Clinical Research Informatics

Year-in-Review

Peter J. Embí, MD, MS, FACP, FACMI, FAIHSI
Chair, Department of Biomedical Informatics
Professor of Biomedical Informatics and Medicine
Senior Vice-President for Research and Innovation

VANDERBILT UNIVERSITY
MEDICAL CENTER

AMIA Informatics Summit
Seattle, WA
March 15, 2023

Disclosures

- **Federal Funding:** NIH: NLM, NCATS; PCORI; AHRQ
- **Academic Consulting:** Yale University, University of Miami, University of Florida, University of Iowa, University of Pittsburgh, University of Colorado, U. Pittsburgh, Rochester U.
- **Study Section:** AHRQ
- **Corporate Institutional Partnerships (no personal compensation):** IBM/Merative, GE Healthcare

Approach to this presentation

- Mixed approach to article identification:
  - Started with structured approach
    - (akin to ACP “update” sessions)
  - Solicited recommendations from colleagues
  - Augment with “what seemed interesting” approach
- Learned a lot from doing this last 12 years
  - Tracked manuscripts throughout the year
  - Intended to spread work out…

    …still worked down to the wire

- So, what was my approach…

Source of Content for Session

- Literature review:
  - Initial search by MESH terms and keywords, including:
    - "Biomedical Research"[Mesh] AND "Informatics"[Mesh]
    - NOT ("computational biology"[mesh] OR "genetics"[mesh])
    - Initial searches resulted in **903** articles
    - Further limits: **648**
  - Additional articles found via:
    - Recommendations from colleagues
    - Other keyword searches using terms like:
      - Clinical Trials, Clinical Research Informatics, Translational, Data Warehouse, Data Network, Recruitment, Learning Health System, Phenotype, etc.
    - Yielding **348** additional
  - From those, I've selected **94** representative papers that I'll present here (*briefly*)

Session caveats

What this is not…

- A systematic review of the literature
- An exhaustive review

What this is…

- My best attempt at *briefly* covering *some* of the representative CRI literature from the past year
- A snap-shot of excellent CRI activity over past year+
- What I thought was particularly notable

Topics

- Grouped **94** articles into several CRI categories (not *all* CRI areas)
  - Data Sharing, Re-Use and RWD
  - CRI Methods and Approaches
  - AI/ML & intersection with CRI
  - Learning Health Systems & Delivery Science
  - CRI Ethics & Policy
- In each category, I'll highlight a **2-4** key articles and then given a quick(er) "shout out" to others
- Conclude with notable events from the past year+

Apologies up front

- I’m CERTAIN I’ve missed a lot of great work
- I’m REALLY SORRY about that

ChatGPT… a little help

write a haiku about me giving a year-in-review keynote to the AMIA Informatics Summit

Year in review talk,
AMIA summit awaits,
Knowledge to impart.

Clinical Research Informatics

Continued focus on CRI

April 2022 Special Issue of JAMIA on Data Warehousing

- Case studies, Network-based articles, Commentaries
- Essential reading for CRI

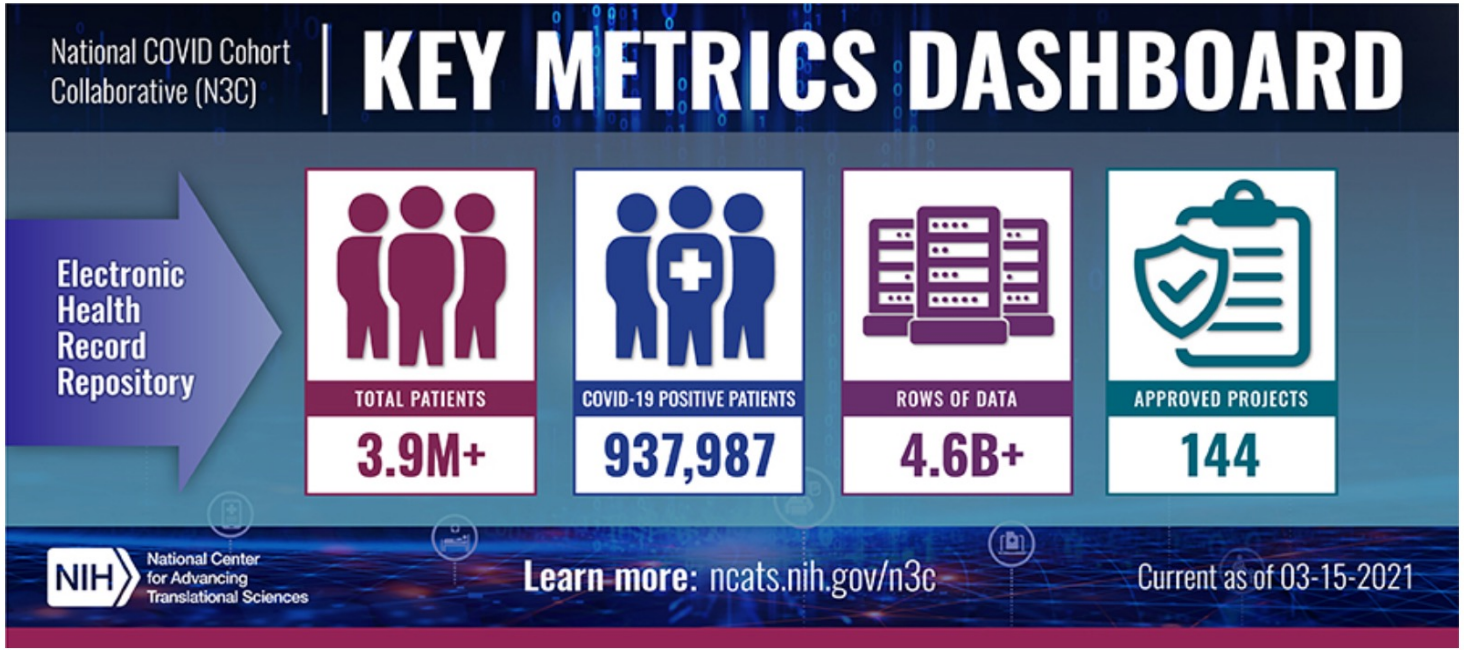
In 2021 YIR:

- **The maturation of clinical research informatics as a subdomain of biomedical informatics** Bakken, S. JAMIA. 2021
- **Review of Clinical Research Informatics** Solomonides, A. Yearbk of Med Info. 2020

Data Sharing, Re-Use & RWD for Research

The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment

Haendel, M, et al. JAMIA. 2020, pub 2021



The National COVID Cohort Collaborative (N3C): Today

National COVID Cohort Collaborative

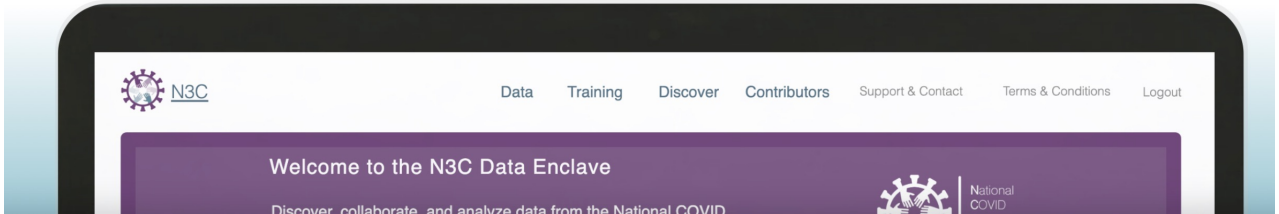
The N3C Data Enclave is a secure platform through which the harmonized clinical data provided by our contributing members is stored. The data itself can only be accessed through a secure cloud portal hosted by NCATS and cannot be downloaded or removed. N3C invites you to begin your journey with the Enclave and join the collaborative efforts of our partners to better understand and address the most pressing COVID-19 clinical questions.

