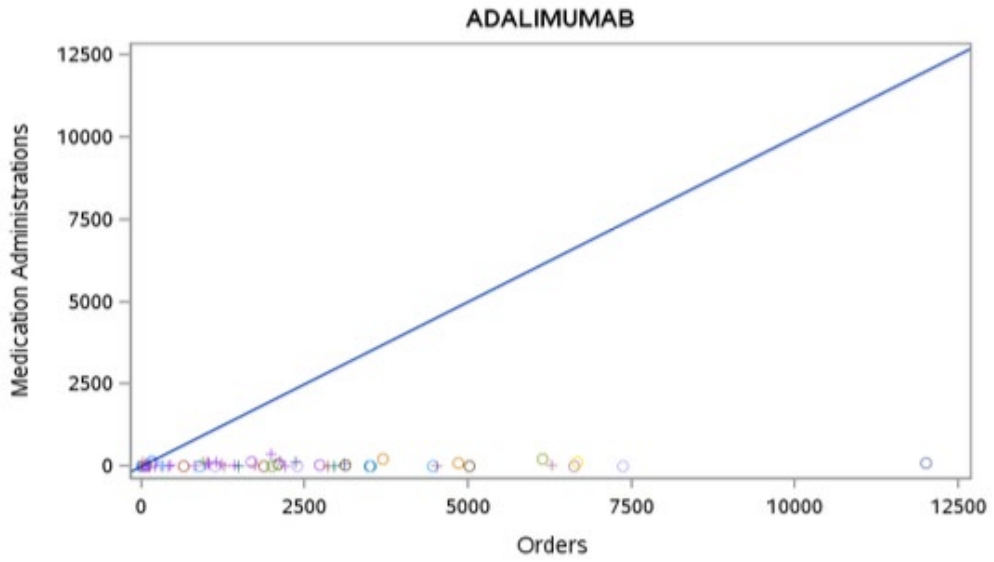
- Generate statistics on number of medication codes/ laboratory tests associated with more than one name within the EHR and vice versa
- Concordance between lab name / medication name (brand and/or ingredient) within the EHR and that derived from the associated code
- Concordance between discrete fields (e.g., lab result unit, medication dose, etc.) and those associated with the associated LOINC / RxNorm code
- Generate characterization of issues by severity (e.g., LOINC code mis-match, combination medication represented by single-ingredient RxNorm code, generic medication represented by brand name, etc.)

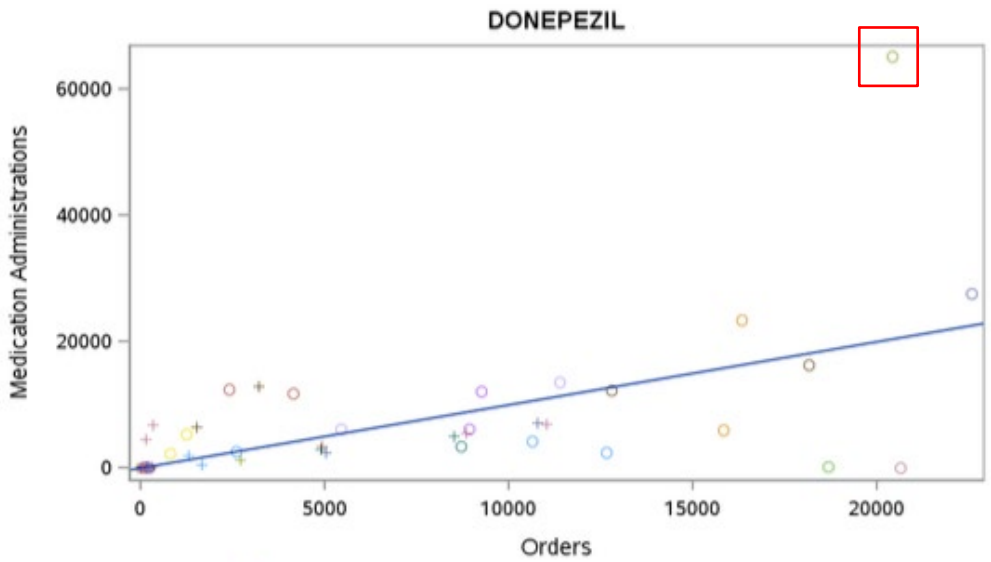
End product:

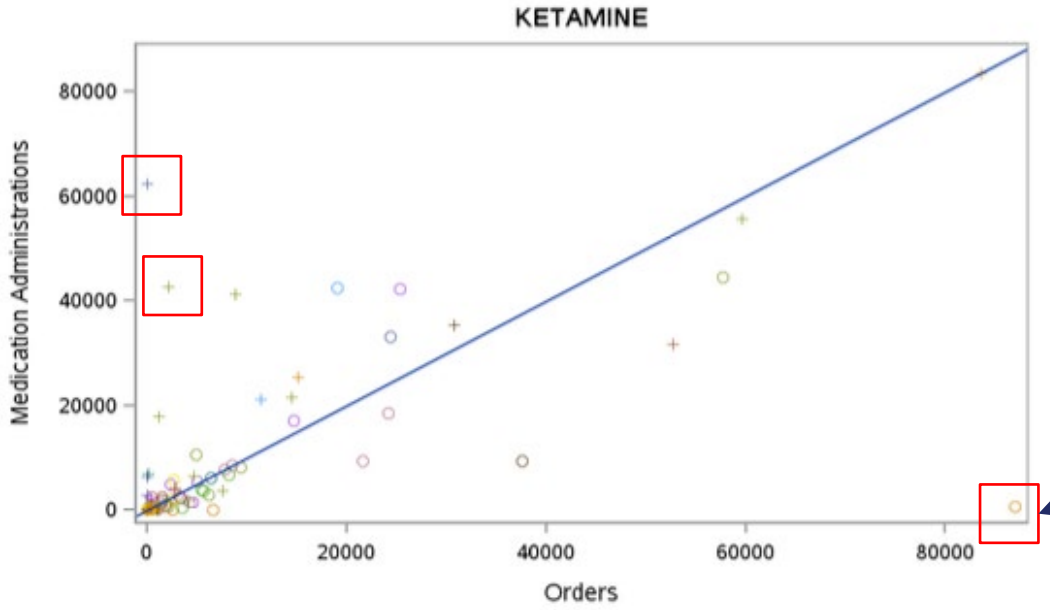
Procedures that can be used to assess mapping of structured EHR domains and a set of statistics on the severity of issues at 2 pilot sites (PCORnet).

Severity	Example issue	Rationale
Critical	(1) Lab test mismatch (incorrect LOINC code) (2) Multi-ingredient drug uses single ingredient RxNorm code (3) Single ingredient drug uses multi-ingredient RxNorm code	(1-3) The LOINC/RxNorm codes that are assigned to these records are incorrect and would not actually represent the test result or exposure to the specified medication.
Major	(1) Ingredient-level RxNorm code utilized when more granular available (single-ingredient drugs only) (2) More granular RxNorm code used than supported by the data	(1) The ingredient is correct, but the other metadata is missing, meaning those records may be excluded if the drug has forms that are not part of an analysis (i.e., topical creams). (2) This example is the inverse – records that should have been excluded were included.
Moderate	(1) Generic medication uses brand name RxNorm code (2) Brand name medication uses a generic-level RxNorm code	(1) Any study that looking for the use of a specific brand of medication will include extra records. (2) Studies that are looking at the use of a specific branded medication will miss records.
Minor	(1) Distribution of lab results is an outlier for a given LOINC.	(1) The test may be only used on specific populations (e.g., inpatients), which may bias results.

Example quality issue – differences based on provenance (orders vs. medication administrations)







Indicates potential outlier

Objective 2: Standardized metrics to generate comparisons based on provenance

General approach:

Develop queries that will support the comparison of records based on provenance – medication orders vs. administrations; billed diagnoses vs. clinician-entered – to identify potential data issues.

Define specific conditions & associated concepts to investigate (e.g., diagnoses, procedures, medications, labs). Look at values within each cohort as well as the population as a whole.

Distribute query package to partner sites to generate summary statistics. Focus of analysis will be within-DataMart comparisons, though cross-DataMart comparisons are also possible.

End product:

Set queries to support cross-domain comparisons within a dataset, at both condition and population-level, along with statistics describing the performance of each at partners sites.

COHORT	PERIOD	CONCEPT	PROVENANCE	NUMBER OF PATIENTS	NUMBER OF RECORDS
COPD	2016	CAD DX	ORDERED		
COPD	2016	CAD DX	BILLED		
COPD	2016	CAD DX	DERIVED (e.g., NLP)		
COPD	2017	CAD DX	ORDERED		
COPD	2017	CAD DX	BILLED		
COPD	2017	CAD DX	DERIVED (e.g., NLP)		
ALL	2016	CAD DX	ORDERED		
ALL	2016	CAD DX	BILLED		
ALL	2016	CAD DX	DERIVED (e.g., NLP)		
ALL	2017	CAD DX	ORDERED		
ALL	2017	CAD DX	BILLED		
ALL	2017	CAD DX	DERIVED (e.g., NLP)		

Diagnoses by provenance for a specific cohort (COPD) and the population as a whole.

COHORT	PERIOD	MEDICATION	ENCOUNTER TYPE	PROVENANCE	NUMBER OF PATIENTS
CKD	2016	LOOP DIURETIC	AMBULATORY	PRESCRIBING	
CKD	2016	LOOP DIURETIC	AMBULATORY	MED_ADMIN	
CKD	2016	LOOP DIURETIC	AMBULATORY	BOTH	
CKD	2016	LOOP DIURETIC	INPATIENT	PRESCRIBING	
CKD	2016	LOOP DIURETIC	INPATIENT	MED_ADMIN	
CKD	2016	LOOP DIURETIC	INPATIENT	BOTH	
ALL	2016	LOOP DIURETIC	AMBULATORY	PRESCRIBING	
ALL	2016	LOOP DIURETIC	AMBULATORY	MED_ADMIN	
ALL	2016	LOOP DIURETIC	AMBULATORY	BOTH	
ALL	2016	LOOP DIURETIC	INPATIENT	PRESCRIBING	
ALL	2016	LOOP DIURETIC	INPATIENT	MED_ADMIN	
ALL	2016	LOOP DIURETIC	INPATIENT	BOTH	

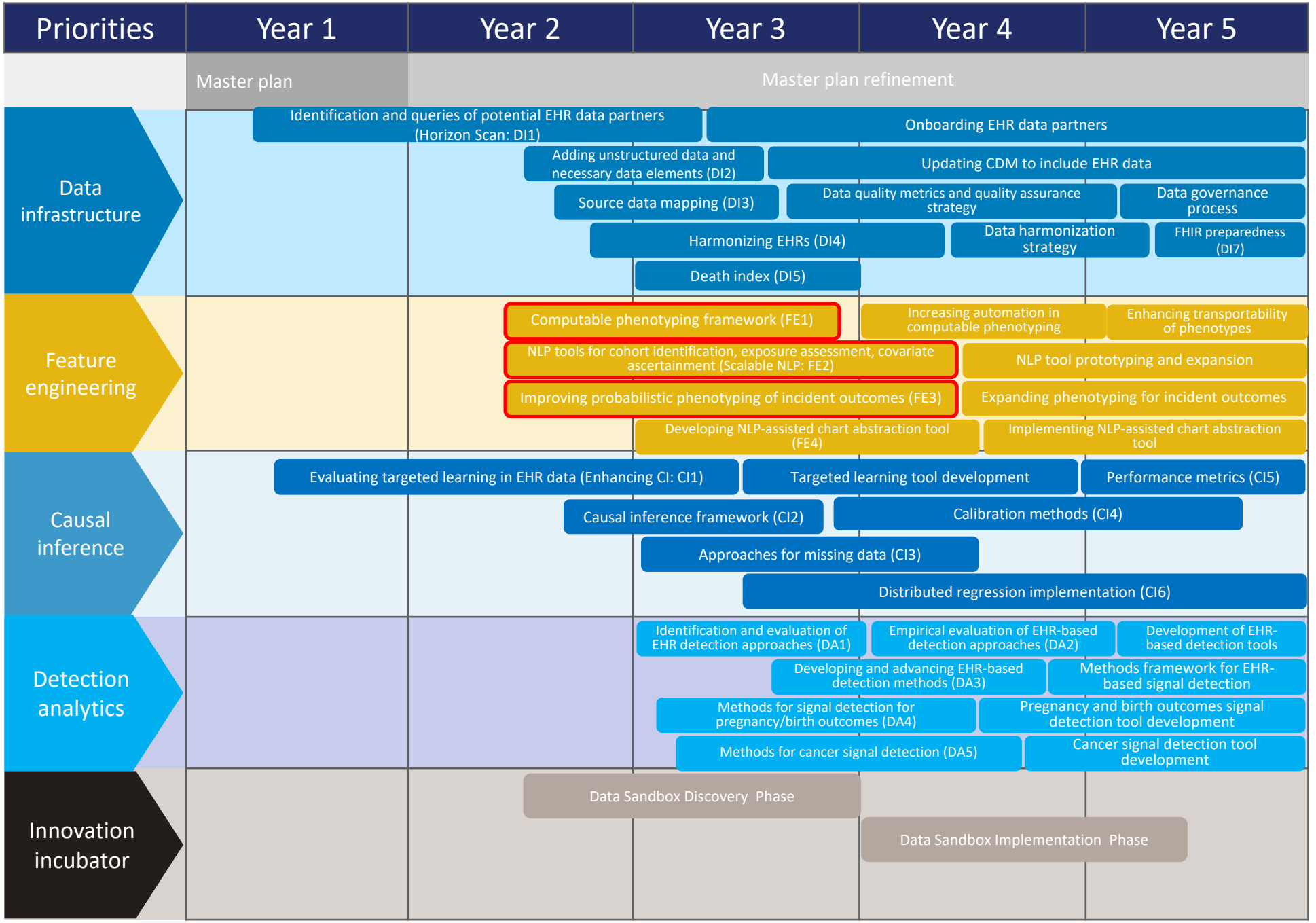
Number of patients with a medication by provenance and encounter type for a specific cohort (CKD) and the population as a whole.

Questions?

FE1: Computable phenotyping framework

FE2: NLP tools for cohort identification, exposure assessment, covariate ascertainment

FE3: Improving probabilistic phenotyping of incident outcomes



Sentinel

Health Outcomes and Covariates for Computable Phenotyping Using EHR Data

Lessons Learned from : *Advancing scalable natural language processing approaches for unstructured electronic health record data*

Workgroup Leads: David S. Carrell, PhD

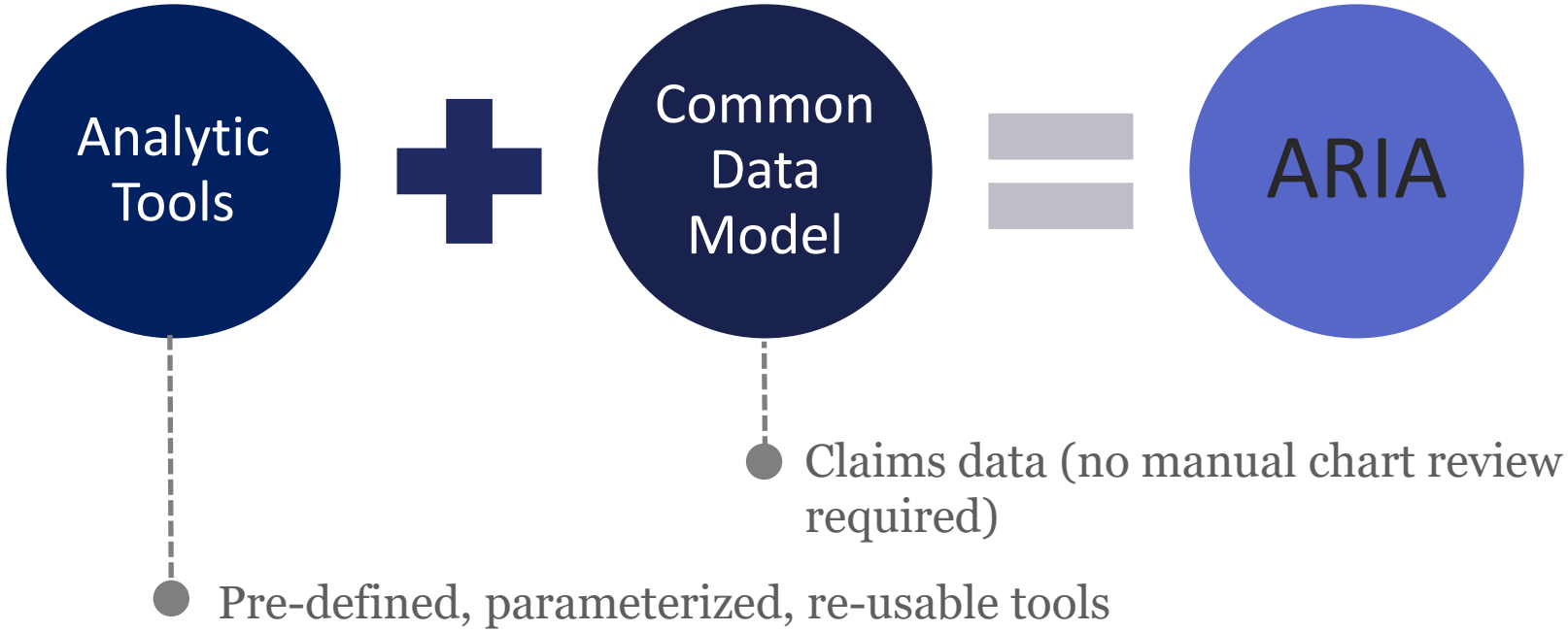
Outline

- **Motivation**
  - Role of computable algorithms in Sentinel
  - Limitations of claims data
  - The promise of using EHR data and machine learning (ML) methods
- **Scalable algorithm *development***
- **Filters in outcome identification**
  - Role in outcome identification
  - Data-driven, high-sensitivity filtering (HSF)

Motivation: Role of computable algorithms in Sentinel

Allow safety issues to be investigated rapidly, at ~low cost

ARIA = the Active Risk Identification and Analysis system

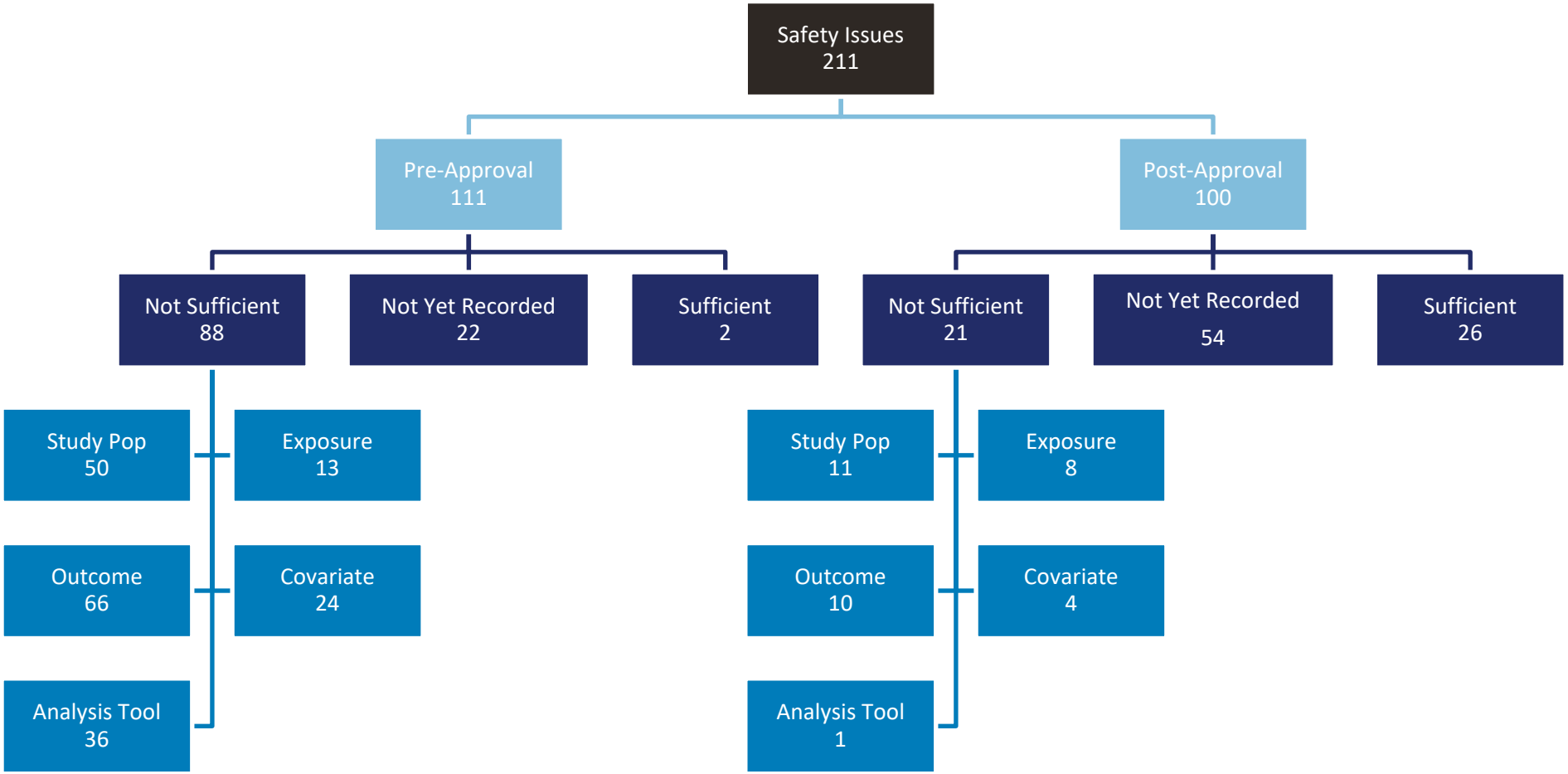


slide courtesy of Michael Nguyen

Motivation: Limitations of structured claims data

Reliance on existing Sentinel data in ARIA analyses has revealed various insufficiencies

Incorporating rich EHR data may overcome some of these insufficiencies.



This slide courtesy of Michael Nguyen

Motivation: Promise of EHR data + ML methods

Accurate identification of some outcomes/covariates requires information only available in EHR data and clinical notes

- Ex. 1: Identification of acute pancreatitis requires labs data (lipase)
- Ex. 2: Key facts for identifying anaphylaxis are absent in claims data but can be extracted from EHRs via natural language processing (NLP)

Relationships between rich features/predictors and outcomes are often nonlinear, making data-driven ML modeling advantageous

- Ex.: Computable algorithms for identifying anaphylaxis based on ML methods consistently outperformed simpler linear models

Scalable algorithm *development*

Efficiency: At reasonable **cost** in a ~short **time** frame

- Cost/time drivers are personnel salaries, gold standard creation

Portability: Easily implemented in diverse real-world settings

- Sharable tools/packages
- Minimal/no local tailoring needed
- Anticipates & accommodates *local* systems & data

Replicability

- Comparable results across settings
- Comparable results across time

Efficiency + Portability + Replicability = Scalable algorithm development

Scalable algorithm development is needed to:

- Keep pace with demand for safety analyses
- Produce results at reasonable cost

Filters: Their role in outcome identification

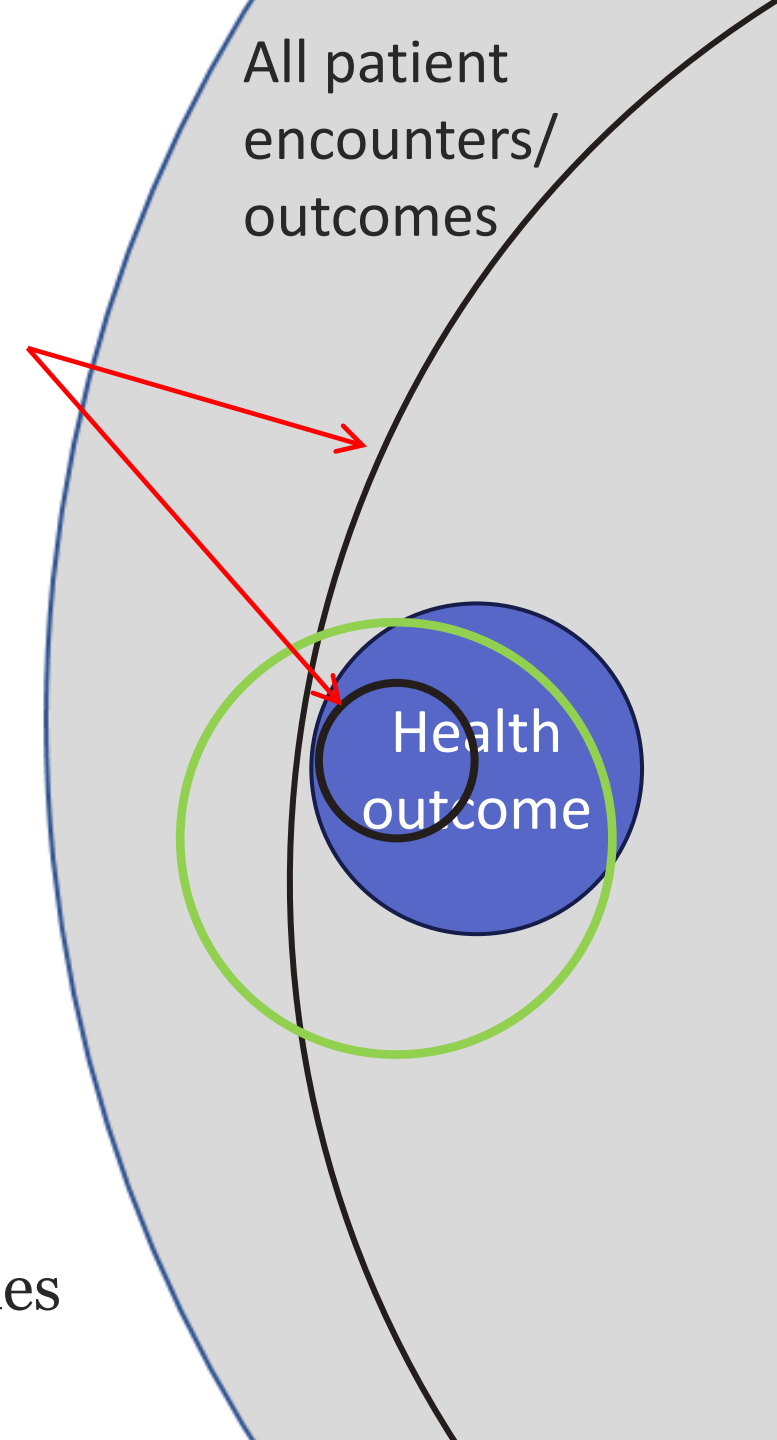
Filters are:

- Expert-specified sets of healthcare data, e.g., diagnosis, procedure, or medication codes
- That *presumptively* identify patients w/ the outcome
- For which true case status will be *determined by a computable algorithm*

Useful filters have:

- Strong face validity
- Simple and generalizable definitions
- High sensitivity (to minimize selection bias)
- Reasonable specificity
  (to limit data collection burden)

- Traditional example: COVID-19-specific ICD-10 dx codes



Filters: Data-driven, high-sensitivity filtering (HSF)

Smith et al. *Data-driven approaches to improve phenotype sensitivity using EHR data*. Under review, ICPE 2022.

Objective:

Improve sensitivity of a "traditional" filter

HSFs use data-driven analytics to identify additional filtering codes:

- To identify patients/events overlooked by simple/traditional filters,
- With modest increase in overall sample size, and
- With reasonable effort (i.e., reusable tool applied to Sentinel data)

How do HSFs work?

1. Divide patients into two groups:
   - *Ever* qualified by the traditional filter
   - *Never* qualified by the traditional filter
2. Identify codes that are ≥10x more common in "Ever" than "Never" patients
3. Manually review and retain identified codes with face validity
4. Add patients/events w/any HSF code to the presumptive patient/event set

Filters: Data-driven, high-sensitivity filtering (HSF)

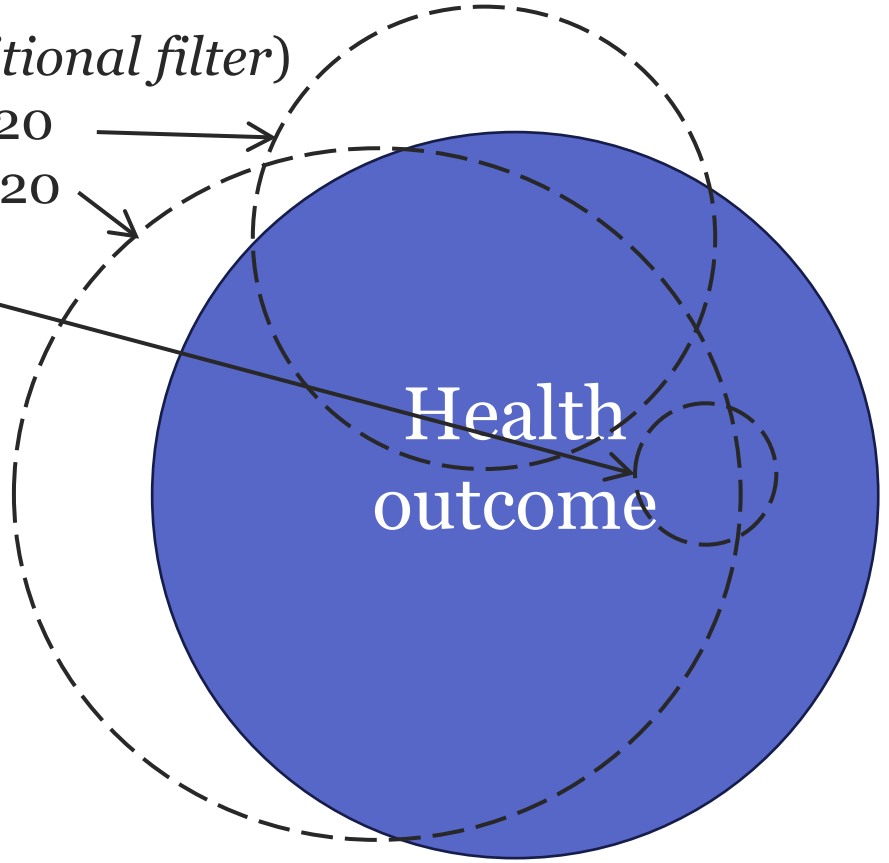
COVID-19-specific dxs [1] (*traditional filter*)

- B9729, COVID-19, pre 4/1/2020
- U07.1, COVID-19, post 4/1/2020
- Z8616, Hx of COVID-19

Filters: Data-driven, high-sensitivity filtering (HSF)

COVID-19-specific dxs [1] (*traditional filter*)

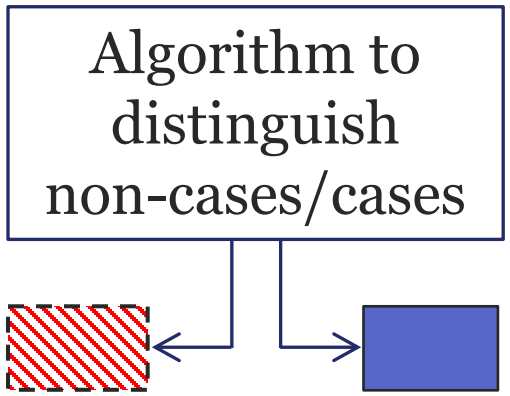
- B9729, COVID-19, pre 4/1/2020
- U07.1, COVID-19, post 4/1/2020
- Z8616, Hx of COVID-19

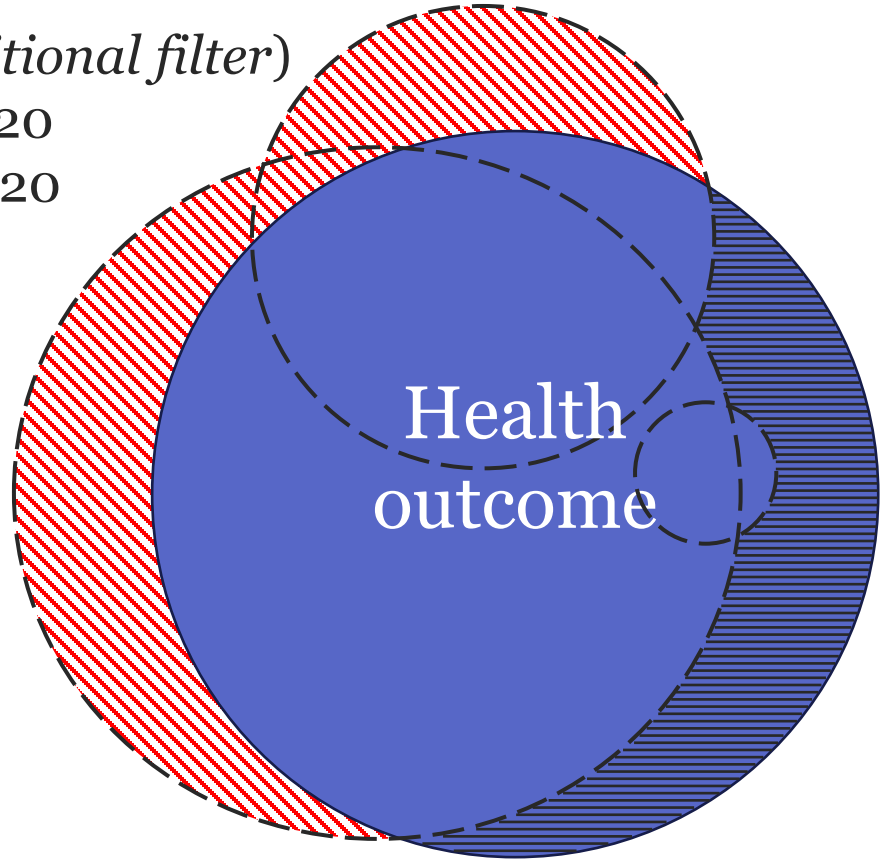


Filters: Data-driven, high-sensitivity filtering (HSF)

COVID-19-specific dxs [1] (*traditional filter*)

- B9729, COVID-19, pre 4/1/2020
- U07.1, COVID-19, post 4/1/2020
- Z8616, Hx of COVID-19



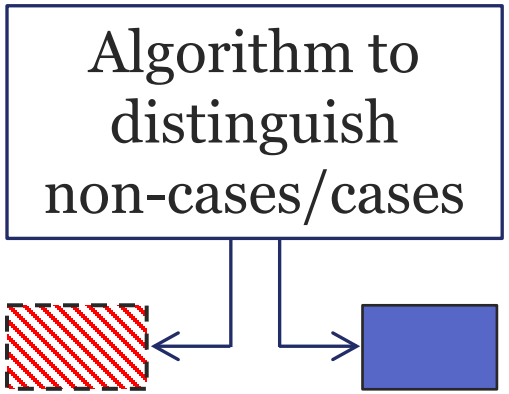


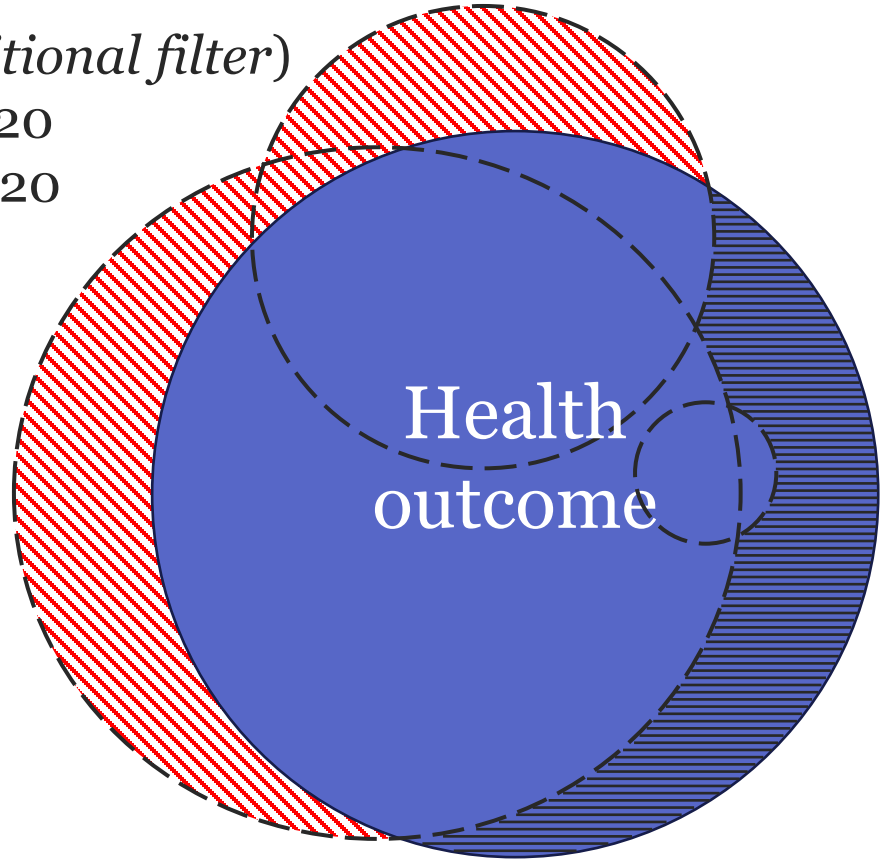
Filter false positives Filter true positives Filter overlooked (lost)

Filters: Data-driven, high-sensitivity filtering (HSF)

COVID-19-specific dxs [1] (*traditional filter*)

- B9729, COVID-19, pre 4/1/2020
- U07.1, COVID-19, post 4/1/2020
- Z8616, Hx of COVID-19





Can HSFs capture overlooked patients?

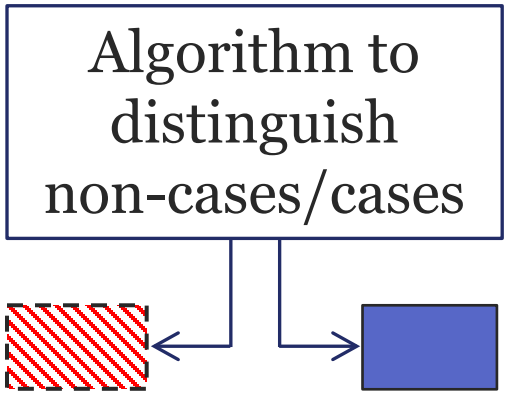
- Other diagnoses?
- Procedures?
- Medications?
- Labs? …

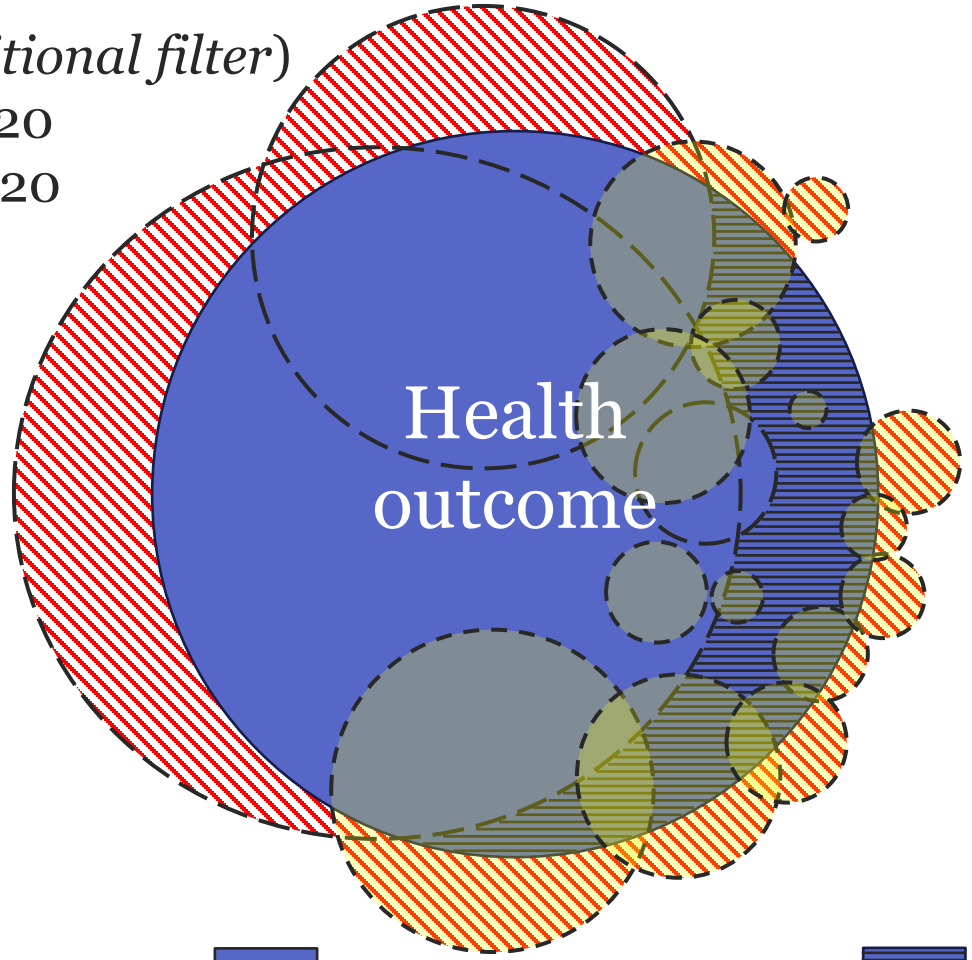
Filter false positives Filter true positives Filter overlooked (lost)

Filters: Data-driven, high-sensitivity filtering (HSF)

COVID-19-specific dxs [1] (*traditional filter*)

- B9729, COVID-19, pre 4/1/2020
- U07.1, COVID-19, post 4/1/2020
- Z8616, Hx of COVID-19





Can HSFs capture overlooked patients?

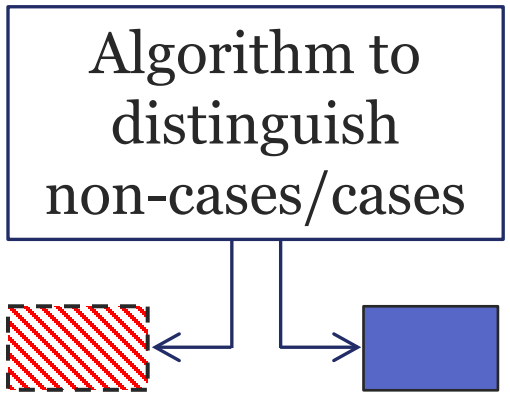
- Other diagnoses?
- Procedures?
- Medications?
- Labs? …

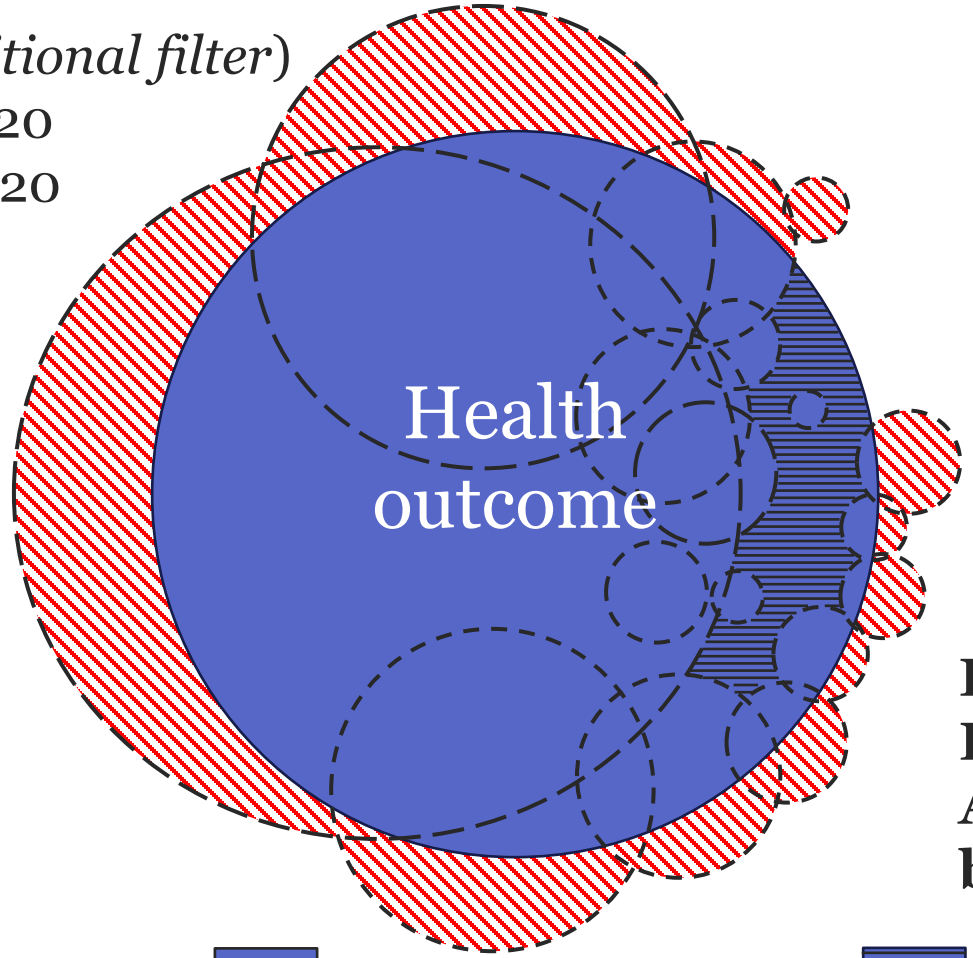
Filter false positives Filter true positives Filter overlooked (lost)

Filters: Data-driven, high-sensitivity filtering (HSF)

COVID-19-specific dxs [1] (*traditional filter*)

- B9729, COVID-19, pre 4/1/2020
- U07.1, COVID-19, post 4/1/2020
- Z8616, Hx of COVID-19





Can HSFs capture overlooked patients?

- Other diagnoses?
- Procedures?
- Medications?
- Labs? …

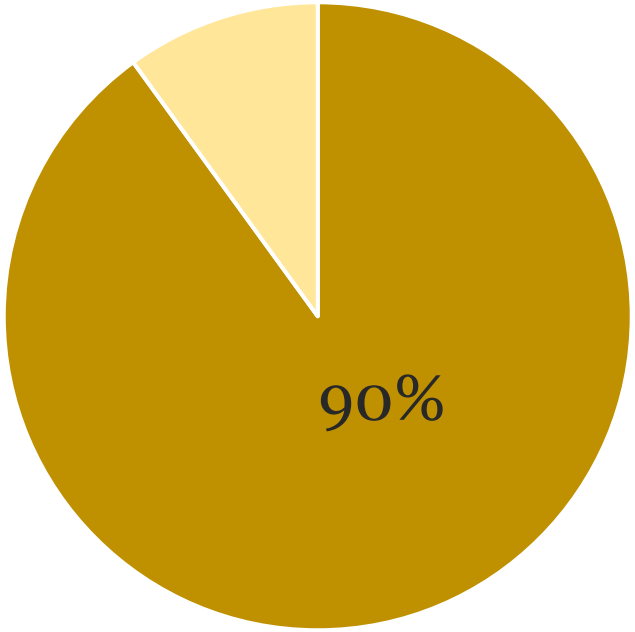
If so …
How many (sensitivity)?
At what cost (data burden)?

Filter false positives | Filter true positives | Filter overlooked (lost)

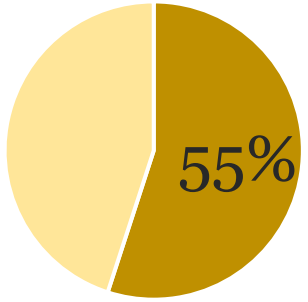
Results: COVID-19 high-sensitivity filtering (HSF)

VU: +13% true cases, +22% pts

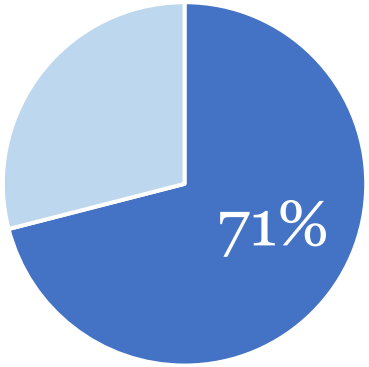
20,951 patients
~90% true case rate



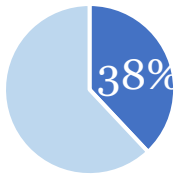
4,566 patients (+22%)
~55% true case rate



KP: +10% true cases, +22% pts



6,847 patients
~71% true case rate



1,482 patients (+22%)
~38% true case rate

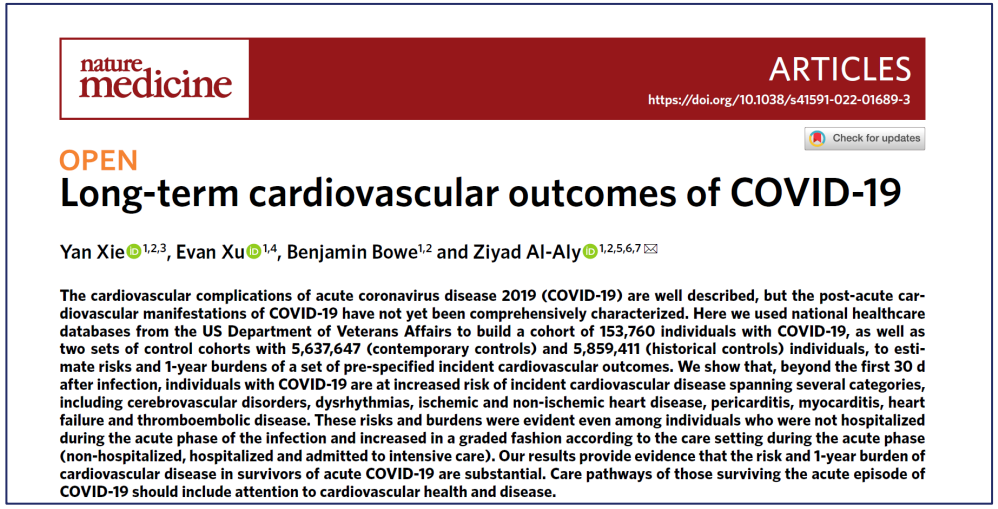
Thank You!

David S. Carrell, PhD
Kaiser Permanente Washington Health Research Institute
Seattle, WA
david.s.carrell@kp.org

Extras

COVID-19 as a covariate in safety studies?

- Nature Medicine
  "… beyond the first 30 d after infection, individuals with COVID-19 are at increased risk of incident cardiovascular disease spanning several categories, including cerebrovascular disorders, dysrhythmias, ischemic and non-ischemic heart disease, pericarditis, myocarditis, heart failure and thromboembolic disease."



nature medicine

ARTICLES

https://doi.org/10.1038/s41591-022-01689-3

Check for updates

OPEN

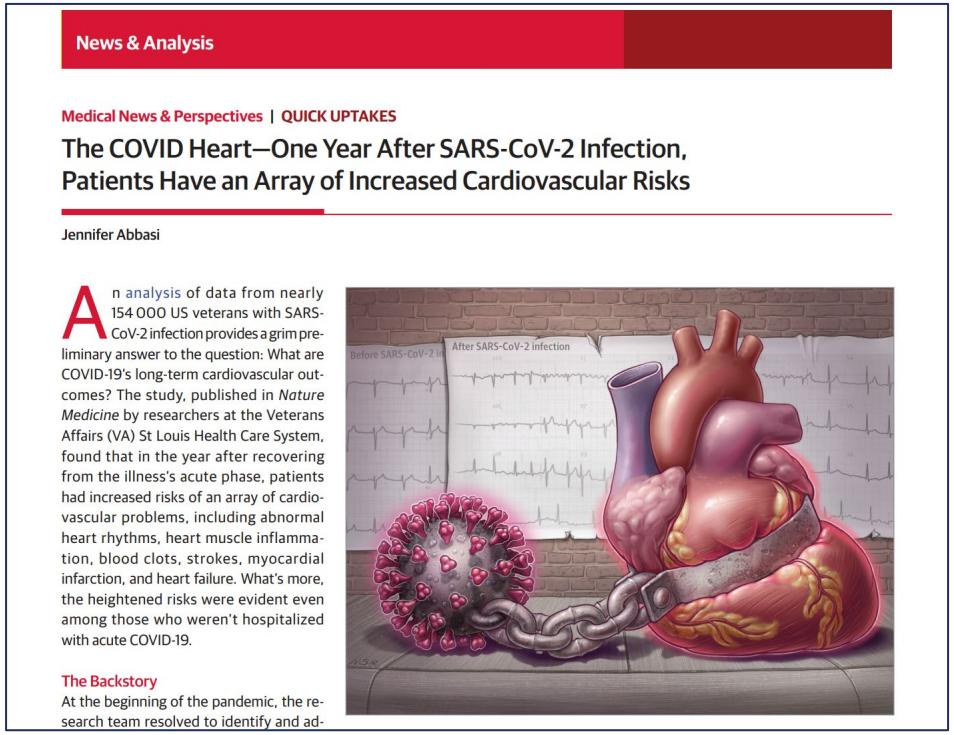
Long-term cardiovascular outcomes of COVID-19

Yan Xie[1,2,3], Evan Xu[1,4], Benjamin Bowe[1,2] and Ziyad Al-Aly[1,2,5,6,7]

The cardiovascular complications of acute coronavirus disease 2019 (COVID-19) are well described, but the post-acute cardiovascular manifestations of COVID-19 have not yet been comprehensively characterized. Here we used national healthcare databases from the US Department of Veterans Affairs to build a cohort of 153,760 individuals with COVID-19, as well as two sets of control cohorts with 5,637,647 (contemporary controls) and 5,859,411 (historical controls) individuals, to estimate risks and 1-year burdens of a set of pre-specified incident cardiovascular outcomes. We show that, beyond the first 30 d after infection, individuals with COVID-19 are at increased risk of incident cardiovascular disease spanning several categories, including cerebrovascular disorders, dysrhythmias, ischemic and non-ischemic heart disease, pericarditis, myocarditis, heart failure and thromboembolic disease. These risks and burdens were evident even among individuals who were not hospitalized during the acute phase of the infection and increased in a graded fashion according to the care setting during the acute phase (non-hospitalized, hospitalized and admitted to intensive care). Our results provide evidence that the risk and 1-year burden of cardiovascular disease in survivors of acute COVID-19 are substantial. Care pathways of those surviving the acute episode of COVID-19 should include attention to cardiovascular health and disease.

Xie, Y., Xu, E., Bowe, B. et al. Long-term cardiovascular outcomes of COVID-19.
Nat Med (Feb. 7, 2022). https://doi.org/10.1038/s41591-022-01689-3

- JAMA
  "Physicians should consider a history of COVID-19 as a cardiovascular disease risk."



News & Analysis

Medical News & Perspectives | QUICK UPTAKES

The COVID Heart—One Year After SARS-CoV-2 Infection, Patients Have an Array of Increased Cardiovascular Risks

Jennifer Abbasi

An analysis of data from nearly 154 000 US veterans with SARS-CoV-2 infection provides a grim preliminary answer to the question: What are COVID-19's long-term cardiovascular outcomes? The study, published in *Nature Medicine* by researchers at the Veterans Affairs (VA) St Louis Health Care System, found that in the year after recovering from the illness's acute phase, patients had increased risks of an array of cardiovascular problems, including abnormal heart rhythms, heart muscle inflammation, blood clots, strokes, myocardial infarction, and heart failure. What's more, the heightened risks were evident even among those who weren't hospitalized with acute COVID-19.

The Backstory

At the beginning of the pandemic, the research team resolved to identify and ad-

Abbasi J. The COVID Heart—One Year After SARS-CoV-2 Infection, Patients Have an Array of Increased Cardiovascular Risks. *JAMA*. Published online March 02, 2022. doi:10.1001/jama.2022.2411

FE2: NLP tools for cohort identification, exposure assessment, covariate ascertainment ("Scalable NLP")

Goal: In two heterogeneous settings develop and validate scalable and reusable NLP tools for leveraging EHR data to address known insufficiencies in existing data and methods to support FDA safety surveillance studies

Progress:

Objective 1: Cohort identification

- Developed and evaluated scalable, replicable approaches to cohort identification in Sentinel safety studies
- Products: ICPE 2022 and AMIA 2022 abstracts (at right)

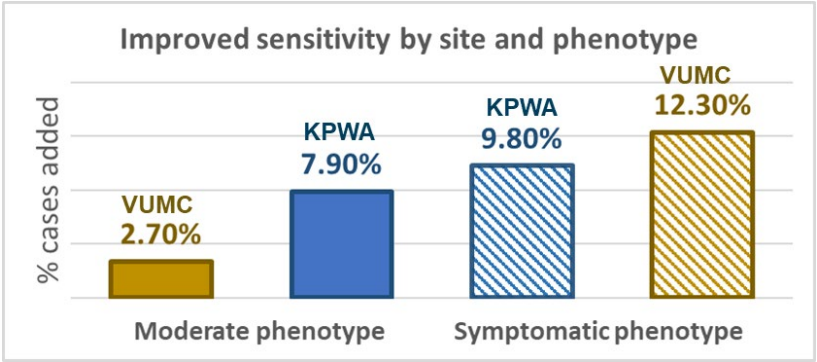
Objective 2: Scalable NLP measures

- Develop, apply and evaluate scalable, replicable methods for NLP-based measurement of exposures, symptoms, and outcomes

Objective 3: Evaluation

- Compare structured data versus NLP for capturing exposures, health outcomes of interest, and covariates

Improving sensitivity of cohort identification



Automated cohort identification model (COVID-19)

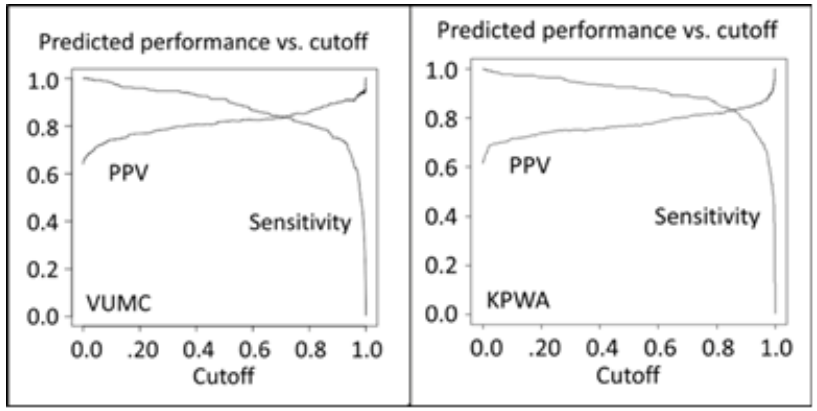


Figure 1. Prediction performance for Silver Standard 4 for symptomatic phenotype: VUMC (left), KPWA (right).

Deliverable for the IC: Manuscript describing key products of this work and a GitHub repository of reusable tools and methods for incorporating scalable NLP into Sentinel safety studies

WG Leads: David Carrell, KPWA, Joshua Smith, VUMC; Timeline: 2/1/2021-1/31/2023

High-sensitivity COVID-19 filter results -- VUMC

VUMC patients identified by COVID-19 "base" and "high-sensitivity" (HSF) filters during study period				
Filter rank	**COVID-19 filter category**	**N patients with this filter**	**N patients with this filter *and no higher rank filters***	**Percent of all patients identified by this filter**
1st	Diagnosis of U07.1 "COVID-19" (base #1)	20,840	20,840	80%
2nd	Any of 5 other COVID-19 diagnoses (base #2)	1,898	111	0.43%
3rd	HSF diagnoses (any of 24)	7,264	3,976	15%
4th	HSF procedures (any of 10)	1198	37	0.14%
5th	HSF medications (any of 4)	473	181	0.70%
6th	HSF problem list in EHR (any of 5)	9,222	892	3.4%
Total			26,037	100%
If we included a 7th filter, PCR+ COVID-19 test (only), 8,825 (+34%) new patients would be added.				

High-sensitivity COVID-19 filter results -- KPWA

KPWA patients identified by COVID-19 "base" and "high-sensitivity" (HSF) filters during study period				
Filter rank	**COVID-19 filter category**	**N patients with this filter**	**N patients with this filter *and no higher rank filters***	**Percent of all patients identified by this filter**
1st	Diagnosis of U07.1 "COVID-19" (base #1)	15,678	15,678	81%
2nd	Any of 5 other COVID-19 diagnoses (base #2)	1,498	166	1%
3rd	HSF diagnoses (any of 24)	5,041	2789	14%
4th	HSF procedures (any of 10)	550	8	0.04%
5th	HSF medications (any of 4)	91	84	0.4%
6th	HSF problem list in EHR (any of 5)	4,845	607	3%
Total			19,332	100%
If we included a 7th filter, PCR+ COVID-19 test (only), 4,737 (+25%) new patients would be added.				

Sentinel

Automated Methods for Developing Computable Phenotypes

Lessons Learned from : *Advancing scalable natural language processing approaches for unstructured electronic health record data*

Workgroup Leads: Joshua C. Smith & David S. Carrell

Phenotyping

Computable phenotype algorithms typically:

- Require time-intensive expert curation and feature engineering
- Require manually-annotated gold- standard training sets
- Result in high cost and limited scalability.

PheNorm, and similar automated approaches:

- Based on natural language processing (NLP), machine learning, and (low-cost) silver-standard training labels
- Have been demonstrated to perform well for various chronic health conditions.

