

# EMERSE Community Meeting

em<sup>ER</sup>rse

2024-Feb-06

<https://project-emerse.org>

# Plan for today

- Welcome and Housekeeping
- Announcements and Updates
- How EMERSE is supporting and enhancing population health cancer research: Nickolas Stabellini
- Discussion: EMERSE NCI U24 Grant
- Open Forum
- Adjourn

# Housekeeping

- *Zoom Meeting*
  - More opportunities for interaction
- Please stay muted, unless you would like to ask a question or make a comment
- Feel free to use the chat function to type questions or provide comments
- We will answer questions throughout
- We will record this meeting, available where prior recordings are located: <https://project-emerse.org/presentations.html>

# A Growing Community

- Registrants from 22 organizations and institutions for today's meeting



# Acknowledgements

- EMERSE has been supported by:
  - NCI ITCR program
  - NCATS/CTSA/MICHR
  - Michigan Medicine
    - Department of Learning Health Sciences
    - Office of Research
    - Health Information Technology & Services

# Updates and Announcements

- EMERSE Research Collaborative
  - Initial research focus: pronoun usage in the EHR
  - Awaiting IRB review (exemption expected)
  - If interested in participating, please contact [EMERSE-team@umich.edu](mailto:EMERSE-team@umich.edu)
  - Next meeting: TBD

# Updates and Announcements

- EMERSE Release, latest version 6.5.2
  - 6.5 has filters
  - 6.5.2. has small change for SQL Server efficiency
- Versions 6.6 (coming soon, maybe next week)
  - Saved/shareable filters
  - Filter restrictions (admin feature)
    - *Kellen McClain to discuss*

# Updates and Announcements

- NLP (integrating NLP into search)
  - Work still ongoing
  - Additional refinements being made
    - Incorporating feedback from some sites
  - Incorporating local/custom NLP (~~BYON model: Bring Your Own NLP~~)
    - **INJECT** → Incorporating NLP Just got Easier to Customize and Tailor
  - Still need to talk to NLM about licensing issues



# Publications

**553**

Papers

**92**

Abstracts

**18**

Other

# Publications – “The largest”

- “...this study represents **one of the largest reported case series** of pts with MBC [metastatic breast cancer] and LMD [leptomeningeal disease]...” [Huppert]
- “...our study provides **the most comprehensive evidence**, to date, that DM [diabetes mellitus] is an important independent marker for amputation risk.” [Schmidt, PMID 38164707]
- “To date, this is the **largest reported series of TEC** [Toxic Erythema of Chemotherapy], a difficult diagnosis given the heterogeneity of clinical presentations and histologic overlap with GVHD and other drug eruptions.” [Cole, PMID 38133527]

# Guest speaker: Nickolas Stabellini

**Stabellini N**, Nazha A, Agrawal N, Huhn M, Shanahan J, Hamerschlak N, Waite K, Barnholtz-Sloan JS, Montero AJ. Thirty-Day Unplanned Hospital Readmissions in Patients With Cancer and the Impact of Social Determinants of Health: A Machine Learning Approach. *JCO Clin Cancer Inform.* 2023 Jul;7:e2200143. doi: 10.1200/CCI.22.00143. PMID: 37463363

**Stabellini N**, Cullen J, Moore JX, Dent S, Sutton AL, Shanahan J, Montero AJ, Guha A. Social Determinants of Health Data Improve the Prediction of Cardiac Outcomes in Females with Breast Cancer. *Cancers (Basel).* 2023 Sep 19;15(18):4630. doi: 10.3390/cancers15184630. PMID: 37760599

**Stabellini N**, Dmukauskas M, Bittencourt MS, Cullen J, Barda AJ, Moore JX, Dent S, Abdel-Qadir H, Kawatkar AA, Pandey A, Shanahan J, Barnholtz-Sloan JS, Waite KA, Montero AJ, Guha A. Social Determinants of Health and Racial Disparities in Cardiac Events in Breast Cancer. *J Natl Compr Canc Netw.* 2023 Jul;21(7):705-714.e17. doi: 10.6004/jnccn.2023.7023. PMID: 37433439

**Stabellini N**, Cullen J, Bittencourt MS, Moore JX, Cao L, Weintraub NL, Harris RA, Wang X, Datta B, Coughlin SS, Garcia J, Shanahan J, Hamerschlak N, Waite K, Fillmore NR, Terris M, Montero AJ, Barnholtz-Sloan JS, Guha A. Allostatic load and cardiovascular outcomes in males with prostate cancer. *JNCI Cancer Spectr.* 2023 Feb 8:pkad005. doi: 10.1093/jncics/pkad005. Epub ahead of print. PMID: 36752520

**Stabellini N**, Cullen J, Cao L, Shanahan J, Hamerschlak N, Waite K, Barnholtz-Sloan JS, Montero AJ. Racial disparities in breast cancer treatment patterns and treatment related adverse events. *Sci Rep.* 2023 Jan 22;13(1):1233. doi: 10.1038/s41598-023-27578-4. PMID: 36683066

**Stabellini N**, Tomlinson B, Cullen J, Shanahan J, Waite K, Montero AJ, Barnholtz-Sloan JS, Hamerschlak N. Sex differences in adults with acute myeloid leukemia and the impact of sex on overall survival. *Cancer Med.* 2022 Nov 23. doi: 10.1002/cam4.5461. Epub ahead of print. PMID: 36419242

# How EMERSE is supporting and enhancing population health cancer research

## **Nickolas Stabellini**

MD/Ph.D. Candidate, Case Western Reserve University

Research collaborator, University Hospitals Cleveland Medical Center, Augusta University, and Hospital Israelita Albert Einstein





# How EMERSE is supporting and enhancing population health cancer research



University  
Hospitals

## Nickolas Stabellini

MD/PhD candidate, Case Western Reserve University School of Medicine & Hospital Israelita Albert Einstein  
Research collaborator, Department of Hematology-Oncology, University Hospitals Cleveland Medical Center  
Research collaborator, Department of Cardiology, Augusta University