Help make science go faster and save lives.

Access the Enclave

23.1B	2.1B	18.3M	7,217,701
Total Rows	*Clinical Observations*	*Persons*	*COVID+ Cases*

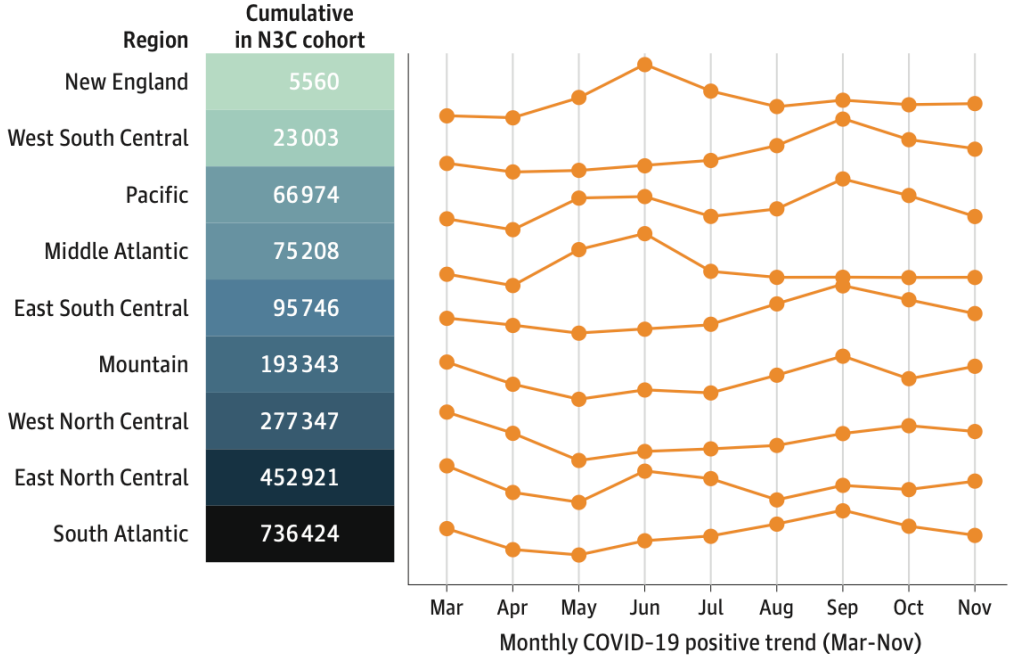
Explore the Full Cohort Dashboard



Clinical Characterization and Prediction of Clinical Severity of SARS-CoV-2 Infection Among US Adults Using Data From the US National COVID Cohort Collaborative

Bennet, T, et al. JAMA Network Open. 2021

Figure 1. Geographic Distribution of Overall SARS-CoV-2–Positive Patients in the US National COVID Cohort Collaborative (N3C) Cohort (N = 1 926 526)



Clinical Characterization and Prediction of Clinical Severity of SARS-CoV-2 Infection Among US Adults Using Data From the US National COVID Cohort Collaborative

Bennet, T, et al. JAMA Network Open. 2021

Figure 2. Comorbidity Distributions of the SARS-CoV-2–Positive Cohort (N = 174 568)

Source	Patients with mild symptoms, %		Hospitalized patients, %			All COVID-19 positive patients, %	All hospitalized patients, %
	No ED (n = 121 078)	ED (n = 21 018)	Moderate (n = 25 907)	Severe (n = 2790)	Mortality/hospice (n = 3775)	(n = 174 568)	(n = 32 472)
Diabetes mellitus	7.1	10.9	24.5	27.4	34.0	11.0	25.9
Renal disease	2.3	3.5	13.6	13.3	25.8	4.8	15.0
Congestive heart failure	1.6	2.5	11.2	11.1	22.1	3.7	12.4
Chronic pulmonary disease	7.0	10.4	16.6	14.0	21.4	9.2	17.0
Peripheral vascular disease	2.6	4.0	10.7	8.7	18.4	4.4	11.5
Stroke	1.7	2.8	8.9	8.4	16.8	3.3	9.8
Cancer	3.0	3.4	9.3	6.7	15.3	4.3	9.8
Dementia	0.5	0.6	4.0	2.7	13.4	1.3	5.0
Myocardial infarction	0.9	1.6	5.4	6.1	11.0	1.9	6.1
Liver disease	2.0	3.2	6.4	6.7	9.1	3.0	6.7
Rheumatologic disease	1.8	2.3	4.0	3.4	4.3	2.3	4.0
Hemiplegia or paraplegia	0.2	0.3	1.8	1.9	3.4	0.6	2.0
Peptic ulcer disease	0.4	0.5	1.3	1.3	2.3	0.6	1.5

Clinical Characterization and Prediction of Clinical Severity of SARS-CoV-2 Infection Among US Adults Using Data From the US National COVID Cohort Collaborative

Bennet, T, et al. JAMA Network Open. 2021

- **Question:** In a US data resource large enough to adjust for multiple confounders, what risk factors are associated with COVID-19 severity and severity trajectory over time, and can machine learning models predict clinical severity?
- **Findings:** Cohort study of 174,568 adults with SARS-CoV-2, 32,472 (18.6%) were hospitalized and 6565 (20.2%) were severely ill, and first-day machine learning models accurately predicted clinical severity. Mortality was 11.6% overall and decreased from 16.4% in March to April 2020 to 8.6% in September to October 2020.
- **Meaning:** These findings suggest that machine learning models can be used to predict COVID-19 clinical severity with the use of an available large-scale US COVID-19 data resource.

Figure 3. Clinical Severity, Age, and Antimicrobial and Immunomodulatory Medication Use Over Time

Long-term use of immunosuppressive medicines and in- hospital COVID-19 outcomes: a retrospective cohort study using data from the National COVID Cohort Collaborative

Anderson, KM, et al. Lancet Digital Medicine. 2022

- **Background**
  - Many individuals take long-term immunosuppressive medications. Few evaluations of effects in hospitalized with COVID-19 compared with non-immunosuppressed individuals.
- **Methods**
  - Retrospective cohort study using data from the National COVID Cohort Collaborative (N3C), between Jan 1, 2020, and June 11, 2021, within 42 health systems.
  - Compared adults with immunosuppressive medications used before admission to adults without long-term immunosuppression.
  - Considered immunosuppression overall, as well as by 15 classes of medication and three broad indications for immunosuppressive medicines.
  - Estimated hazard ratio (HR) for the risk of invasive mechanical ventilation, with the competing risk of death. We used Cox proportional hazards models to estimate HRs for in-hospital death. Models adjusted using doubly robust propensity score methodology.

Long-term use of immunosuppressive medicines and in- hospital COVID-19 outcomes: a retrospective cohort study using data from the National COVID Cohort Collaborative
Anderson, KM, et al. Lancet Digital Medicine. 2022

- **Findings**
  - 222,575 met the inclusion criteria (mean age 59 years; 50% male/female).
  - The most common comorbidities were diabetes (23%), pulmonary disease (17%), and renal disease (13%).
  - 16,494 (7%) patients had long-term immunosuppression with medications for diverse conditions, including
    - Rheumatological disease (33%), solid organ transplant (26%), or cancer (22%).
  - In the propensity score matched cohort, immunosuppression was associated with a reduced risk of invasive ventilation (HR 0·89, 95% CI 0·83–0·96) and there was no overall association between long-term immunosuppression and the risk of in-hospital death.
  - None of the 15 medication classes examined were associated with an increased risk of invasive mechanical ventilation.
  - Although there was no statistically significant association between most drugs and in-hospital death, increases were found with rituximab for rheumatological disease (1·72, 1·10–2·69) and for cancer (2·57, 1·86–3·56)…

Long-term use of immunosuppressive medicines and in- hospital COVID-19 outcomes: a retrospective cohort study using data from the National COVID Cohort Collaborative

Anderson, KM, et al. Lancet Digital Medicine. 2022

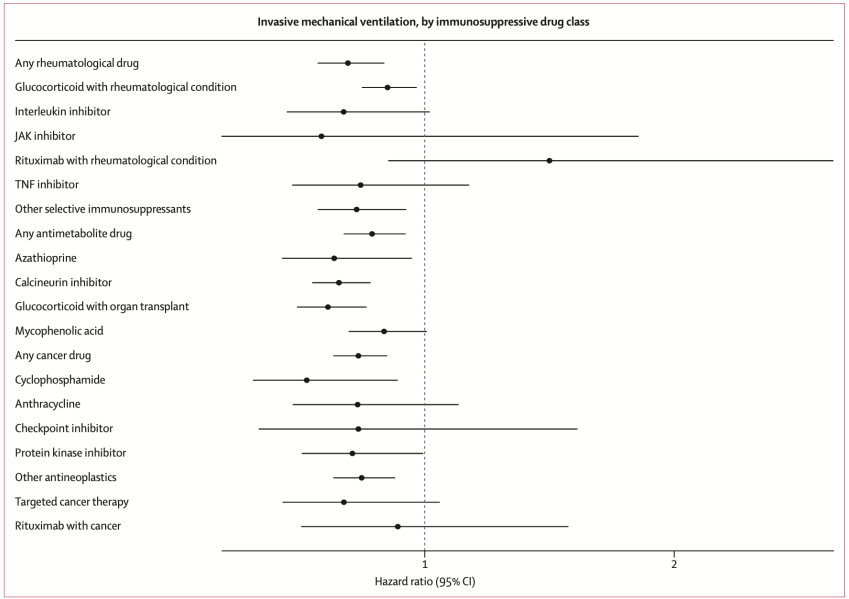


Figure 1: **Association between long-term immunosuppression and invasive mechanical ventilation, by medication class**
Analyses were done in the propensity score matched cohort, with doubly robust adjustment for any remaining covariate imbalances after matching.

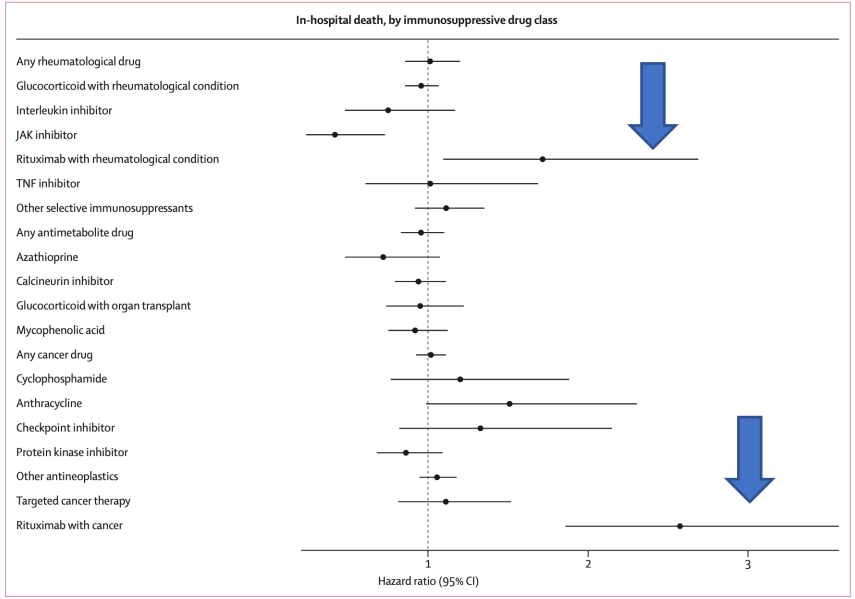


Figure 2: **Association between long-term immunosuppression and in-hospital death, by medication class**
Analyses were done in the propensity score matched cohort, with doubly robust adjustment for any remaining covariate imbalances after matching.

Long-term use of immunosuppressive medicines and in- hospital COVID-19 outcomes: a retrospective cohort study using data from the National COVID Cohort Collaborative

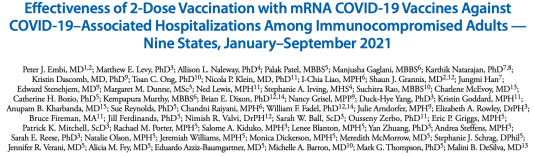
Anderson, KM, et al. Lancet Digital Medicine. 2022

- Main conclusions:
  - Except for rituximab, there was no increased risk of mechanical ventilation or in-hospital death for the rheumatological, antineoplastic, or antimetabolite therapies examined. AND
  - Sample size was large enough to consider separately a variety of drug classes with distinct mechanisms of action, including the targeting of B-cell versus T-cell mediated immunity.
- Implications of available evidence: Adds to growing evidence on the overall safety of most long-term immunosuppressive medications against the backdrop of continued COVID-19-related morbidity and mortality…. And risks of poor outcomes among individuals who are hospitalized with COVID-19.

- **CRI Implication: This analysis was possible because of aggregation of EHR-derived data including pre-hospitalization medication records, and demonstrates growing strengths of such large data sets for RWD analyses**

Effectiveness of COVID-19 mRNA Vaccines Against COVID-19–Associated Hospitalizations Among Immunocompromised Adults During SARS-CoV-2 Omicron Predominance — VISION Network, 10 States, December 2021–August 2022
Britton, A, et al. MMWR. 2022

- **Question:** Data on vaccine effectiveness (VE) of monovalent COVID-19 vaccines among persons with immunocompromising conditions (IC) since the emergence of the Omicron variant in December 2021 are limited.
- **Approach:** In the multi-state (10) VISION Network, monovalent 2-, 3-, and 4-dose mRNA VE against COVID-19–related hospitalization were estimated among adults with immunocompromising conditions (IC) hospitalized with COVID-19–like illness, using a test-negative design comparing odds of previous vaccination among persons with a positive or negative test result (case-patients and control-patients) for SARS-CoV-2.
- **Findings:** December 16, 2021–August 20, 2022, overall VE against COVID-19–related hospitalization among adults with IC hospitalized for COVID-like illness during Omicron predominance was 36% ≥14 days after dose 2, 69% 7–89 days after dose 3, and 44% ≥90 days after dose 3.
- Restricting analysis to later periods when Omicron BA.2/BA.2.12.1 and BA.4/BA.5 were predominant and 3-dose recipients were eligible to receive a fourth dose, VE was 32% ≥90 days after dose 3 and 43% ≥7 days after dose 4.


Effectiveness of 2-Dose Vaccination with mRNA COVID-19 Vaccines Against COVID-19–Associated Hospitalizations Among Immunocompromised Adults — Nine States, January–September 2021

Peter J. Embi, MD[1,2]; Matthew E. Levy, PhD[3]; Allison L. Naleway, PhD[4]; Palak Patel, MBBS[5]; Manjusha Gaglani, MBBS[6]; Karthik Natarajan, PhD[7,8]; Kristin Dascomb, MD, PhD[9]; Toan C. Ong, PhD[10]; Nicola P. Klein, MD, PhD[11]; I-Chia Liao, MPH[4]; Shaun J. Grannis, MD[2,12]; Jungmi Han[7]; Edward Stenehjem, MD[3]; Margaret M. Dunne, MSc[3]; Ned Lewis, MPH[11]; Stephanie A. Irving, MHS[4]; Suchitra Rao, MBBS[10]; Charlene McEvoy, MD[13]; Catherine H. Bozio, PhD[5]; Kempapura Murthy, MBBS[6]; Brian E. Dixon, PhD[12,14]; Nancy Grisel, MPP[9]; Duck-Hye Yang, PhD[3]; Kristin Goddard, MPH[11]; Anupam B. Kharbanda, MD[15]; Sue Reynolds, PhD[3]; Chandni Raiyani, MPH[6]; William F. Fadel, PhD[12,14]; Julie Arndorfer, MPH[4]; Elizabeth A. Rowley, DrPH[3]; Bruce Fireman, MA[11]; Jill Ferdinands, PhD[3]; Nimish R. Valvi, DrPH[11]; Sarah W. Ball, ScD[3]; Ousseny Zerbo, PhD[11]; Eric P. Griggs, MPH[4]; Patrick K. Mitchell, ScD[3]; Rachael M. Porter, MPH[3]; Salome A. Kidulo, MPH[3]; Lenee Blanton, MPH[3]; Yan Zhuang, PhD[3]; Andrea Steffens, MPH[3]; Sarah E. Reese, PhD[3]; Natalie Olson, MPH[3]; Jeremiah Williams, MPH[3]; Monica Dickerson, MPH[3]; Meredith McMorrow, MD[3]; Stephanie J. Schrag, DPhil[3]; Jennifer R. Verani, MD[3]; Alicia M. Fry, MD[3]; Eduardo Azziz-Baumgartner, MD[3]; Michelle A. Barron, MD[16]; Mark G. Thompson, PhD[3]; Malini B. DeSilva, MD[15]

P. J. Embi, M. E. Levy, A. L. Naleway, P. Patel, M. Gaglani, K. Natarajan, et al. **Effectiveness of 2-Dose Vaccination with mRNA COVID-19 Vaccines Against COVID-19-Associated Hospitalizations Among Immunocompromised Adults - Nine States, January-September 2021.** MMWR. 2021 Vol. 70 Issue 44

Effectiveness of COVID-19 mRNA Vaccines Against COVID-19–Associated Hospitalizations Among Immunocompromised Adults During SARS-CoV-2 Omicron Predominance — VISION Network, 10 States, December 2021–August 2022
Britton, A, et al. MMWR. 2022

- **Implications:** VE among IC patients during Omicron was moderate even after a 3-dose monovalent primary series or booster dose. So… they might benefit from updated bivalent vaccine booster doses that target recently circulating Omicron sub-variants…

- **CRI Implications:** Large RWD networks like VISION are enabling studies that were heretofore not possible, even as they have limitations

Summary

What is already known about this topic?

COVID-19 vaccine effectiveness (VE) data among immunocompromised persons during SARS-CoV-2 Omicron variant predominance are limited.

What is added by this report?

Among immunocompromised adults hospitalized with a COVID-like illness, 2-dose monovalent mRNA COVID-19 vaccine VE against COVID-19–associated hospitalization during Omicron predominance was 36%. VE was 67% ≥7 days after a third dose during BA.1 predominance but declined during BA.2/BA.2.12.1 and BA.4/BA.5 predominance to 32% ≥90 days after dose 3 and 43% ≥7 days after dose 4.

What are the implications for public health practice?

Monovalent COVID-19 vaccine protection among persons with immunocompromising conditions during Omicron predominance was moderate after a 3-dose primary series or booster dose. Persons with immunocompromising conditions might benefit from updated bivalent boosters that target circulating BA.4/BA.5 sublineages.

Effectiveness of COVID-19 mRNA Vaccination in Preventing COVID-19–Associated Hospitalization Among Adults with Previous SARS-CoV-2 Infection — United States, June 2021–February 2022

Plumb, ID, et al. MMWR. 2022

- Question; Vaccine effectiveness among patients previously infected with SARS-CoV-2.
- A test-negative design to estimate effectiveness of COVID-19 mRNA vaccines in preventing subsequent COVID-19–associated hospitalization among adults with previous positive test or diagnosis of COVID-19.
- The analysis used data from **Epic's Cosmos**, an electronic health record (EHR)–aggregated data set, and compared vaccination status of 3,761 case-patients (positive NAAT result associated with hospitalization) with 7,522 matched control-patients (negative NAAT result).

Summary

What is already known about this topic?

Persons with previous SARS-CoV-2 infection have some protection against reinfection leading to hospitalization, but there is limited evidence regarding the additional benefit of vaccination among these persons.

What is added by this report?

Among persons with previous infection, COVID-19 mRNA vaccination provided protection against subsequent COVID-19–associated hospitalization. Estimated vaccine effectiveness against reinfection leading to hospitalization during the Omicron-predominant period was approximately 35% after dose 2, and 68% after a booster dose.

What are the implications for public health practice?

To prevent COVID-19–associated hospitalization, all eligible persons should stay up to date with vaccination, including those with previous SARS-CoV-2 infection.

Effectiveness of COVID-19 mRNA Vaccination in Preventing COVID-19–Associated Hospitalization Among Adults with Previous SARS-CoV-2 Infection — United States, June 2021–February 2022

Plumb, ID, et al. MMWR. 2022

- After previous SARS-CoV-2 infection:
  - Estimated vaccine effectiveness (VE) against COVID-19–associated hospitalization was 47.5% (95% CI = 38.8%–54.9%) after 2 vaccine doses and 57.8% (95% CI = 32.1%–73.8%) after a booster dose during the Delta-predominant period (June 20–December 18, 2021), and
  - 34.6% (95% CI = 25.5%–42.5%) after 2 doses and 67.6% (95% CI = 61.4%–72.8%) after a booster dose during the Omicron-predominant period (December 19, 2021–February 24, 2022).
- Vaccination provides protection, higher if boosted.
- **CRI Implication: With over 135M people in the dataset then, and now over 183M, datasets like these are enabling studies heretofore not possible**

Summary

What is already known about this topic?

Persons with previous SARS-CoV-2 infection have some protection against reinfection leading to hospitalization, but there is limited evidence regarding the additional benefit of vaccination among these persons.

What is added by this report?

Among persons with previous infection, COVID-19 mRNA vaccination provided protection against subsequent COVID-19–associated hospitalization. Estimated vaccine effectiveness against reinfection leading to hospitalization during the Omicron-predominant period was approximately 35% after dose 2, and 68% after a booster dose.

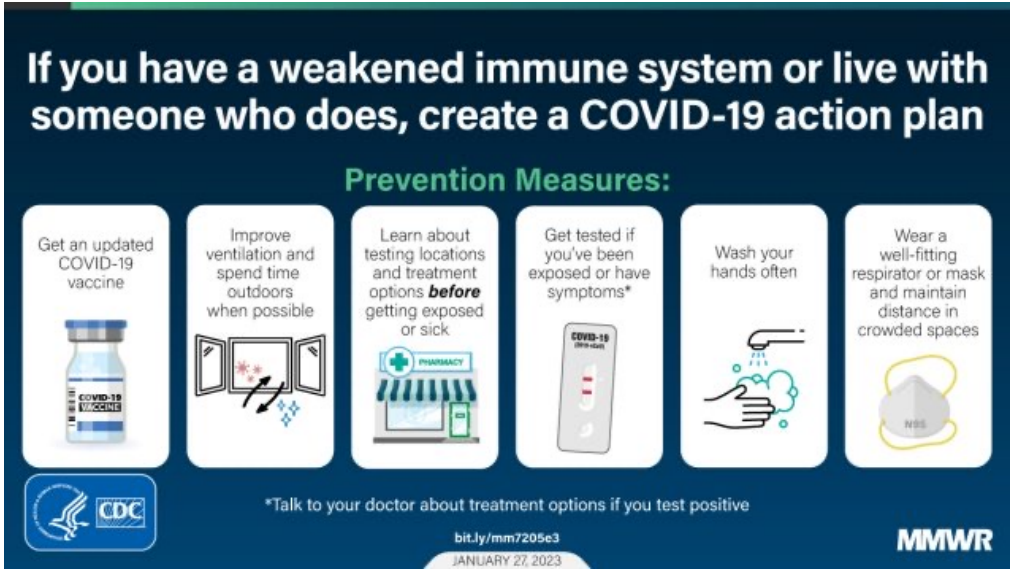
What are the implications for public health practice?

To prevent COVID-19–associated hospitalization, all eligible persons should stay up to date with vaccination, including those with previous SARS-CoV-2 infection.

Evidence to inform prevention and care from many large data-sharing networks.

Not possible without efforts from our CRI community

Informatics saves lives!



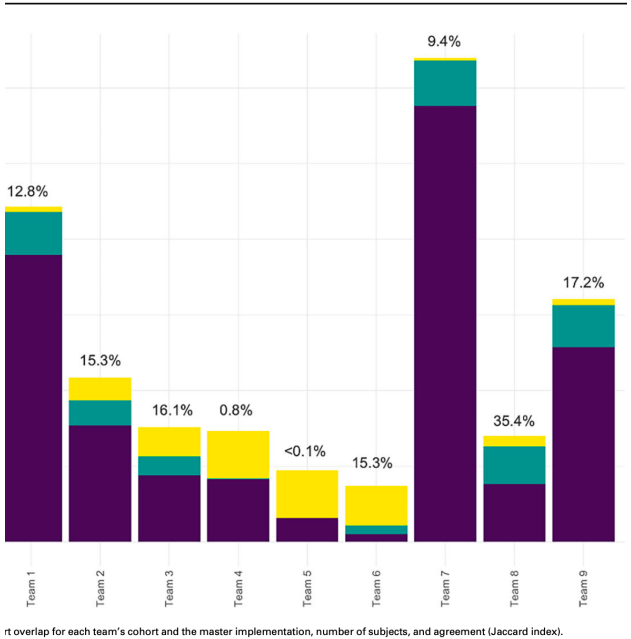
Reproducible variability: assessing investigator discordance across 9 research teams attempting to reproduce the same observational study
Ostropolets A, et al. (OHDSI group). JAMIA. 2023

- **Objective:** Observational studies can impact patient care but must be robust and reproducible. Non-reproducibility is primarily caused by unclear reporting of design choices and analytic procedures. This study aimed to:
  - (1) assess how the study logic described in an observational study could be interpreted by independent researchers and
  - (2) quantify the impact of interpretations' variability on patient characteristics.
- **Materials and Methods:** Nine teams of highly qualified researchers reproduced a cohort from a study by Albogami et al. The teams were provided the clinical codes and access to the tools to create cohort definitions such that the only variable part was their logic choices. They executed teams' cohort definitions against the database and compared the number of subjects, patient overlap, and patient characteristics.

Reproducible variability: assessing investigator discordance across 9 research teams attempting to reproduce the same observational study

Ostropolets A, et al. (OHDSI group). JAMIA. 2023

- **Results:** On average, only 40% match on inclusion criteria interpretation compared with master, with at least 4 deviations per team.
- Median agreement was 9.4%. Teams' cohorts significantly differed from the master implementation by at least 2 baseline characteristics, and most of the teams differed by at least 5.
- Influence of different choices on patient characteristics
  - "We observed **high variation in cohort size** from having **one-third** of the master implementation patient count **to having 10 times the cohort** size (2159–63,619 subjects compared to 6196 subjects in the master implementation).
  - Not surprisingly, agreement between the master cohort and the teams' implementations also varied greatly (Figure 4)."



rt overlap for each team's cohort and the master implementation, number of subjects, and agreement (Jaccard index).

Reproducible variability: assessing investigator discordance across 9 research teams attempting to reproduce the same observational study Ostropolets A, et al. (OHDSI group). JAMIA. 2023

- **Conclusions:** Independent research teams attempting to reproduce the study based on its free-text description alone produce different implementations that vary in the population size and composition. Sharing analytical code supported by a common data model and open-source tools allows reproducing a study unambiguously thereby preserving initial design choices.

- **CRI Implications: Poorly kept secret in our field that if you ask three teams, you'll likely get three different results. Reproducibility is a major need and problem. This is real progress to quantify, measure and (hopefully) call attention to and address this issue.**

Maturity in enterprise data warehouses for research operations: Analysis of a pilot study

B. M. Knosp, D. A. Dorr and T. R. Campion. Journal of Clinical and Translational Science 2023

- Enterprise data warehouses for research (EDW4R) effort – represents a critical component of all NIH-CTSA hubs
- EDW4R operations have unique needs that require specialized skills and collaborations across multiple domains which limit the ability to apply existing models of information technology (IT) performance.
- Because of this uniqueness, *developed a new EDW4R maturity model* based on prior qualitative study of operational practices for supporting EDW4Rs at CTSA hubs.
- In a pilot study, respondents from fifteen CTSA hubs completed the novel EDW4R maturity index survey by rating 33 maturity statements across 6 categories using a 5-point Likert scale.

Maturity in enterprise data warehouses for research operations: Analysis of a pilot study

B. M. Knosp, D. A. Dorr and T. R. Campion. Journal of Clinical and Translational Science 2023

- Of the six categories, workforce as most mature (4.17 [3.67–4.42]) and relationship with enterprise IT as the least mature (3.00 [2.80–3.80]).
- Implication: baseline quantitative measure of EDW4R functions across fifteen CTSA hubs. The maturity index may be useful those leading an EDW4R
- **CRI Implication: Another important step in the "maturity" of our field by focusing on ability to consistently measure and assess our progress, and where we still need to improve in core functions.**
- **Adds to maturing maturity models**

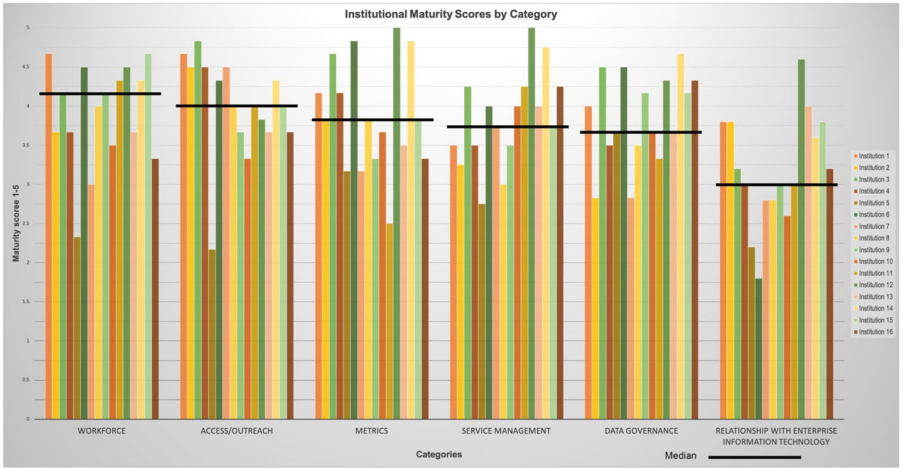


Fig. 2. Maturity scores across all six categories colored by institution.

Data Sharing, Re-Use, and RWD: Other notable papers

- B. Martin, P. E. DeWitt, S. Russell, A. Anand, K. R. Bradwell, C. Bremer, et al. **Characteristics, Outcomes, and Severity Risk Factors Associated With SARS-CoV-2 Infection Among Children in the US National COVID Cohort Collaborative.** JAMA Netw Open 2022 Vol. 5 Issue 2 Pages e2143151
- E. R. Pfaff, A. T. Girvin, D. L. Gabriel, K. Kostka, M. Morris, M. B. Palchuk, et al. **Synergies between centralized and federated approaches to data quality: a report from the national COVID cohort collaborative.** J Am Med Inform Assoc 2022 Vol. 29 Issue 4 Pages 609-618
- J. A. Thomas, R. E. Foraker, N. Zamstein, J. D. Morrow, P. R. O. Payne, A. B. Wilcox, et al. **Demonstrating an approach for evaluating synthetic geospatial and temporal epidemiologic data utility: results from analyzing >1.8 million SARS-CoV-2 tests in the United States National COVID Cohort Collaborative (N3C).** J Am Med Inform Assoc 2022 Vol. 29 Issue 8 Pages 1350-1365
- W. He, K. G. Kirchoff, R. R. Sampson, K. K. McGhee, A. M. Cates, J. S. Obeid, et al. **Research Integrated Network of Systems (RINS): a virtual data warehouse for the acceleration of translational research.** J Am Med Inform Assoc 2021 Vol. 28 Issue 7 Pages 1440-1450

Data Sharing, Re-Use, and RWD: Other notable papers

- A. H. Ramirez, K. A. Gebo and P. A. Harris. **Progress With the All of Us Research Program: Opening Access for Researchers.** JAMA 2021 Vol. 325 Issue 24 Pages 2441-2442
- A. H. Ramirez, L. Sulieman, D. J. Schlueter, A. Halvorson, J. Qian, F. Ratsimbazafy, et al. **The All of Us Research Program: Data quality, utility, and diversity.** Patterns (N Y) 2022 Vol. 3 Issue 8 Pages 100570
- L. Sulieman, R. M. Cronin, R. J. Carroll, K. Natarajan, K. Marginean, B. Mapes, et al. **Comparing medical history data derived from electronic health records and survey answers in the All of Us Research Program.** J Am Med Inform Assoc 2022 Vol. 29 Issue 7 Pages 1131-1141

- Beyond informatics science, clinical research leveraging all of us…

Data Sharing, Re-Use, and RWD: Other notable papers

- All of US data enabling many studies:
- H. Master, J. Annis, S. Huang, J. A. Beckman, F. Ratsimbazafy, K. Marginean, et al. **Association of step counts over time with the risk of chronic disease in the All of Us Research Program.** Nat Med 2022 Vol. 28 Issue 11 Pages 2301-2308
- A. S. Perry, J. S. Annis, H. Master, M. Nayor, A. Hughes, A. Kouame, et al. **Association of longitudinal activity measures and diabetes risk: an analysis from the NIH All of Us Research Program.** J Clin Endocrinol Metab 2022
- N. Soley, S. Song, N. Flaks-Manov and C. O. Taylor. **Risk for Poor Post-Operative Quality of Life Among Wearable Use Subgroups in an All of Us Research Cohort.** Pac Symp Biocomput 2023 Vol. 28 Pages 31-42
- E. B. Lee, W. Hu, K. Singh and S. Y. Wang. **The Association among Blood Pressure, Blood Pressure Medications, and Glaucoma in a Nationwide Electronic Health Records Database.** Ophthalmology 2022 Vol. 129 Issue 3 Pages 276-284

Data Sharing, Re-Use, and RWD: Other notable papers

- K. Kostka, T. Duarte-Salles, A. Prats-Uribe, A. G. Sena, A. Pistillo, S. Khalid, et al. **Unraveling COVID-19: A Large-Scale Characterization of 4.5 Million COVID-19 Cases Using CHARYBDIS.** Clin Epidemiol 2022 Vol. 14 Pages 369-384
  - Global, multi-center view to describe trends in COVID-19 progression, management and evolution over time. By characterizing baseline variability in patients and geography, *work provides critical context that may otherwise be misconstrued as data quality issues. This is important for studies on adverse events.*
- Li X, Ostropolets A, Makadia R, Shoaibi A, Rao G, Sena AG, Martinez-Hernandez E, Delmestri A, Verhamme K, Rijnbeek PR, Duarte-Salles T, Suchard MA, Ryan PB, Hripcsak G, Prieto-Alhambra D. **Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study.** BMJ. 2021 Jun 14;373:n1435. doi: 10.1136/bmj.n1435.
  - The study by OHDSI consortium that the European Medicines Agency used to turn the AstraZeneca vaccine back on in March 2021; paper published later that year.
- C. Reyes, A. Pistillo, S. Fernandez-Bertolin, M. Recalde, E. Roel, D. Puente, et al. **Characteristics and outcomes of patients with COVID-19 with and without prevalent hypertension: a multinational cohort study.** BMJ Open 2021 Vol. 11 Issue 12 Pages e057632

Data Sharing, Re-Use, and RWD: Other notable papers

- O. Magen, J. G. Waxman, M. Makov-Assif, R. Vered, D. Dicker, M. A. Hernan, et al. **Fourth Dose of BNT162b2 mRNA Covid-19 Vaccine in a Nationwide Setting.** N Engl J Med 2022 Vol. 386 Issue 17 Pages 1603-1614
  - One example of rapid-turn around, RWD enabled evidence generation from largest health system in Israel
- M. W. Tenforde, Z. A. Weber, K. Natarajan, N. P. Klein, A. B. Kharbanda, E. Stenehjem, et al. **Early Estimates of Bivalent mRNA Vaccine Effectiveness in Preventing COVID-19-Associated Emergency Department or Urgent Care Encounters and Hospitalizations Among Immunocompetent Adults - VISION Network, Nine States, September-November 2022.** MMWR Morb Mortal Wkly Rep 2022 Vol. 71 Issue 5152 Pages 1616-1624
  - Another example from CDC-VISION networks
- W. Tu, P. Zhang, A. Roberts, K. S. Allen, J. Williams, P. Embi, et al. **SARS-CoV-2 Infection, Hospitalization, and Death in Vaccinated and Infected Individuals by Age Groups in Indiana, 2021–2022.** Am J Public Health 2023 Vol. 113 Issue 1 Pages 96-104

CRI Methods and Approaches

Automated production of research data marts from a canonical fast healthcare interoperability resource (FHIR) data repository: applications to COVID-19 research
L. A. Lenert, et al.
JAMIA. 2021

- Objective: Motivated by COVID-19 pandemic, need for timely data from the healthcare systems for research.
- Ramped up our need to contribute to research data consortia requiring frequent updating and sharing of electronic health record (EHR) data in different common data models (CDMs) to create multi-institutional databases for research.
- Traditionally, each CDM has had a custom pipeline for extract, transform, and load operations for production and incremental updates of data feeds to the networks from raw EHR data.
- Demands of COVID-19 research for timely data required faster responses. New approaches needed.
- **(I'd argue, they were before, but another great example of necessity being the mother of invention)**

Automated production of research data marts from a canonical fast healthcare interoperability resource data repository: applications to COVID-19 research

L. A. Lenert, et al. JAMIA. 2021

- **Methods:** FHIR data model as a canonical data model and automated transformation of clinical data to:
  - Patient-Centered Outcomes Research Network (PCORnet) and
  - Observational Medical Outcomes Partnership (OMOP) CDMs
  - for data sharing and research collaboration on COVID-19.
- **Results:** FHIR data resources could be transformed to operational PCORnet and OMOP CDMs with minimal production delays through a combination of real-time and postprocessing steps, leveraging the FHIR data subscription feature.
- **Conclusions:** The approach leverages evolving standards for the availability of EHR data developed to facilitate data exchange under the 21st Century Cures Act and could greatly enhance the availability of standardized data- sets for research.
- **CRI Implications:** Great example of new approaches that we needed anyway, and created during the crisis that was the pandemic. Current approaches weren't efficient, crises make change necessary.
- **Never waste a crisis!**

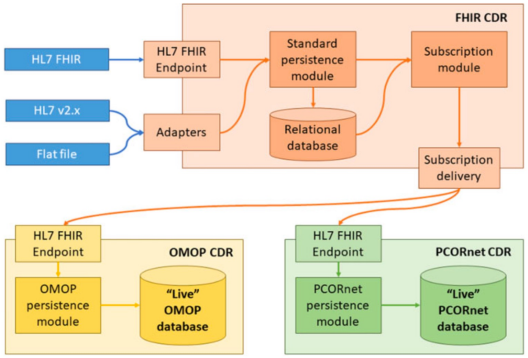


Figure 1. Adapting a FHIR CDR for real-time ETL to OMOP and PCORnet CDMs.

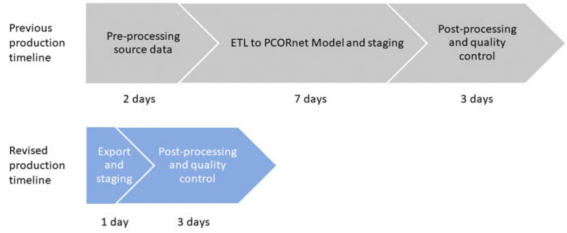


Figure 3. Production timelines for PCORnet database release.

Assessing the impact of privacy-preserving record linkage on record overlap and patient demographic and clinical characteristics in PCORnet(R), the National Patient-Centered Clinical Research Network

K. Marsolo, et al. J Am Med Inform Assoc. 2023

- OBJECTIVE: This article describes the implementation of a privacy-preserving record linkage (PPRL) solution across PCORnet(R), the National Patient-Centered Clinical Research Network.
- MATERIAL AND METHODS: Using a PPRL solution from Datavant, we quantified the degree of patient overlap across the network and report a de-duplicated analysis of the demographic and clinical characteristics of the PCORnet population.

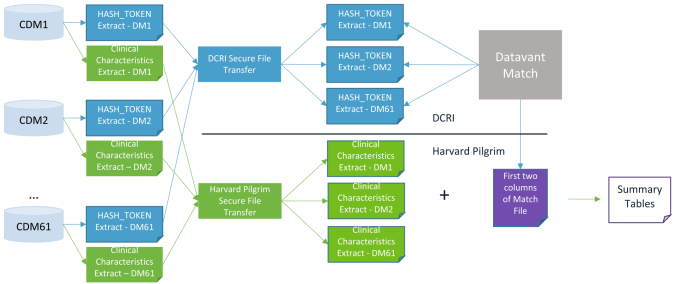


Figure 1. Flow of data from each PCORnet Data Mart to the Coordinating Center. The HASH_TOKEN extract goes to DCRI for processing and then is matched using the Datavant software. The match IDs are then sent to Harvard Pilgrim for linking with the clinical characteristics for further processing and analysis.

Assessing the impact of privacy-preserving record linkage on record overlap and patient demographic and clinical characteristics in PCORnet(R), the National Patient-Centered Clinical Research Network

K. Marsolo, et al. J Am Med Inform Assoc. 2023

- RESULTS: There were approximately 170M patient records across the responding Network Partners, with approximately 138M (81%) of those corresponding to a unique patient. 82.1% of patients were found in a single partner and 14.7% were in 2. The percentage overlap between Partners ranged between 0% and 80% with a median of 0%. Linking patients' electronic health records with claims increased disease prevalence in every clinical characteristic, ranging between 63% and 173%.
- DISCUSSION: The overlap between Partners was variable and depended on timeframe. However, patient data linkage changed the prevalence profile of the PCORnet patient population.
- CONCLUSIONS: This project was one of the largest linkage efforts of its kind and demonstrates the potential value of record linkage. Linkage between Partners may be most useful in cases where there is geographic proximity between Partners, an expectation that potential linkage Partners will be able to fill gaps in data, or a longer study timeframe.
- **CRI Implications**: **PPRL is increasingly important as large data-sharing efforts and repositories (central or federated) are created**

Evaluating automated electronic case report form data entry from electronic health records

A. C. Cheng, et al. Journal of Clinical and Translational Science 2023

- BACKGROUND: Many clinical trials leverage real-world data. Typically, these data are manually abstracted from electronic health records (EHRs) and entered into electronic case report forms (CRFs), a time and labor-intensive process that is also error-prone and may miss information. Automated transfer of data from EHRs to eCRFs has the potential to reduce data abstraction and entry burden as well as improve data quality and safety.
- METHODS: We conducted a test of automated EHR-to-CRF data transfer for 40 participants in a clinical trial of hospitalized COVID-19 patients. We determined which coordinator-entered data could be automated from the EHR (coverage), and the frequency with which the values from the automated EHR feed and values entered by study personnel for the actual study matched exactly (concordance).

Evaluating automated electronic case report form data entry from electronic health records

A. C. Cheng, et al. Journal of Clinical and Translational Science 2023

- **RESULTS:** The automated EHR feed populated 10,081/11,952 (84%) coordinator-completed values. For fields where both the automation and study personnel provided data, the values matched exactly 89% of the time. Highest concordance was for daily lab results (94%), which also required the most personnel resources (30 minutes per participant). In a detailed analysis of 196 instances where personnel and automation entered values differed, both a study coordinator and a data analyst agreed that 152 (78%) instances were a result of data entry error.
- **CONCLUSIONS: An automated EHR feed has the potential to significantly decrease study personnel effort while improving the accuracy of CRF data.**

Table 1. *Coverage of FHIR to complete data filled by coordinator by CRF for 40 participants*

CRF name	# Fields in source project	# Fields with FHIR data mapped	# Values entered by coordinator	# Values entered by FHIR	% Values entered by FHIR	Potential hours saved with FHIR*
Demographics	13	9	349	120	34	0.5
Eligibility criteria	23	20	920	800	87	3.3
COVID-19 testing/ vaccination	4	4	160	159	99	0.7
Daily inpatient form	14	9	2040	1270	62	5.3
Vital signs	17	14	3433	2946	86	12.3
Clinical labs	29	29	5050	4785	95	19.9
Total	100	85	11,952	10,080	84	42

*Based on Nordo *et al.* mean of 15 seconds per value entered by coordinator [9].
CRF, case report form; FHIR, Fast Healthcare Interoperability Resources; COVID-19, coronavirus disease of 2019.

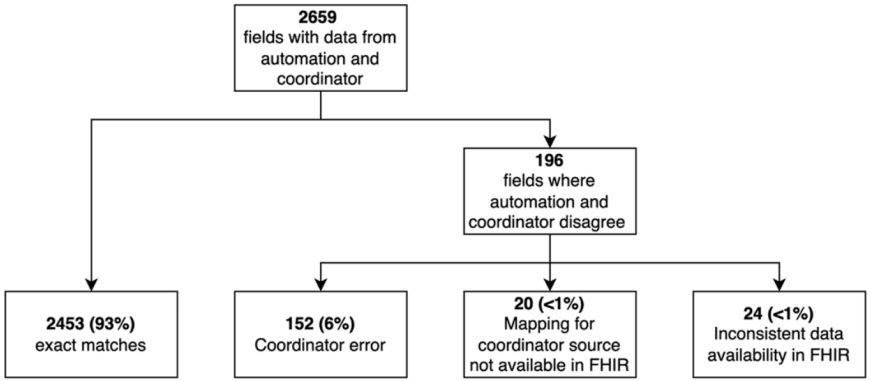


Fig. 1. Summary of data concordance for the first 10 participants in the trial at Vanderbilt University Medical Center. FHIR, Fast Healthcare Interoperability Resources.

CRI Methods:
Other notable papers

- C. Yan, Y. Yan, Z. Wan, Z. Zhang, L. Omberg, J. Guinney, S. Mooney, B. Malin. **A Multifaceted benchmarking of synthetic electronic health record generation models.** Nature Communications 2022
- R. Foraker, A. Guo, J. Thomas, N. Zamstein, P. R. Payne, A. Wilcox, et al. **The National COVID Cohort Collaborative: Analyses of Original and Computationally Derived Electronic Health Record Data.** J Med Internet Res 2021 Vol. 23 Issue 10 Pages e30697
- J. A. Thomas, R. E. Foraker, N. Zamstein, J. D. Morrow, P. R. O. Payne, A. B. Wilcox, et al. **Demonstrating an approach for evaluating synthetic geospatial and temporal epidemiologic data utility: results from analyzing >1.8 million SARS-CoV-2 tests in the United States National COVID Cohort Collaborative (N3C).** J Am Med Inform Assoc 2022 Vol. 29 Issue 8 Pages 1350-1365
- C. Blacketer, F. J. Defalco, P. B. Ryan and P. R. Rijnbeek. **Increasing trust in real-world evidence through evaluation of observational data quality.** J Am Med Inform Assoc 2021 Vol. 28 Issue 10 Pages 2251-2257
- W. Xia, M. Basford, R. Carroll, E. W. Clayton, P. Harris, M. Kantacioglu, et al.. **Managing reidentification risks while providing access to the All of Us research program.** J Am Med Inform Assoc 2023

CRI Methods:
Other notable papers

- M. G. Crowson, D. Moukheiber, A. R. Arévalo, B. D. Lam, S. Mantena, A. Rana, et al. **A systematic review of federated learning applications for biomedical data.** PLOS Digit Health 2022 Vol. 1 Issue 5 Pages e0000033
- S. N. Duda, N. Kennedy, D. Conway, A. C. Cheng, V. Nguyen, T. Zayas-Caban, et al. **HL7 FHIR-based tools and initiatives to support clinical research: a scoping review.** J Am Med Inform Assoc 2022 Vol. 29 Issue 9 Pages 1642-1653
- H. Liu, S. Carini, Z. Chen, S. Phillips Hey, I. Sim and C. Weng. **Ontology-based categorization of clinical studies by their conditions.** J Biomed Inform 2022 Vol. 135 Pages 104235
- Y. Peng, E. Henke, I. Reinecke, M. Zoch, M. Sedlmayr and F. