



Yale SCHOOL OF MEDICINE

Biomedical Informatics and Data Science

Making Sense of EHR Data in IMPACT-MH

Hua Xu PhD

April 29th, 2026

IMPACT-MH EHR Session

Outline

- 01 EHR Data Representation**
- 02 EHR Phenotyping**
- 03 IMPACT-MH Approaches for EHR Data**



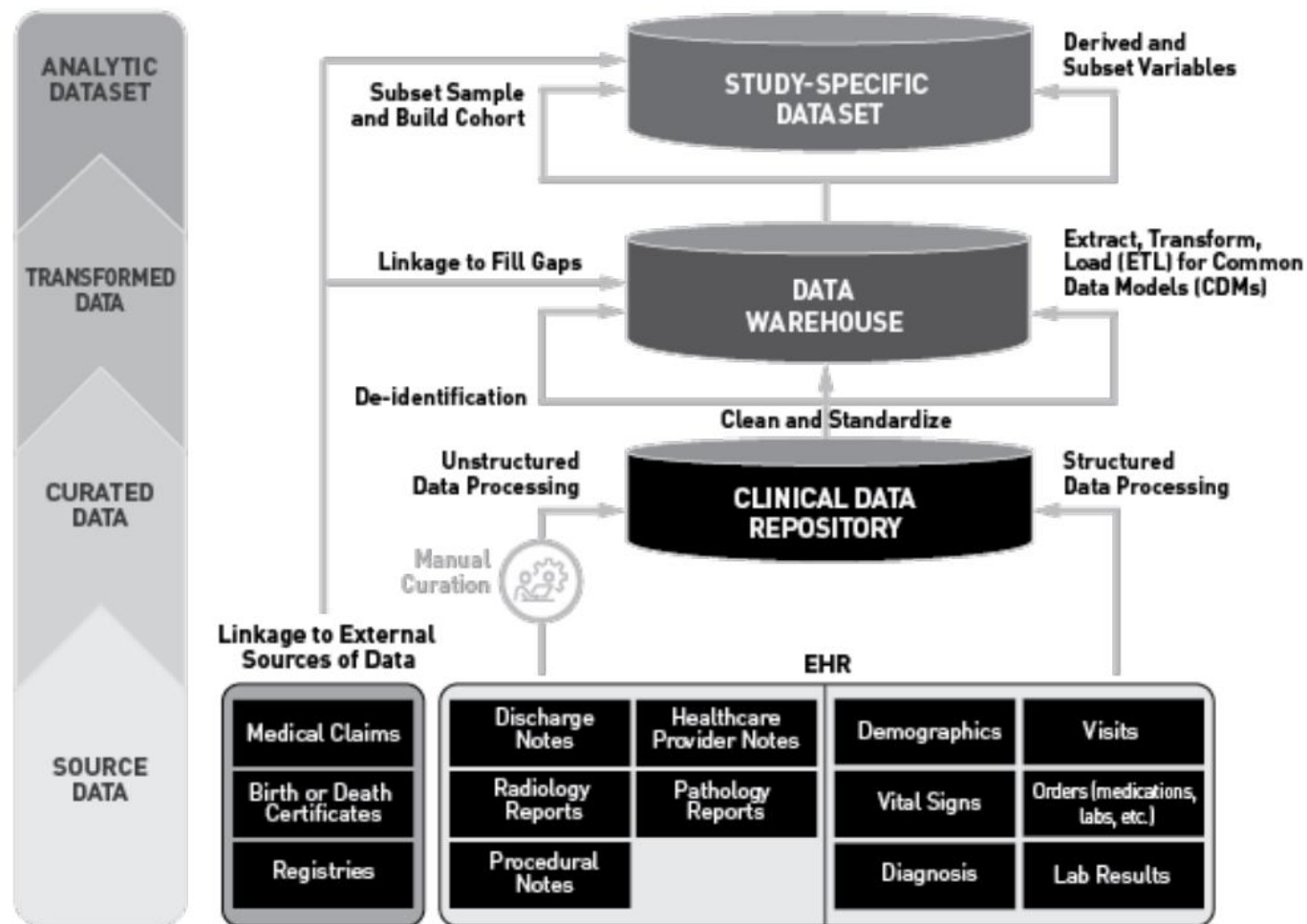
EHR Data Representation

Electronic Health Records (EHRs) for Clinical Studies

- EHRs (and linked data) becomes an enabling resource for Real World Evidence generation



FDA RWD Guidance 2021 *

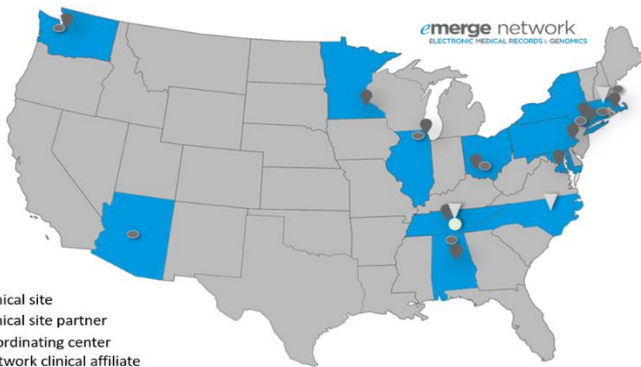


* Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products – Draft Guidance by FDA, September 2021

Large Clinical Data Networks for Research

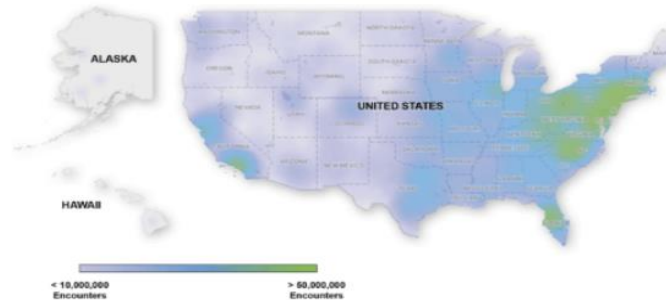


EMERGE SITES



>47 Million in the Network

PCORnet represents data from everyday healthcare encounters with more than 47 million people across the U.S. each year.



Participants at a Glance

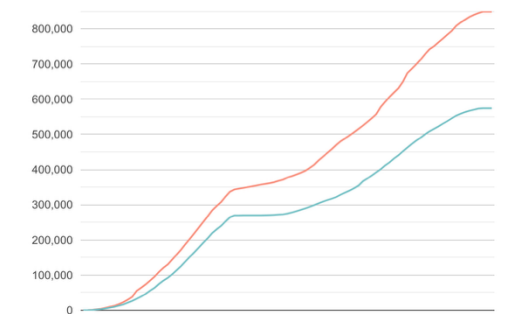
849,000+
Participants

574,000+
Participants who have completed initial steps of the program

Enrollment Numbers

This graph represents participants who have consented to join the program and those who have completed all initial steps of the program. The initial steps are consenting, agreeing to share electronic health records, completing the first three surveys, providing physical measurements, and donating at least one biospecimen to be stored at the biobank.

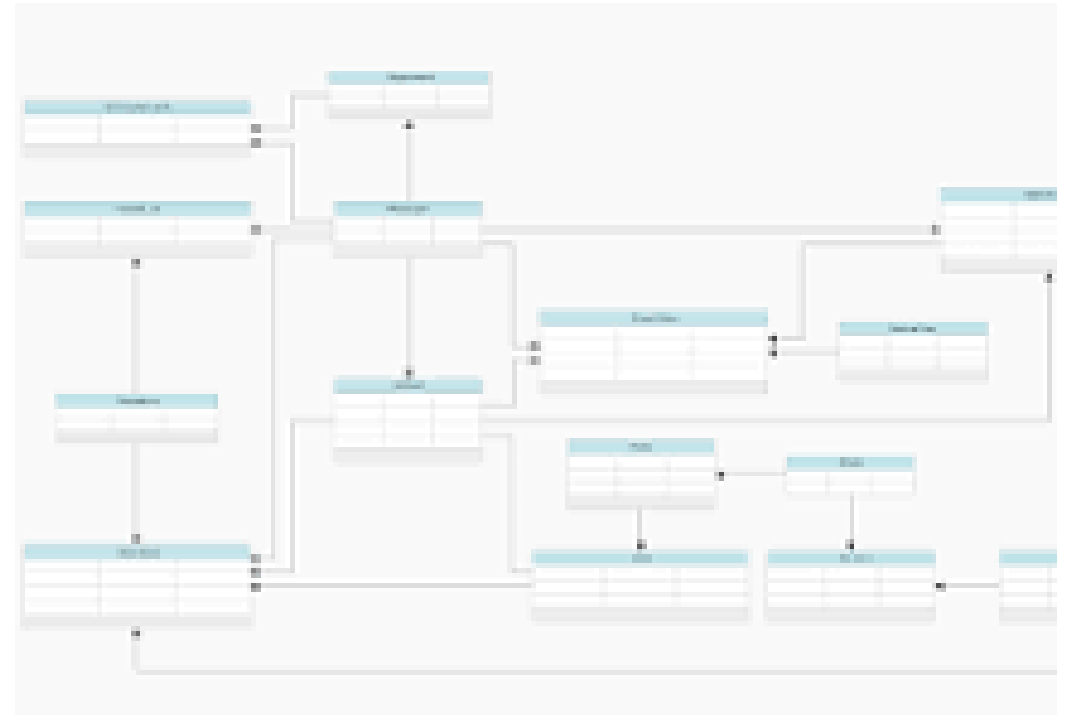
The following numbers are approximated to protect participants' privacy. Numbers are updated as of February 2, 2025.