# The focus of our research group



## Population Health in Oncology



Outcomes

Predictors

Patterns

# The focus of our research group

HEALTH DISPARITIES

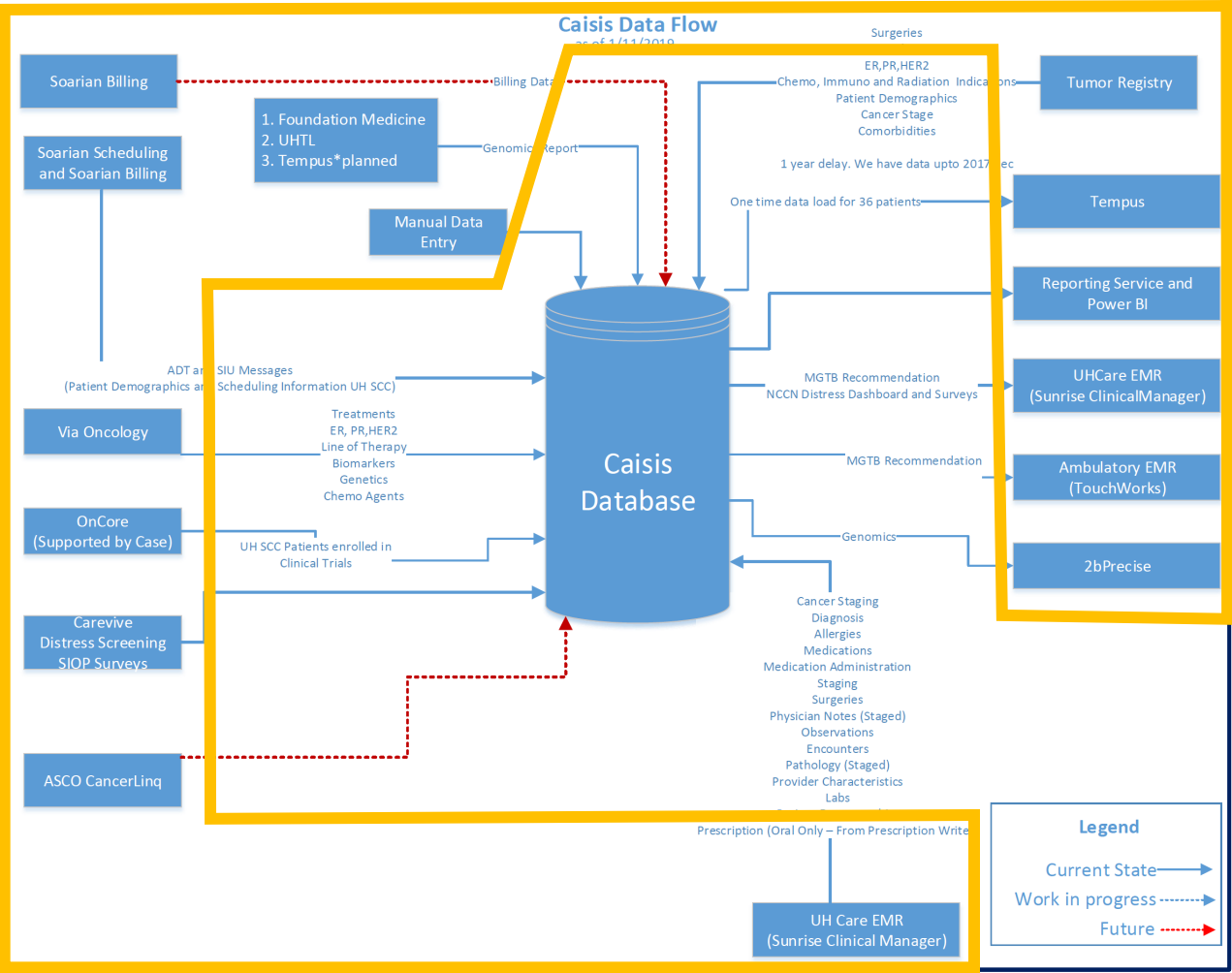
Sex differences

Racial disparities



Source: CDC

# RESTROSPECTIVE DATA ("BIGDATA")



LexisNexis® Health Care  
RISK SOLUTIONS

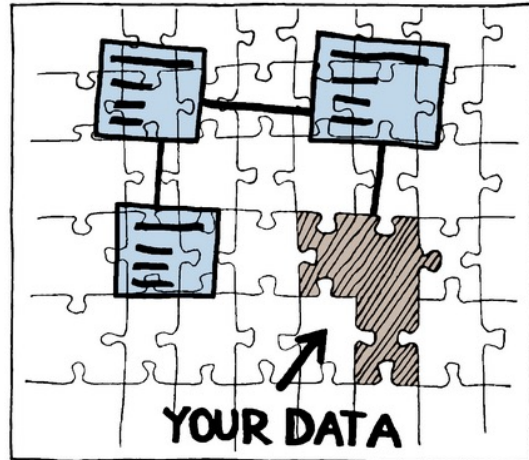
**Social Determinants of Health**

Statistical approaches

AI



# However...



No available structured field to fill the information in the EHR



FREE-TEXT FIELD



Other reasons

# Free-text field problem



**Marcio S Bittencourt** @MBittencourtMD · Jul 2, 2021

ASRAZENECA; ASREAZENECA; ASROGENICA; ASRTAZENECA;  
ASRTRAZENECA; ASSSTRAZENECA; ASTARAZENECA; ASTAZENECA;  
ASTREAZANECA; ASTREAZENECA; ASTREZENACA; ASTREZENECA;  
ASTREZENECAR; ASTROGENICA; ASTROZEMICA;  
ASTROZENCA; ASTROZENECA; ASTRSZENECA; ASTRZANECA; ASTRZEN;  
ASTRZENECA; ASTRZENIA

2 13 187



**Marcio S Bittencourt** @MBittencourtMD · Jul 2, 2021

ASTTRAEN; ASTUAZENECA; ATRASZENECA; ATRAZANECA;  
ATRAZENCA; ATRAZENECA; AXTRAZENECA; AZ; AZASGTRENICA;  
AZRAZENECA; AZTN; AZTREZENECA.

3 10 185

Stage

AML prognosis

Gleason prostate cancer

Prostate cancer histology



FREE-TEXT FIELD

TERMS

LIST OF PATIENTS

STRUCTURED INFO



## Search Option: Wildcards

Wildcards help with finding concepts where you may not be able to specify the exact details/phrasing/spelling of the terms.

For example, sometimes you may want to look for a term where the ending of the word may be variable. Suppose you want to look for terms related to **hypertension**. If you just search for **hypertension** you may miss other variations. Using the wildcard symbol, the asterisk (\*), you can overcome this problem. Simply type in something like **hyperten\*** and you will be able to match terms such as **hypertension**, **hypertensive**, **hypertensives**, etc. This would also match potential misspellings such as **hypertenion**.

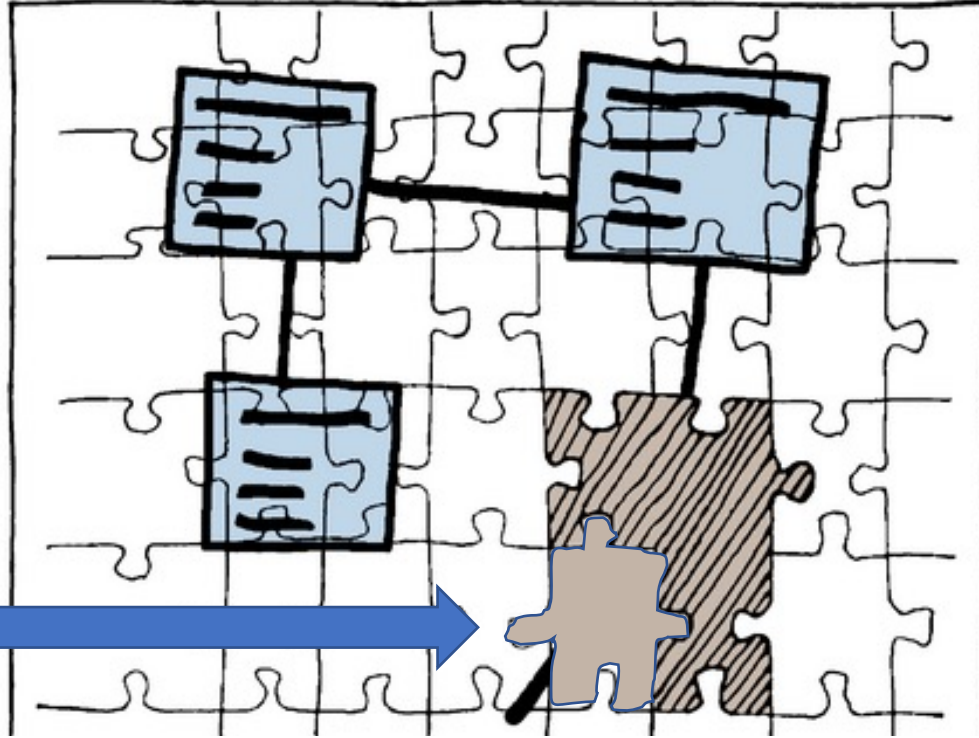
In addition to adding a wildcard to the end of a word, you can also add it to the middle of a word. If you search for **hyper\*ia** you will match terms such as **hypercapnia**, **hypernatremia**, and others. In this example, the search term with the embedded wildcard will look for words that start with **hyper** and end with **ia**.

## Synonyms

The *Synonyms* feature in EMERSE can help you conduct more thorough searches by providing suggestions for additional terms or phrases that are related to the terms you entered. This is sometimes referred to as "query expansion". An important aspect of the *Synonyms* is that you have full control over what is included in your search. *Synonyms* are essentially system-provided suggestions about additional words or phrases that you may want to include in your searches, based on the words/phrases you have already entered. You can incorporate the suggestions into your own search terms, either with the *Quick Terms* feature or the *Save Terms* feature. Because these are system-wide suggestions, these terms are available to all users in the system. The term *Synonym* is used loosely here. In fact, these can include many types of related terms that might be useful for a search.



**emrse**  
ELECTRONIC MEDICAL RECORD SEARCH ENGINE



# EMERSE

Name	Description	Owner	Created	Modified	Last Used	↓	Patient Count
INTERMEDIATE_AML	INTERMEDIATE	Nickolas Stabellini	05/22/2022	05/22/2022	02/02/2024		178
FAVORABLE_AML	FAVORABLE	Nickolas Stabellini	05/06/2022	05/22/2022	05/22/2022		589
Gleason 8	Gleason 8	Nickolas Stabellini	05/06/2022	05/06/2022	05/06/2022		988
Gleason 10	Gleason 10	Nickolas Stabellini	05/06/2022	05/06/2022	05/06/2022		785
Gleason9	Gleason9	Nickolas Stabellini	05/06/2022	05/06/2022	05/06/2022		791
Gleason 7	Gleason7	Nickolas Stabellini	05/06/2022	05/06/2022	05/06/2022		1,821
Gleason 6	Gleason 6	Nickolas Stabellini	05/06/2022	05/06/2022	05/06/2022		3,213
Gleason	Gleason	Nickolas Stabellini	05/06/2022	05/06/2022	05/06/2022		6,628
unfavorable_aml	unfavorable	Nickolas Stabellini	05/22/2022	05/22/2022			11
Prostate Adenocarcinoma	Adenocarcinoma	Nickolas Stabellini	05/06/2022	05/06/2022			5,737

## scientific reports

### OPEN Racial disparities in treatment patterns related adverse events

Nickolas Stabellini<sup>1,2,3,4</sup>, Jennifer Cullen<sup>4,5</sup>, Lifen Cao<sup>6</sup>, Nelson Hamerschlag<sup>7</sup>, Kristin Waite<sup>8</sup>, Jill S. Barnholtz-Sloan<sup>7,8</sup>

#### Methods

The study setting was the University Hospitals (UH) Seidman Cancer Center (UHSCC). All patient data were obtained from the UH data repository based on the CAISIS platform, which consists of an open-source, web-based cancer data management system composed by disparate sources of cancer patient data (ie, Soarian, NGS Labs, Sunrise Clinical Manager, Tumor Registry, Via Oncology, OnCore, MosaiQ, PRO tools, and others).<sup>25-27</sup> All patient records were de-identified, and all analyses were performed in accordance with relevant guidelines and regulations, respecting the Declaration of Helsinki. Informed consent was approved by the University Hospitals of Cleveland Institutional Review Board (IRB). All data were subsequently complemented with electronic health record (EHR) information captured via EMERSE (Electronic Medical Record Search Engine) to obtain the most accurate and complete information per patient, avoiding high missi-

#### Methods

#### Data Source

The study setting was the University Hospitals Seidman Cancer Center (UHSCC). All patient data were obtained from the UHSCC data repository on the basis of the CAISIS platform, which consists of an open-source, web-based cancer data management system that integrates disparate sources of cancer patient data (ie, Soarian, NGS Labs, Sunrise Clinical Manager, Tumor Registry, Via Oncology, OnCore, MosaiQ, PRO tools, and others).<sup>25-27</sup> All patient records were deidentified, and the study was approved by the University Hospitals (UH) of Cleveland Institutional Review Board. All the data were complemented with electronic health record (EHR) information captured via Electronic Medical Record Search Engine to obtain the most accurate and complete information per patient.<sup>28</sup>

OPEN ACCESS | ORIGINAL REPORTS | July 18, 2023



## Thirty-Day Unplanned Hospital Readmissions in Patients With Cancer and the Impact of Social Determinants of Health: A Machine Learning Approach

Authors: Nickolas Stabellini, BS, Aziz Nazha, MD, Nikita Agrawal, BA, Merilys Huhn, BA, John Shanahan, BA, Nelson Hamerschlag, MD, PhD, Kristin Waite, PhD, Jill S. Barnholtz-Sloan, PhD, and Alberto J. Montero, MD, MBA

[AUTHORS INFO & AFFILIATIONS](#)

Jill S. Barnholtz-Sloan<sup>7,8</sup>

University Hospitals (UH) data were obtained from the CAISIS platform, which consists of an open-source, web-based cancer data management system composed by disparate sources of cancer patient data (ie, Soarian, NGS Labs, Sunrise Clinical Manager, Tumor Registry, Via Oncology, OnCore, MosaiQ, PRO tools, and others).<sup>25-27</sup> All patient records were deidentified, and the study was approved by the University Hospitals (UH) of Cleveland Institutional Review Board (IRB). All data were subsequently complemented with electronic health record (EHR) information captured via EMERSE (Electronic Medical Record Search Engine) to obtain the most accurate and complete information per patient.<sup>28</sup>



# Thank you

 [Nickolas.Stabellini@UHhospitals.org](mailto:Nickolas.Stabellini@UHhospitals.org)

 @nickolastab

 Nickolas Stabellini



# Administration, Architecture and Technology

## EMERSE NCI U24 Grant

➤ 5 Years: 10/2023 – 9/2028

Table 4. Project Timeline		YEAR 1 10/1/23 – 9/30/24				YEAR 2 10/1/24 – 9/30/25				YEAR 3 10/1/25 – 9/30/26				YEAR 4 10/1/26 – 9/30/27				YEAR 5 10/1/27 – 9/30/28			
Aim	Sub-Aim	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1	1.1 Standards-compliant APIs/DeepPhe																				
	1.2 Timeline visualizations																				
	1.3 Data extraction workflow editor																				
	1.4 Optical character recognition (OCR)																				
2	2.1 Incorporate NLP as data layers																				
	2.2 Partner with 2 NLP platform experts cTAKES team: Boston Children's MedTagger team: Mayo Clinic																				
	2.3 Build generalizable NLP indexer/import City of Hope Moffitt Cancer Center University of California – San Francisco																				
3	3.1 Develop network framework with sites Columbia University University of Cincinnati University of Kentucky University of Virginia																				
	3.2 Data standards for network queries																				
	3.3 Advanced auditing of network searches																				
	3.4 Limit searches on special-status patients																				
4	4.1 Assess EMERSE user perceptions																				
	4.2 Solicit perspectives from site leads/others																				
	4.3 Impact on scientific/clinical outcomes Enhancements based on user suggestions Bug fixes, security updates Documentation/training updates; Outreach																				



# Administration, Architecture and Technology

**AIM 1.1:** Build a standards-compliant, secure application program interface (API) and test it through an integration with the ITCR-supported DeepPhe (Cancer Deep Phenotyping) software.

*Collaborators: Harry Hochheiser/DeepPhe Team @ University of Pittsburgh*

---

**AIM 1.2:** Build timeline data visualizations.

---

**AIM 1.3:** Develop a novel, easy-to-use, self-service Data Extraction Workflow (DEW) editor to help users extract data from standardized EHR note templates (e.g., synoptic reports)

---

**AIM 1.4:** Develop an optical character recognition (OCR) pipeline

# Administration, Architecture and Technology

**AIM 2.1:** Incorporate NLP as data layers - ontology-based named-entity recognition (NER), negation, uncertainty, and experimenter.

---

**AIM 2.2:** Partner with 2 NLP-platform development teams (cTAKES and MedTagger) to ensure that EMERSE can incorporate novel innovations being developed in the NLP community.

*Collaborators: Guergana Savova, Boston Children's; Hongfang Liu, The University of Texas Health Science Center at Houston*

---

**AIM 2.3:** Partner with 3 teams developing localized NLP solutions to build a generalizable NLP importer/indexer to accommodate the various technologies used by our partner institutions.

*Collaborators: City of Hope, Moffitt Cancer Center, UCSF*

# Administration, Architecture and Technology

**AIM 3.1:** Work with 4 partner sites to develop a generalizable network participation framework.

*Collaborators: Columbia, University of Cincinnati, University of KY, UVA*

---

**AIM 3.2:** Incorporate data standards to enable more granular network search filtering.

---

**AIM 3.3:** Develop advanced auditing of networked searches to ensure that queries do not attempt to re-identify individual patients

---

**AIM 3.4:** Add the ability to exclude special-status patients from local or networked searches.

# Administration, Architecture and Technology

**AIM 4.1:** Assess the perceptions of EMERSE among users.

*Collaborator: Kai Zheng, UCI*

---

**AIM 4.2:** Solicit the perspectives of EMERSE site leadership and other relevant personnel, focusing on security and regulatory issues.

*Collaborator: Kai Zheng, UCI*

---

**AIM 4.3:** Measure the impact of EMERSE on scientific and clinical outcomes based on the peer-reviewed literature and user interviews:

1. Survey of first and last authors of newly published studies that identified EMERSE in their methods section
2. Interviews of 3 individuals each from 5 sites (N≈15) who use EMERSE for clinical/quality improvement work

*Collaborator: Kai Zheng, UCI*

# Wrap up

- Open Discussion
  - Questions?
  - Comments?
  - Best Practices?
- THANK YOU
- Next Meeting: TBD (May/June 2024)
- Please complete meeting survey
  - [https://umich.qualtrics.com/jfe/form/SV\\_5amRnshITVIJEd8](https://umich.qualtrics.com/jfe/form/SV_5amRnshITVIJEd8)

