Childhood Cancer Data Initiative Webinar Series

Unlock the Power of Unstructured Data with Electronic Medical Record Search Engine (EMERSE)

David Hanauer, M.D., M.S.

NIH NATIONAL CANCER INSTITUTE

January 14, 2025

Agenda

1. *Background*
2. *EMERSE Solution*
3. *EMERSE Value*
4. *EMERSE Capabilities and Features*
5. *Getting EMERSE*
6. *CCDI Example*
7. *Q&A*

Today's Speaker

David Hanauer, M.D., M.S.

Associate Professor, University of Michigan

Background

EMERSE is for Free Text Data

EMERSE is for this…	**…not this**	
Unstructured Data (free text)	*Structured Data*	
Mrs. Jones is a 56-year-old female with a history of HTN, hypercholesterolemia, and T2DM who comes to the clinic today with a 3-day h/o dizziness and a severe headache on the left side.	WBC: Cholesterol: Weight: AST: ALT:	5.6 182 67.4 30 52

Free Text Notes Are Important for Research

- 80% of electronic health record (EHR) data are unstructured free text
- Numerous studies have shown the importance of free text for accurate and complete data extraction
- Most medical centers lack tools for free text, especially tools that are used by non-technical experts

EMERSE Solution

The EMERSE Solution

- A system “for the people”
- Users search the EHR notes on their own
- Don’t need to wait in a queue for an analyst or a data scientist
- Easy-to-use for non-technical users
- Unlike with some EHRs, EMERSE can search across all notes and all patients at once
- Continuous refinements for 19 years

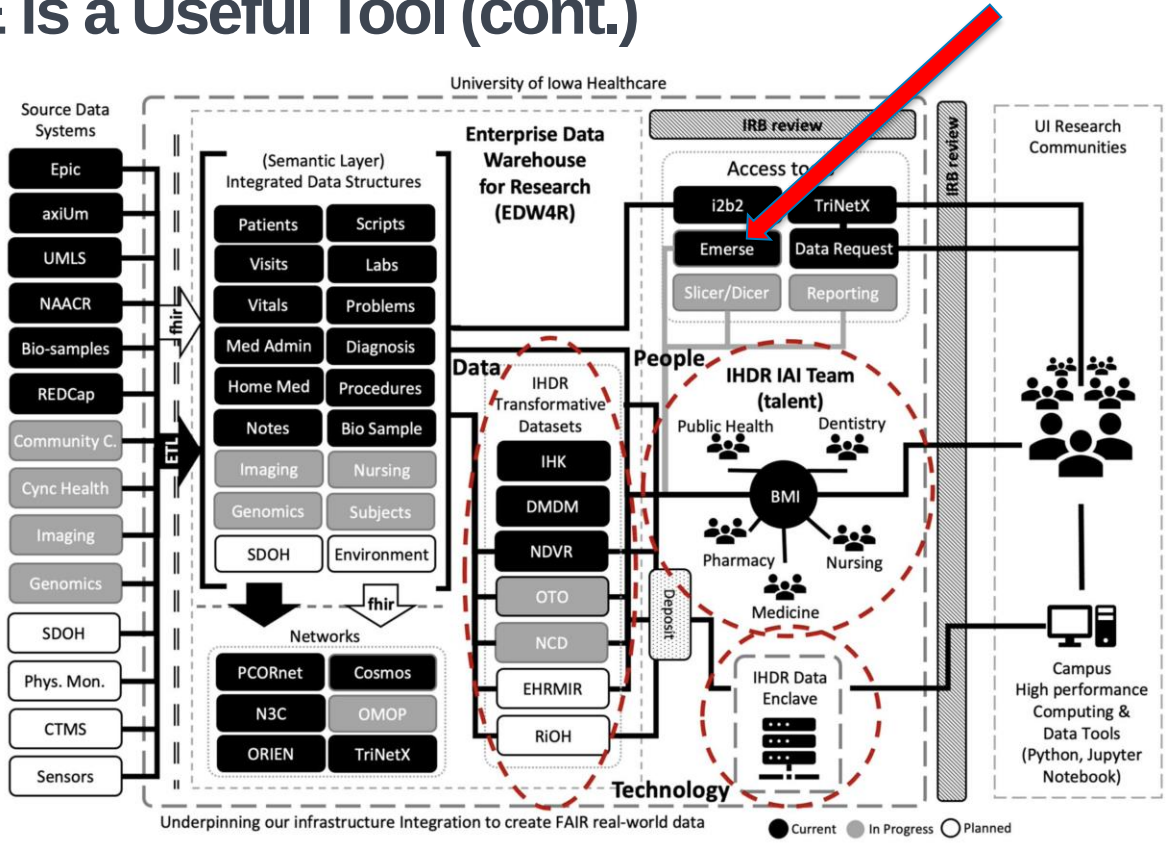
The EMERSE Solution (cont.)

- A system “for IT teams”
- Enterprise grade, easy to support
- Configurable with granular roles/privileges to control access (e.g., all patients vs. some patients)
- Data are kept secure within a centralized, audited system at your own site
- No need to download/store the data elsewhere

EMERSE Is a Useful Tool

EMERSE will likely be one of many tools you need to build and support your project(s)

EMERSE is a Useful Tool (cont.)



https://pmc.ncbi.nlm.nih.gov/articles/PMC10873835/

EMERSE Value

We Think EMERSE Is Really Good

- **We think it is better** than other tools that cost money
- **We've been laser-focused** on just this one feature: *search*
- **We've studied it:** "New users of the EMERSE system are...highly satisfied with the interface and have highly positive perceptions of its expected utility in their work"
- **Our users tell us:** "EMERSE is an absolute gem. Thank you to the team that created this powerful research tool."

We Think EMERSE Is Really Good (cont.)

- **Publications mention it:** “The tool avoids the pitfalls of diagnostic inaccuracy seen with tools querying on ICD and billing codes...” [PMID 36114099]
- **Top-tier medical centers use it:** University of Michigan, Harvard University – Dana Farber Cancer Center, Columbia University Cancer Center, University of North Carolina – Chapel Hill, University of California – San Francisco, and more

EMERSE Capabilities and Features

EMERSE Can Support Multiple Research Activities

- **Finding cohorts based on what is mentioned in notes:** disease, drugs, symptoms, adverse events, etc.
- **Search through existing patient lists:** just copy-paste in a list of medical record numbers (MRNs) and start searching
- **Highlighting search terms in clinical notes** to support chart reviews

EMERSE Supports Research

- ~700 peer-reviewed papers and abstracts
  - Full list: https://project-emerse.org/publications.html

EMERSE Is Highly Refined

- **It's fast:** up to 7,000 times faster than searching text in a traditional database
- **It provides many query expansion options:** cancer → tumor, neoplasm, carcinoma, malignancy, etc.
- **It provides multiple management features** such as saved patient lists, saved search terms, multiple filtering options, sharing lists and terms among team members, and more

Natural Language Processing (NLP) Capabilities

- NLP is built in
  - Named entity recognition/mapping to the Unified Medical Language Systems (UMLS) concept unique identifiers (CUIs)
  - Negation
  - Subject vs. non-subject
  - Uncertainty
  - History of
- Sites can incorporate their own NLP annotations

Getting EMERSE

Cost

- Software $0.00
- Installation Guidance $0.00
- Access to Documentation $0.00
- Training (within reason) $0.00
- Synonyms dataset (optional) $0.00*

**for research use within EMERSE*

The Fine Print

- Sites must install the software and operate it themselves
- Installation can be local or cloud
- Site pay their own local/cloud installation costs
- Our team does not have access to anyone else’s data

EMERSE Team

- Contact us to schedule a time with your team for:
  - Discussions about usage strategies
  - Training
  - Live demonstrations (for abstractors, IT teams, etc.)
  - Technical/implementation details
- Email: **EMERSE-team@umich.edu**

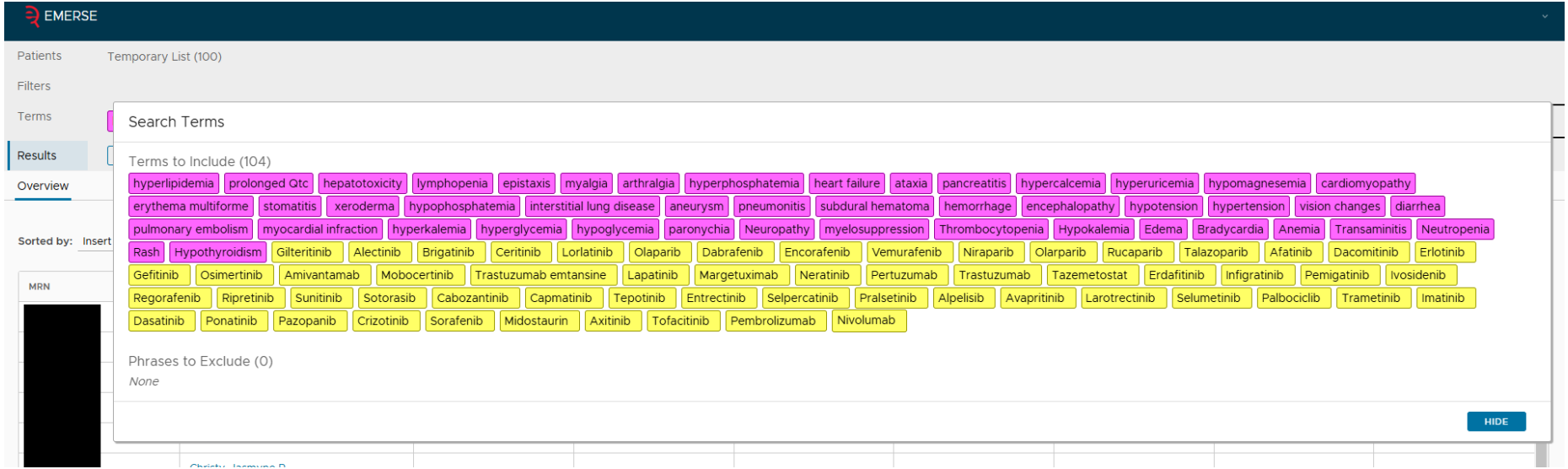
EMERSE Community Meeting

- Online via Zoom; open to everyone
- Next meeting: **Monday, February 10, 2025, 1:00–2:00 p.m. ET**
- Register at: https://bit.ly/emerse-community-feb-2025

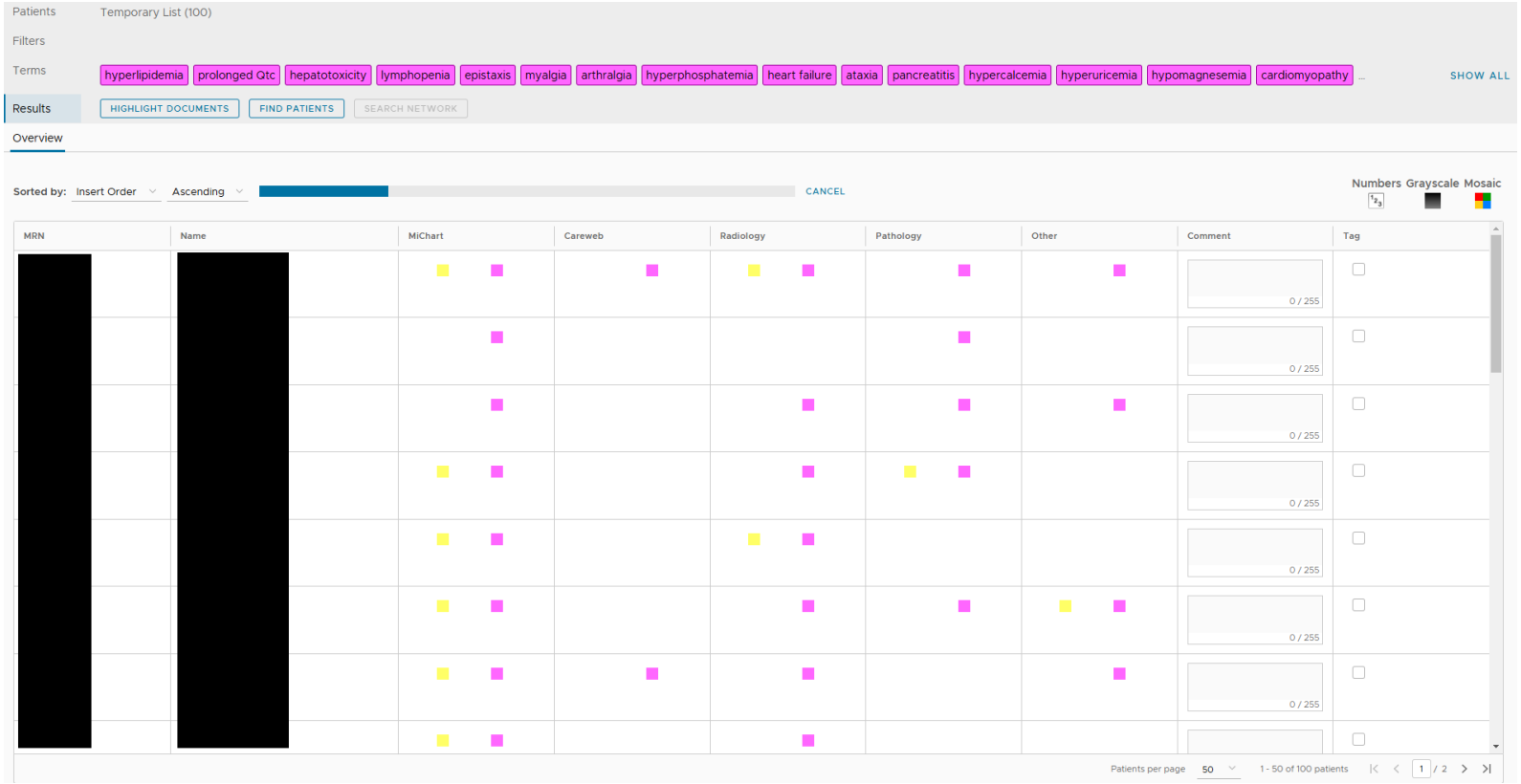
CCDI Example

CCDI Program Example

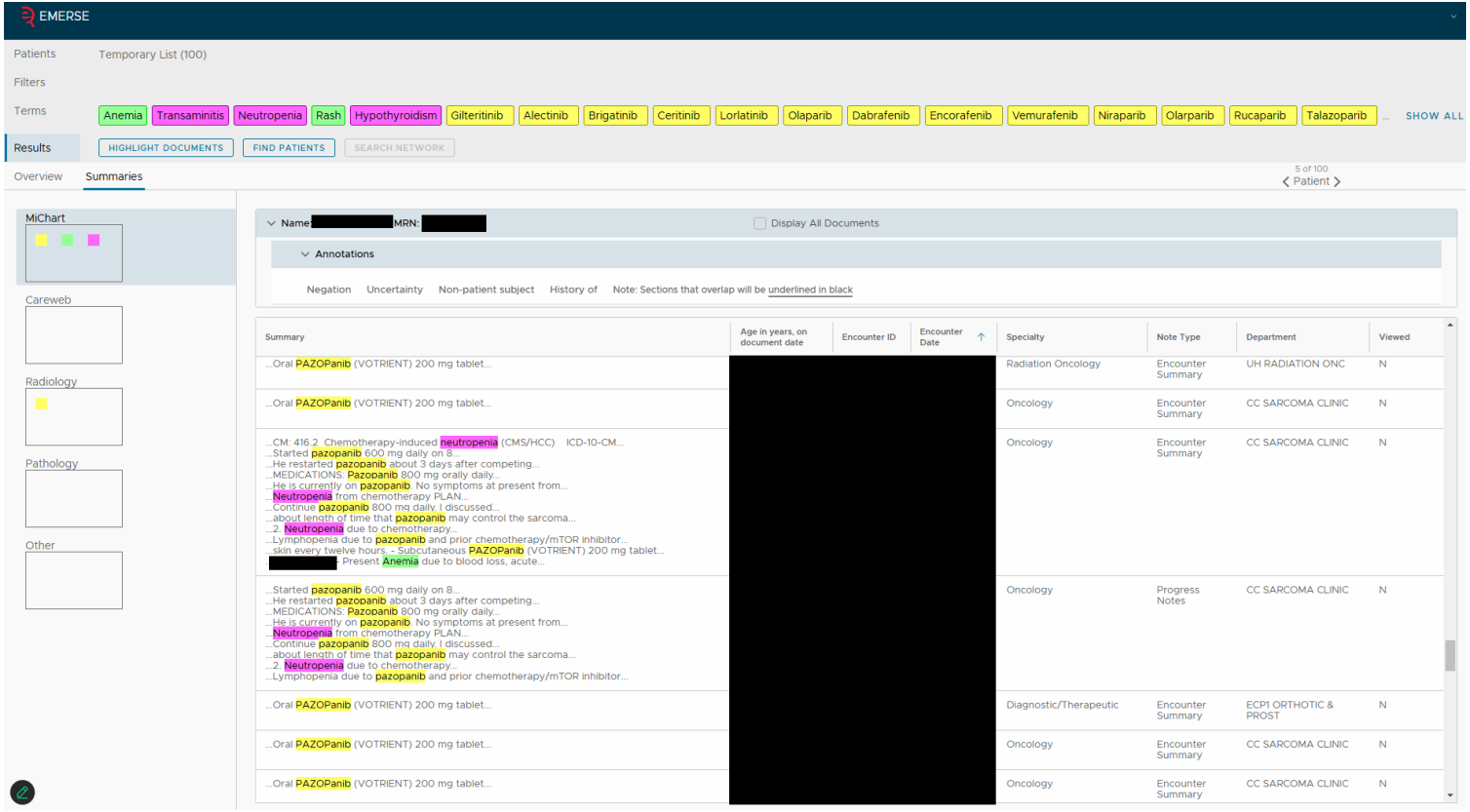
- Search for multiples drugs and toxicities



CCDI Program Example



CCDI Program Example



Live Demo

- No real names
- No PHI
- Publicly available
- “Documents” are abstracts and case reports

Q&A

How You Can Engage with CCDI

Learn about CCDI and subscribe to our monthly newsletter:
cancer.gov/CCDI

Access CCDI data and resources:
ccdi.cancer.gov

Questions? Email us at:
NCIChildhoodCancerDataInitiative@mail.nih.gov

Thank you for attending!

cancer.gov

cancer.gov/espanol