The Complexity of Raw EHR Data

Why Research is Difficult

- **Proprietary Architectures:** Systems like Epic utilize complex backend structures (e.g., Clarity/Chronicles) with **tens of thousands of tables**.
- **Operational Focus:** Designed for transaction processing, billing, and clinical documentation, not for analytical or longitudinal queries.
- **Data Fragmentation:** Information is siloed across disparate modules for pharmacy, lab, and inpatient care, making record linkage a major barrier



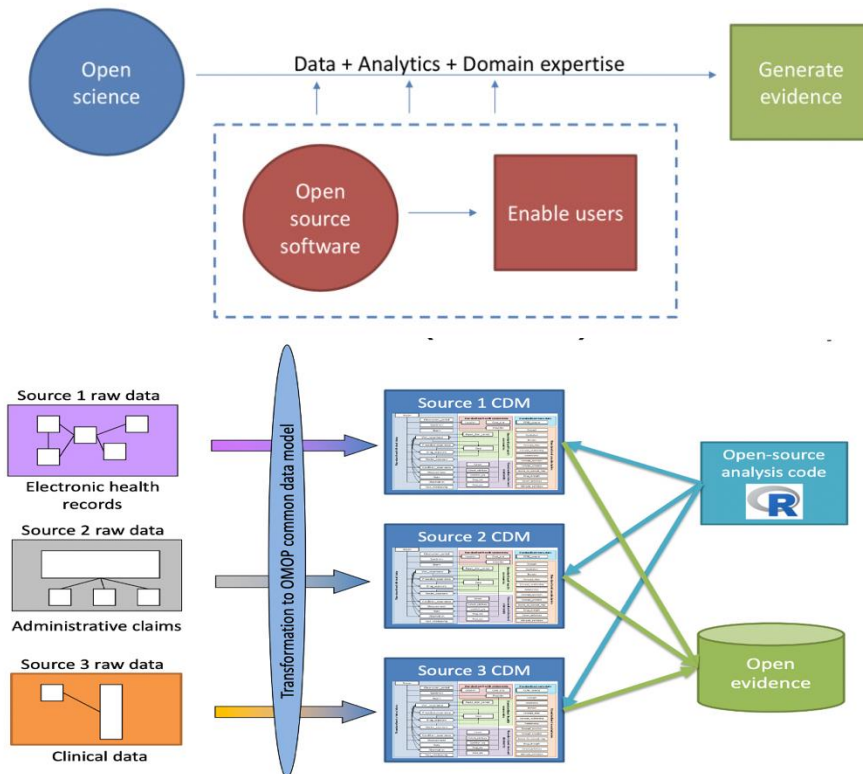
Standardizing with Common Data Models (CDM)

To overcome raw data complexity, fragmented EHR data is transformed into standardized **Common Data Models (CDMs)** for cross-institutional research.

Common Data Model	Philosophy & Architecture	Typical Research Application
OMOP CDM (OHDSI)	Person-centric model using standard global vocabularies (SNOMED, RxNorm).	Large-scale federated analytics and drug safety studies.
i2b2	"Informatics for Integrating Biology and the Bedside." Centered on star schema facts.	Hospital-level cohort discovery and clinical trial recruitment.
PCORnet CDM	Derived from Mini-Sentinel. Pragmatic, table-based model for rapid querying.	Patient-centered outcomes and comparative effectiveness studies.

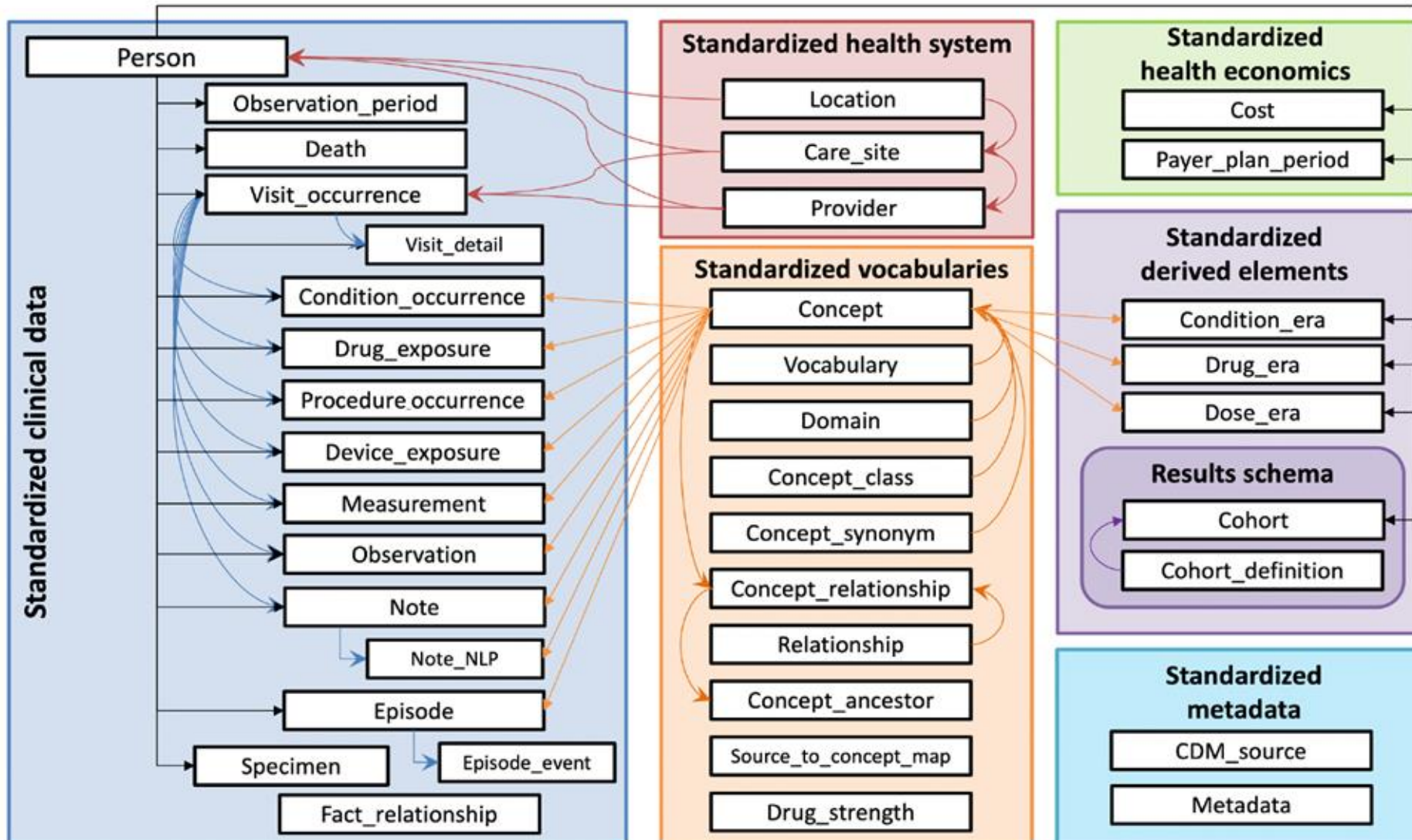
The Observational Health Data Sciences and Informatics (OHDSI) Consortium

- A multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics



OHDSI Collaborators:	OHDSI Data Network:
<ul style="list-style-type: none"> • >2,770 researchers in academia, industry and government • >21 countries 	<ul style="list-style-type: none"> • >133 databases from 18 countries • 1.9 billion patient records (duplicates) • ~369 million non-US patients

OMOP Common Data Model



Converting EHR data to OMOP CDM: The ETL Process

The transition from raw data to a Common Data Model follows a rigorous **Extract, Transform, and Load (ETL)** framework to ensure semantic accuracy.





Yale SCHOOL OF MEDICINE

Biomedical Informatics and Data Science

EHR Phenotyping

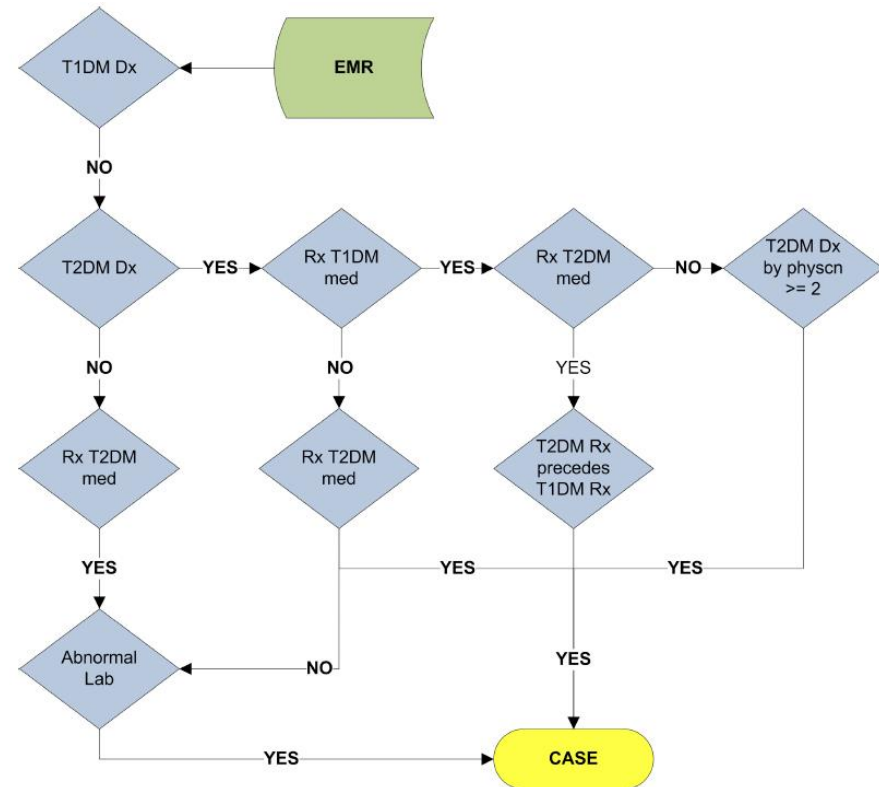
EHR Phenotyping

Computable Phenotype: A clinical condition or trait ascertainable via computerized query to an EHR system.

It involves identifying patient cohorts using clinical footprints such as:

- ICD Diagnosis Codes
- Lab Tests & Vital Signs
- Medication Records
- Narrative Clinical Notes (NLP)

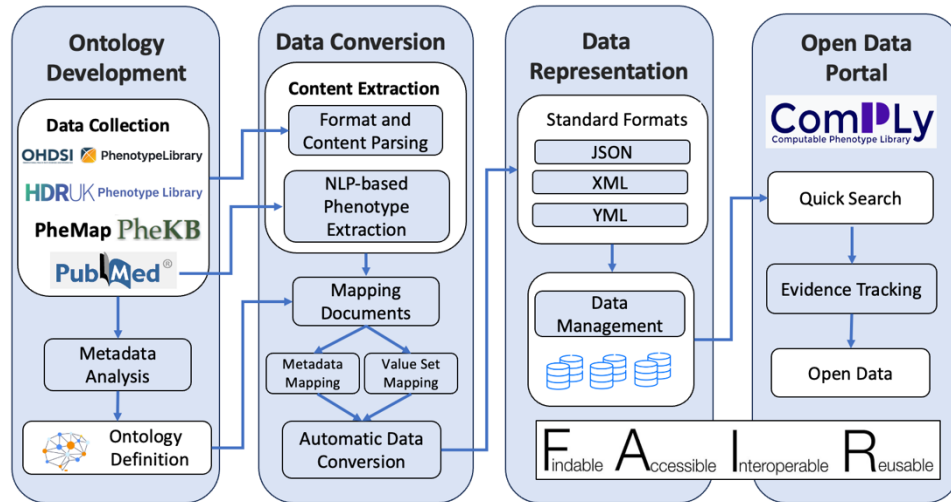
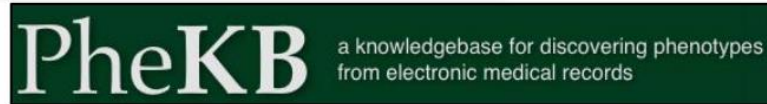
Identifying patients based on their digital health profiles across time.



Ref: eMERGE project: Type 2 Diabetes Mellitus Electronic Medical Record Case and Control Selection Algorithms



Existing Phenotyping Libraries



Challenges of EHR Phenotyping



Data Quality

EHR data is optimized for billing, leading to discrepancies between ICD codes and true clinical status.



Missing Data

Information is "missing not at random." Absence of evidence is not always evidence of absence.



Heterogeneity

Inconsistent institutional workflows and coding practices make algorithms difficult to port across sites.

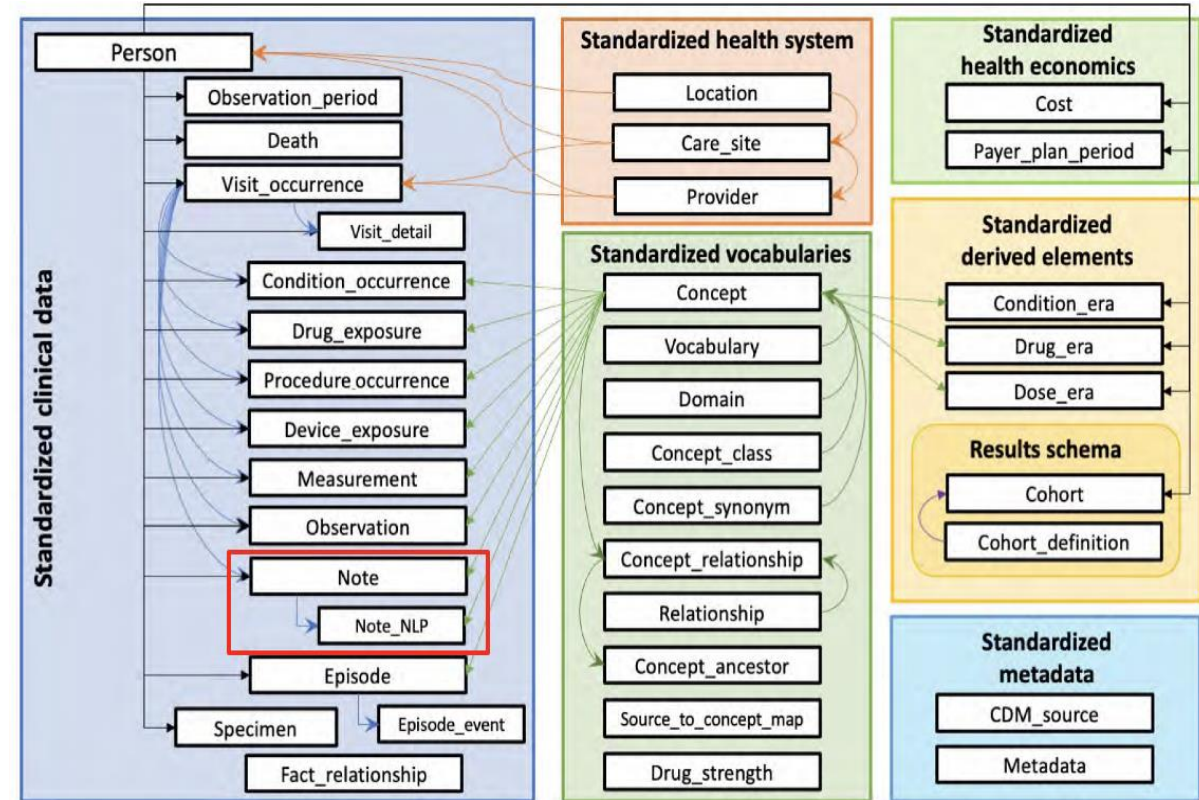
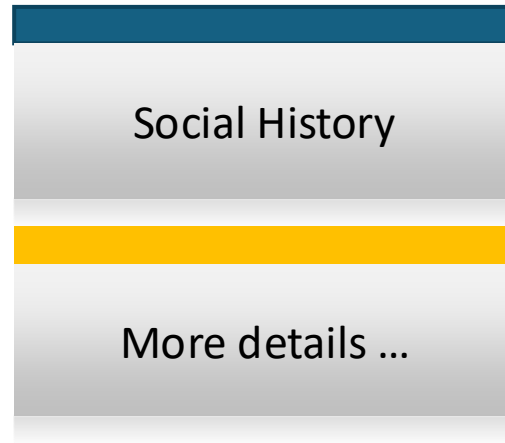
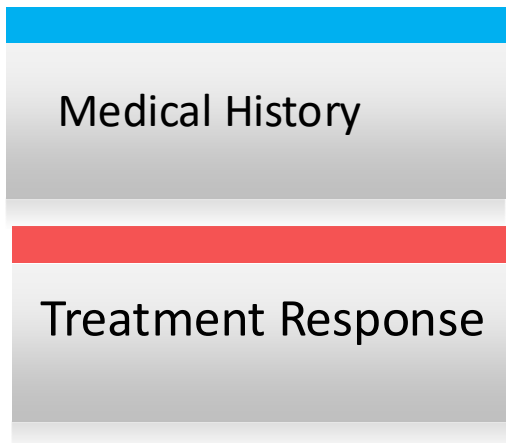
Textual Documents in EHRs

Admit 10/23

Medical History: 71 yo woman h/o DM, HTN, Dilated CM/CHF, Afib s/p embolic event, chronic diarrhea, admitted with SOB. CXR pulm edema. Rx'd Lasix.

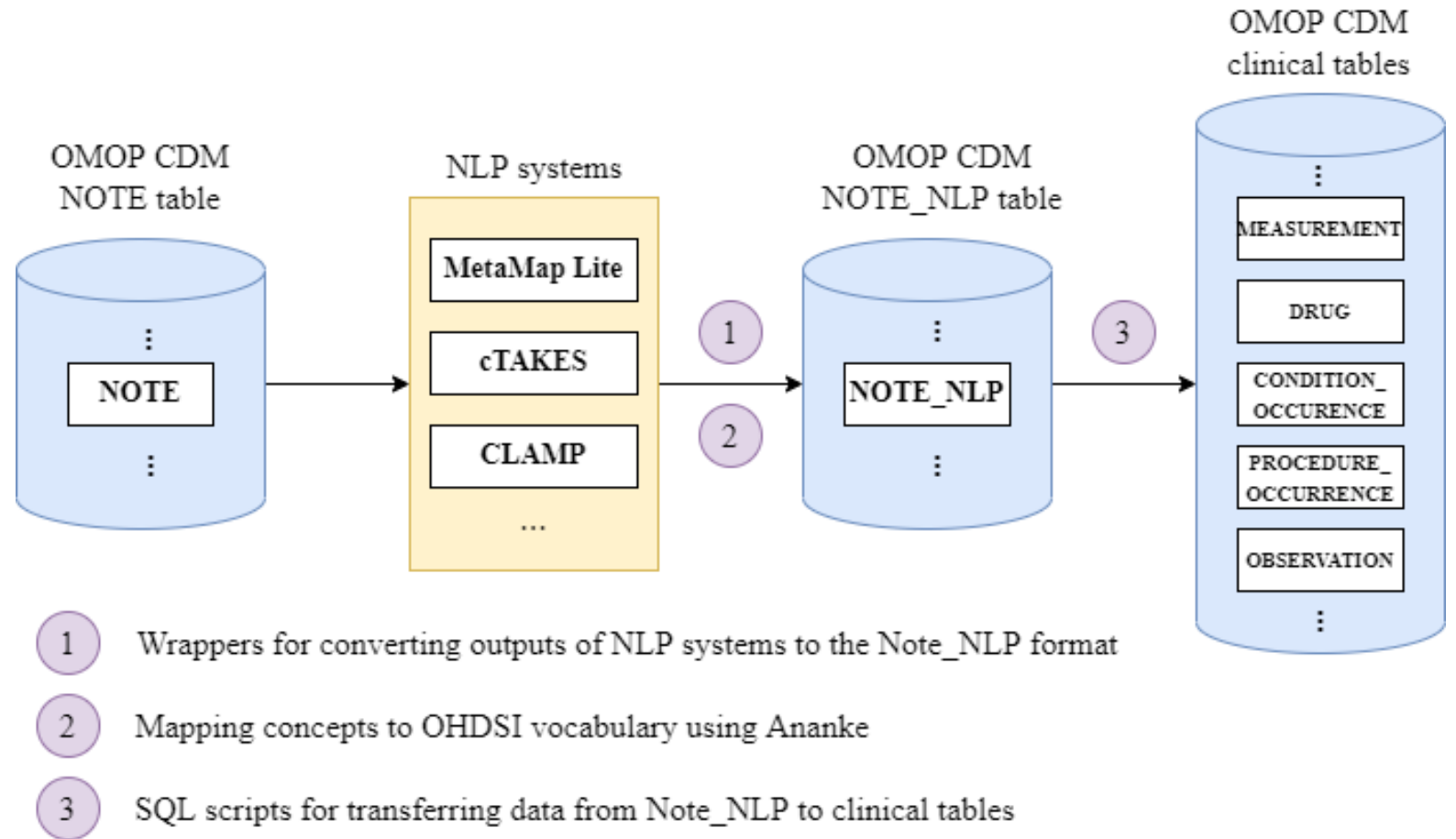
Social History: PT isolates to self in her apartment.

Meds Lasix 40mg IVP bid, ASA, Coumadin 5, Prinivil 10, glucophage 850 bid, glipizide 10 bid, immodium prn



Workflow for Textual Data in CDM

- Run natural language processing (NLP) systems to process textual notes in NOTE table
- Convert NLP system output into NOTE_NLP table
- Transfer concepts from NOTE_NLP to clinical tables in CDM



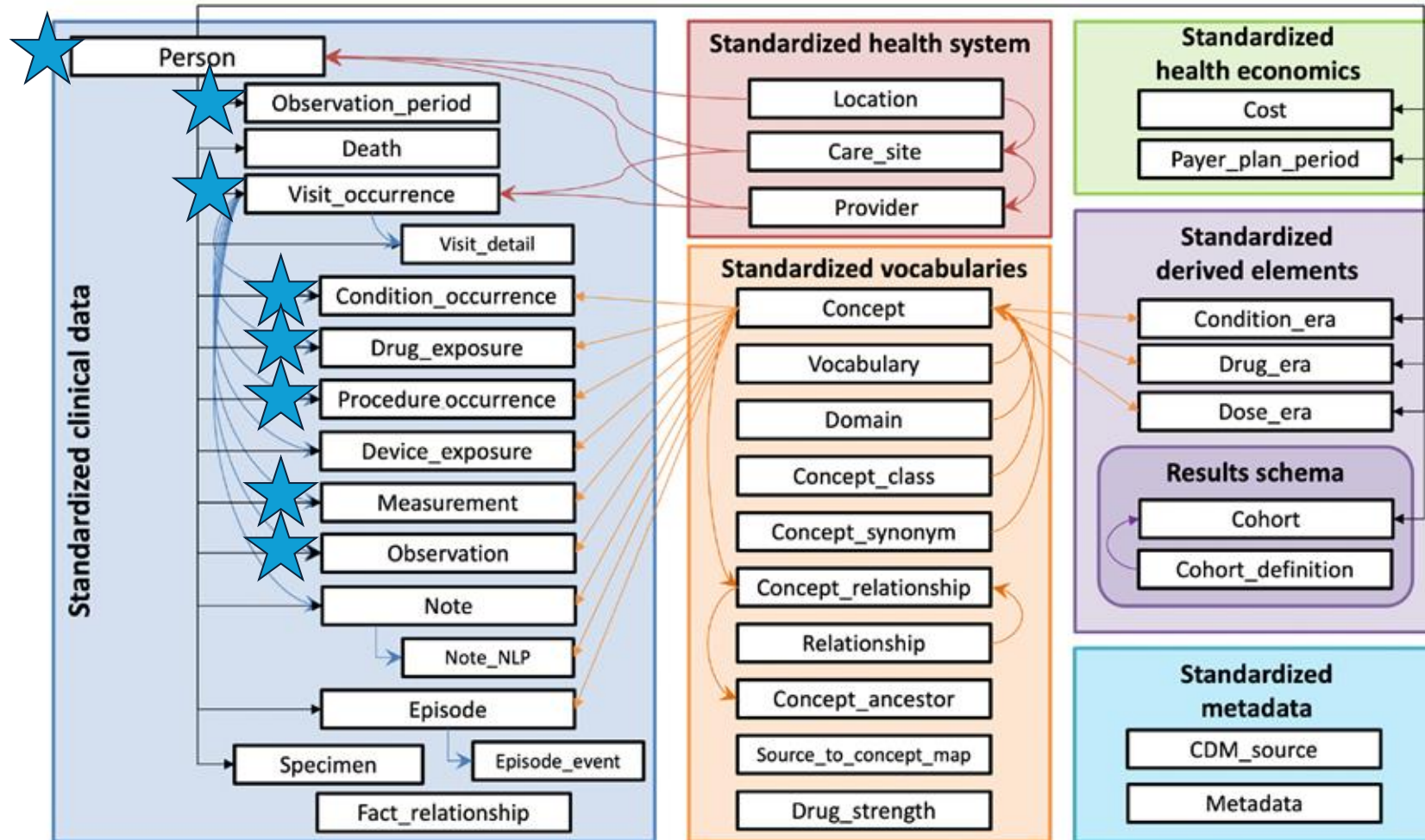
Keloth VK et al. Representing and utilizing clinical textual data for real world studies: An OHDSI approach. J Biomed Inform. 2023 Jun;142:104343. doi: 10.1016/j.jbi.2023.104343.



IMPACT-MH Approaches for EHR Data

- Data Structures and Workflow
- Diverse Tools
- Guidance and Support

Data Structures – Simplified OMOP CDM



★ Recommended 8 core tables

DCC Suggested NDA Data Structure – impact_ehr01

NDA required elements



ElementName	DataType	Size	Required	ElementDescription	ValueRange
subjectkey	GUID		Required	The NDAR Global Unique Identifier (GUID) for research subject	NDAR*
src_subject_id	String	45	Required	Subject ID how it's defined in lab/project	
interview_date	Date		Required	convert_to_omop_date; Date on which the original EHR data was convert to OMOP in attached files MM/DD/YYYY	
interview_age	Integer		Required	Age in months at the time of the interview/test/sampling/imaging/convertingtiOMOP	0::1440
sex	String	20	Required	Sex of subject at birth	M;F; O; NR
Person_manifest	File		Recommended	Manifest file for Patient Demographics data	
Condition_manifest	File		Recommended	Manifest file for Condition data	
Drug_manifest	File		Recommended	Manifest file for Medication data	
Procedure_manifest	File		Recommended	Manifest file for Procedure data	
Visit_manifest	File		Recommended	Manifest file for Visit/Encounter data	
Observation_manifest	File		Recommended	Manifest file for Observation data	
Observation_period_manifest	File		Recommended	Manifest file for Observation Period data	
Measurement_manifest	File		Recommended	Manifest file for Measurement data	
Visit_detail_manifest	File		Optional	Manifest file for Visit Detail data	
Death_manifest	File		Optional	Manifest file for Death data	
Device_exposure_manifest	File		Optional	Manifest file for Device Exposure data	
Specimen_manifest	File		Optional	Manifest file for Specimen data	
Note_manifest	File		Optional	Manifest file for clinical Note data	
Note_nlp_manifest	File		Optional	Manifest file for Note NLP data	
Fact_relationship_manifest	File		Optional	Manifest file for Fact Relationship data	
Location_manifest	File		Optional	Manifest file for Location data	
Care_site_manifest	File		Optional	Manifest file for Care Site data	
Provider_manifest	File		Optional	Manifest file for Provider data	
Payer_plan_period_manifest	File		Optional	Manifest file for Payer Plan Period data	
Cost_manifest	File		Optional	Manifest file for Cost data	
Drug_era_manifest	File		Optional	Manifest file for Drug Era data	
Dose_era_manifest	File		Optional	Manifest file for Dose Era data	
Condition_era_manifest	File		Optional	Manifest file for Condition Era data	
Episode_manifest	File		Optional	Manifest file for Episode data	
.....					

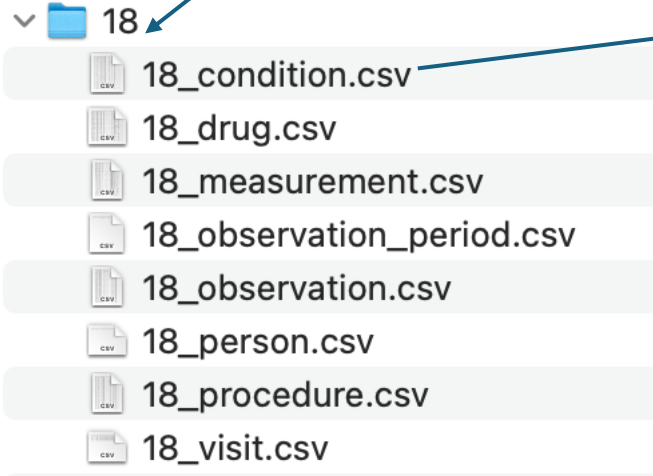
OMOP CDM 5.4



EHR Data Example – file structure

8 records of patients for NDA data submission

subjectKey	src_subject_id	interview_date	interview_age	sex	Person_manifest	Condition_manifest	Drug_manifest	Procedure_manifest	Visit_manifest	Observation_manifest	Observation_period_manifest	Measurement_manifest
89f3081d-3deb-	18	4/24/26	35	F	18/18_person.csv	18/18_condition.csv	18/18_drug.csv	18/18_procedure.csv	18/18_visit.csv	18/18_observation.csv	18/18_observation_period.csv	18/18_measurement.csv
a3dfc07e-2461-	37	4/24/26	26	M	37/37_person.csv	37/37_condition.csv	37/37_drug.csv	37/37_procedure.csv	37/37_visit.csv	37/37_observation.csv	37/37_observation_period.csv	37/37_measurement.csv
1647ccfd-d121-	38	4/24/26	24	M	38/38_person.csv	38/38_condition.csv	38/38_drug.csv	38/38_procedure.csv	38/38_visit.csv	38/38_observation.csv	38/38_observation_period.csv	38/38_measurement.csv
dafcf4f-7fcb-4e	47	4/24/26	74	F	47/47_person.csv	47/47_condition.csv	47/47_drug.csv	47/47_procedure.csv	47/47_visit.csv	47/47_observation.csv	47/47_observation_period.csv	47/47_measurement.csv
a9796a73-216c-	50	4/24/26	23	F	50/50_person.csv	50/50_condition.csv	50/50_drug.csv	50/50_procedure.csv	50/50_visit.csv	50/50_observation.csv	50/50_observation_period.csv	50/50_measurement.csv
8cda4faa-8a7d-	55	4/24/26	64	M	55/55_person.csv	55/55_condition.csv	55/55_drug.csv	55/55_procedure.csv	55/55_visit.csv	55/55_observation.csv	55/55_observation_period.csv	55/55_measurement.csv
41dae394-269a-	59	4/24/26	29	F	59/59_person.csv	59/59_condition.csv	59/59_drug.csv	59/59_procedure.csv	59/59_visit.csv	59/59_observation.csv	59/59_observation_period.csv	59/59_measurement.csv
a5602725-ac45-	66	4/24/26	33	M	66/66_person.csv	66/66_condition.csv	66/66_drug.csv	66/66_procedure.csv	66/66_visit.csv	66/66_observation.csv	66/66_observation_period.csv	66/66_measurement.csv


 A file explorer view showing a folder named '18' containing several CSV files: 18_condition.csv, 18_drug.csv, 18_measurement.csv, 18_observation_period.csv, 18_observation.csv, 18_person.csv, 18_procedure.csv, and 18_visit.csv. A blue arrow points from the '18' folder to the '18_condition.csv' file, and another blue arrow points from the '18_condition.csv' file to the 'condition_occurrence_id' column of the table below.

condition_occurrence_id	person_id	condition_concept_id	condition_start_date	condition_start_datetime	condition_end_date	condition_end_datetime	condition_type_concept_id
446479	18	434610	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821
446480	18	443597	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821
446481	18	432867	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821
446482	18	433811	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821
446483	18	197320	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821
446484	18	4152384	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821
446485	18	316139	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821
446486	18	321052	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821
446487	18	4288310	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821
446488	18	439696	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821
446489	18	80809	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821
446490	18	434005	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821
446491	18	4233565	7/12/72	7/12/72 13:50	7/19/72	7/19/72 0:05	32821
446492	18	433736	7/12/72	7/12/72 13:50	7/19/72	7/19/72 0:05	32821
446493	18	37170728	7/12/72	7/12/72 13:50	7/19/72	7/19/72 0:05	32821
446494	18	440674	7/12/72	7/12/72 13:50	7/19/72	7/19/72 0:05	32821
446495	18	80809	7/12/72	7/12/72 13:50	7/19/72	7/19/72 0:05	32821
446496	18	36713576	7/12/72	7/12/72 13:50	7/19/72	7/19/72 0:05	32821
446497	18	444044	7/12/72	7/12/72 13:50	7/19/72	7/19/72 0:05	32821
446498	18	193242	7/12/72	7/12/72 13:50	7/19/72	7/19/72 0:05	32821
446499	18	443327	7/12/72	7/12/72 13:50	7/19/72	7/19/72 0:05	32821

EHR Data Example – data standardization

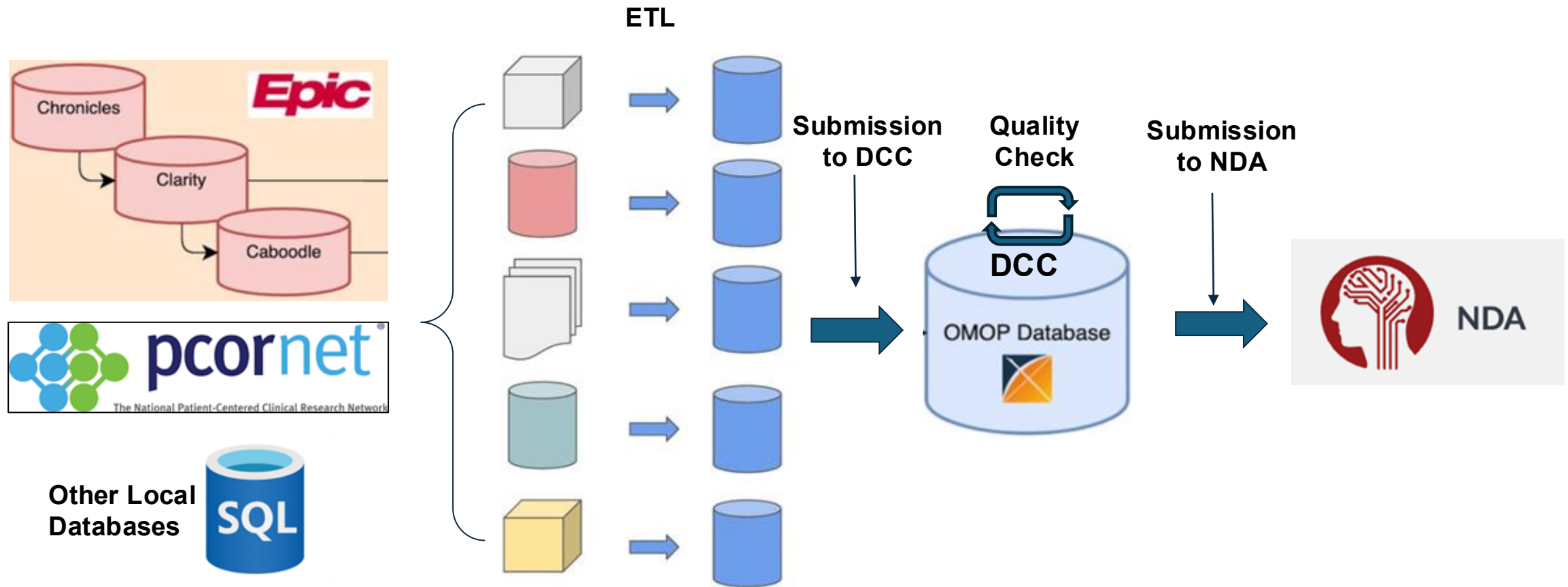
8 records of patients for NDA data submission

condition_occurrence_id	person_id	condition_concept_id	condition_start_date	condition_start_datetime	condition_end_date	condition_end_datetime	condition_type_concept_id								
446479	18	434610	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821								
446480	18	443597	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821								
446481	18	432867	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821								
446482	18	433811	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821								
446483	18	197320	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821								
446484	18	4152384	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821								
446485	18	316139	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821								
446486	18	321052	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821								
446487	18	4288310	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821								
446488	18	439696	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821								
446489	18	80809	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821								
446490	18	434005	1/30/70	1/30/70 11:15	2/1/70	2/1/70 13:00	32821								
446491	18	4233565	condition_end_date	condition_end_datetime	condition_type_concept_id	stop_reason	provider_id	visit_occurrence_id	visit_detail_id	condition_source_value	condition_source_concept_id				
446492	18	433736	2/1/70	2/1/70 13:00	32821			45695		E875	35207097				
446493	18	37170728	2/1/70	2/1/70 13:00	32821			45695		N183	35209276				
446494	18	440674	2/1/70	2/1/70 13:00	32821			45695		E785	35207065				
446495	18	80809	2/1/70	2/1/70 13:00	32821			45695		N1330	45572699				
446496	18	36713576	2/1/70	2/1/70 13:00	32821			45695		N179	35209273				
446497	18	444044	2/1/70	2/1/70 13:00	32821			45695		I350	35207733				
446498	18	193242	2/1/70	2/1/70 13:00	32821			45695		I509	35207793				
446499	18	443327	2/1/70	2/1/70 13:00	32821			45695		I739	35207860				
			2/1/70	2/1/70 13:00	32821			45695		I6529	45601050				
			2/1/70	2/1/70 13:00	32821			45695		I130	35207673				
			2/1/70	2/1/70 13:00	32821			45695		M069	35208753				
			2/1/70	2/1/70 13:00	32821			45695		E6601	45600659				
			7/19/72	7/19/72 0:05	32821			45696		E43	35206967				
			7/19/72	7/19/72 0:05	32821			45696		E669	35207024				
			7/19/72	7/19/72 0:05	32821			45696		L89152	45601241				
			7/19/72	7/19/72 0:05	32821			45696		M109	35208756				
			7/19/72	7/19/72 0:05	32821			45696		M069	35208753				
			7/19/72	7/19/72 0:05	32821			45696		N130	37200876				
			7/19/72	7/19/72 0:05	32821			45696		N170	35209269				
			7/19/72	7/19/72 0:05	32821			45696		K631	35208315				
			7/19/72	7/19/72 0:05	32821			45696		K611	35208301				

table continue 

Standard OMOP Concept ID

EHR Data Workflow



Tools

- ETL
 - Athena – concept mapping
 - AutoETL using LLMs (under development)
- NLP
 - De-identification
 - Kiwi – information extraction
- Phenotyping
 - Echo – AI agents for phenotyping (under development)
- Federated Data Analysis
 - PDA

Athena – OHDSI Standard Vocabularies Search Platform

The screenshot displays the Athena search interface. At the top, there is a green header with the Athena logo and navigation buttons for SEARCH, DOWNLOAD, and a user profile (Na Hong). Below the header is a search bar with the keyword 'aspirin'. The main content area shows a list of search results with columns for ID, CODE, NAME, CLASS, CONCEPT, VALIDITY, DOMAIN, and VOCAB. The results are filtered to show 15 items, with a total of 687,839 items available. The results list various clinical findings related to aspirin, such as '1 o'clock position on mammogram' and '1 year examination abnormal - for observation'.

DOMAIN	ID	CODE	NAME	CLASS	CONCEPT	VALIDITY	DOMAIN	VOCAB
	3188503	2250001000004107	(sn) only sentinel nodes evaluated	Clinical Finding	Standard	Valid	Condition	Nebraska Lexicon
	4240068	369787008	0-5 mitoses per 10 HPF (score = 1)	Clinical Finding	Standard	Valid	Condition	SNOMED
	4031867	129772004	1 o'clock position on mammogram	Clinical Finding	Standard	Valid	Condition	SNOMED
	4288932	396447006	1 or more mitotic figure per mm2	Clinical Finding	Standard	Valid	Condition	SNOMED
	4015559	170259009	1 year examination abnormal - for observation	Clinical Finding	Standard	Valid	Condition	SNOMED
	4016214	170261000	1 year examination abnormal - on treatment	Clinical Finding	Standard	Valid	Condition	SNOMED
	4014878	170260004	1 year examination abnormal - referred	Clinical Finding	Standard	Valid	Condition	SNOMED
	4227224	420829009	1+ pitting edema	Clinical Finding	Standard	Valid	Condition	SNOMED
	4016490	170110001	10 day examination abnormal - for observation	Clinical Finding	Standard	Valid	Condition	SNOMED
	4016491	170112009	10 day examination abnormal - on treatment	Clinical Finding	Standard	Valid	Condition	SNOMED
	4016188	170111002	10 day examination abnormal - referred	Clinical Finding	Standard	Valid	Condition	SNOMED
	4031244	129781005	10 o'clock position on mammogram	Clinical Finding	Standard	Valid	Condition	SNOMED

KIWI - An LLM-based Clinical Information Extraction System

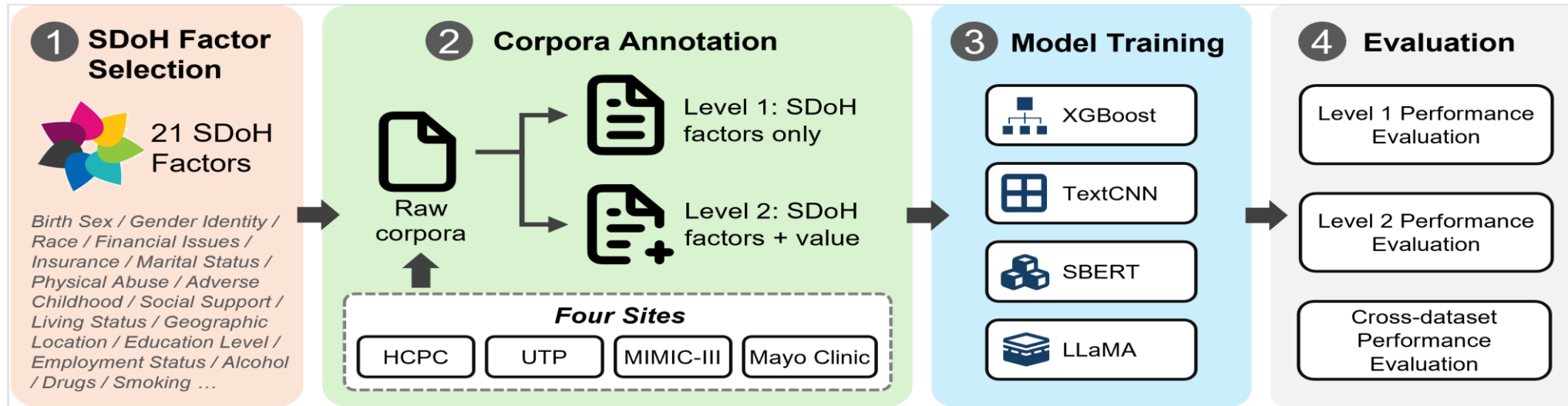


- Provide both LLaMA and BERT models for clinical information extraction
- Offer general and disease specific pipelines
- Available as a docker image for easy installation

<https://kiwi.clinicalnlp.org/>

A screenshot of the KIWI Live Demo web interface. The browser address bar shows "kiwi.clinicalnlp.org". The page has a navigation bar with "KIWI", "Live Demo", "Download", and "Support". The main heading is "Kiwi Live Demo" with a sub-heading: "An advanced medical NLP tool designed to automatically extract information from medical texts." Below this is a "Message Kiwi" section with a text input area containing a sample medical text and a "Submit" button. The "Result" section displays the same text with various entities and relationships highlighted in red and green, such as "Problem", "Treatment", "Body Location", "Temporal", "Course", and "Negation". The footer contains the copyright notice: "© Copyright Clinical NLP Lab. All Rights Reserved".

LLMs for Extracting Social Determinants of Health



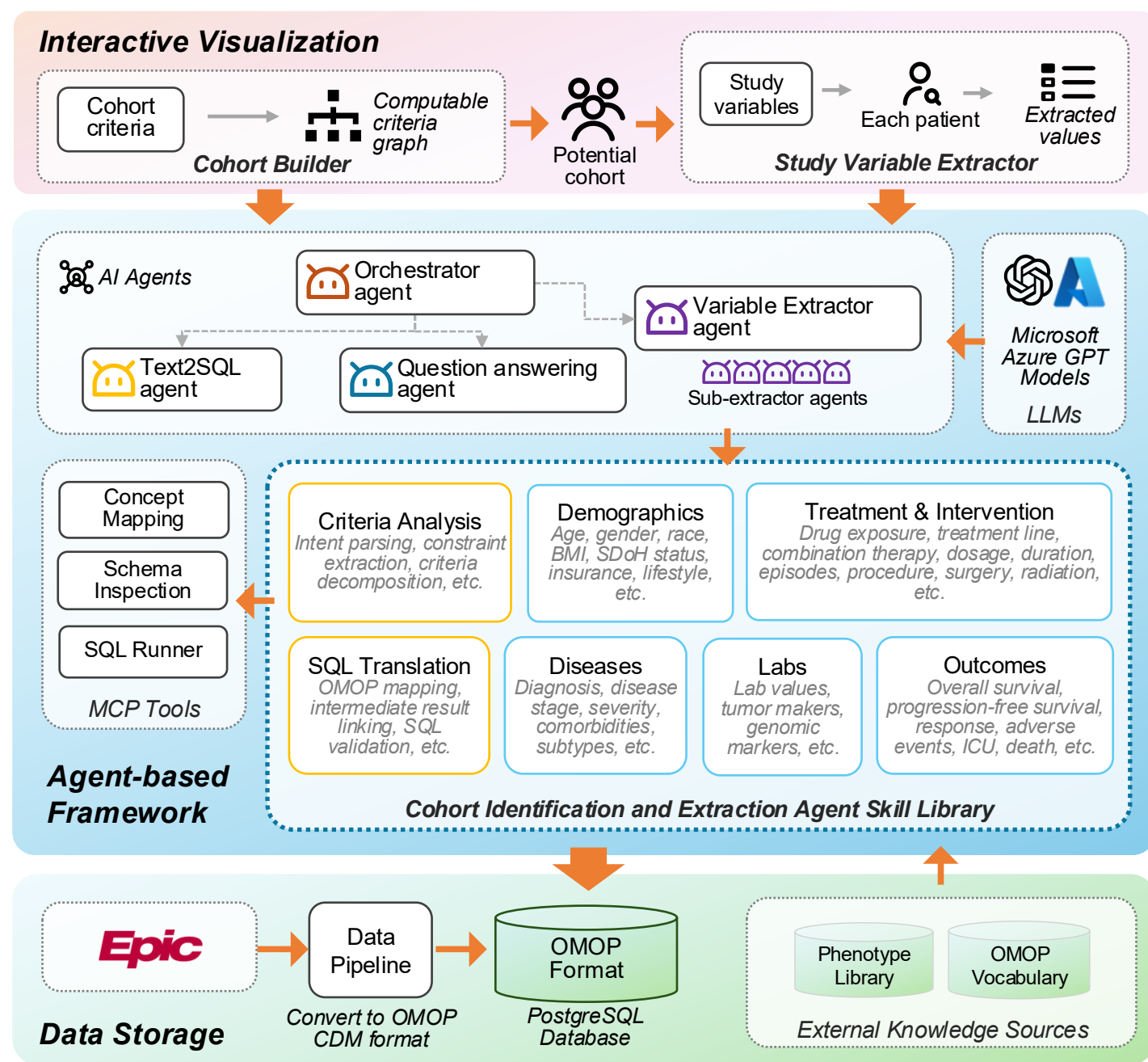
SDoH	Example
Geographic location	Pt born and <u>raised in Rio Grande, Mexico.</u> Location-Raised Location-Born
Sex, Race	The patient is an 80-year-old <u>WF</u> Race-Caucasian Sex-Female
Employment status	Pt has been <u>unemployed for past year ...</u> EmploymentUnemployed-Present
Social support	... with her peers and <u>has a good social support network</u> SocialSupport-Strong
Isolation	He <u>has been very isolative</u> and refuses to ... Isolation-Yes
Food insecurity	... said he <u>couldn't afford to eat balanced meals.</u> FoodInsecure-Yes

Dataset	XGBoost	TextCNN	Sent. BERT	LLaMA
SDoH factors only				
HCPC	0.907/0.803/0.851	0.895/0.781/0.834	0.880/0.858/0.869	0.941/0.913/ 0.927
UTP	0.982/0.935/0.958	0.980/0.927/0.952	0.979/0.948/0.963	0.990/0.979/ 0.984
MIMIC-III	0.887/0.780/0.830	0.841/0.732/0.782	0.890/0.821/0.854	0.934/0.840/ 0.883
Mayo	0.852/0.799/0.825	0.823/0.734/0.775	0.892/0.672/0.766	0.953/0.938/ 0.945
SDoH factors + values				
HCPC	0.821/0.690/0.750	0.824/0.569/0.673	0.826/0.751/0.786	0.903/0.869/ 0.886
UTP	0.946/0.880/0.912	0.889/0.815/0.850	0.957/0.882/0.918	0.982/0.932/ 0.956
MIMIC-III	0.802/0.649/0.717	0.737/0.430/0.543	0.805/0.674/0.734	0.877/0.801/ 0.837
Mayo	0.795/0.711/0.750	0.770/0.572/0.656	0.878/0.629/0.732	0.935/0.901/ 0.918

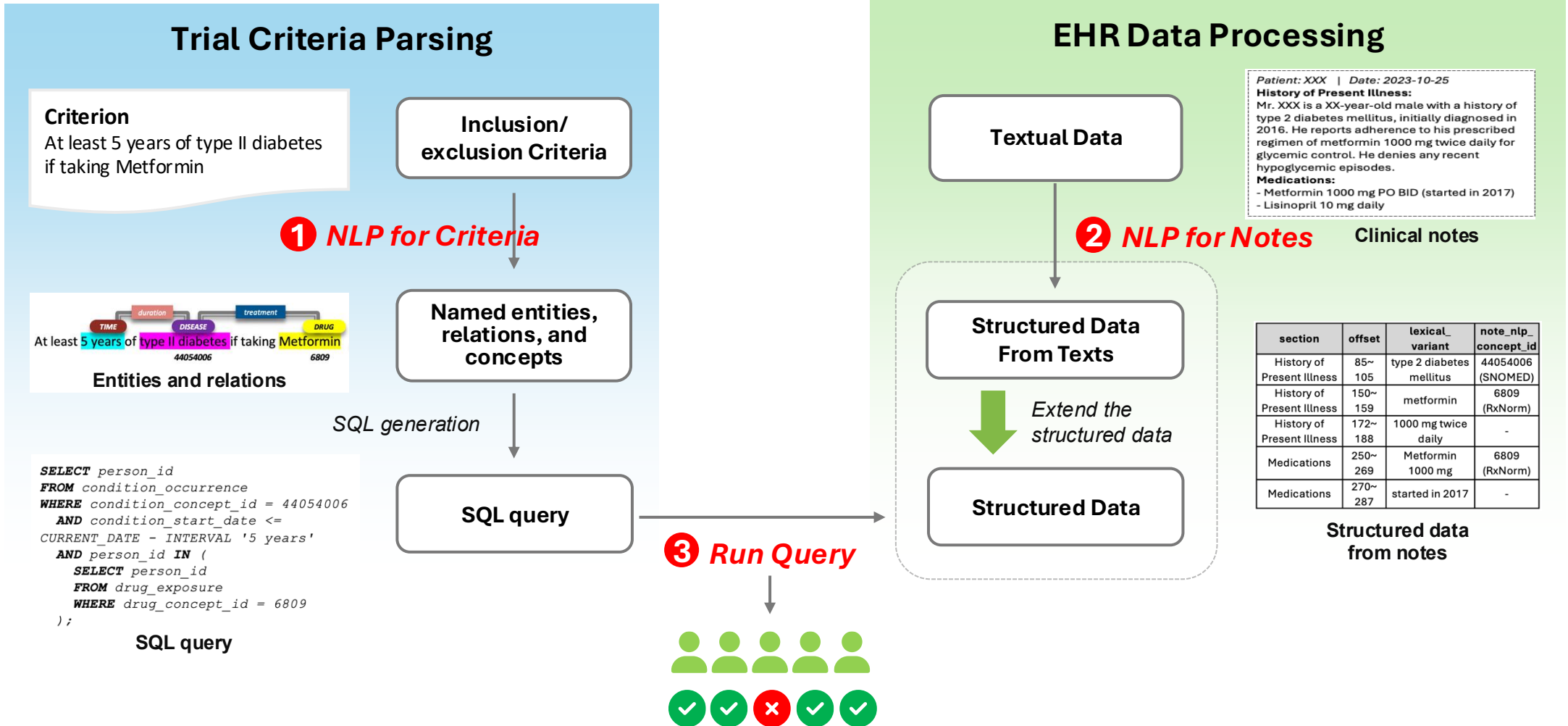
Keloth, V.K., Selek, S., Chen, Q. et al. Social determinants of health extraction from clinical notes across institutions using large language models. npj Digit. Med. 8, 287 (2025)

ECHO

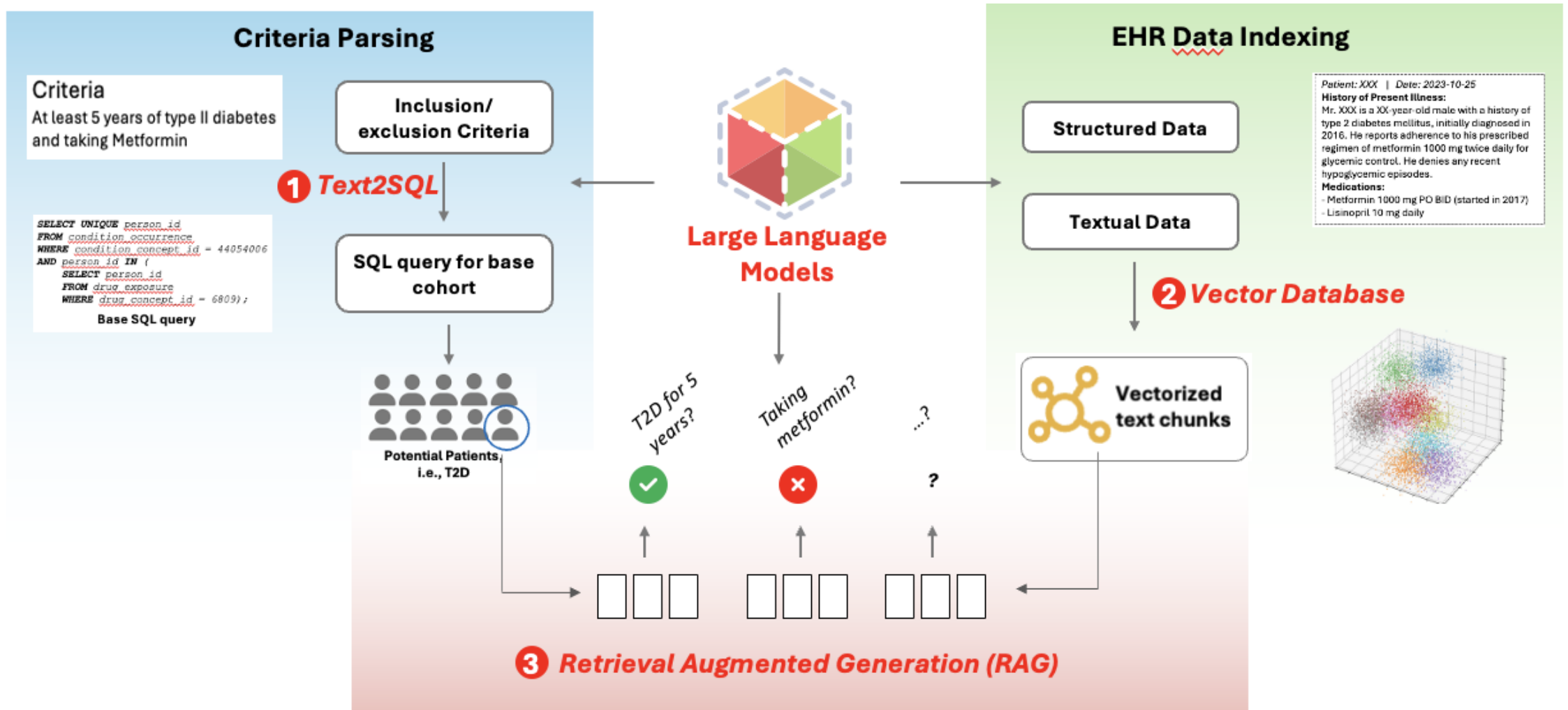
- Decompose complex tasks into reusable components
- Parallel multiple agents
- Skill library
 - Cohort Identification Skills:
 - Criteria Analysis
 - SQL Translation
 - Study Variable Extraction Skills:
 - Demographics
 - Diseases / Phenotypes
 - Treatment & Intervention
 - Labs & Biomarkers
 - Outcomes



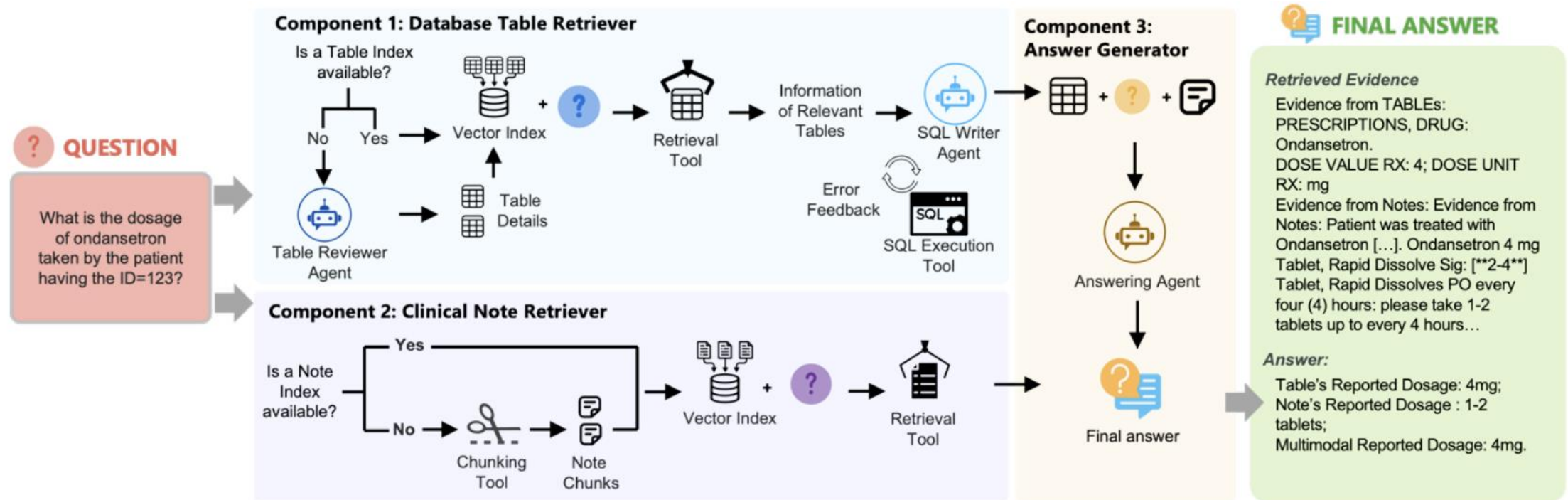
Current NLP-based Phenotyping Approaches



A New Architecture for LLM-based EHR Phenotyping



Approach Overview



Framework

- (1) Structured data retrieval – text2sql
- (2) Unstructured notes retrieval – vector database
- (3) Answer generation by LLM

Results

Setting	Backbone Model	Accuracy
MultimodalQA	ClinicalBERT	0.6108
LLM only	GPT-4o	0.3993
	Llama3-70B-Instruct	0.1021
	Llama3.1-70B-Instruct	0.3686
RAG-EHR (w/o training the answer generator)	GPT-4o	0.5104
	Llama3-70B-Instruct	0.3378
	Llama3.1-70B-Instruct	0.4702
RAG-EHR (fine-tuned answer generator)	Llama3-70B-Instruct	0.7978
	Llama3.1-70B-Instruct	0.8132

Dataset	Structured QA	Unstructured QA	Multimodal QA
DrugEHRQA	5,559	5,693	27,795

Bardhan, Jayetri, et al. "Drugehrqa: A question answering dataset on structured and unstructured electronic health records for medicine related queries." *arXiv preprint arXiv:2205.01290* (2022).

Echo – Cohort Definition

Echo | Adult ICU Patients with Major Depressive Disorder

Emily Johnson | [Logout](#)

Cohort Definition >

Refresh Add Criteria

Criteria Type

Inclusion

Definition

ICU patient

Cancel

Refresh Add Criteria + AND ↔ OR ⌵ At Least Target Cohort Run All

Console

Info | **Result** | **logs**

Short Name:
Only the first ICU stay f...

Definition:
Only the first ICU stay for each patient

SQL : SQL generate

```

1 WITH all_icu AS (
2   -- ICU stays > 24
3   hours from
4   visit_occurrence level
5   SELECT vo.person_id,
6         vo.visit_occurrence_id,
7         vo.visit_start_date
8 )
9 
```

Save
Execute SQL

```

SELECT DISTINCT
  p.person_id
FROM
  omop.person p
WHERE
  p.person_id IN (
    SELECT
      person_id
    FROM
      omop.visit_occurrence
    WHERE
      visit_concept_id =
    UNION
    SELECT
      person_id
    FROM
      omop.visit_detail
    WHERE
      visit_detail_concept_id =
  )
  AND p.person_id IN (
    SELECT
      co.person_id
    FROM
      omop.condition_occurrence co
    JOIN omop.concept_name cna
    ON cna.concept_id = co.condition_concept_id
    WHERE
      cna.ancestor_concept_id =
  )
    
```

Version: 1.0.0

Echo – Study Variable Definition

Echo | Adult ICU Patients with Major Depressive Disorder

Cohort Definition > **Study Variables** > Chart Review

Study Variables

Demographics – Baseline demographic and communication-related characteristics of the patient.

+ Add Study Variable + Add Category

+ Add

Short Name

Insurance type

Definition

Insurance coverage or payer category documented at the index date

Cancel Create

```
---
name: insurance-type-at-index-date
description: >
  Use this skill whenever you need to determine, assign, classify, or validate a patient's
  (or member's) insurance type as of a specific index date. Triggers include: any task
  involving "insurance at index date", "coverage type at enrollment", "payer type at
  diagnosis date", "insurance status at baseline", "primary payer at index", or any
  request to look up, derive, or categorize health insurance (commercial, Medicare,
  Medicaid, dual-eligible, uninsured, etc.) relative to a point-in-time reference date.
---
```

Insurance Type at Index Date

This skill guides the classification of a patient's (or member's) insurance type – the primary payer category in effect on or around a specified **index date**.

Key Concepts

Term	Definition
---	---
Index date	The anchor date for a patient – often first diagnosis, first prescription, cohort entry, or study enrollment date.
Insurance type	The payer category in effect at the index date (e.g., Commercial, Medicare FFS, Medicare Advantage, Medicaid, Dual-Eligible, Uninsured/Self-Pay).
Eligibility span	A row in an enrollment/eligibility table with a start date and end date representing a continuous period of coverage under a given plan or payer.
Look-back / look-forward window	A tolerance window around the index date (e.g., ±30 days) used when no coverage record exactly overlaps the index date. Always confirm with the user whether a window is appropriate.

Agent Skills

Echo – Variable List

- Define categories and variables
- Execute to generate data matrix

Echo | Adult ICU Patients with Major Depressive Disorder | Emily Johnson | Logout

≡ Cohort Definition > **Study Variables** > Chart Review

Study Variables(8)

- Demographics** – Baseline demographic and communication-related characteristics of the patient.
 - Insurance type** – Insurance coverage or payer category documented at the index date.
 - Marital status** – Marital or partnership status documented in the medical record.
 - Primary language** – Preferred or primary language used for clinical communication.
 - + Add Study Variable + Add Category
- Depression Assessment** – Standardized depression severity assessments documented in the clinical record.
 - HAMD-17** – HAMD-17 score or assessment documenting clinician-rated depression severity.
 - PHQ-9** – PHQ-9 score or assessment documenting depressive symptom severity.
 - + Add Study Variable + Add Category
- Social Determinants of Health** – Social and economic factors that may affect patient stability, support, and access to care.
 - Economic status** – Financial hardship, limited resources, or difficulty affording healthcare or daily needs.
 - Housing instability / homelessness** – Housing insecurity, homelessness, shelter use, or lack of stable residence.
 - Social support** – Availability of family, caregiver, or community support.
 - + Add Study Variable + Add Category

Extraction Overview

Statistics

- 8** Study Variables
- 347** Patients

Extract All Data
Extract 8 study variables in 347 patients.

[▶ Extract All](#)

Version: 1.0.0 [GitHub](#)

Echo – Chart Review



Echo

Adult ICU Patients with Major Depressive Disorder

Emily Johnson | Logout

☰ Cohort Definition > 🔍 Study Variables > **Chart Review**

#	Patient	Insurance ty...	Marital stat...	Primary lang...	HAMD-17	PHQ-9	Econ...
1	9007412000	Medicare	SINGLE	English	23 (Very severe)	7 (Mild)	Stable
2	9007412037	Medicare	SINGLE	English	7 (No depression)	15 (Moderately severe)	Financ
3	9007412074	Private	SEPARATED	English	22 (Severe)	14 (Moderate)	Stable
4	9007412111	Medicaid	DIVORCED	English	19 (Severe)	20 (Severe)	Stable
5	9007412148	Medicaid	MARRIED	Spanish	21 (Severe)	17 (Moderately severe)	Stable
6	9007412185	Private	MARRIED	Russian	20 (Severe)	13 (Moderate)	Stable
7	9007412222	Medicare Advantage	WIDOWED	English	24 (Very severe)	22 (Severe)	Stable
8	9007412259	Medicare Advantage	MARRIED	English	21 (Severe)	19 (Moderately severe)	Financ
9	9007412296	Self-pay	MARRIED	English	8 (Mild)	15 (Moderately severe)	Stable
10	9007412333	Medicare	MARRIED	Russian	4 (No depression)	24 (Severe)	Stable
11	9007412370	Dual-eligible	SINGLE	English	4 (No depression)	20 (Severe)	Unknot
12	9007412407	Dual-eligible	DIVORCED	Mandarin	3 (No depression)	16 (Moderately severe)	Unknot
13	9007412444	Medicare	SINGLE	English	5 (No depression)	20 (Severe)	Stable
14	9007412481	Medicare	MARRIED	English	24 (Very severe)	22 (Severe)	Financ
15	9007412518	Self-pay	SINGLE	English	6 (No depression)	11 (Moderate)	Financ
16	9007412555	Medicare Advantage	SEPARATED	Spanish	23 (Very severe)	18 (Moderately severe)	Financ
17	9007412592	Medicaid	MARRIED	Russian	18 (Moderate)	22 (Severe)	Stable

« < 1 2 3 4 5 > » 50 1-50 of 347

Patient 9007412000 variables extracted — Collapse All

- Demographics** (3 variables)
 - 1 INSURANCE TYPE
Medicare Evidence
 - 2 MARITAL STATUS
SINGLE Evidence
 - 3 PRIMARY LANGUAGE
English Evidence
- Depression Assessment** (2 variables)
 - 4 HAMD-17
23 (Very severe) Evidence
 - 5 PHQ-9
7 (Mild) Evidence
- Social Determinants of Health** (3 variables)
 - 6 ECONOMIC STATUS
Stable Evidence
 - 7 HOUSING INSTABILITY / HOMELESSNESS
Unknown Evidence

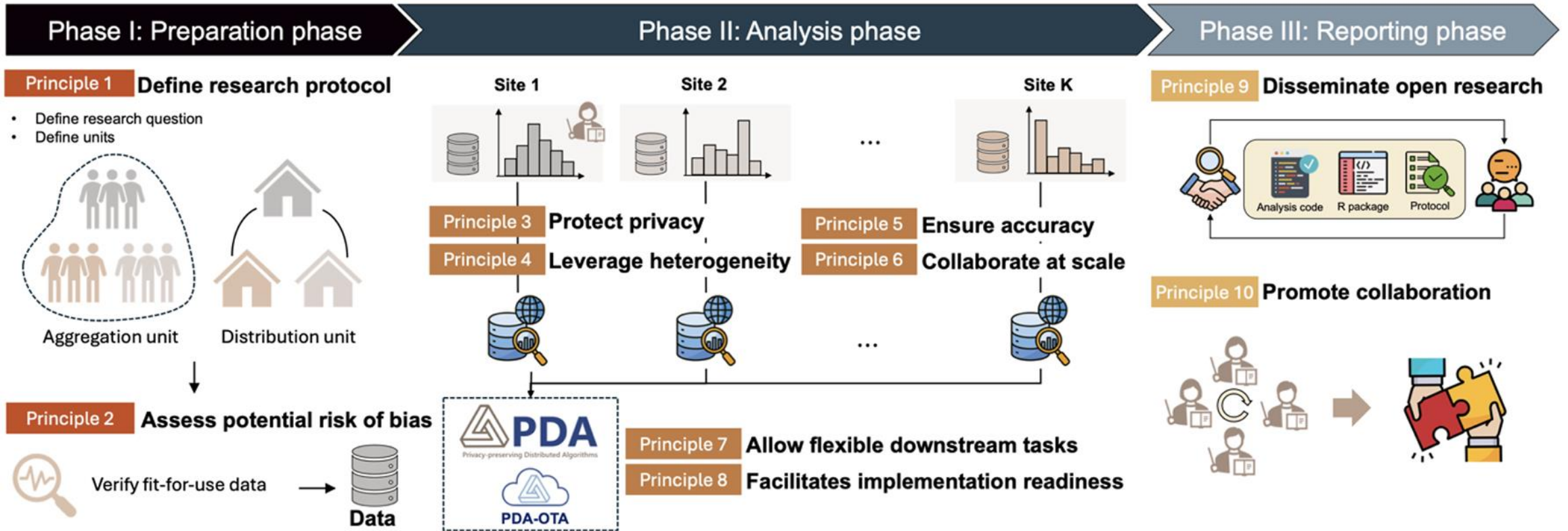
Patient 9007412000 variables extracted — Collapse All

- Demographics** (3 variables)
 - 1 INSURANCE TYPE
Medicare Evidence
Structured data consistently indicates 'Medicare' under 'Insurance' across admission records. source note_ids: ['235710168256366211', '891570543975374130', '547447919694382049']
 - 2 MARITAL STATUS
SINGLE Evidence
Marital status recorded as 'SINGLE' in structured demographics, with no conflicting documentation in clinical notes. source note_ids: ['3789219599995140']
 - 3 PRIMARY LANGUAGE
English Evidence
Preferred language documented as 'English' in the observation table, consistent across multiple encounters. source note_ids: ['280288767787985386', '936148143523993305']

Version: 1.0.0



Federated Analysis of Multi-Site EHR Data - PDA





Yale SCHOOL OF MEDICINE

Biomedical Informatics and Data Science

Thank you! Questions?

hua.xu@yale.edu