We evaluated PheNorm for use with *acute* conditions (COVID-19)

- PheNorm currently being applied to acute pancreatitis in another IC project

Rationale for exploring automating phenotyping methods

Scalability

- Manual approach is burdensome/slow, requires substantial expertise

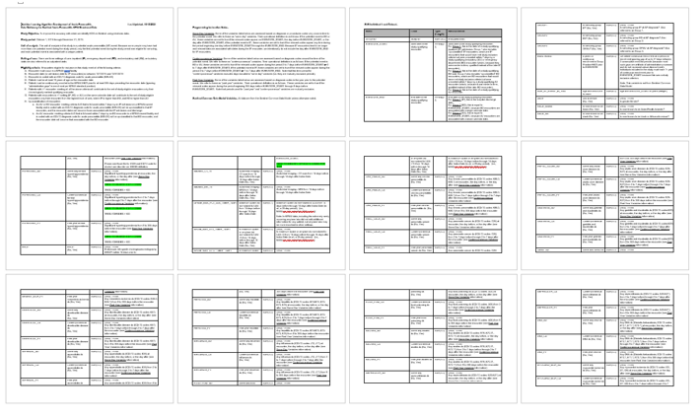
Replicability

- Reduced operator-dependence

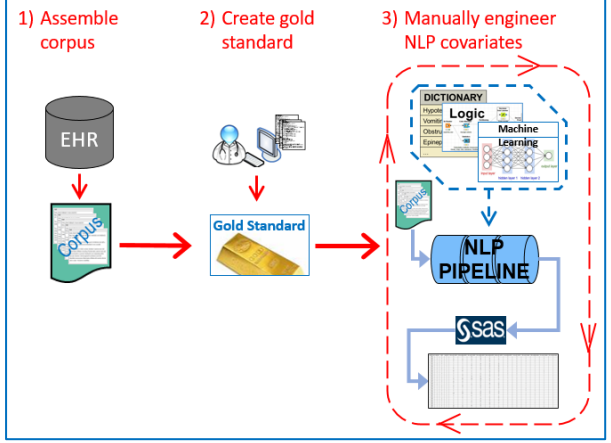
Hybrid solutions?

- PheNorm → PheCAP → blended methods?

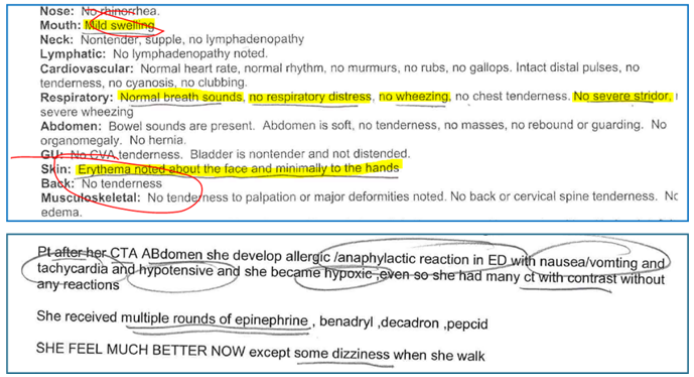
Manually curated structured covariates



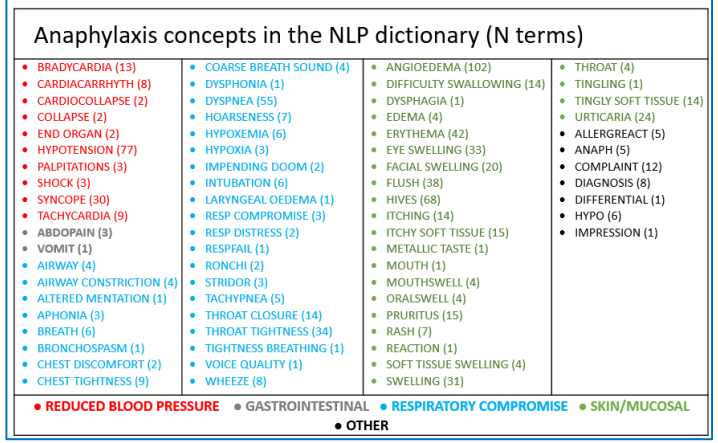
Manual NLP Process



Manual chart review by experts to discover relevant NLP covariates



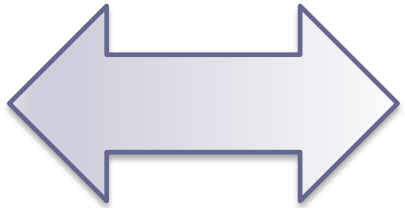
Manually Curate NLP Dictionaries



Rationale for exploring automating phenotyping methods

Continuum of development approaches

Manual development
- *Expert*-driven
- *Manual* engineering
- Heavy reliance on *gold standard labels*
- Substantial operator dependence
- Slow



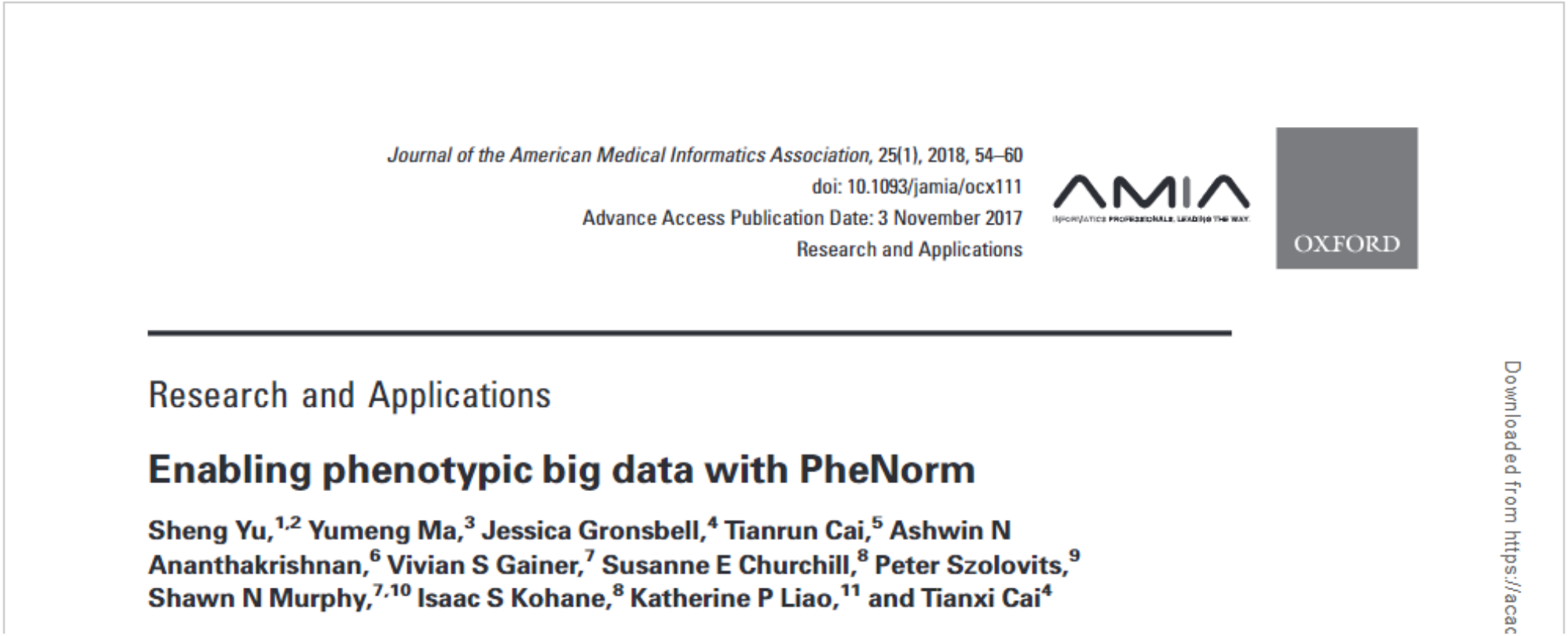
Automated development
- Data-driven
- Automated engineering
- Heavy reliance on silver standard labels
- Reduced operator dependence
- Fast

- Automated feature engineering (AFEP)[1]
- Surrogate-assisted feature extraction (SAFE)[2]
- Phenotype algorithm normalization (PheNorm)[3]
- Phenotyping common approach (PheCAP)[4]

1. Yu et al. Toward high-throughput phenotyping: unbiased automated feature extraction and selection from knowledge sources. JAMIA 2015
2. Yu et al. Surrogate-assisted feature extraction for high-throughput phenotyping. JAMIA 2017
3. Yu et al. Enabling phenotypic big data with PheNorm. JAMIA 2018
4. Zhang et al. High-throughput phenotyping with EMR data using a common semi-supervised approach (PheCAP). Nature Protocols. 2019

Automated modeling: PheNorm

Sheng Yu, Yumeng Ma, Jessica Gronsbell, Tianrun Cai, Ashwin N Ananthakrishnan, Vivian S Gainer, Susanne E Churchill, Peter Szolovits, Shawn N Murphy, Isaac S Kohane, Katherine P Liao, Tianxi Cai. **Enabling phenotypic big data with PheNorm**. J Am Med Inform Assoc. **2018** Jan 1;25(1):54-60.



Journal of the American Medical Informatics Association, 25(1), 2018, 54–60
doi: 10.1093/jamia/ocx111
Advance Access Publication Date: 3 November 2017
Research and Applications

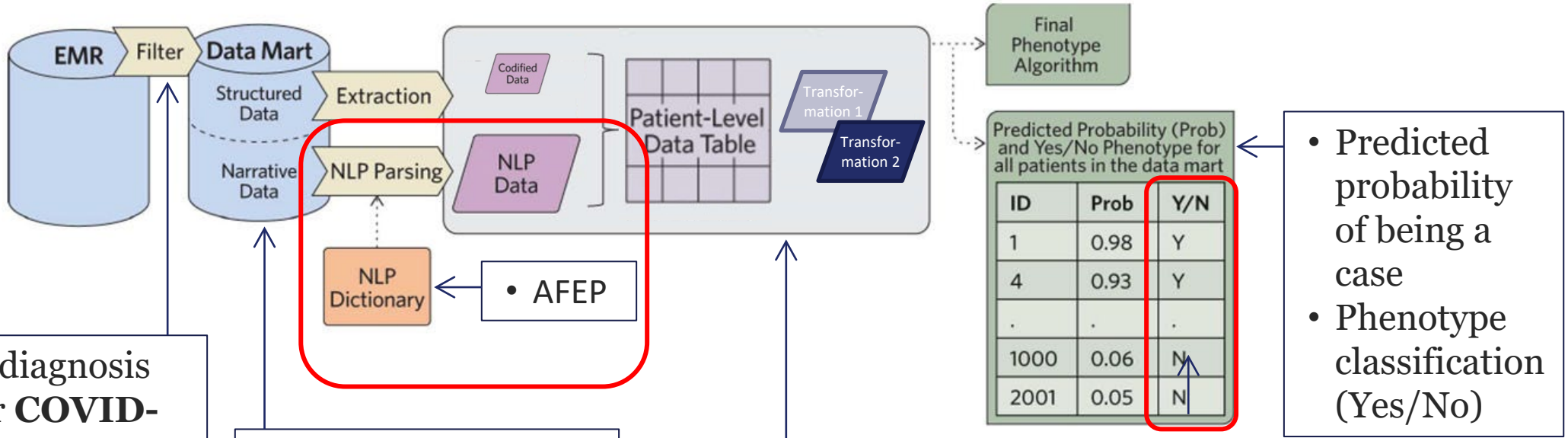
Research and Applications

Enabling phenotypic big data with PheNorm

Sheng Yu,[1,2] Yumeng Ma,[3] Jessica Gronsbell,[4] Tianrun Cai,[5] Ashwin N Ananthakrishnan,[6] Vivian S Gainer,[7] Susanne E Churchill,[8] Peter Szolovits,[9] Shawn N Murphy,[7,10] Isaac S Kohane,[8] Katherine P Liao,[11] and Tianxi Cai[4]

Overview of PheNorm/PheCap

Zheng et al. High-throughput phenotyping with electronic medical record data using a common semi-supervised approach (PheCAP). Nat protocols. 2019 Dec;14(12):3426-3444. doi: 10.1038/s41596-019-0227-6. Epub 2019 Nov 20.



ICD-10 diagnosis codes for **COVID-19** and other diagnosis codes (HSF*)
- Cases & non-cases

- Index date
- Catchment period
- Define **silver** label & counts of DX codes

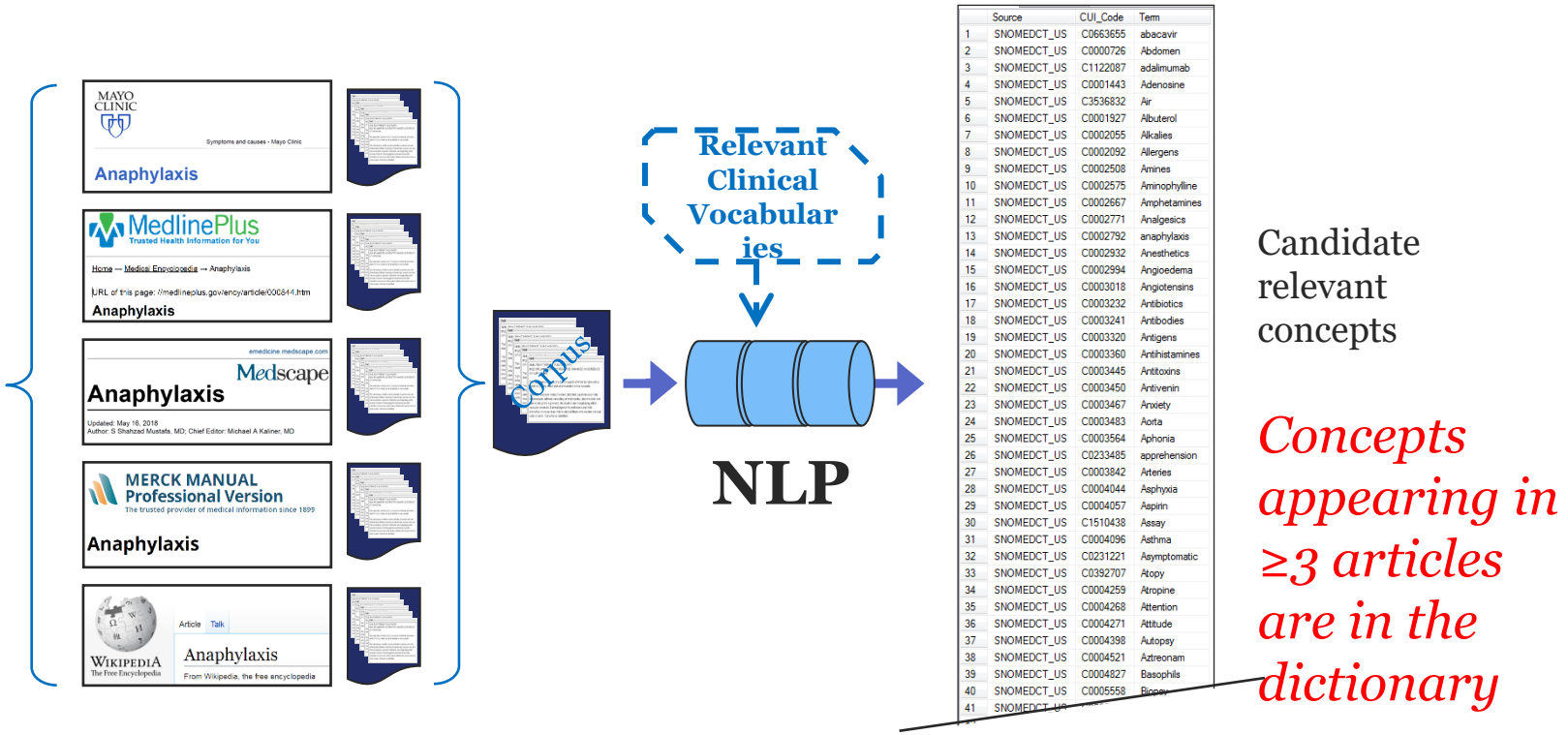
- Feature selection
- Transformations

- Predicted probability of being a case
- Phenotype classification (Yes/No)

- Evaluate using gold-labeled data (not included)

Automating NLP dictionary creation (AFEP)

5 clinical knowledge base articles on a topic



Candidate relevant concepts

Concepts appearing in ≥3 articles are in the dictionary

Yu et al. Toward high-throughput phenotyping: unbiased automated feature extraction and selection from knowledge sources. JAMIA 2015

Running PheNorm

AFEP Dictionary

- 159 CUIs extracted from 6 articles on COVID-19

Data/text catchment Period

- Index date +/-30 days

Input Data

- KPWA: 143,584 notes from 8,329 patients
- VUMC: Approximately 1.1 million notes from 24,355 patients

Process notes using MetaMapLite

- Transform counts of each NLP-extracted concept from the AFEP dictionary into input vectors for PheNorm

Running PheNorm

Silver Standard Labels

1. **Structured Label** – count of days with U07.1 diagnosis code (COVID-19)
2. **Structured Label** – counts of six COVID-related CUIs
3. **NLP Label** – Cumulative count of "COVID-19" mentions in patients' charts
4. **NLP Label** – number of days (KPWA) or notes (VUMC) in which a COVID-19 concepts was mentioned in patients charts

- Apply PheNorm, **evaluate**

COVID-19 Phenotype

Evidence of COVID-19 infection

Definite or highly probable infection

- Lab data or clinical note indicates patient was PCR-positive ***or***
- Assertion the patient has COVID-19 in a free text statement ***or***
- Strong evidence of proximal exposure and serologic evidence of prior infection

Probable or possible infection

- Patient symptoms are consistent with a diagnosis of COVID-19
- Absence of an explicit *alternative* diagnosis and/or absence of a statement that a non-COVID-19 cause is more likely
- Strong evidence of proximal exposure

Unlikely infection

- Explicit *alternative* diagnosis or statement that a non-COVID-19 cause is more likely
- Absence of symptoms consistent with a diagnosis of COVID-19 *and* absence of lab data or clinical note indicating a positive PCR test

Not infected

- No indication in the EHR of infection [i.e., symptoms, exposure, and/or labs/serology] during the relevant time window) EHR appears to thoroughly document the patient's care during the relevant time window

Insufficient Information

- EHR appears not to be a reasonably complete source of documentation about the patient's care during the relevant time window

Severity of illness scale (NIH)

SEVERITY LEVEL	SIGN/SYMPTOM
Asymptomatic	No symptoms
Mild	Fever (>=100.4F)
	Cough
	Sore throat
	Malaise/fatigue
	Headache
	Muscle pain
	Nausea
	Vomiting
	Diarrhea
	Loss of sense of taste or smell
Moderate	Shortness of breath (SpO2 >=94%)
	Dyspnea (SpO2 >=94%)
	Abnormal chest imaging (SpO2 >=94%)
Severe	SpO2 <94%
	PaO2/FiO2* <300 mm Hg
	Respiratory freq >30 breaths/min
	Lung infiltrates >50%
Critical	Respiratory failure
	Septic shock
	Multiple organ dysfunction

COVID-19 phenotype chart review results

Gold standard chart review results by study site and COVID-19 phenotype definition				
Study site	**COVID-19 phenotype definition**	**Chart review result**	**Number of charts**	**Percent of charts**
VUMC (N=483)	Moderate+ severity	Non-case	334	69%
		Case	149	**31%**
	Mild+ severity	Non-case	188	39%
		Case	295	**61%**
KPWA (N=437)	Moderate+ severity	Non-case	315	72%
		Case	122	**28%**
	Mild+ severity	Non-case	168	38%
		Case	269	**62%**

Chart samples were stratified to represent all filter types (not a random sample of all eligible charts)

PheNorm Results – Moderate+ Phenotype

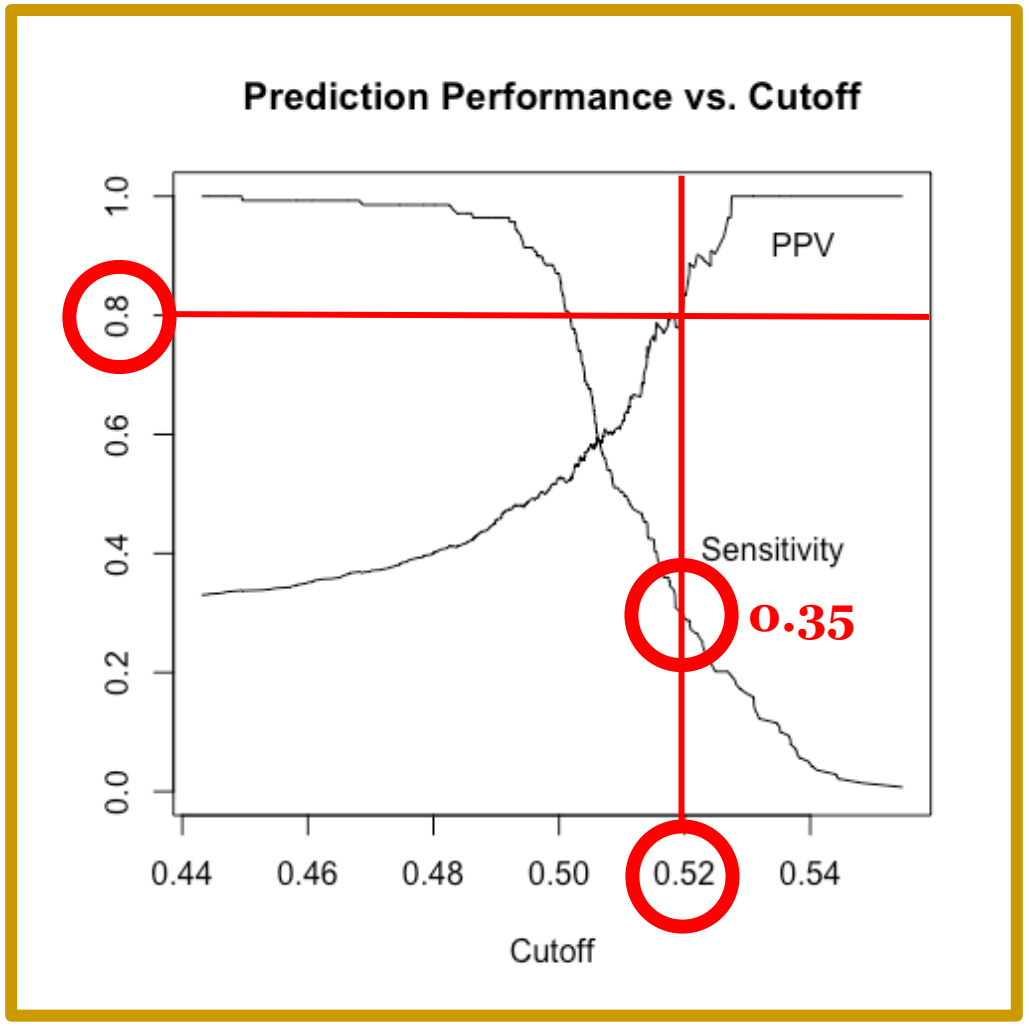
Site	Silver Standard	Phenotype	AUC	Sensitivity at PPV=0.8
KPWA	1 - U07.1 Days	Moderate+	0.700	0.07
VUMC	1 - U07.1 Days	Moderate+	0.814	0.29
KPWA	2 - Six-CUI Days	Moderate+	0.695	0.05
VUMC	2 - Six-CUI Days	Moderate+	0.841	0.47
KPWA	3 - COVID Mentions	Moderate+	0.674	0.00
VUMC	3 - COVID Mentions	Moderate+	0.775	0.29
KPWA	4A - CUI Days	Moderate+	0.695	0.00
VUMC	4B - CUI Notes	Moderate+	0.768	0.27

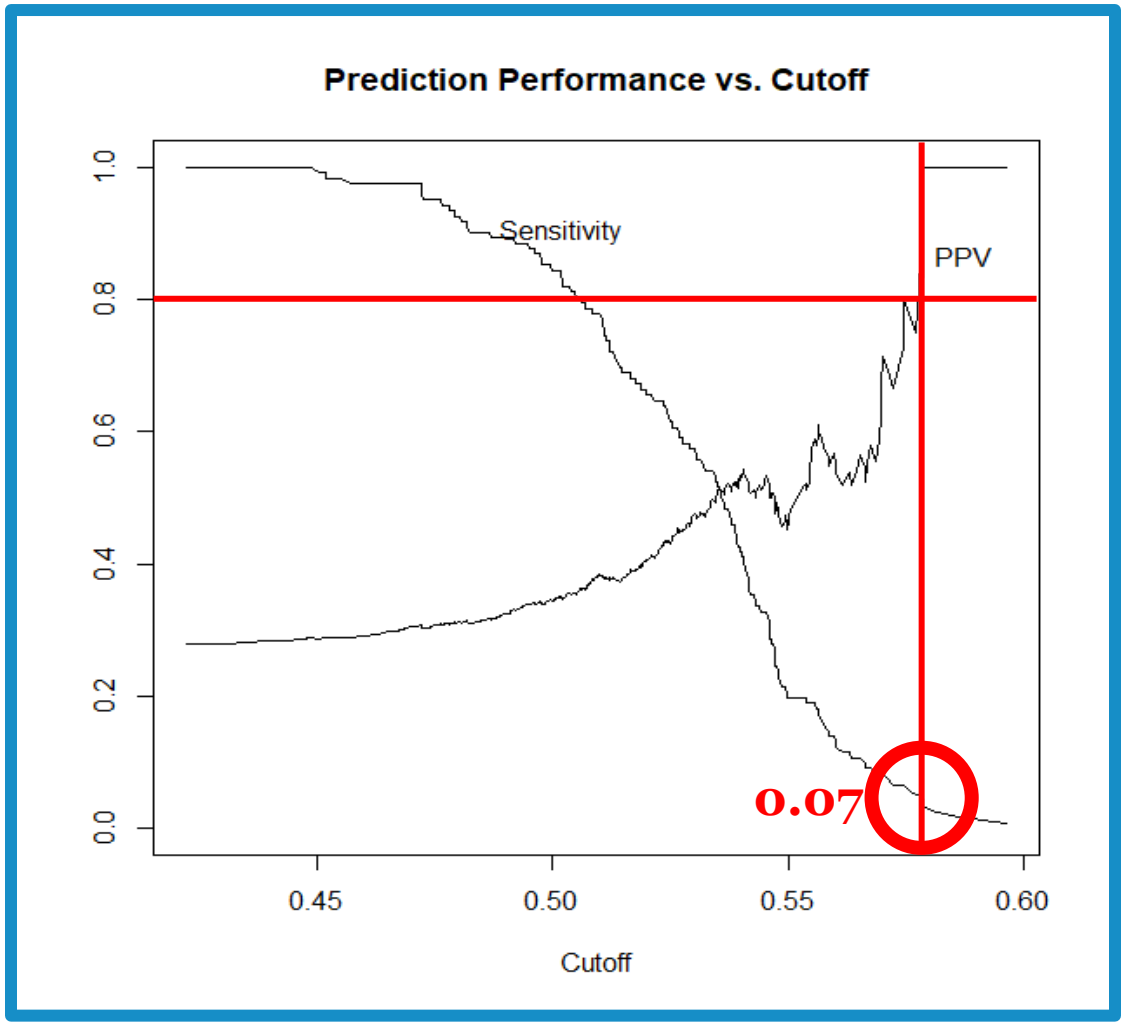
PheNorm Results – Symptomatic COVID-19

Site	Silver Standard	Phenotype	AUC	Sensitivity at PPV=0.8
KPWA	1 - U07.1 Days	Symptomatic	0.773	0.89
VUMC	1 - U07.1 Days	Symptomatic	0.901	0.99
KPWA	2 - Six-CUI Days	Symptomatic	0.766	0.88
VUMC	2 - Six-CUI Days	Symptomatic	0.899	0.95
KPWA	3 - COVID Mentions	Symptomatic	0.864	0.98
VUMC	3 - COVID Mentions	Symptomatic	0.887	0.94
KPWA	4A - CUI Days	Symptomatic	0.892	0.98
VUMC	4B - CUI Notes	Symptomatic	0.875	0.95

Prediction Performance

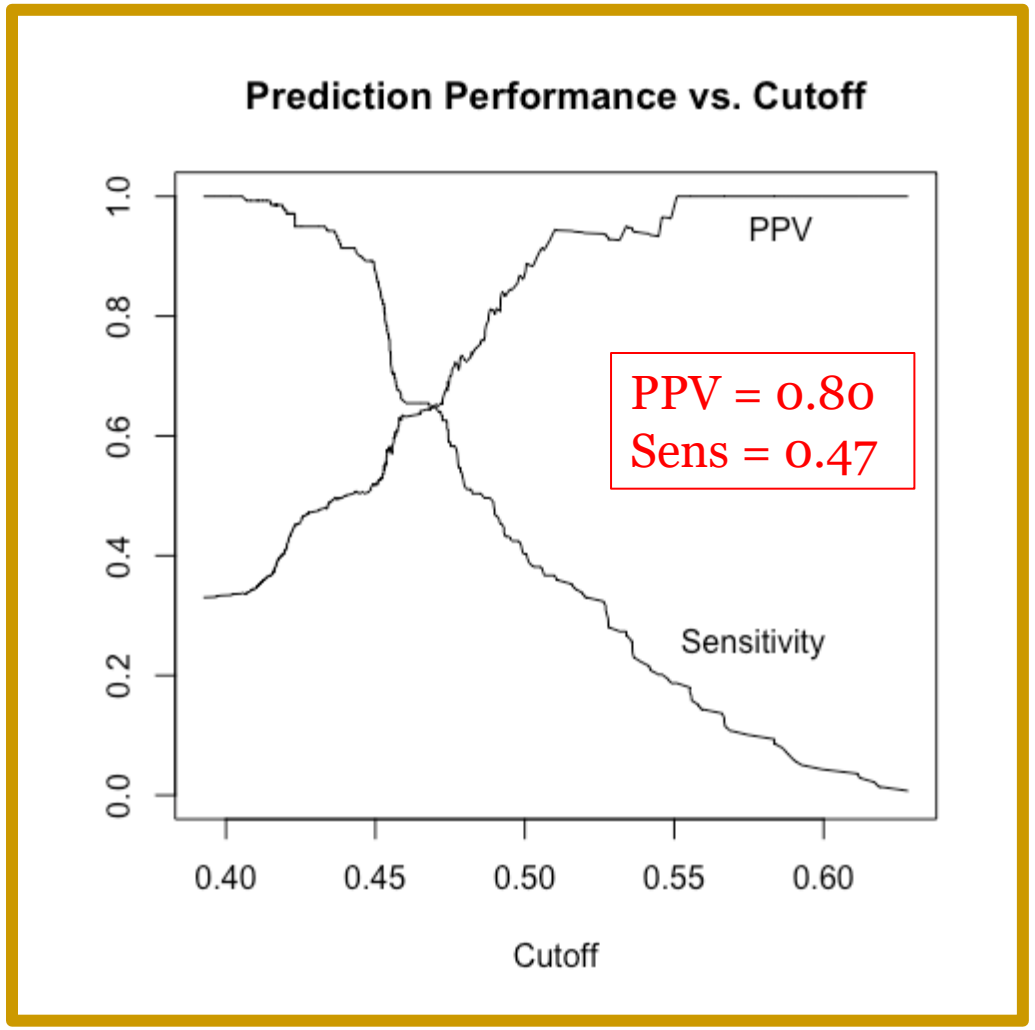
Moderate+ phenotype, Silver #1 – U07.1 Days

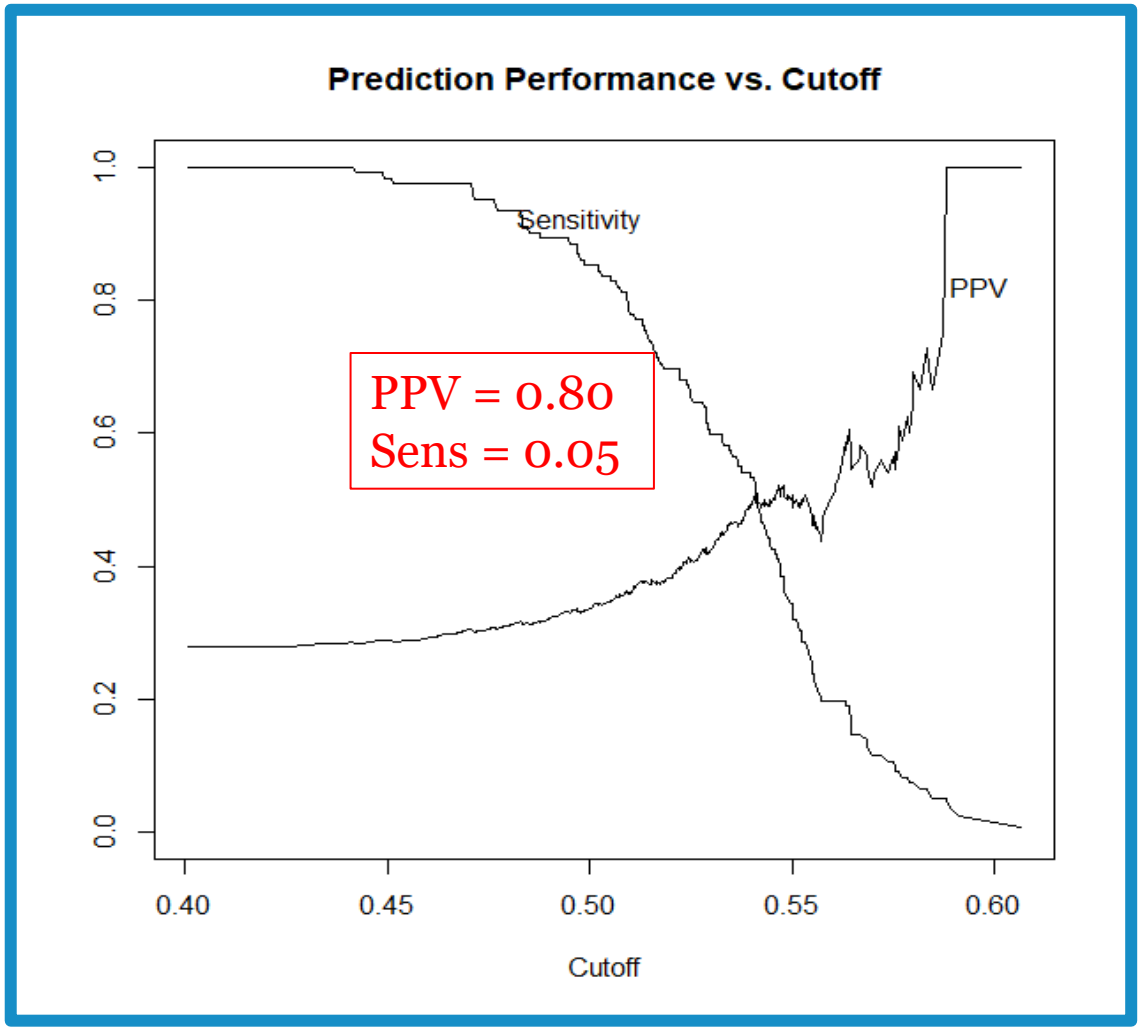




Prediction Performance

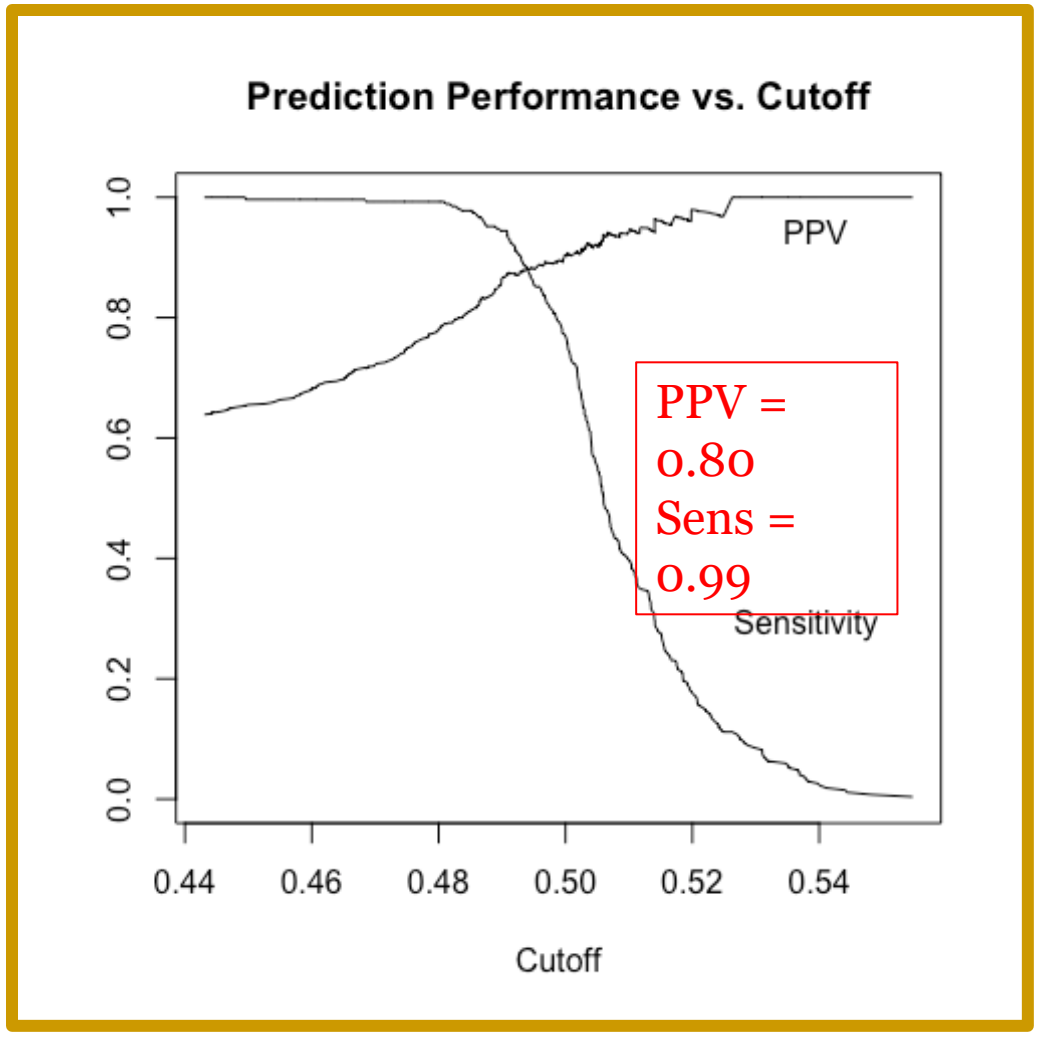
Moderate+ phenotype, **Silver #2 – "Six-CUI" Days**

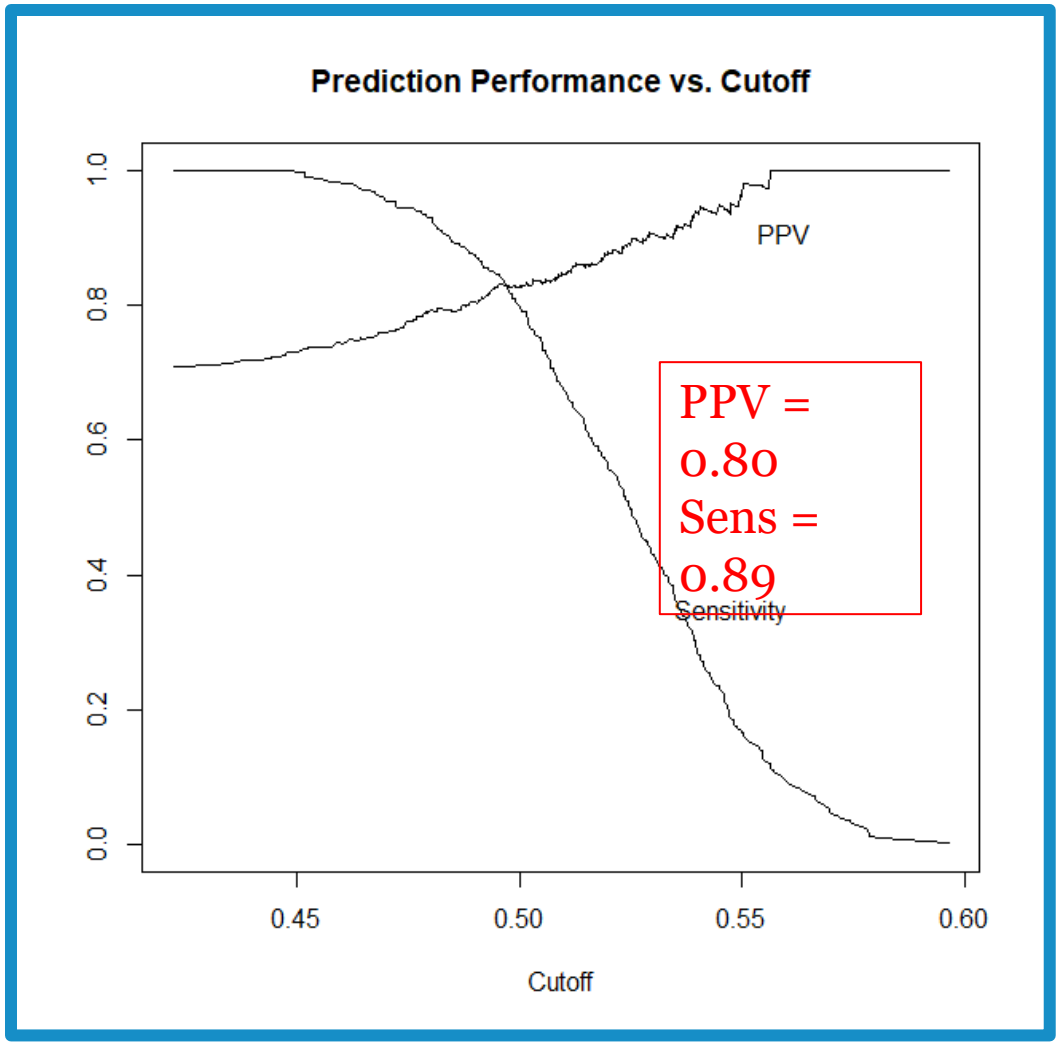




Prediction Performance

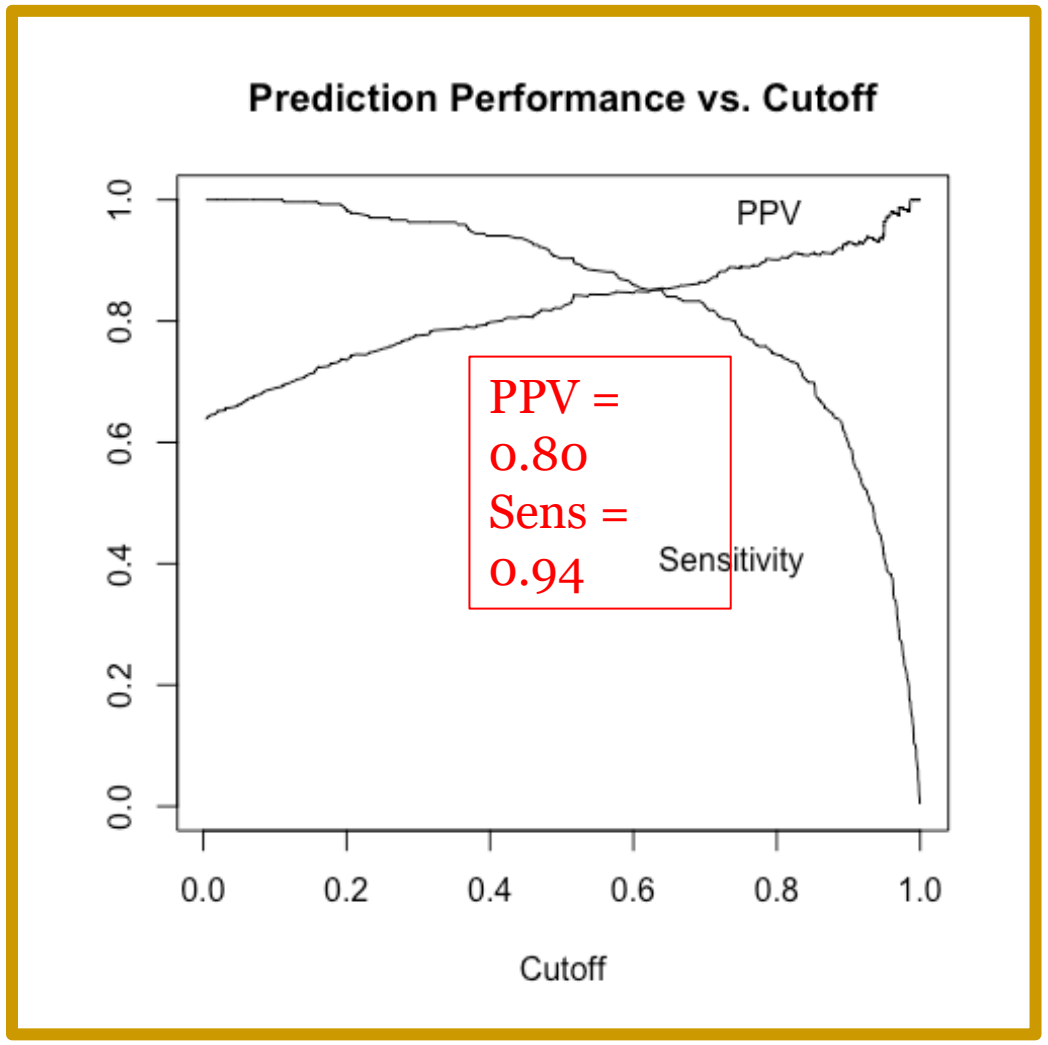
Mild+ phenotype, Silver #1 – U07.1 Days

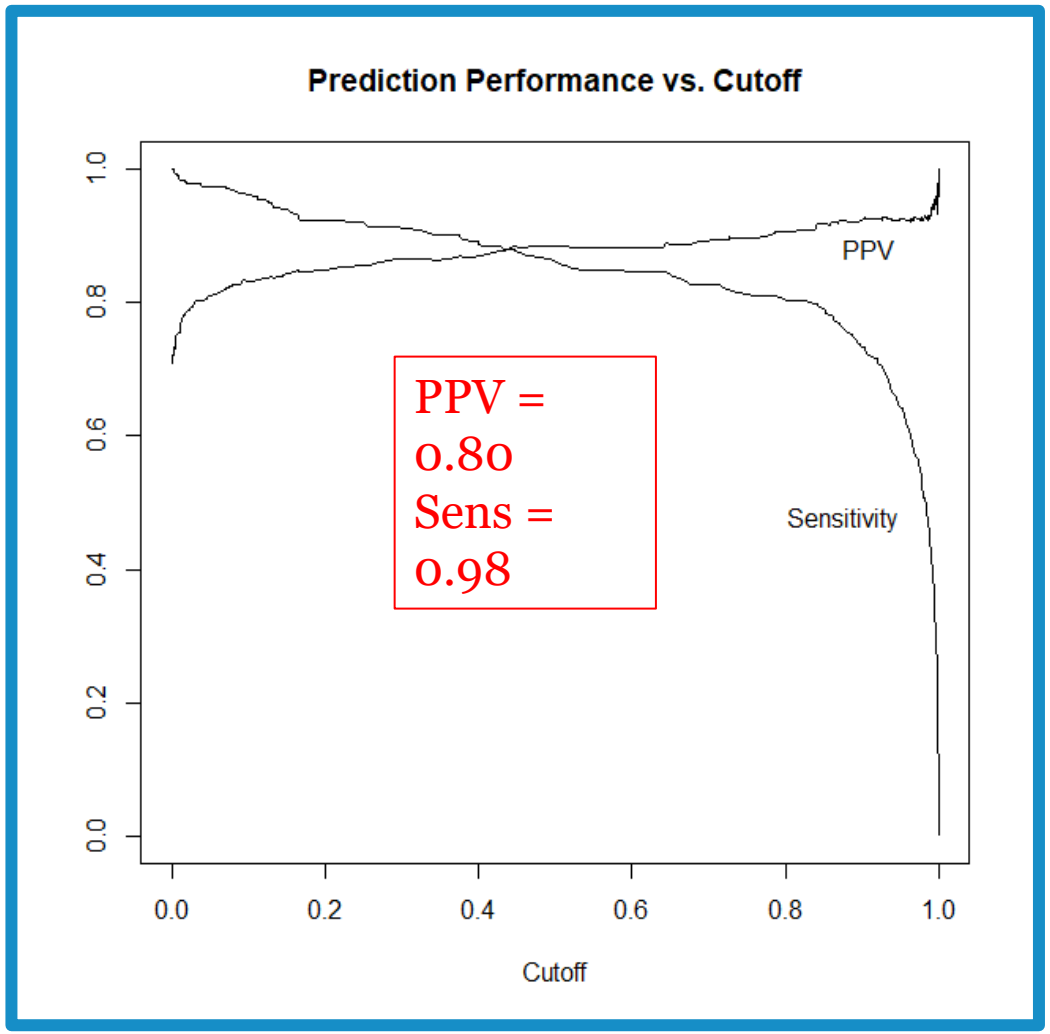




Prediction Performance

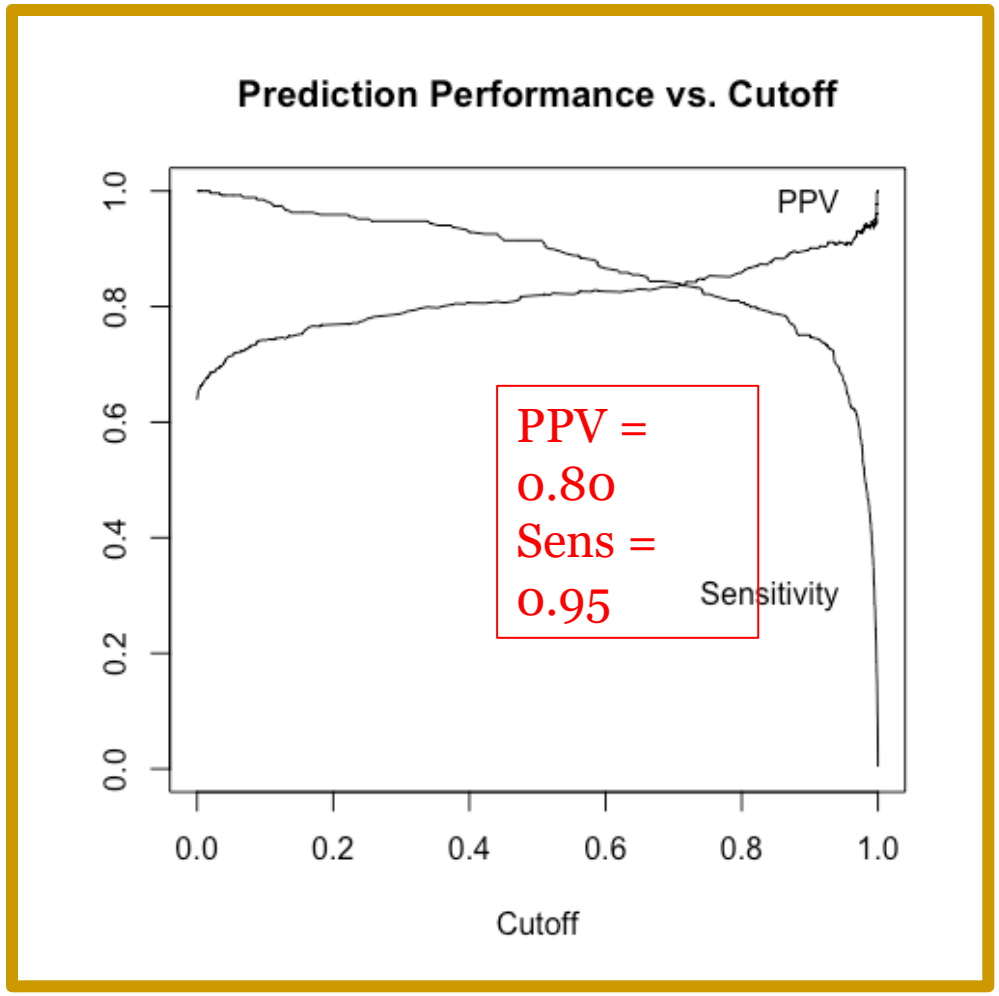
Mild+ phenotype, **Silver #3** – COVID-19 Mentions

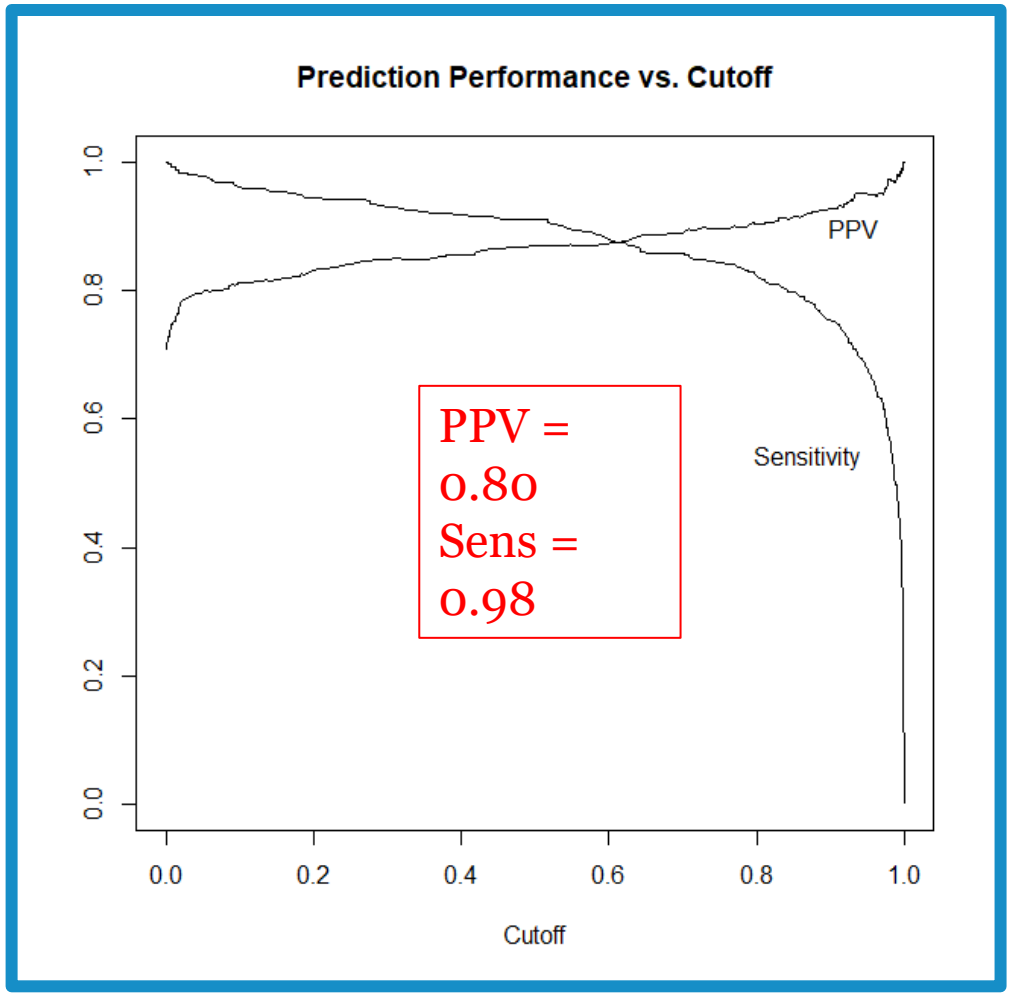




Prediction Performance

Mild+ phenotype, **Silver #4** – COVID Notes / COVID Days





Take-home messages

- **Relevance to Sentinel safety surveillance**
  - *Relatively modest effort* was needed to implement this approach
  - *Replication* in (two) heterogeneous settings was straightforward
  - May be relevant for both chronic and acute health conditions
- **Performance of automated models**
  - "Fit" between ***silver label*** and phenotype definition appears important
  - "Fit" between ***source data*** and *phenotype definition* appears important (e.g., inpatient data needed for moderate+ severity)
  - When performance is less than desirable, automated approaches may still be a useful **starting point** for model development
- **Hybrid approaches – automated and manually-curated features**
  - PheCap and Multimodal Automated Phenotyping (MAP)

More information

Data-driven automated classification algorithms for acute health conditions: Applying PheNorm to COVID-19 disease

- Abstract submitted for AMIA 2022 Annual Symposium

Joshua Smith, PhD (VUMC) joshua.smith@vumc.org

David Carrell, PhD (KPWA) david.s.carrell@kp.org

Sentinel

Large-scale Phenotyping With Natural Language Processing

Cosmin Adrian Bejan, PhD

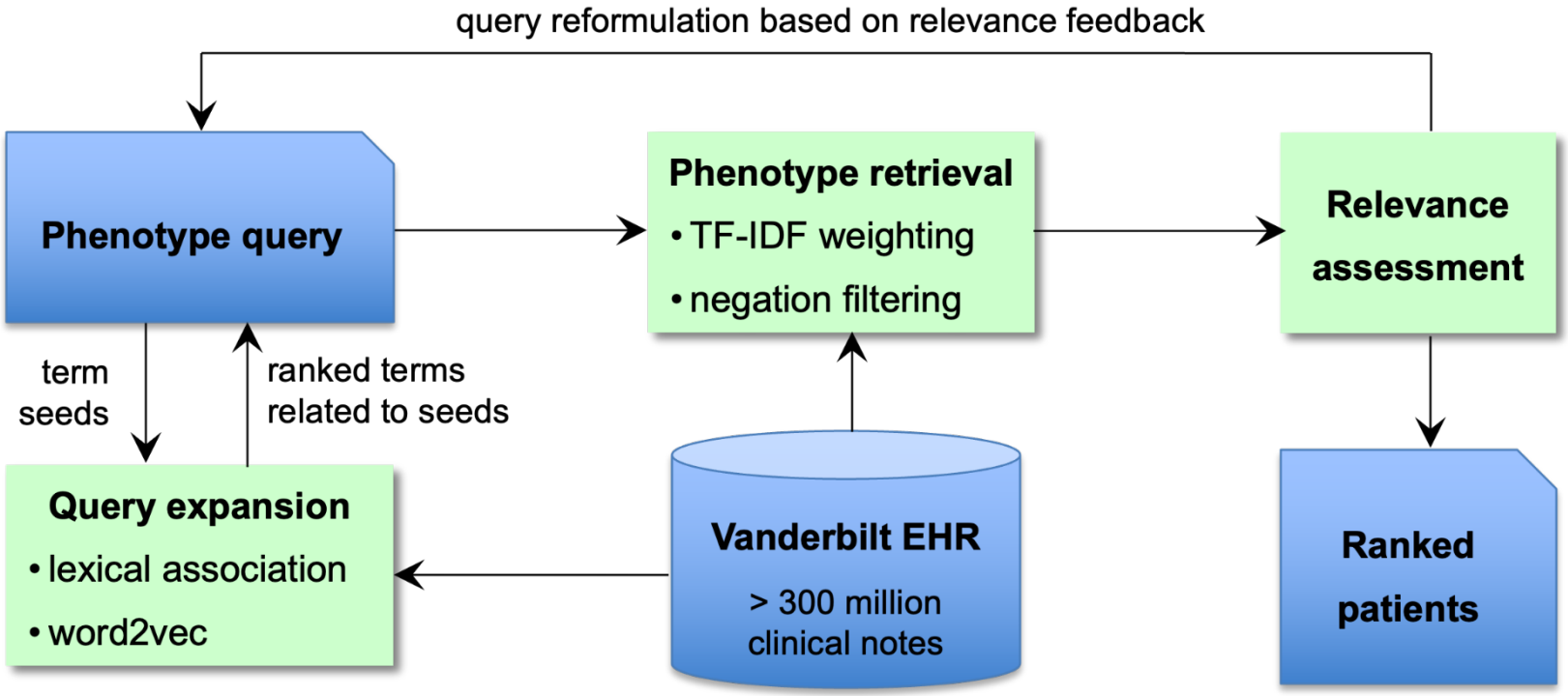
Department of Biomedical Informatics

VANDERBILT UNIVERSITY MEDICAL CENTER

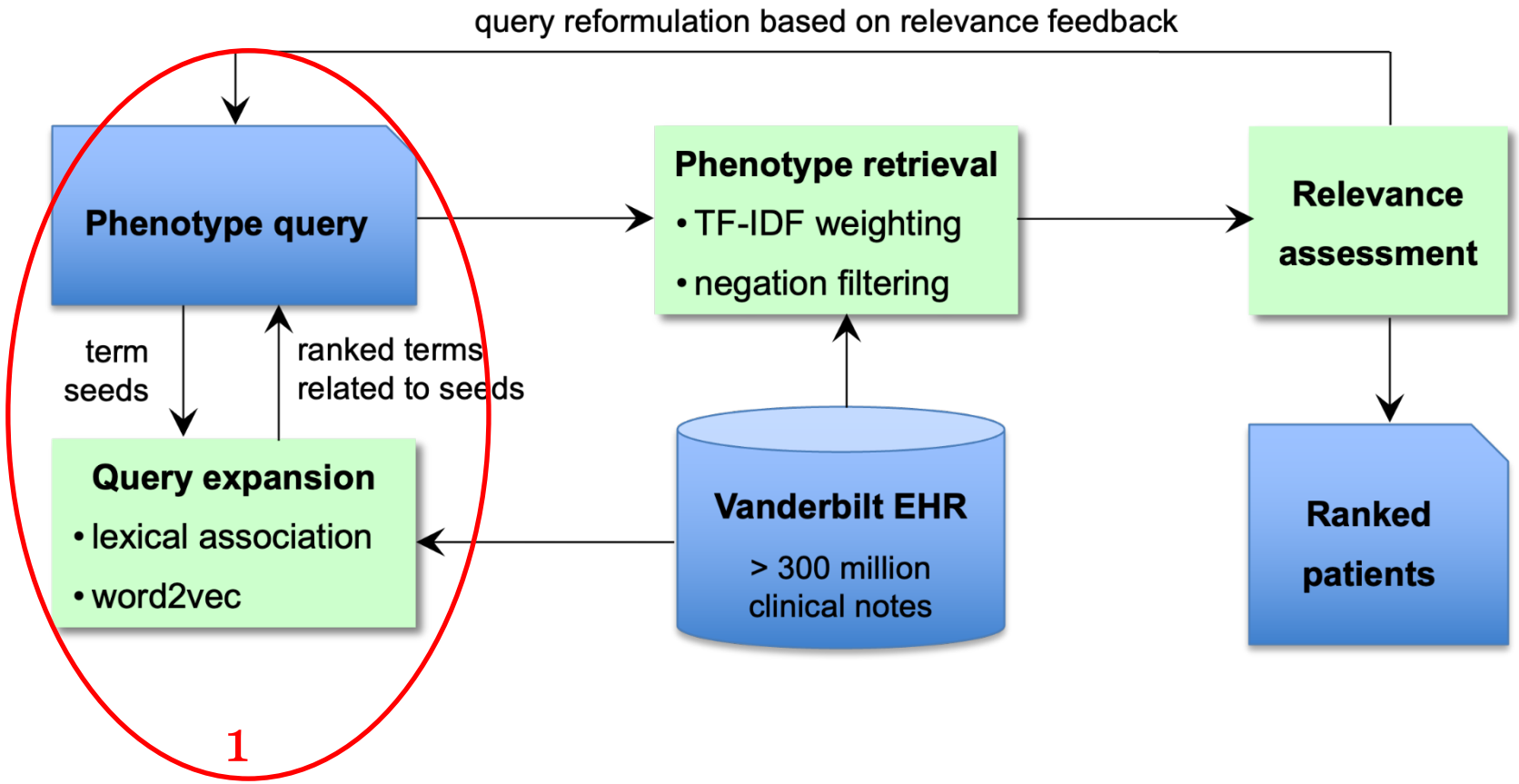
Desiderata for NLP-based phenotyping

- Improve phenotype identification based on structured data
- Analyze large volumes of clinical notes
- Data-driven generation of phenotype profiles
- Minimize the amount of chart review
- Generalize across phenotypes
- Replicate across EHR repositories

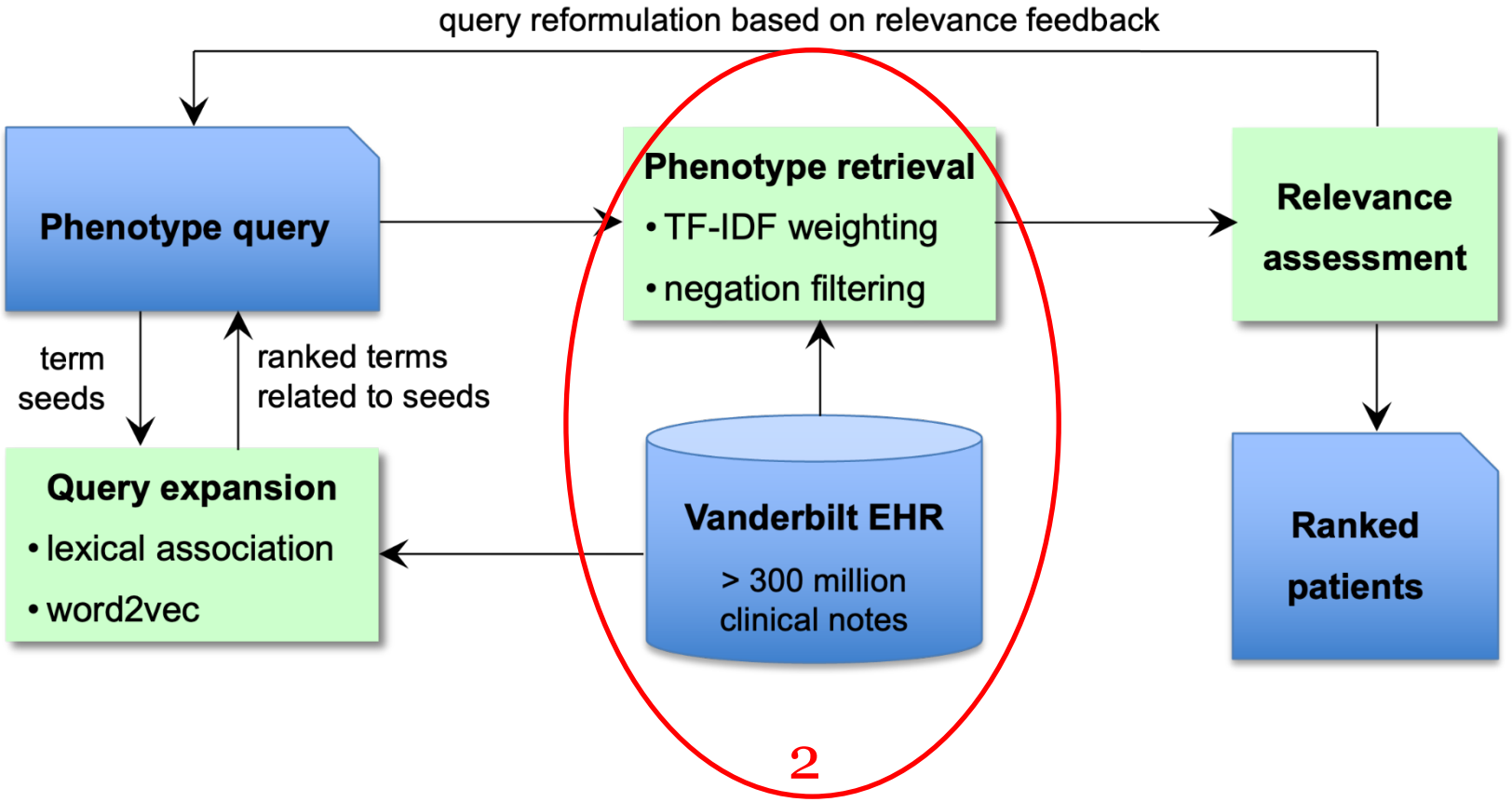
Proposed NLP system architecture



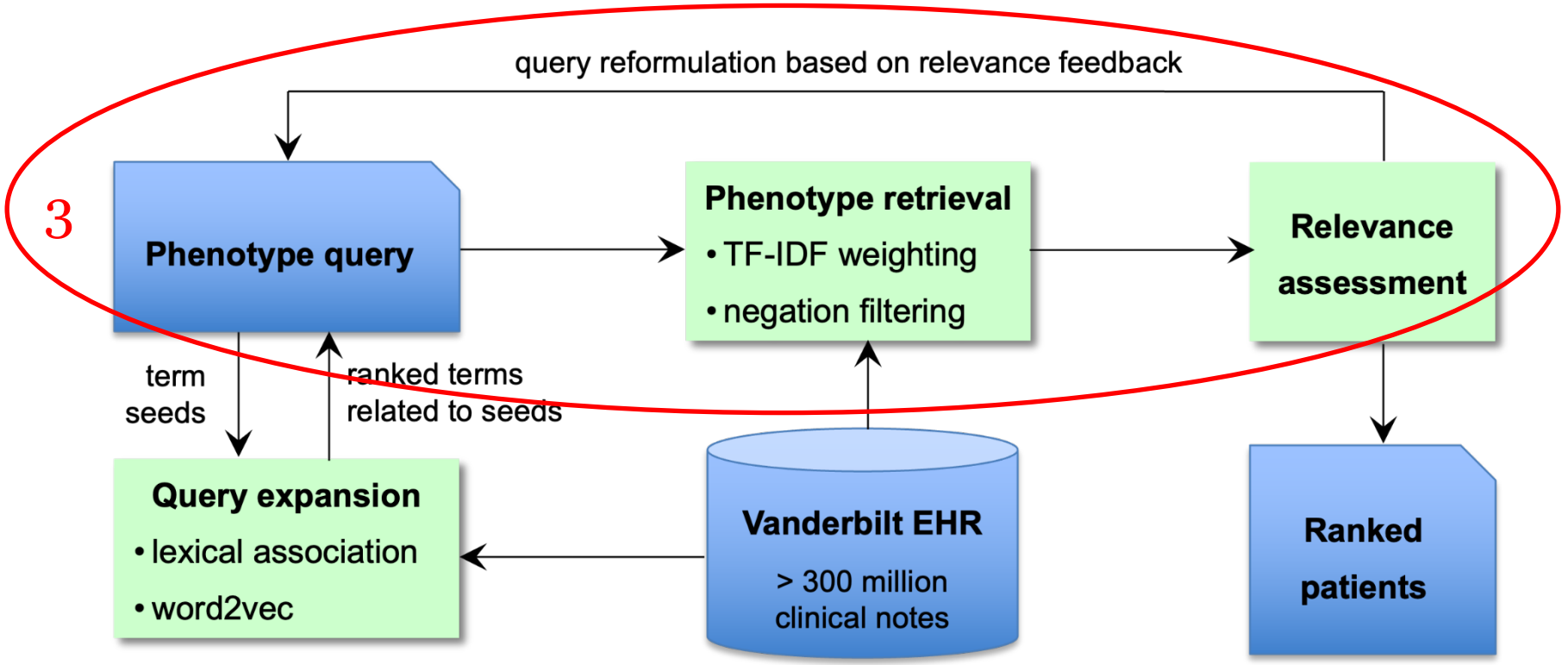
Proposed NLP system architecture



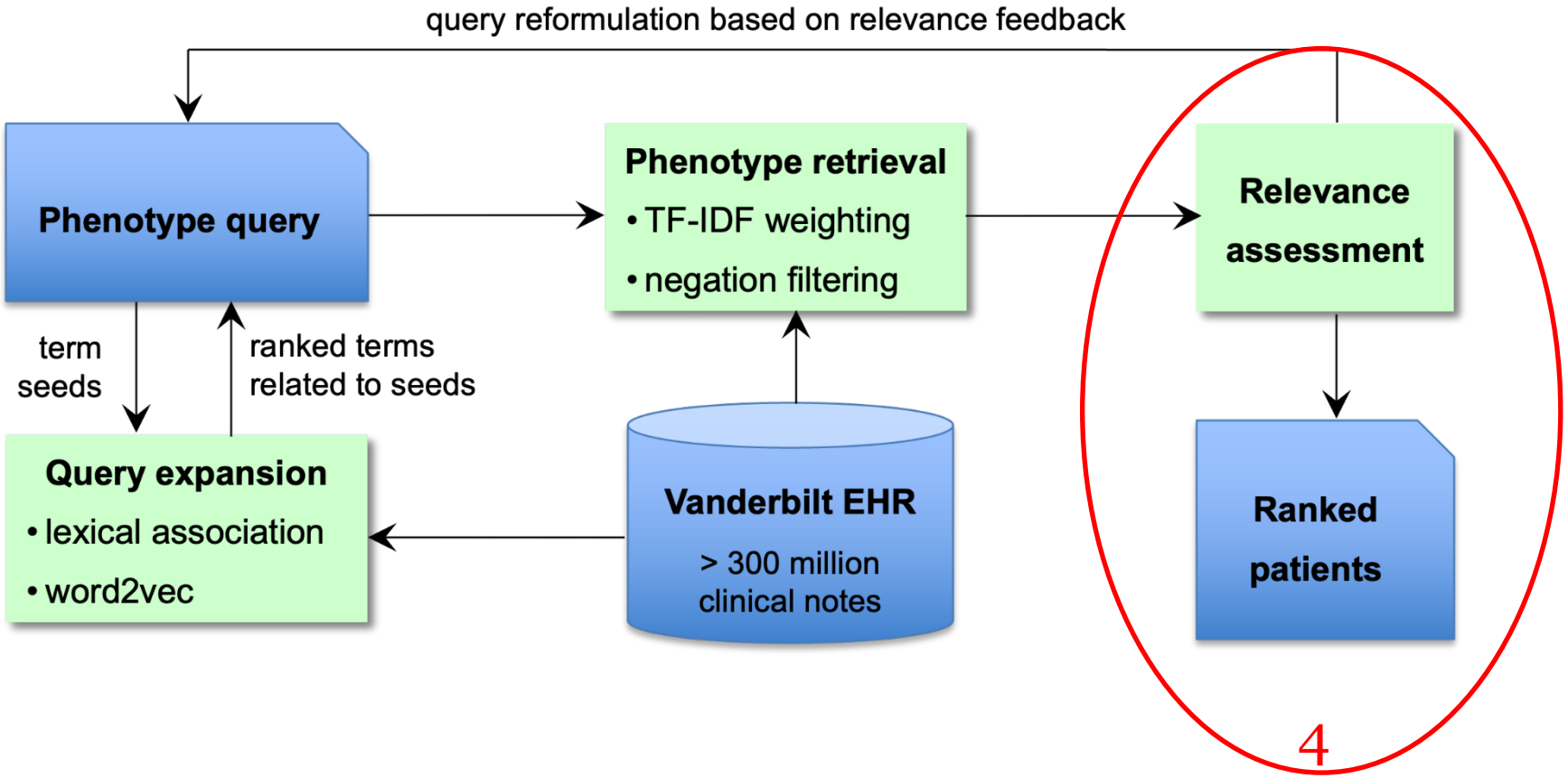
Proposed NLP system architecture



Proposed NLP system architecture



Proposed NLP system architecture



Applications

Social determinants of health

- Homelessness (VUMC)
- Adverse Childhood Experiences (VUMC)
- Homelessness (OHSU)
- Social Isolation (OHSU)
- Financial Insecurity (OHSU)
- Chronic Stress (OHSU)

(Bejan et al., *JAMIA* 2018)

(Dorr, Bejan et al., *MedInfo* 2019)

Suicide phenotypes

- Suicidal Ideation (VUMC)
- Suicide Attempt (VUMC)
- Suicide Attempt - incidence

(Bejan et al., *medRxiv* 2022)

(Walsh et al., *submitted*)

Data-driven methods for extracting phenotype profiles

Homelessness

	rank[cosine(homeless+homelessness, *w*)]		rank[cosine(homeless, *w*)]		rank[cosine(homelessness, *w*)]	
	context size=5	**context size=15**	**context size=5**	**context size=15**	**context size=5**	**context size=15**
1	homeless	homeless	prison	jail	polysubstance	homeless
2	homelessness	homelessness	jail	homelessness	homeless	polysubstance
3	prison	methamphetamine	ex-wife	prison	sober	methamphetamine
4	sober	polysubstance	girlfriend	girlfriend	abuser	sober
5	jail	jail	homelessness	sober	schizophrenia	schizo-affective
6	abuser	sober	abuser	methamphetamine	methamphetamine	schizophrenia
7	prostitution	prison	live-in	dui	abuse/dependence	prostitute
8	polysubstance	prostitute	sober	imprisoned	poly-substance	jail
9	ex-wife	ecstasy	ex-husband	ex-husband	prostitution	overdoses
10	ex-husband	mtmhi	fiancee	burglary	multi-substance	mtmhi

Suicide phenotypes

	suicide+ suicidal		suicide		suicidal	
	context size = 5	context size = 15	context size = 5	context size = 15	context size = 5	context size = 15
1	suicide	suicide	self-harm	manic	ideation	ideation
2	suicidal	suicidal	suicidal	ideation	homicidal	homicidal
3	ideation	ideation	paranoid	suicidal	ideations	ideations
4	homicidal	homicidal	homicide	self-harm	paranoia	suicidality
5	self-harm	ideations	ideation	suicided	suidical	paranoia
6	ideations	manic	suicide/homicide	mania	suicidality	suidical
7	paranoia	self-harm	self-mutilation	homicidal	self-harm	delusional
8	paranoid	mania	paranoia	s/h	delusional	self-harm
9	suidical	suicidality	self-harm--plan	self-mutilation	paranoid	thoughts
10	suicidality	paranoia	manic	ptsd	suicidial	mania

Building phenotype queries (I)

ACE

- child abuse
- sexual abuse
- child neglect
- childhood trauma
- child protective service
- physical abuse
- psychological abuse
- verbal abuse
- poverty
- food insecurity
- cps supervisor
- cps report
- cps worker
- cps investigation

Homelessness

- homeless
- homelessness
- shelter
- unemployed
- jobless
- incarceration

Building phenotype queries (II)

- Suicidal Ideation

suicid(al|e) idea(tion|s)*

suicid(al|e) thought(s)*

thought(s)* of suicide

(wish|wishes|intent|intend|intends|plans) to commit suicide

(want|wish) (s|ing|es)* to die

(thoughts|think|want|wish) (s|ing|es)* (of|to|about) (take|end) (ing)* (my|his|her|their) (own)* life

(thoughts|think|want|wish) (s|ing|es)* (of|to|about)
(kill|shot|shoot|hang|poison|asphyxiate|asphyxiat|mutilate|mutilat|harm|overdose|overdos|cut|cutt|gas|gass|slash)
(ing)* (myself|himself|herself|themself)

(thoughts|think|want|wish) (s|ing|es)* (of|to|about) (slit|slitt|cut|cutt|slash) (ing)* (my|his|her|their|the)*
(wrist|arm|throat)

feel(s|ing) (very)* suicidal

(thoughts|think|want|wish) (s|ing|es)* (of|to|about) (jump|jumping) off (a|the|interstate|my|his|her|their)*
(bridge|building|balcony|window|roof)

(thoughts|think|want|wish) (s|ing|es)* (of|to|about) (jump|jumping) out of (a|the)* moving (vehicle|car)

(thoughts|think|want|wish) (s|ing|es)* (of|to|about) (jump|jumping) from a moving (vehicle|car)

(thoughts|think|want|wish) (s|ing|es)* (of|to|about) (jump|jumping) out of (his|her|the|a)* (\d+) (nd|rd|th)
(floor|story|balcony|window)

(thoughts|think|want|wish) (s|ing|es)* (of|to|about) (jump|jumping) in front of a (car|truck|train|vehicle)

(thoughts|think|want|wish) (s|ing|es)* (of|to|about) (jump|jumping) into interstate

(thoughts|think|want|wish) (s|ing|es)* (of|to|about) (jump|jumping) out of (a|the|his|her)* (window|balcony)

Building phenotype queries (III)

- Suicide Attempt

  suicid(al|e) attempt

  suicid(al|e) ideation and attempt

  (attempted|committed) suicide

  (try|tried|tries|trying|attempted|attempts|attempting) (of|to) (take|end) (ing)* (my|his|her|their) (own)* life

  (try|tried|tries|trying|attempted|attempts|attempting) (of|to) (kill|shot|shoot|hang|poison|asphyxiate|asphyxiat|mutilate|mutilat|harm|overdose|overdos|cut|cutt|gas|gass|slash) (ing)* (myself|himself|herself|themself)

  (try|tried|tries|trying|attempted|attempts|attempting) (of|to) (slit|slitt|cut|cutt|slash) (ing)* (my|his|her|their|the)* (wrist|arm|throat)

  (try|tried|tries|trying|attempted|attempts|attempting) (of|to) (jump|jumping) off (a|the|interstate|my|his|her|their)* (bridge|building|balcony|window|roof)

  (try|tried|tries|trying|attempted|attempts|attempting) (of|to) (jump|jumping) out of (a|the)* moving (vehicle|car)

  (try|tried|tries|trying|attempted|attempts|attempting) (of|to) (jump|jumping) from a moving (vehicle|car)

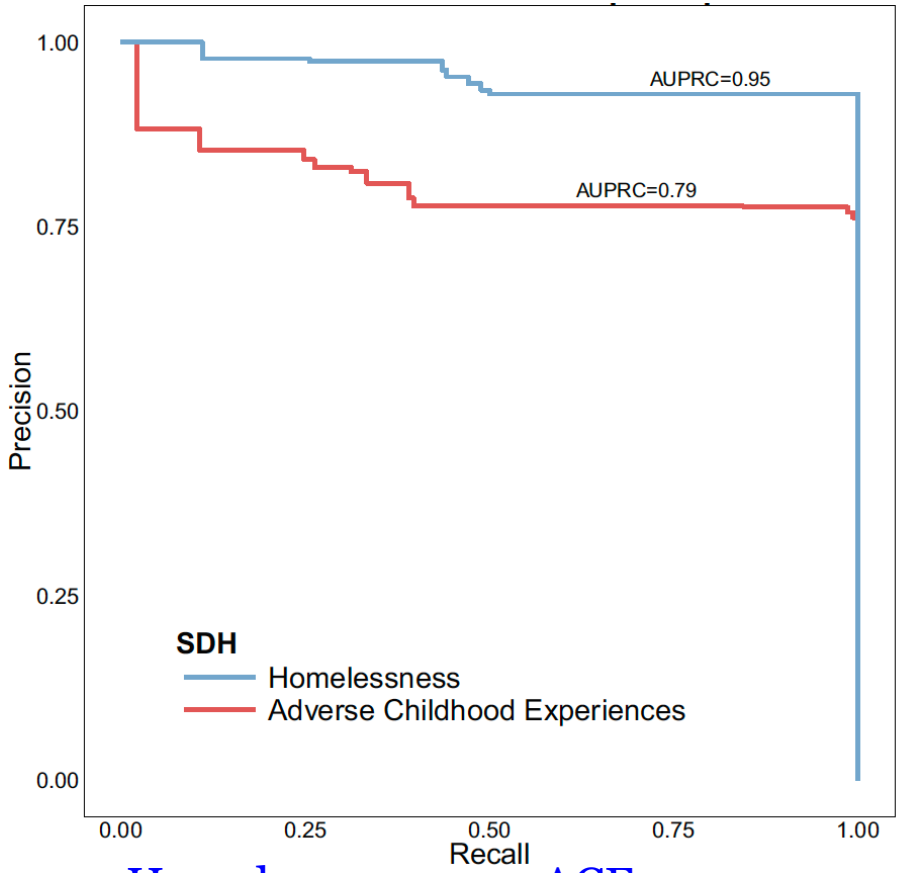
  (try|tried|tries|trying|attempted|attempts|attempting) (of|to) (jump|jumping) out of (his|her|the|a)* (\d+) (nd|rd|th) (floor|story|balcony|window)

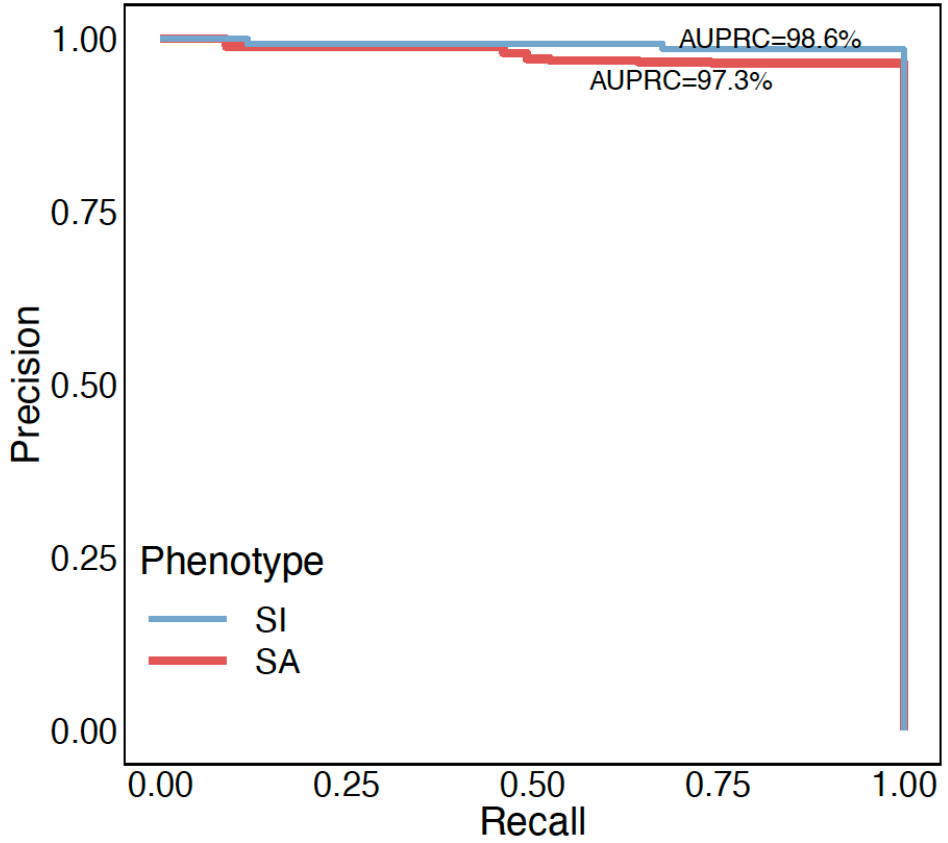
  (try|tried|tries|trying|attempted|attempts|attempting) (of|to) (jump|jumping) in front of a (car|truck|train|vehicle)

  (try|tried|tries|trying|attempted|attempts|attempting) (of|to) (jump|jumping) into interstate

  (try|tried|tries|trying|attempted|attempts|attempting) (of|to) (jump|jumping) out of (a|the|his|her)* (window|balcony)

Patient retrieval evaluation (Top K)





Homelessness

- P@185 = 93%
- N=35,220

ACE

- P@185 = 76%
- N=27,861

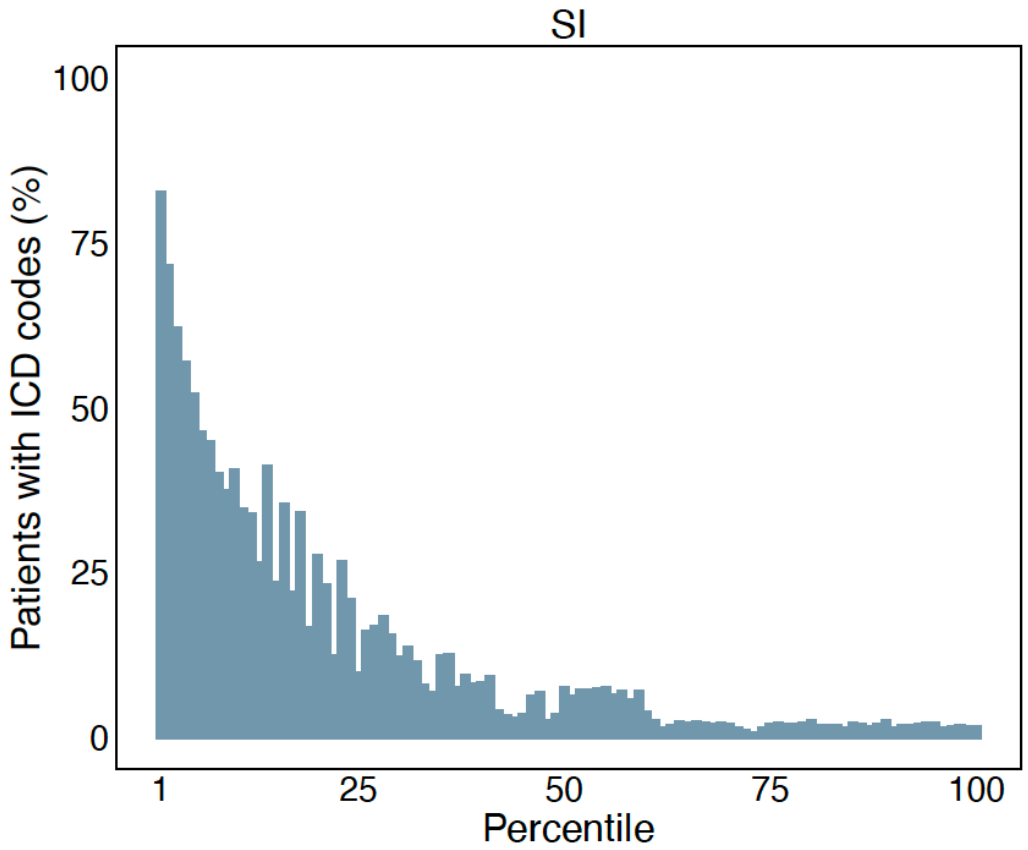
Suicidal Ideation

- P@200 = 98.5%
- N=187,047

Suicide Attempt

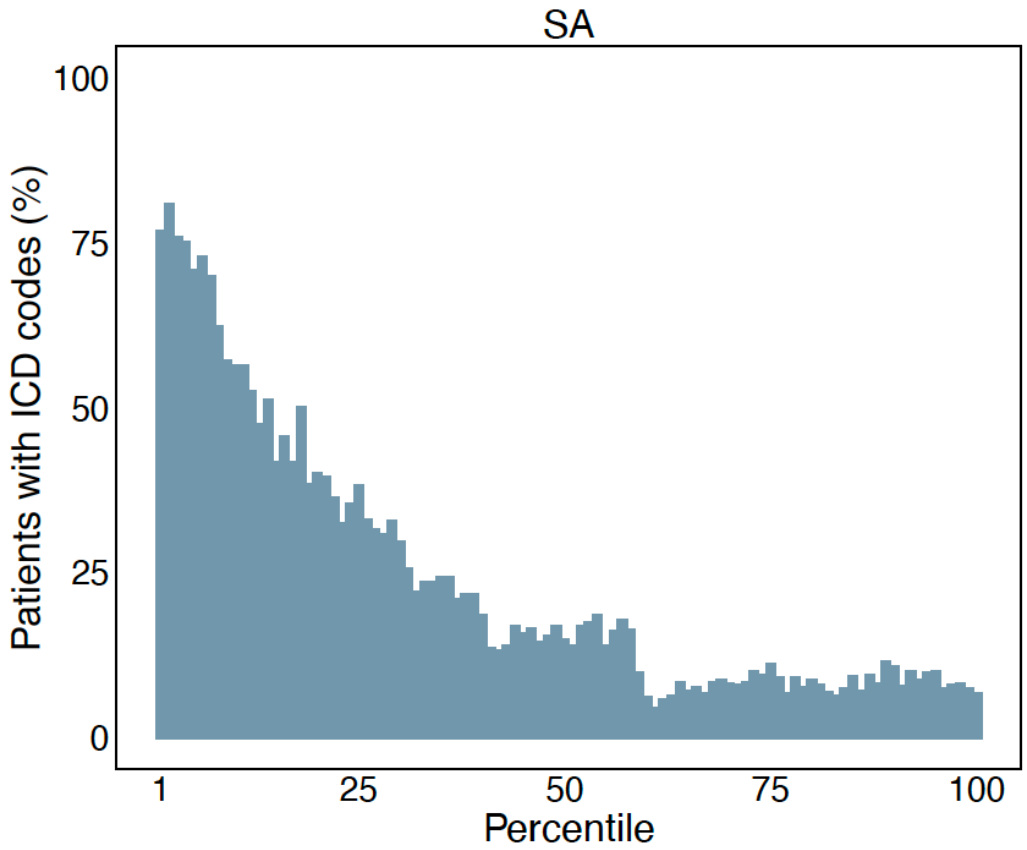
- P@200 = 96.5%
- N=52,738

ICD-based identification of suicide phenotypes



Suicidal Ideation

- P(ICD10CM) = 96%



Suicide Attempt

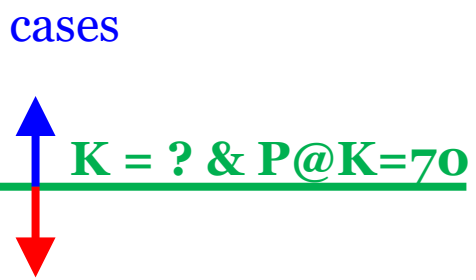
- P(ICD10CM) = 85%

From ranking to classification

Rank	Patient ID	NLP score	Case label	K	P@K
1	.	4,717	1	1	100
2	.		1	2	100
2	.		1	3	100
3	.		0	4	75
4	.		?	5	?
.	.		?	6	?
.	.		?	.	?
.	.		1	.	?
.	.		?	.	?
Nrank	.		?	N	?

From ranking to classification

Rank	Patient ID	NLP score	Case label	K	P@K
1	.	4,717	1	1	100
2	.		1	2	100
2	.		1	3	100
3	.		0	4	75
4	.		?	5	?
.	.		?	6	?
.	.		?	.	?
.	.		1	.	?
.	.		?	.	?
Nrank	.		?	N	?



From ranking to classification

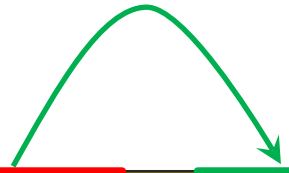
Rank	Patient ID	NLP score	Case label	K	P@K	P(NLP)
1	.	4,717	1	1	100	1
2	.		1	2	100	1
2	.		1	3	100	1
3	.		0	4	75	.99
4	.		?	5	?	.
.	.		?	6	?	.
.	.		?	.	?	.
.	.		1	.	?	.
.	.		?	.	?	.
Nrank	.		?	N	?	0

From ranking to classification

Rank	Patient ID	NLP score	Case label	K	P@K	P(NLP)
1	.	4,717	1	1	100	1
2	.		1	2	100	1
2	.		1	3	100	1
3	.		0	4	75	.99
4	.		?	5	?	.
.	.		?	6	?	.
.	.		?	.	?	.
.	.		1	.	?	.
.	.		?	.	?	.
Nrank	.		?	N	?	0

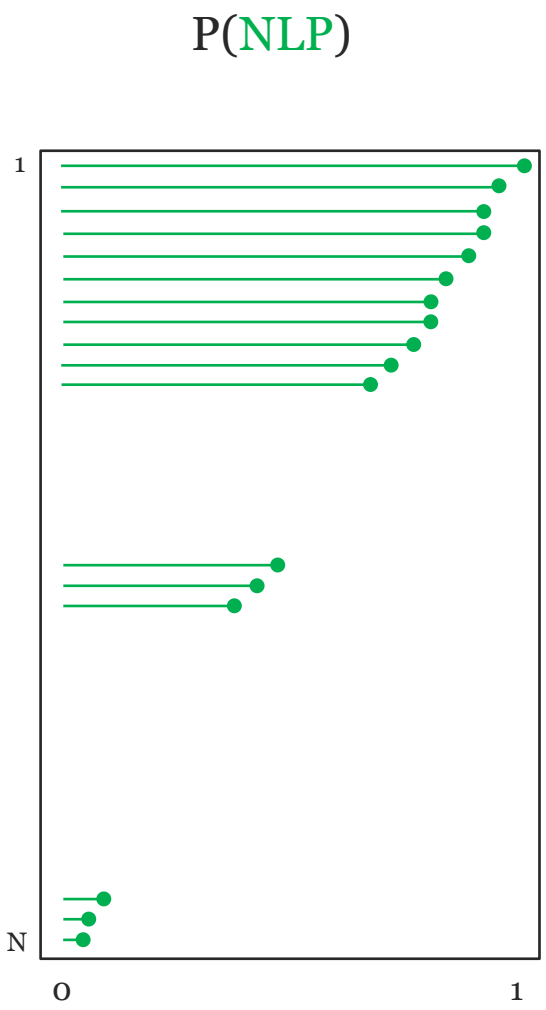
u ~ Uniform(0, 1)

From ranking to classification



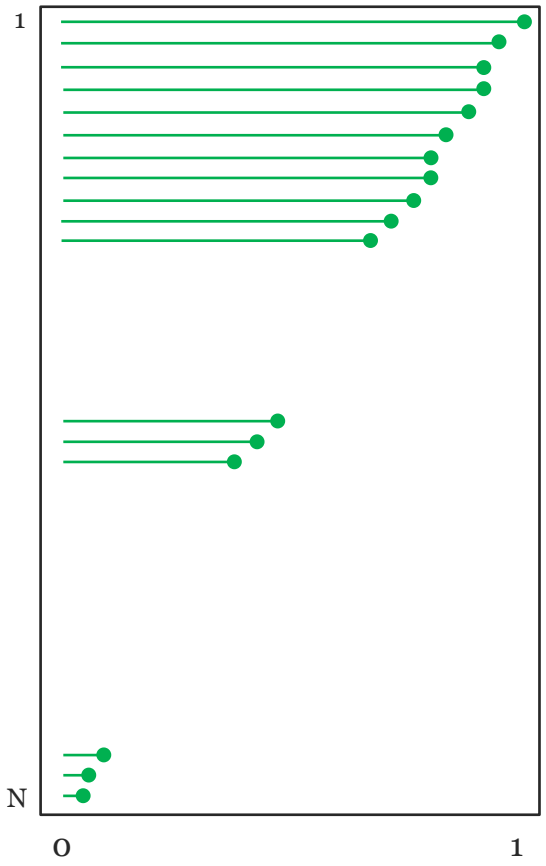
Rank	Patient ID	NLP score	Case label	K	P@K	P(NLP)
1	.	4,717	1	1	100	1
2	.		1	2	100	1
2	.		1	3	100	1
3	.		0	4	75	.99
4	.		?	5	?	.
.	.		?	6	?	.
.	.		?	.	?	.
.	.		1	.	?	.
.	.		?	.	?	.
Nrank	.		?	N	?	0

Probabilistic labeling of cases

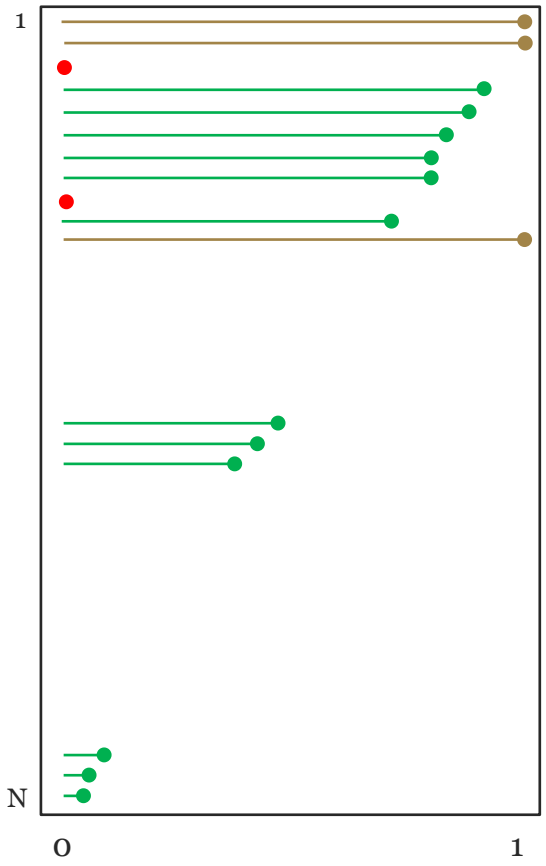


Probabilistic labeling of cases

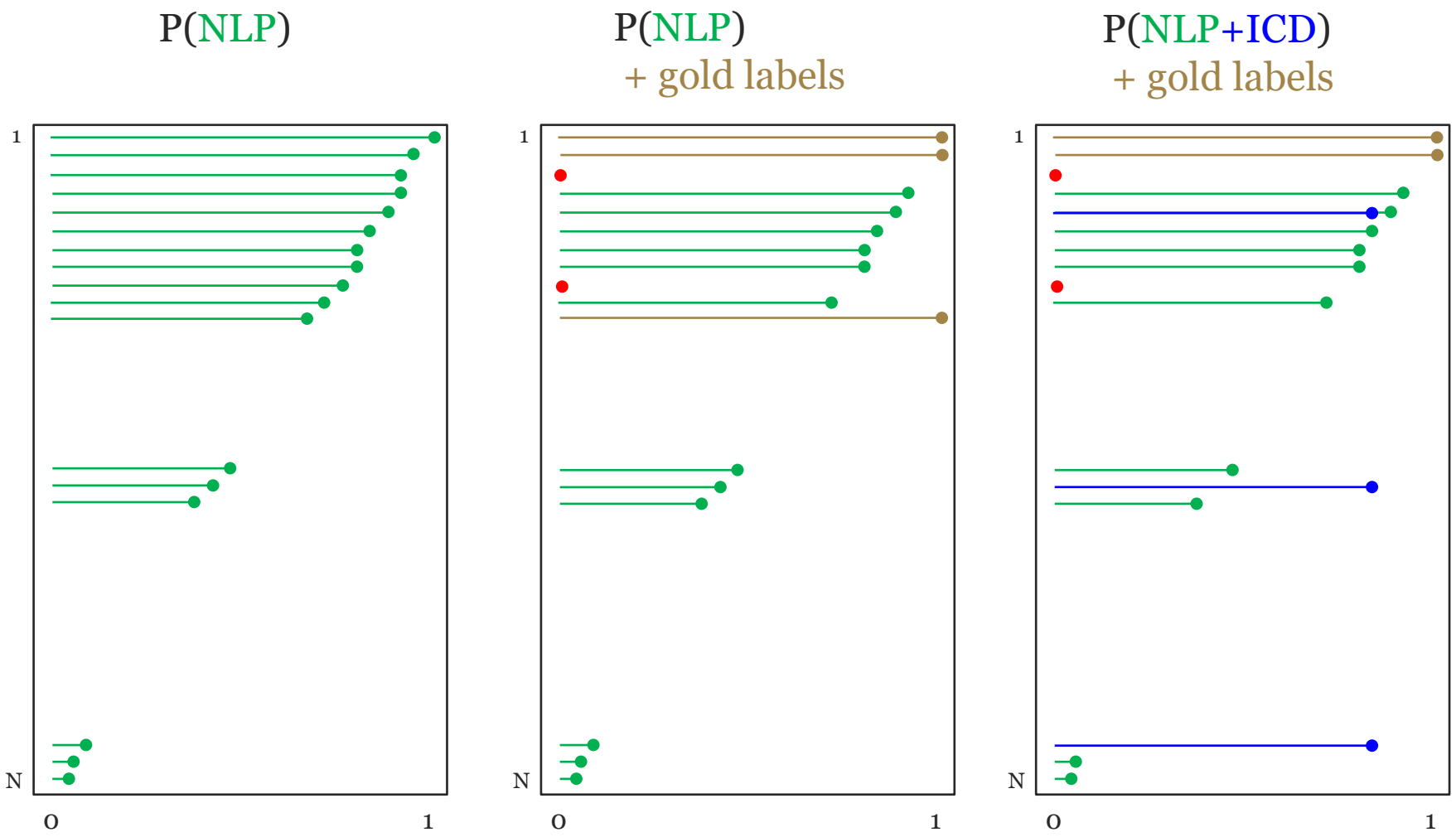
P(NLP)



P(NLP)
+ gold labels



Probabilistic labeling of cases

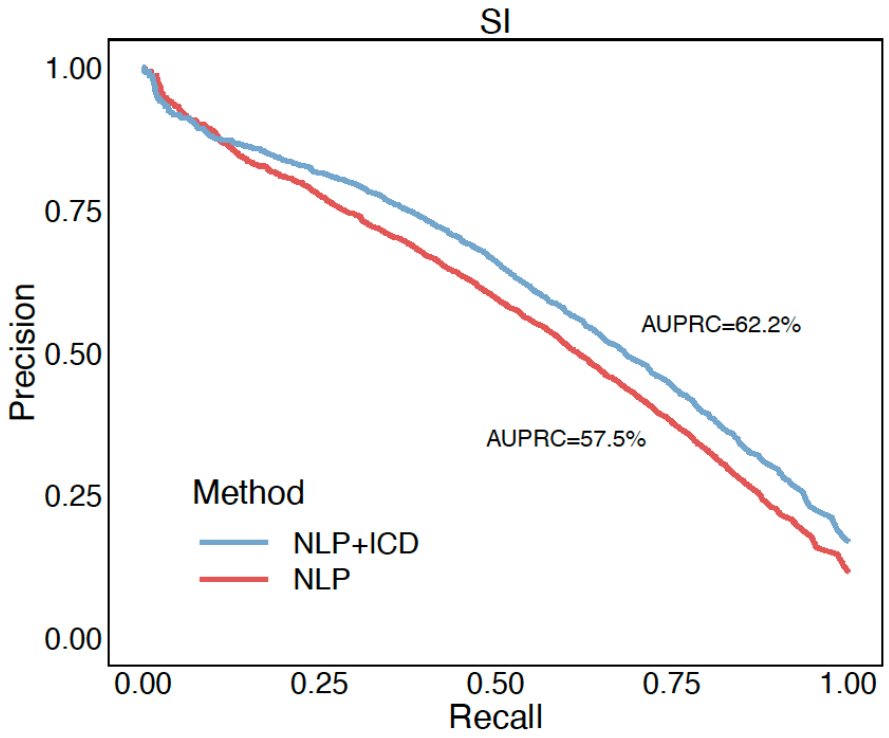


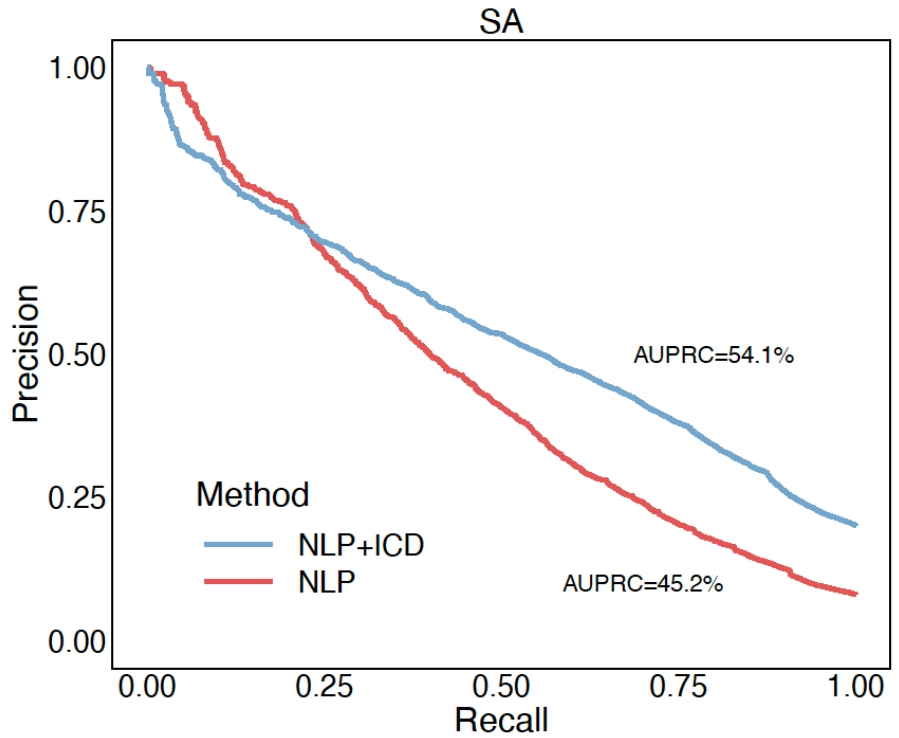
Classification of suicide phenotypes

AUPRC improvement based on negation detection:

- Suicidal ideation: 2.3% (NLP), 3.7% (NLP+ICD)
- Suicide attempt: 0.7% (NLP), 1.2% (NLP+ICD)

NLP vs. NLP+ICD



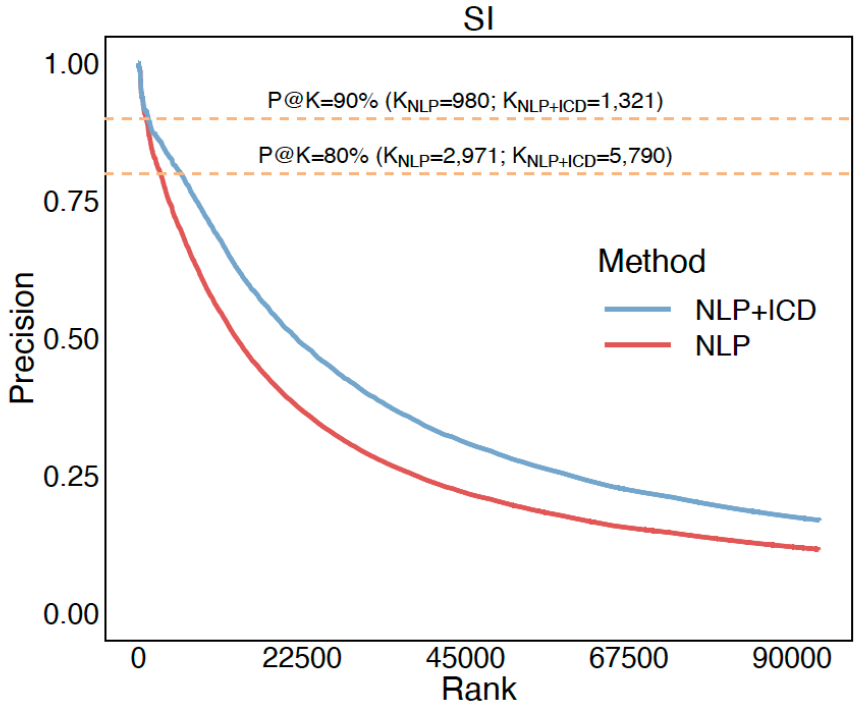


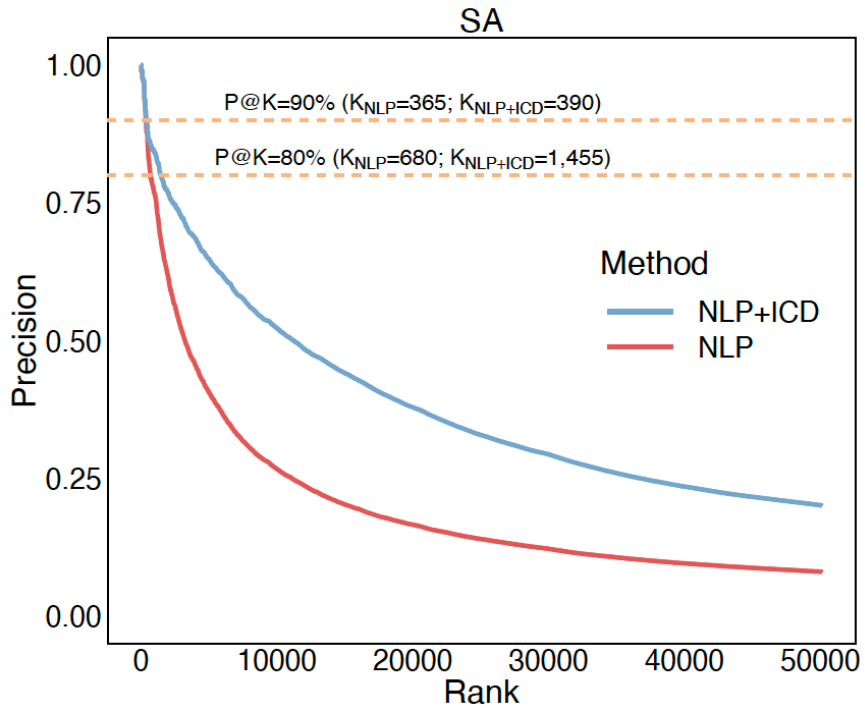
Classification of suicide phenotypes

AUPRC improvement based on negation detection:

- Suicidal ideation: 2.3% (NLP), 3.7% (NLP+ICD)
- Suicide attempt: 0.7% (NLP), 1.2% (NLP+ICD)

NLP vs. NLP+ICD





From prevalence to incidence

Phenotype: suicide attempt

Retrieval: “day of notes”

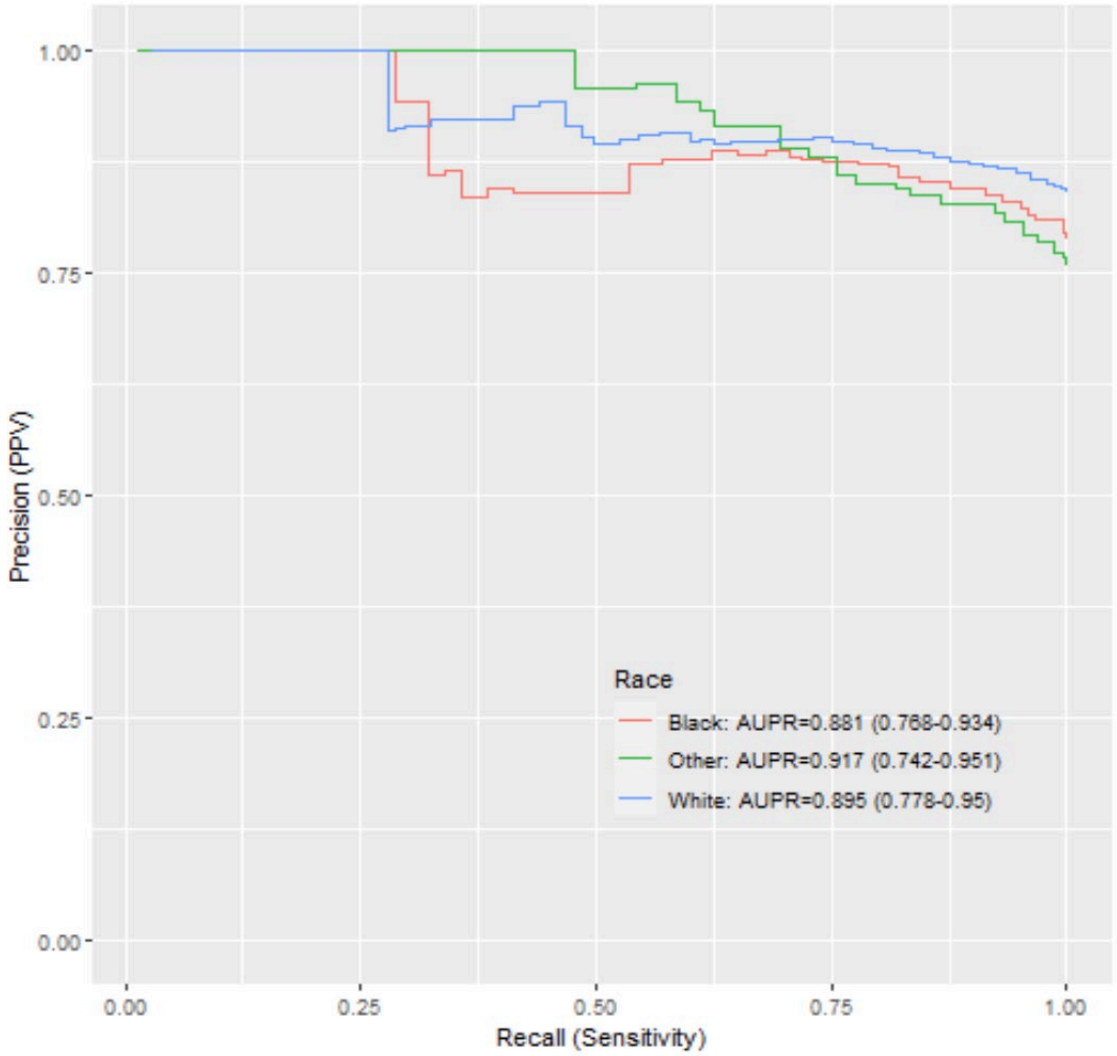
Output: <patient, day>

Weighted sampling of charts

Double chart review

Results:

- 263,403 <patient, day> retrieved
- 3,566 reviewed charts
- AUPRC range: 0.88-0.92
- Good inter-rater agreement (K=.89)



Conclusions

- Scalable NLP system for extracting low-prevalence (under-coded and under-reported) phenotypes from EHR
- Proved the generalizability of the method over multiple phenotypes
- Showed replication of results across two EHR repositories
- Data-driven generation of phenotype profiles leveraging unsupervised learning
- Extraction of phenotype cases with high precision
- Diagnostic coding and NLP yield optimal ascertainment
- Demonstrated the feasibility of the method for identifying incidents of suicide attempt

Acknowledgements

VUMC:

- Katelyn Robinson
- Michael Ripperger
- Drew Wilimitis
- Ryan Ahmed
- JooEun Kang
- Theodore J. Morley
- Jhansi Kolli
- John (Jack) Angiolillo
- Douglas Conway
- Robertson Nash
- Jana K. Shirey-Rice
- Loren Lipworth
- Gabriella Papa
- Robert M. Cronin
- Jill Pulley
- Sunil Kripalani
- William Stead
- Shari Barkin
- Kevin B. Johnson
- Joshua C. Denny

VUMC (cont.):

- Michael Matheny
- Aileen Wright
- Qingxia Chen
- Daniel Fabbri
- Douglas M. Ruderfer
- Colin G. Walsh

OHSU:

- David Dorr
- Ana Quinones
- Christie Pizzimenti
- Sumeet Singh
- Matt Storer

Harvard:

- Sruthi Adimadhyam
- Shamika More
- Adee Kennedy

FDA:

- Andy D. Mosholder
- Sai H. Dharmarajan
- Danijela Stojanovic

KPWA:

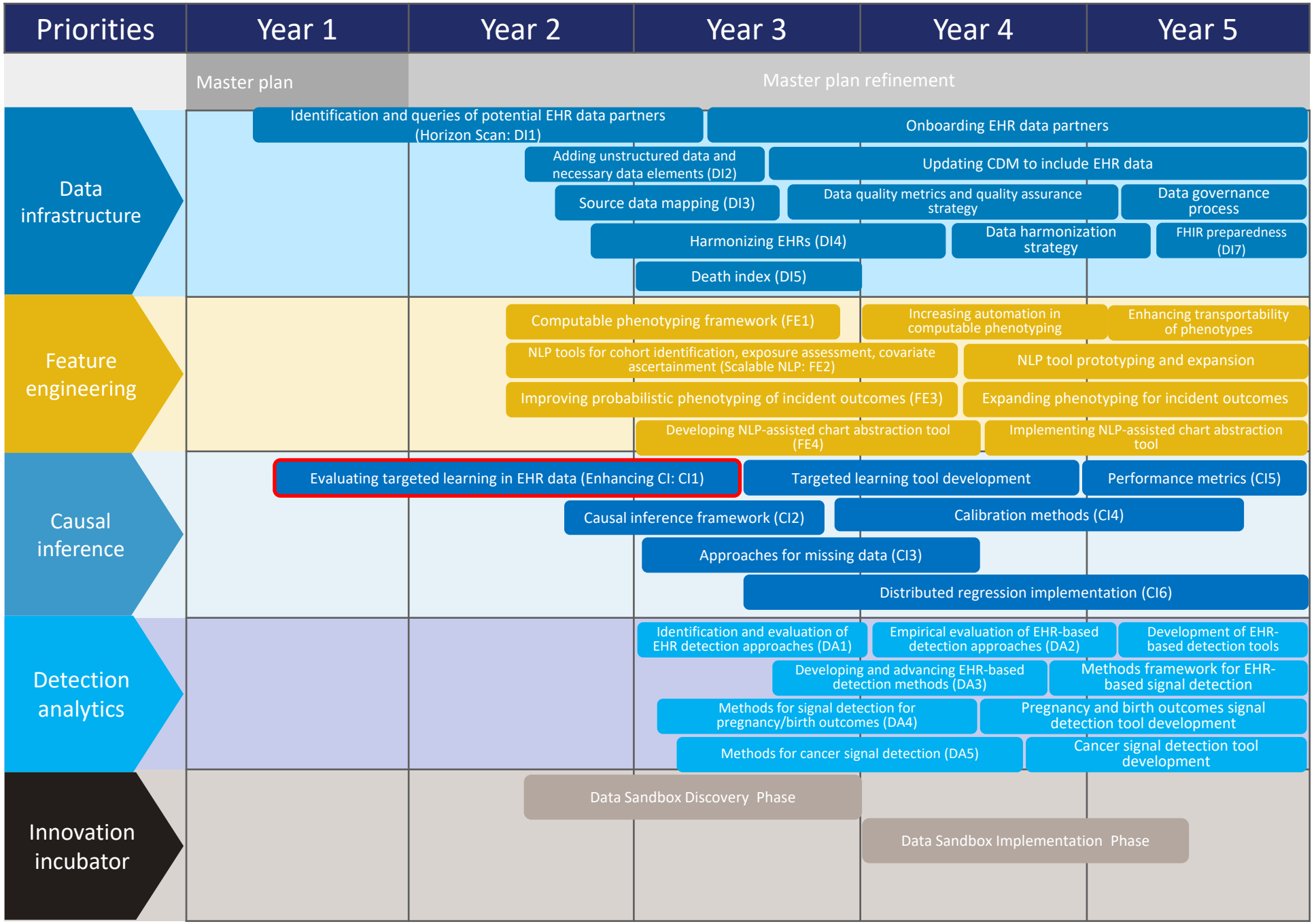
- David Carrell

Funding :

- U54 MD010722
- R01 LM010685
- R01 GM103859
- UL1 TR000445
- CDRN-1306-04869
- R01 MH121455
- R01 MH116269
- R01 MH118233
- FDA

Questions?

CI1: Enhancing Causal Inference in the Sentinel System



Sentinel

Enhancing Causal Inference in the Sentinel System

Leveraging unstructured electronic health records for large-scale confounding control in real-world evidence studies

Richard Wyss, PhD, MSc

Background

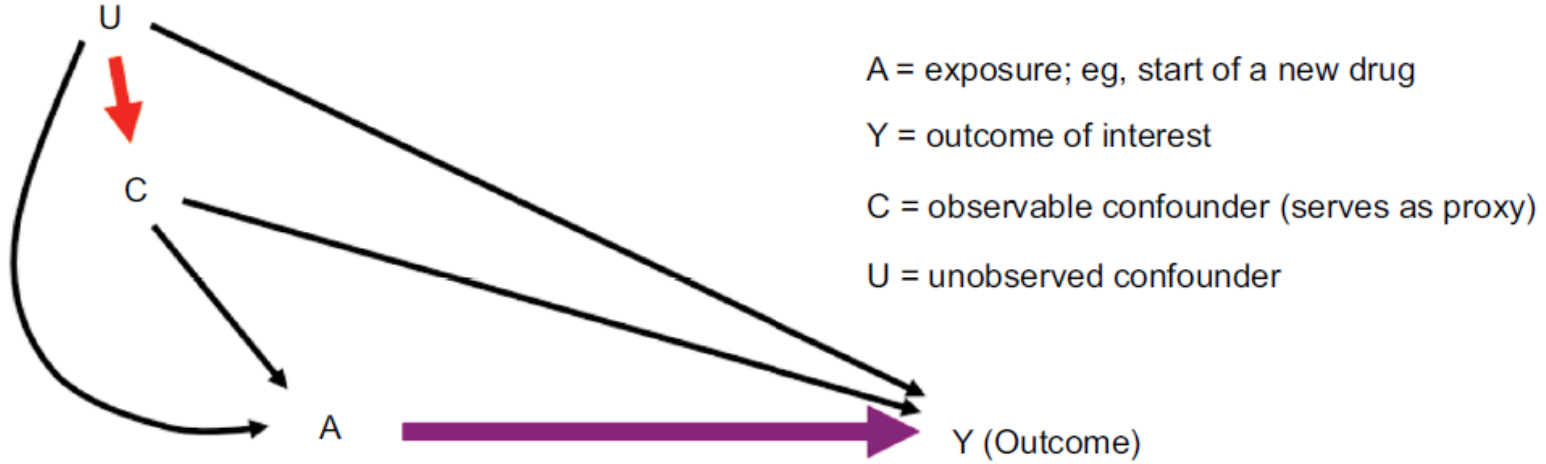
Background: Challenges for Confounding Control in RWE Studies

- Confounding arising from non-randomized treatment choices remains a fundamental challenge for extracting valid evidence to help guide treatment and regulatory decisions.

- Standard tools for confounding adjustment have typically relied on adjusting for a limited number of investigator specified variables.
  - Adjusting for investigator-specified variables alone is often inadequate
    - Some confounders are unknown at the time of drug approval
    - Many confounders are not directly measured in routine-care databases.

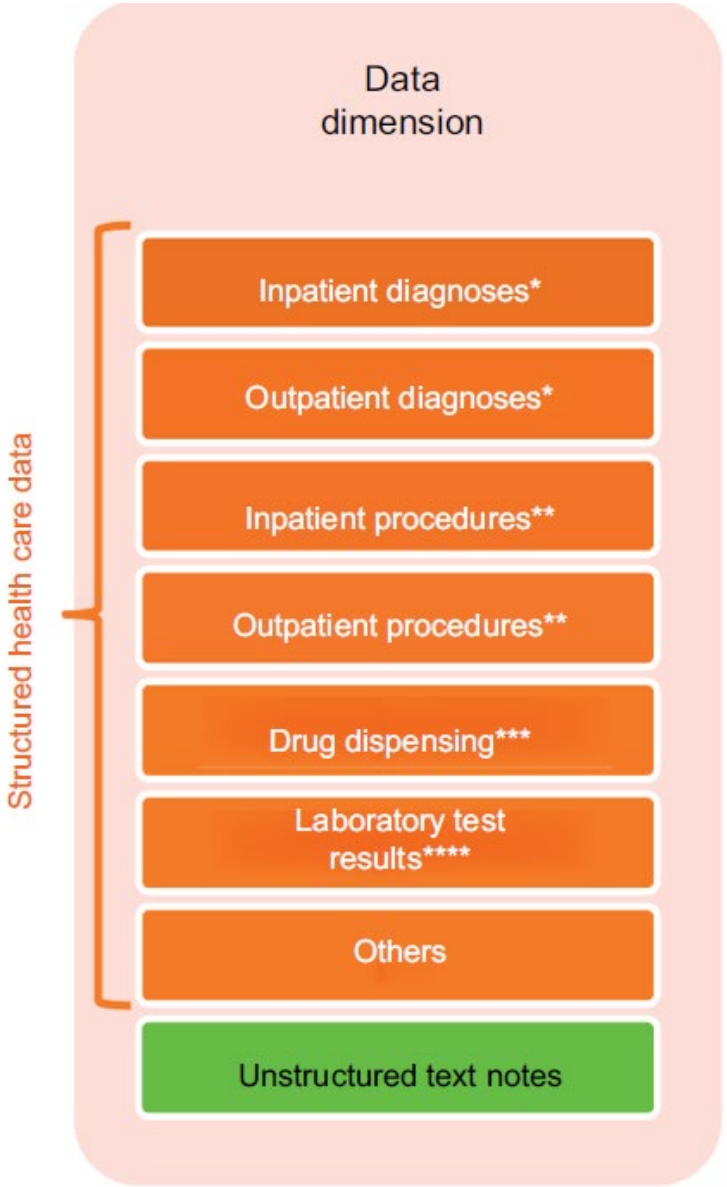
Background: Proxy Confounder Adjustment

- Healthcare databases may be understood and analyzed as a high-dimensional set of "proxy" factors that indirectly describe the health status of patients (Schneeweiss 2009, 2017).



Unobserved confounder	Observable proxy measurement	Coding examples
Very frail health	Use of oxygen canister	CPT-4
Sick but not critical	Code for hypertension during a hospital stay	ICD-9, ICD-10
Health-seeking behavior	Regular check-up visit; regular screening examinations	ICD-9, CPT-4, #PCP visits

Background: High-Dimensional Proxy Confounder Adjustment



- How to identify/generate proxy variables for adjustment?
  - High-dimensional propensity score (Schneeweiss 2009)
    - Does not require data pre-processing
  - OMOP approach:
    - Pre-process data into a common data model then use machine learning algorithms for variable selection (e.g., Lasso)
- **Current approaches for generating proxy variables for confounder adjustment do not leverage information from unstructured EHR text notes.**

Background: Leveraging Unstructured Electronic Health Records for Large-Scale Proxy Adjustment.

- NLP tools turn free-text notes from EHR data into structured features that can supplement confounding adjustment.
  - However, traditional applications are difficult to scale for large-scale proxy adjustment.

- **Project Objective 3 (use of NLP-generated information from unstructured data):** To explore if unsupervised NLP can be used to generate high-dimensional sets of features from free-text notes for improved large-scale proxy confounding control
  - **Aim 1:** To use scalable applications of NLP to generate structured features from high-dimensional data for large-scale proxy adjustment.
    - leverages work from RO1 (Josh Lin, PI; Richie Wyss, Co-PI; Sebastian Schneeweiss, Co-PI)
  - **Aim 2: To better understand what machine learning tools for confounder selection perform well for large-scale proxy adjustment in ultra high-dimensional RWE studies.**

Methods

Methods: Data Source for Generating Cohort Studies

- Mass General Brigham (MGB) Research Patient Data Registry (RPDR)
  - The electronic health records (EHR) of all the patients aged 65 and above identified in the Mass General Brigham (MGB) Research Patient Data Registry (RPDR) were linked to Medicare claims data

- Linked RPDR-Medicare claims were used to generate 3 cohort studies comparing different classes of medications (details on later slide).
  - Purpose: case studies for evaluating and testing various methods for NLP feature generation for ultra high-dimensional proxy confounder adjustment.

Methods: Using NLP to Generate Structured Features.

- We used 'bag-of-words' to generate features for the top 20,000 most prevalent terms from free-text notes.
  - Very common, simple, and flexible NLP approach
  - Measures the frequency (occurrence) of words within a document
    - Order and structure of words in the document is discarded.
    - The model is only concerned with whether words occur in the document, not where in the document or in relation to other words
- Each word count is then a feature that can be used for modeling

Methods: Study Cohorts

Table 1. Study Cohorts

		Total N			# Baseline Covariates		
No.	Description	Study Population	Treatment (%)	Outcome (%)	Investigator Specified	Claims Codes	EHR features
1.	High vs low intensity statin with an outcome of major cardiac events	3,529	1,244 (35.3)	138 (3.9)	39	18,409	20,017
2.	Oral anti-coagulants vs non-use with an outcome of stroke and major bleeding	9,571	5,991 (62.6)	158 (1.7)	39	19,517	20,051
3.	High vs. low dose PPI with an outcome of peptic ulcer complications	20,862	7,108 (34.1)	234 (1.1)	39	28,041	20,025

Methods: How to best identify confounder information in ultra high-dimensional real-world data?

- Predictive performance did not improve when modeling the outcome, but does this mean that there is no additional confounder information in EHR generated variables?

- Begin by considering various methods for confounder selection
  - Focus on lasso-based approaches
    - Regular Lasso
    - Outcome adaptive lasso
    - Collaborative controlled lasso
    - Outcome highly-adaptive lasso

Methods: How to make objective decisions on which modeling approach is best?

- **Cannot use actual study with estimated effects to make modeling decisions**
- Recent papers have proposed using synthetic control studies to help assess validity of alternative causal inference models and tailor analyses to the given study (Alaa & Van Der Scharr 2019; Schuler et al. 2017; Athey S et al. 2019; Bahamyirou A., et al. 2018; Schuemie MJ, et al. 2018; Petersen et al. 2012)
  - Provides an objective assessment of validity and model selection.
  - A common theme is that they use a variation of 'plasmode simulation' (Franklin et al. 2014).

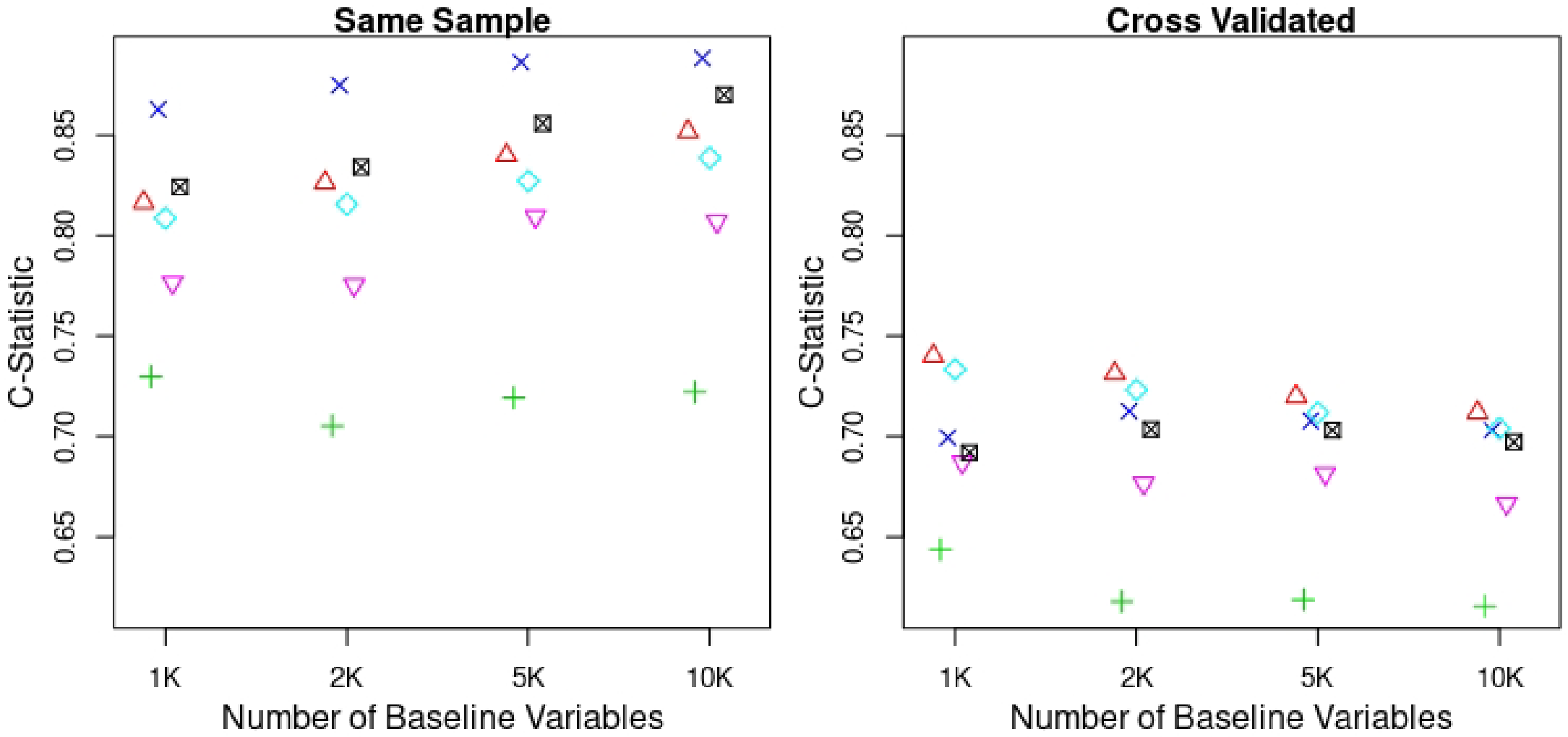
Variation of the parametric bootstrap where we bootstrap from the original study population, but simulate some aspects of the data structure while leaving other features of the data unchanged.
Typically, we set the outcome data aside (outcome blind data), then simulate the outcome while leaving baseline covariates and treatment status unchanged.
Try to generate synthetic control outcomes (and treatment) that mimic as closely as possible the observed confounding structure in the study cohort.
Will be inexact, but close approximations can be useful for testing robustness and validity of causal inference methods for the study at hand.

Confounder Selection & Propensity Score Models	
Lasso PS Models	Description
Standard Lasso	Lasso modeling treatment assignment with penalty factor (lambda) that optimizes CV treatment prediction
CTMLE Lasso w/ predictions	Collaborative controlled lasso—Lasso modeling treatment assignment but uses ctmle to choose penalty factor. We include initial predictions for the counterfactual outcomes using an outcome lasso model.
CTMLE Lasso w/ no predictions	Collaborative controlled lasso—Lasso modeling treatment assignment but uses ctmle to choose penalty factor. We did not include initial predictions for the counterfactual outcomes (only included treatment in the initial outcome model).
Outcome Adaptive Lasso (OAL)	adaptive lasso modeling treatment assignment with a penalty factor set by user. We assigned a penalty of 0 for all variables selected by the outcome lasso and a penalty of 1 for all other variables (i.e., we forced variables selected by outcome lasso into the lasso model for treatment).
CTMLE OAL w/ predictions	Collaborative controlled outcome adaptive lasso with initial predictions for the counterfactual outcomes
CTMLE OAL w/ no predictions	Collaborative controlled outcome adaptive lasso with no initial predictions for the counterfactual outcomes (initial outcome model includes only treatment)

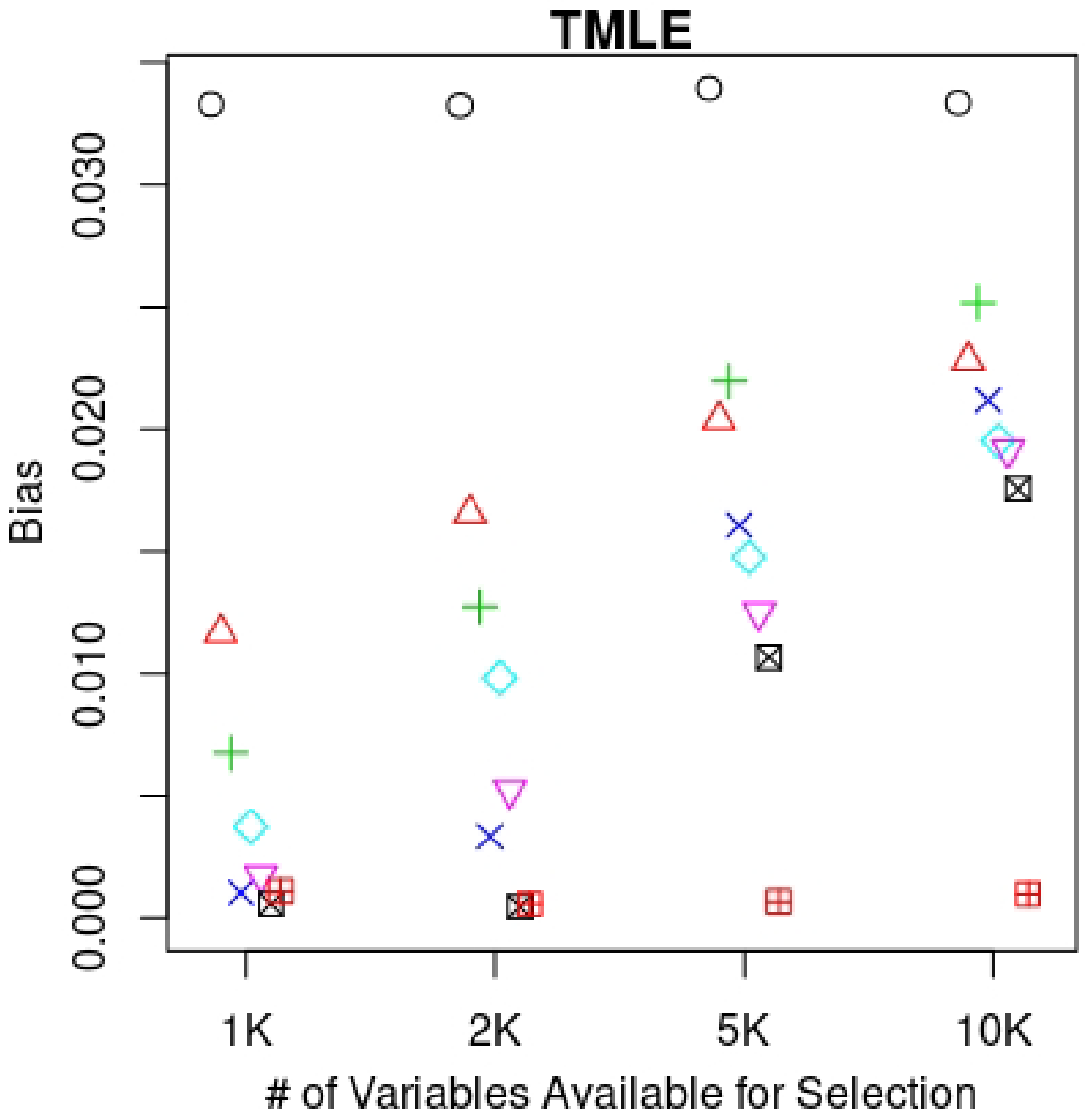
- **For each PS model, we estimated the treatment effect using Targeted Maximum Likelihood Estimation (TMLE) that included initial predictions from an outcome lasso model and PS weighting**

Simulation Results

Selected Simulation Results for Prediction


Same Sample
Cross Validated
C-Statistic
0.65
0.70
0.75
0.80
0.85
1K
2K
5K
10K
Number of Baseline Variables
Method 1: Uses PS selected by Lasso with optimizing CV prediction
Method 2: Uses PS selected by CTMLE Lasso with initial outcome predictions
Method 3: Uses PS selected by CTMLE Lasso with no initial outcome predictions
Method 4: Uses PS selected by Adaptive Lasso optimizing CV prediction
Method 5: Uses PS selected by CTMLE Adaptive Lasso with initial outcome predictions
Method 6: Uses PS selected by CTMLE Adaptive Lasso without initial outcome predictions

Selected Simulation Results for Bias



Lambda Selection for Lasso PS Model

- Unadjusted
- PS Model 1: Traditional Lasso
- PS Model 2: CTMLE Lasso with predictions
- PS Model 3: CTMLE Lasso no predictions
- PS Model 4: Outcome Adaptive Lasso (OAL)
- PS Model 5: CTMLE OAL with predictions
- PS Model 6: CTMLE OAL no predictions
- Oracle: includes all confounders

General points for discussion

- Selecting models based on collaborative learning improved bias reduction even though predictive performance declined.
  - Outcome adaptive lasso with collaborative selection generally performed best.
  - Some degree of overfitting is beneficial for confounding control when using Machine Learning to data-adaptively select (model) high-dimensional sets of variables

- Bias increased as the number of spurious variables available for selection increased.
- Bias can result from two sources
  1. Lasso model not selecting confounding variables
  2. Even when lasso selects confounders there can still be regularization bias (Chernozhukov 2018).

- Use relaxed lasso to reduce regularization bias in sparse high-dimensional data (Meinshausen 2007).

Relaxed lasso

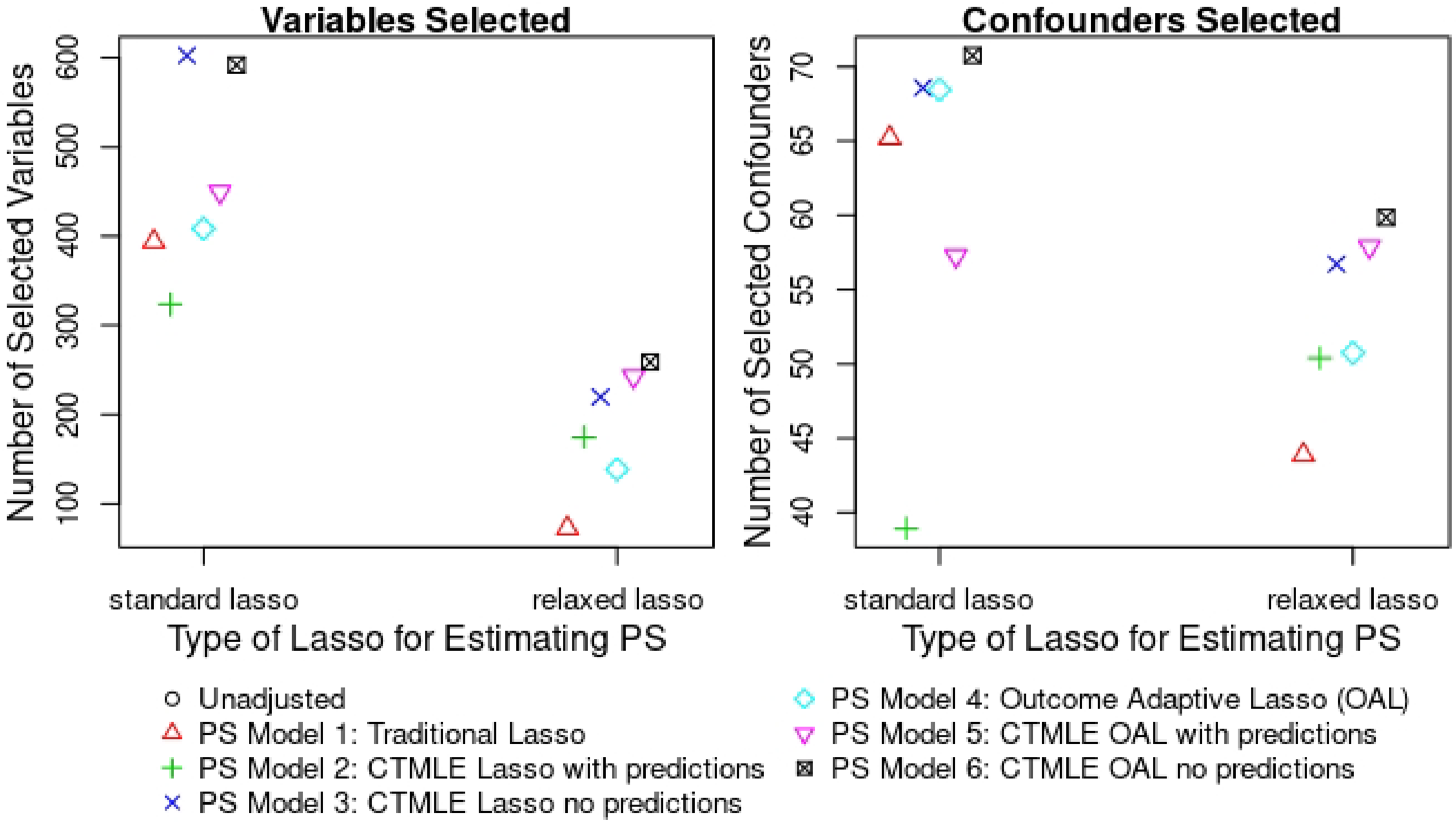
Use relaxed lasso to reduce regularization bias (Meinshausen 2007).

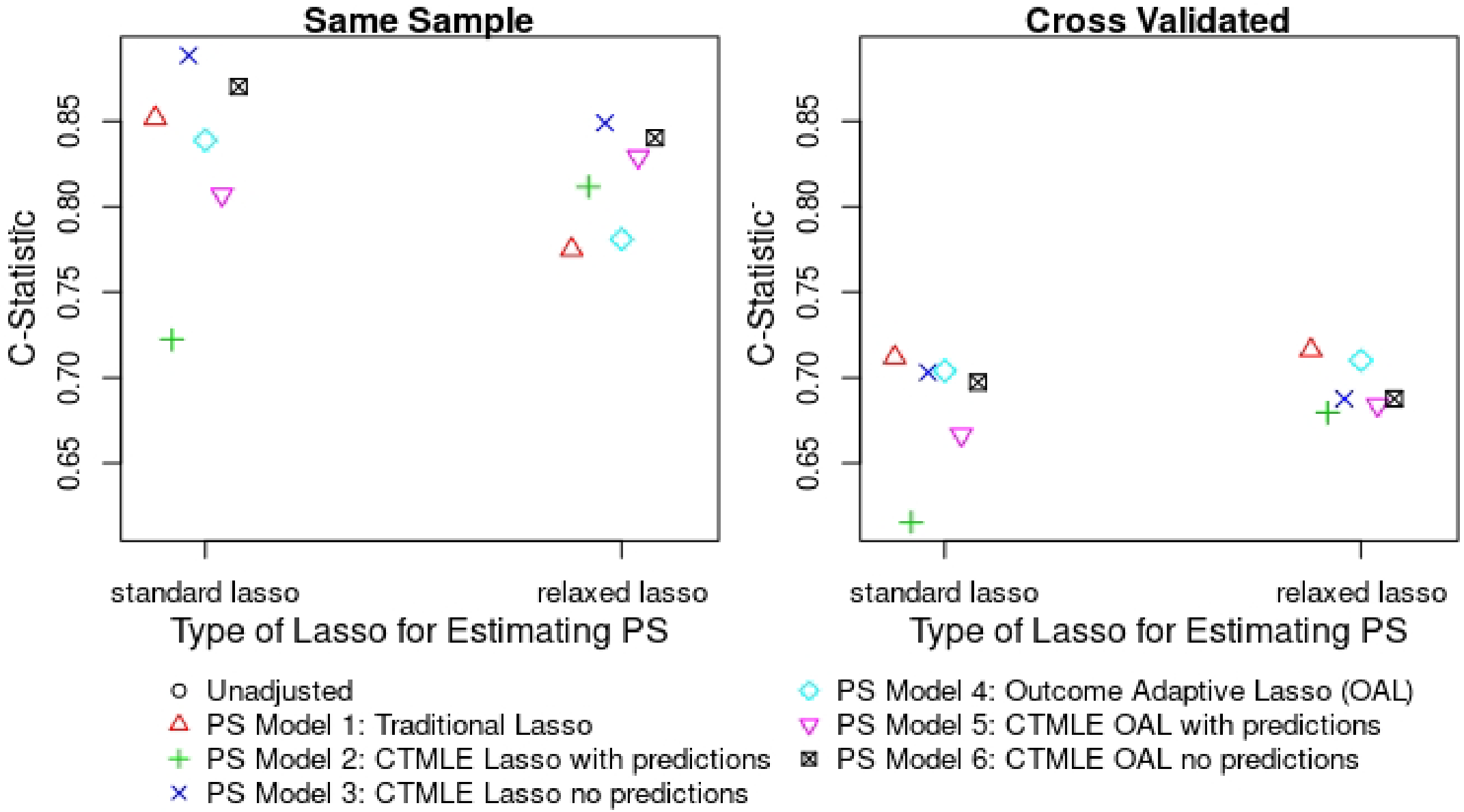
- Runs regularized regression twice:
  1. First runs lasso to select lambdas to control variable selection (which variables are selected for each lambda);
  2. Second step runs regularized regression again for each set of variables selected by each lambda with less penalization to control shrinkage level of coefficients. The shrinkage penalization in the second step can be selected using Cross Validation.

- *'Idea of the relaxed lasso is to take the lasso fitted object and then for each lambda, refit the variables in the active set with either no penalization or less penalization. This gives the "relaxed" fit'*. (Hastie & Tibshirani 2021)

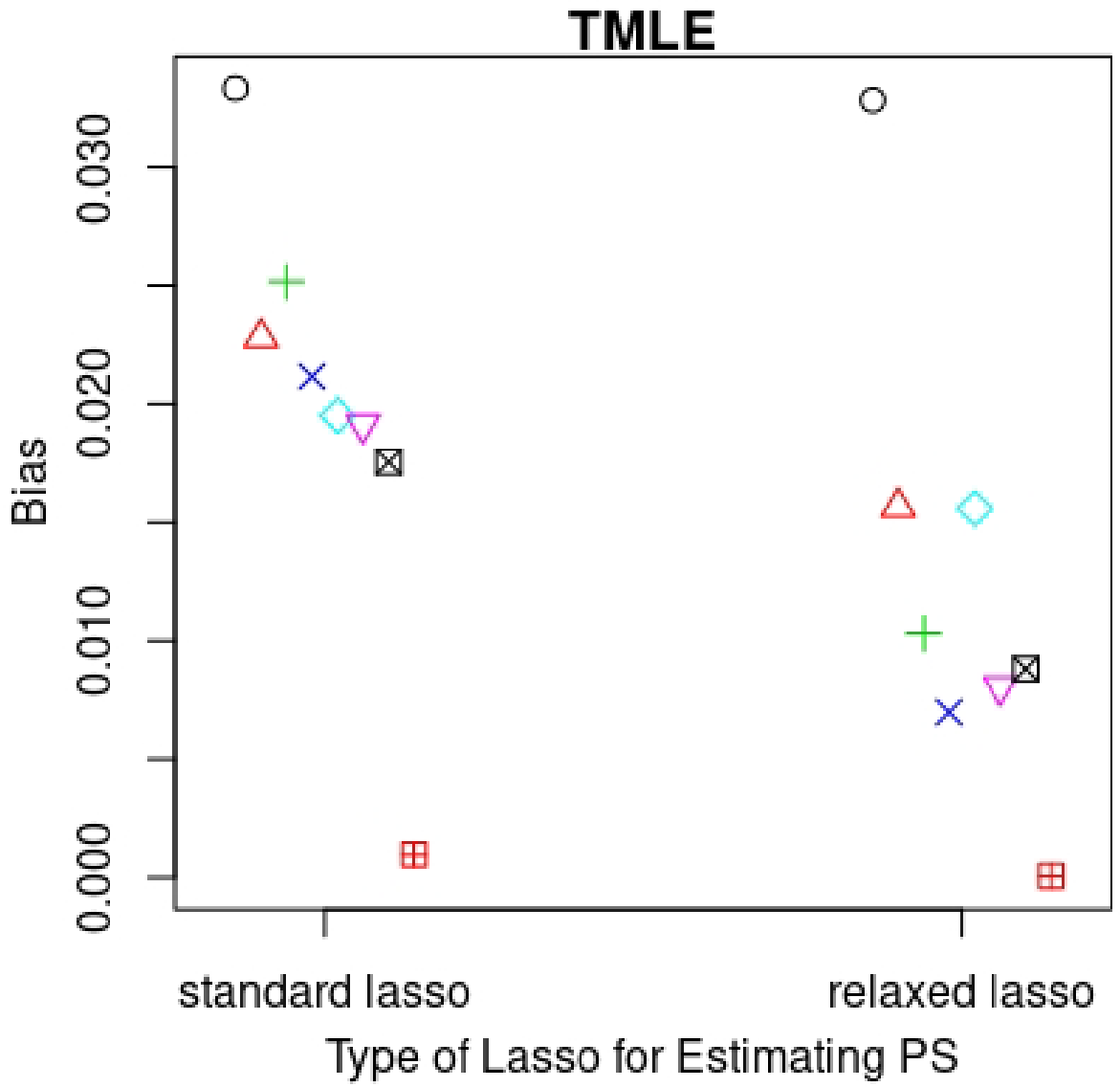
- Relaxed lasso can often improve predictive performance by fitting more parsimonious models with less penalization in sparse high-dimensional data (Meinhausen 2007).

Selected Simulation Results for Variable Selection and Prediction with Relaxed Lasso


Variables Selected
Number of Selected Variables
100
200
300
400
500
600
standard lasso
relaxed lasso
Type of Lasso for Estimating PS
Confounders Selected
Number of Selected Confounders
40
45
50
55
60
65
70
standard lasso
relaxed lasso
Type of Lasso for Estimating PS
Unadjusted
PS Model 1: Traditional Lasso
PS Model 2: CTMLE Lasso with predictions
PS Model 3: CTMLE Lasso no predictions
PS Model 4: Outcome Adaptive Lasso (OAL)
PS Model 5: CTMLE OAL with predictions
PS Model 6: CTMLE OAL no predictions


Same Sample
Cross Validated
C-Statistic
0.65
0.70
0.75
0.80
0.85
standard lasso
relaxed lasso
Type of Lasso for Estimating PS
Unadjusted
PS Model 1: Traditional Lasso
PS Model 2: CTMLE Lasso with predictions
PS Model 3: CTMLE Lasso no predictions
PS Model 4: Outcome Adaptive Lasso (OAL)
PS Model 5: CTMLE OAL with predictions
PS Model 6: CTMLE OAL no predictions

Selected Simulation Results for Bias with Relaxed Lasso



- Unadjusted
- PS Model 1: Traditional Lasso
- PS Model 2: CTMLE Lasso with predictions
- PS Model 3: CTMLE Lasso no predictions
- PS Model 4: Outcome Adaptive Lasso (OAL)
- PS Model 5: CTMLE OAL with predictions
- PS Model 6: CTMLE OAL no predictions
- Oracle: includes all confounders

Discussion

General Points for Discussion after running 'relaxed' lasso

- Relaxed lasso reduced bias in effect estimate compared with standard lasso

- Selecting models based on collaborative learning still improved bias reduction at the expense of predictive performance.
  - Outcome adaptive lasso with collaborative selection generally performed best.
  - Some degree of overfitting is beneficial for confounding control when using Machine Learning to data-adaptively select (model) high-dimensional sets of variables

- Still some bias with large numbers of variables
  - May need large samples to use ML to identify confounders in sparse high-dimensional data.

Future work/next step is to apply top performing models from simulations to empirical studies

Research team

Food and Drug Administration
- Hana Lee, PhD, MS
- Sarah Dutcher, PhD, MS

Brigham and Women's Hospital
- Richard Wyss, PhD, MS
- Rishi Desai, PharmD, PhD
- Josh Lin, MD, PhD
- Shirley Wang, PhD, MS
- Yinzhu Jin, MS, MPH
- Shamika More, MS
- Luke Zabotka, BA

DPM/HPHCI/Operations Center
- Darren Toh, ScD

Kaiser Washington/University of Washington
- Jennifer Nelson, PhD

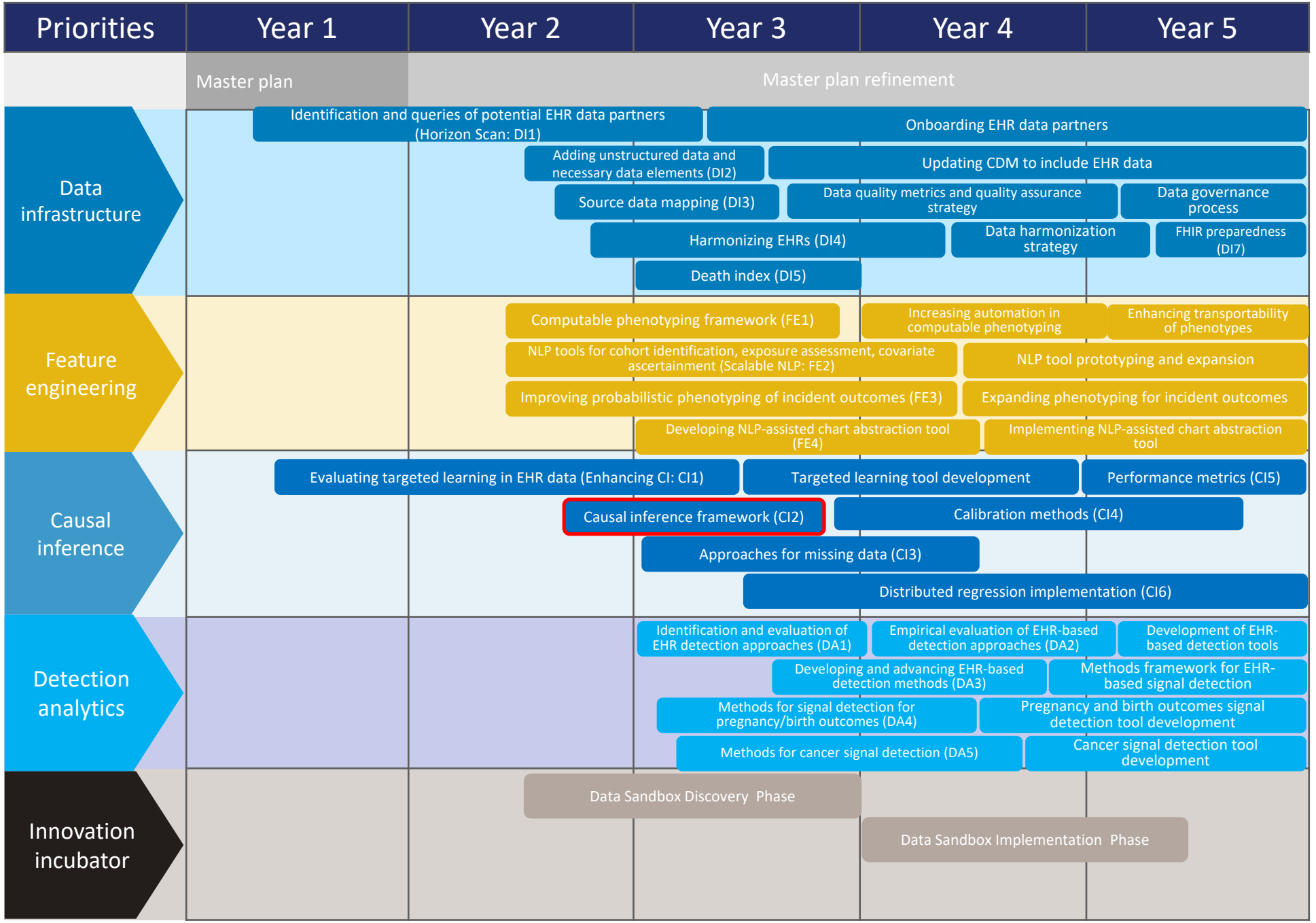
University of California, Berkeley
- Mark van der Laan, PhD
- Lars van der Laan

University of Michigan
- Xu Shi, PhD

Putnam Data Sciences
- Susan Gruber, PhD, MS, MPH

Questions?

CI2: A causal inference framework for Sentinel



A Causal Inference Framework for Sentinel

Rishi J Desai, PhD,
Assistant Professor, Division of Pharmacoepidemiology and Pharmacoeconomics,
Brigham and Women's Hospital, Harvard Medical School,
Boston, MA

Background and motivation

Why do we need another framework?

Quality assessment tools

RESEARCH METHODS AND REPORTING

OPEN ACCESS

ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions

Jonathan AC Sterne,[1] Miguel A Hernán,[2] Barnaby C Reeves,[3] Jelena Savović,[1,4] Nancy D Berkman,[5] Meera Viswanathan,[6] David Henry,[7] Douglas G Altman,[8] Mohammed T Ansari,[9] Isabelle Boutron,[10] James R Carpenter,[11] An-Wen Chan,[12] Rachel Churchill,[13] Jonathan J Deeks,[14] Asbjørn Hróbjartsson,[15] Jamie Kirkham,[16] Peter Jüni,[17] Yoon K Loke,[18] Theresa D Pigott,[19] Craig R Ramsay,[20] Deborah Regidor,[21] Hannah R Rothstein,[22] Lakhbir Sandhu,[23] Pasqualina L Santaguida,[24] Holger J Schünemann,[25] Beverly Shea,[26] Ian Shrier,[27] Peter Tugwell,[28] Lucy Turner,[29] Jeffrey C Valentine,[30] Hugh Waddington,[31] Elizabeth Waters,[32] George A Wells,[33] Penny F Whiting,[34] Julian PT Higgins[35]

RESEARCH

The GRACE Checklist for Rating the Quality of Observational Studies of Comparative Effectiveness: A Tale of Hope and Caution

Nancy A. Dreyer, PhD, MPH; Priscilla Velentgas, PhD; Kimberly Westrich, MA; and Robert Dubois, MD

Reporting tools

RESEARCH METHODS AND REPORTING

OPEN ACCESS

Check for updates

The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE)

Sinéad M Langan,[1] Sigrún AJ Schmidt,[2] Kevin Wing,[1] Vera Ehrenstein,[2] Stuart G Nicholls,[3,4] Kristian B Filion,[5,6] Olaf Klungel,[7] Irene Petersen,[2,8] Henrik T Sorensen,[2] William G Dixon,[9] Astrid Guttmann,[10,11] Katie Harron,[12] Lars G Hemkens,[13] David Moher,[3] Sebastian Schneeweiss,[14] Liam Smeeth,[1] Miriam Sturkenboom,[15] Erik von Elm,[16] Shirley V Wang,[14] Eric I Benchimol[10,17,18]

RESEARCH METHODS AND REPORTING

OPEN ACCESS

Check for updates

STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies

Shirley V Wang,[1] Simone Pinheiro,[2] Wei Hua,[2] Peter Arlett,[3,4] Yoshiaki Uyama,[5] Jesse A Berlin,[6] Dorothee B Bartels,[7] Kristijan H Kahler,[9] Lily G Bessette,[1] Sebastian Schneeweiss[1]

Best practices

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

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

ENCePP
European Network of Centres for Pharmacoepidemiology and Pharmacovigilance

EMA/95098/2010 Rev.9

The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Guide on Methodological Standards in Pharmacoepidemiology
(Revision 9)

Misc: Highly specific or focusing on parts of the process

Journal of the American Medical Informatics Association, 27(8), 2020, 1331–1337
doi: 10.1093/jamia/ocaa103
Perspective

AMIA

Perspective

Principles of Large-scale Evidence Generation and Evaluation across a Network of Databases (LEGEND)

Martijn J. Schuemie[1,2], Patrick B. Ryan[1,3], Nicole Pratt[4], RuiJun Chen[3,5], Seng Chan You[6], Harlan M. Krumholz[7], David Madigan[8], George Hripcsak[3,9], and Marc A. Suchard[2,10]

Clinical Pharmacology & Therapeutics

REVIEW | Open Access

The Structured Process to Identify Fit-for-purpose Data (SPIFD): A data feasibility assessment framework

Nicolle M Gatto, Ulka B Campbell, Emily Rubinstein, Ashley Jaksa, Pattra Mattox, Jingping Mo, Robert F Reynolds

First published: 30 October 2021 | https://doi-org.ezp-prod1.hul.harvard.edu/10.1002/cpt.2466

Why do we need another framework?

What do we have?

- Various tools exist in the literature for quality assessment, reporting, and describing best practices for pharmacoepidemiologic research

What don't we have?

- None of these tools offer a general framework to guide decision making at various steps along the way

Vision for a framework to guide principled investigations using non-randomized, secondary data

- The Sentinel Innovation Center is developing a causal inference framework proposing *a stepwise process that systematically considers key choices* with respect to design and analysis that influence the validity of studies conducted with non-randomized, secondary data
- A standardized "industrial" process that will be outlined in this framework will serve as *a guide to inform the conduct of non-randomized secondary database studies* of drug-outcome evaluation
- Key considerations to meet the FDA need of informing regulatory decision making based on such investigations
  - Limit variations across investigators by outlining a general process
  - Focus on repeatability of the process
  - Written and endorsed by independent experts

A draft of the proposed framework

A working draft of a causal inference framework for Sentinel

Study planning

Inference

STEP 1

Well defined research question in the target trial framework specifying PICOTS

STEP 2

Determining fit-for-purpose data source

STEP 3

Refining target trial parameters and translate to RWE study parameters

STEP 4

Feasibility assessment

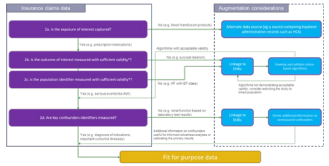
STEP 5

Robustness evaluations

STEP 6

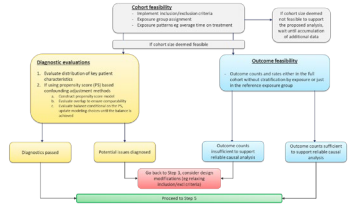
Inferential analysis

See Figure for Step 2

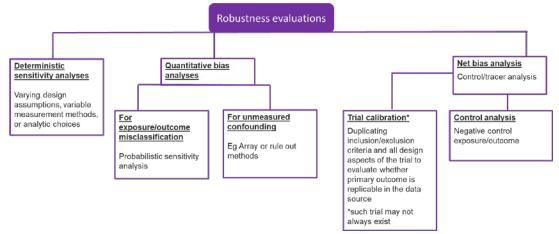


See Figure for Step 3

See Figure for Step 4



See Figure for Step 5

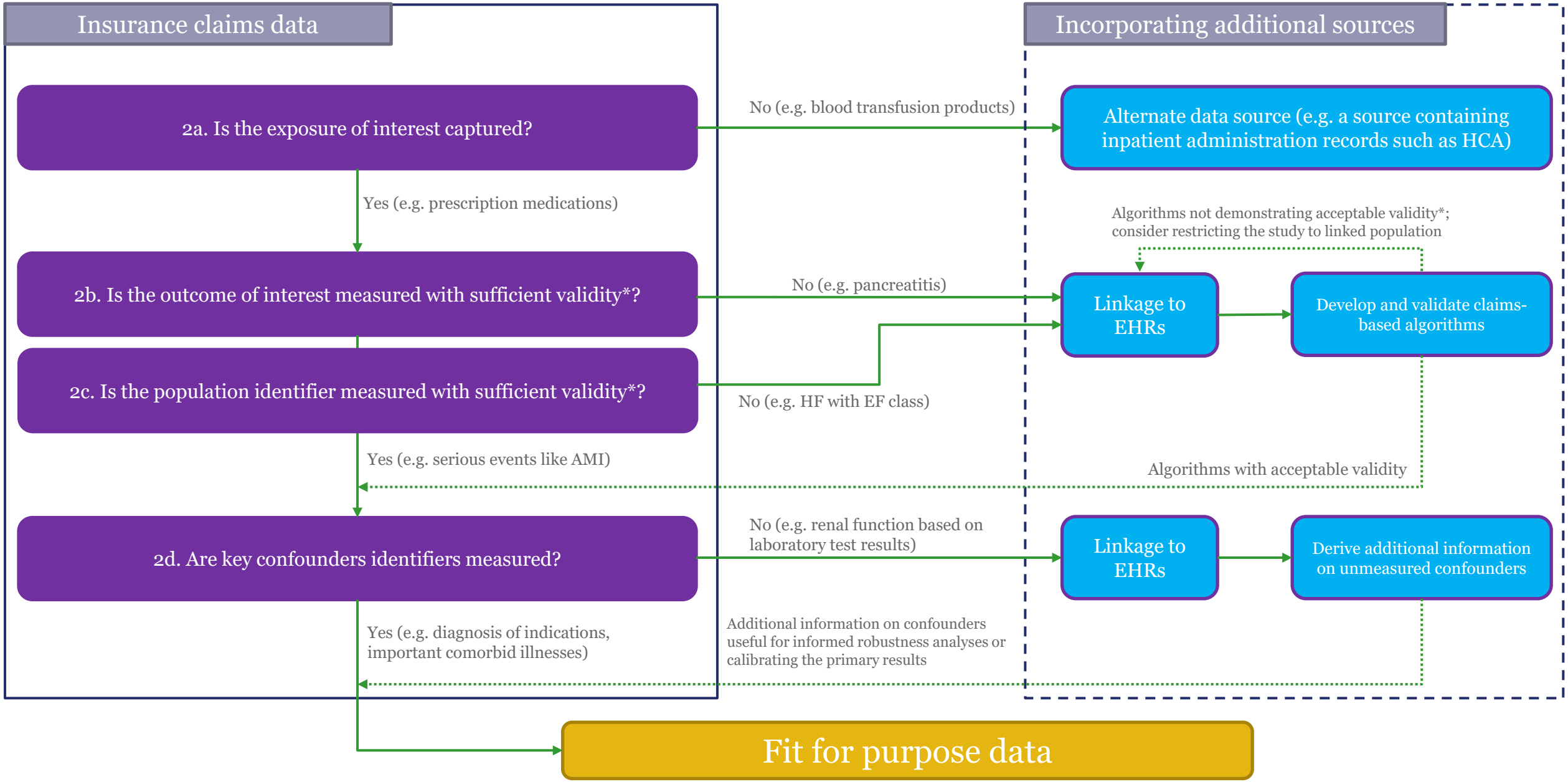


Step 1: Well defined research question in the target trial framework specifying PICOTS

- First and non-negotiable step in any framework that intends to generate causal inference from observed data
- Target trial framework, which is conceptualized as envisioning a hypothetical prospective randomized controlled trial, provides a useful and practical device to sharply define a causal question of interest
- Explicit identification of the following key study parameters
  - patient population (P)
  - the intervention (I) specifying the medical product under investigation,
  - a comparator group (C)
  - the outcome (O) along with an appropriate time horizon (T)
  - setting (S) where the study is implemented

Hernan. *AJPH*. 2018; 08:616-619
Hernan and Robins. *Am J Epidemiol*. 2016;183:758-764
Guidance from the AHRQ's Evidence-based Practice Centers Program

Step 2: Determining fit-for-purpose data sources



* Validity as demonstrated by parameters including PPV, sensitivity, specificity for binary outcomes; proportion missing for continuous outcomes; and accurate onset for time to event outcomes and availability of long-term follow-up data for latent outcomes

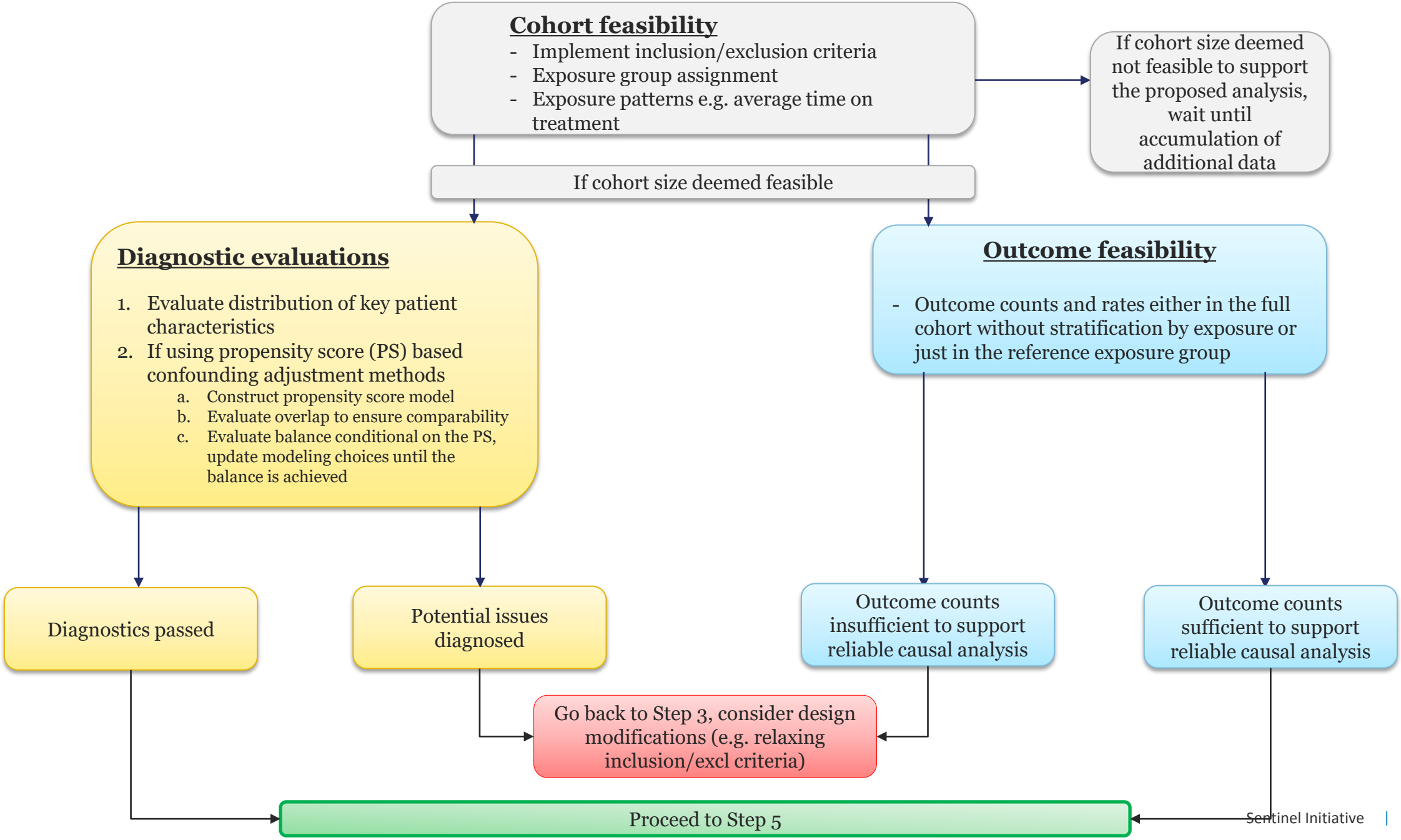
Step 3: Refining target trial parameters[1] and translate to RWE study parameters

(Using a hypothetical example case study of SGLT2 inhibitors and the risk of genital infection in a claims-EHR linked data source)

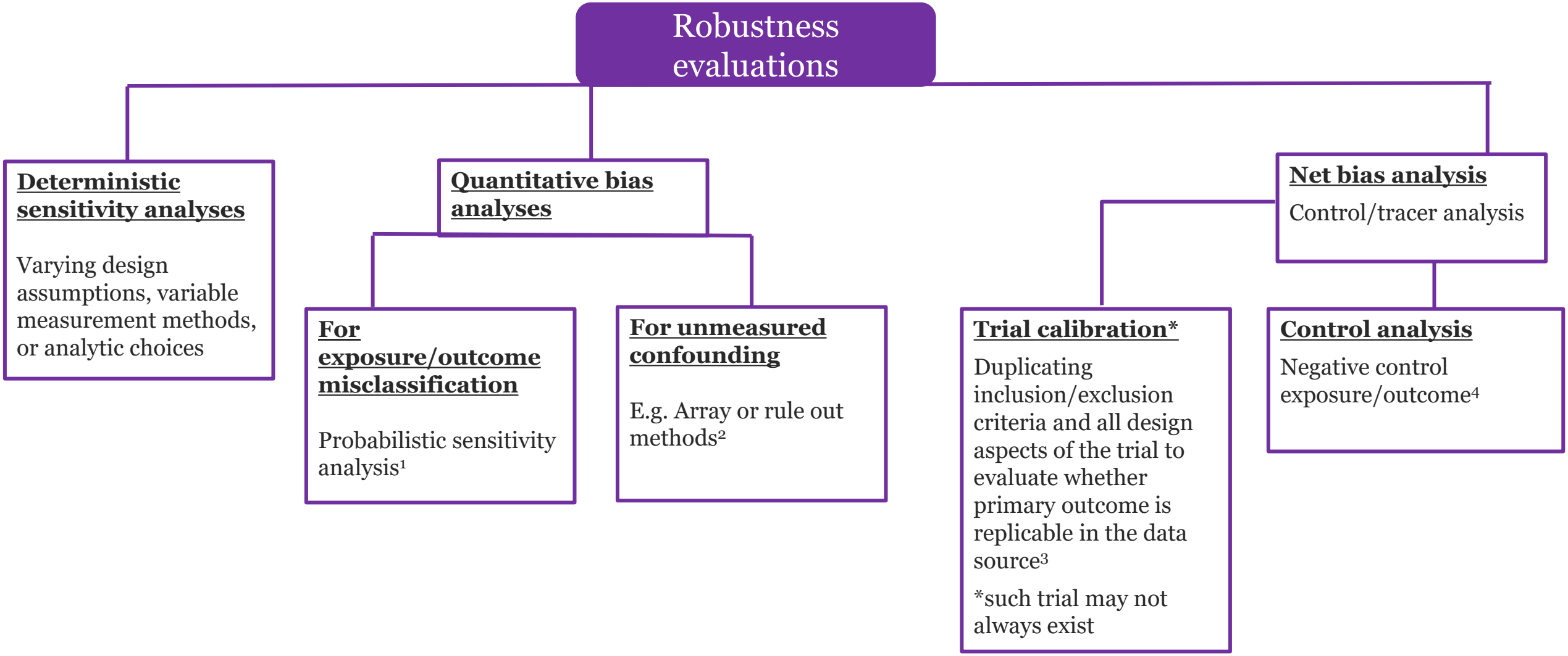
Element	Ideal trial	RWE translation
Exposure ("treatment strategies")	Randomly assigned initiation of SGLT2i (canagliflozin, dapagliflozin, empagliflozin) versus a DPP4 inhibitors	First prescription dispensing of SGLT2i (canagliflozin, dapagliflozin, empagliflozin) or DPP4 inhibitors identified based on pharmacy claims
Eligibility (assessed at baseline, prior to time 0)	Patients aged 18 years or older, with type 2 diabetes mellitus, and no use of study medications before randomization	Observability related: continuous enrollment for 12 months and >80% mean capture proportion[2] in EHRs before study medication initiation
		Treatment related: No prior use of study medications
		Indication related: Diagnosis of type 2 diabetes based on diagnosis codes or HbA1c results
		Other: Age 18 or older
Follow-up start (Time 0)	At randomization	At prescription dispensing
Follow-up end	1-year post-randomization unless patients are lost to follow-up or die or have the outcome	Earliest of the outcome, death, insurance disenrollment, or 1-year post initiation
Primary outcome	Hospitalization for genital infections	Hospitalization for genital infections assessed based on primary discharge diagnosis codes
Baseline covariates	-	Demographics, diabetes severity related variables including micro and macrovascular complications and laboratory test results such as HbA1c and serum creatinine, comorbid conditions, comedications, markers for healthy behavior and healthcare utilization
Causal estimand	Intent-to-treat (ITT)	Observational analogue of ITT
Statistical analysis	A Cox proportional hazards model	Adjustment of baseline confounding with propensity score matching followed by an outcome analysis using a Cox proportional hazards model
Subgroup analyses	Stratified by gender	Same as ideal trial

1Hernan and Robins. *Am J Epidemiol.* 2016;183:758-764; Hernan. *N Engl J Med.* 2021;385(15):1345-1348
2 Lin et al. *Epidemiology* 2018;29: 356–363

Step 4: Feasibility analysis



Step 5: Pre-specification of robustness evaluations



1 Fox et al. International Journal of Epidemiology 2005;34:1370–1376
2 Schneeweiss. Pharmacoepiemiology Drug Saf 2006; 15: 291–303
3 Khosrow-Khavar et al. Annals Rheum Dis. 2022
4 Lipsitch et al. Epidemiology 2010;21: 383–388

Summary and next steps

- Continuing to fine tune the framework steps
- Conducting a demonstration project to highlight how decisions are made at each step along the way and walk users through the steps based on a realistic case-example
- The goal is dissemination of this framework in peer-reviewed publication by early next year

Questions?

Closing remarks

- Through initiatives such as those discussed today, Sentinel Innovation Center is making strides in helping to achieve the FDA's vision of a Medical Data Enterprise with a query-ready system containing >10 million EHR lives
- Key research needs have been identified and ongoing research projects are addressing some salient challenges presented by EHRs in 4 key domains
  - Data infrastructure
  - Feature engineering
  - Causal inference
  - Detection analytics
- Highly interdisciplinary research work being conducted at the Innovation Center involving experts in the fields of epidemiology, informatics, medicine, and statistics, will generate unique insights regarding meaningful use of EHRs for clinical research and provide practical solutions

npj | Digital Medicine

www.nature.com/npjdigitalmed

PERSPECTIVE OPEN

Check for updates

Broadening the reach of the FDA Sentinel system: A roadmap for integrating electronic health record data in a causal analysis framework

Rishi J. Desai[1]✉, Michael E. Matheny[2], Kevin Johnson[2], Keith Marsolo[3], Lesley H. Curtis[3], Jennifer C. Nelson[4], Patrick J. Heagerty[5], Judith Maro[6], Jeffery Brown[6], Sengwee Toh[6], Michael Nguyen[7], Robert Ball[7], Gerald Dal Pan[7], Shirley V. Wang[1], Joshua J. Gagne[1,8] and Sebastian Schneeweiss[1]

The Sentinel System is a major component of the United States Food and Drug Administration's (FDA) approach to active medical product safety surveillance. While Sentinel has historically relied on large quantities of health insurance claims data, leveraging longitudinal electronic health records (EHRs) that contain more detailed clinical information, as structured and unstructured features, may address some of the current gaps in capabilities. We identify key challenges when using EHR data to investigate medical product safety in a scalable and accelerated way, outline potential solutions, and describe the Sentinel Innovation Center's initiatives to put solutions into practice by expanding and strengthening the existing system with a query-ready, large-scale data infrastructure of linked EHR and claims data. We describe our initiatives in four strategic priority areas: (1) data infrastructure, (2) feature engineering, (3) causal inference, and (4) detection analytics, with the goal of incorporating emerging data science innovations to maximize the utility of EHR data for medical product safety surveillance.

npj Digital Medicine (2021)4:170 ; https://doi.org/10.1038/s41746-021-00542-0

Innovation Center collaborating organizations

Lead sites:

KAISER PERMANENTE®
Kaiser Permanente Washington Health Research Institute

VANDERBILT UNIVERSITY MEDICAL CENTER

BRIGHAM HEALTH
BWH BRIGHAM AND WOMEN'S HOSPITAL

SCHOOL OF PUBLIC HEALTH
UNIVERSITY of WASHINGTON

Duke Clinical Research Institute

Thank you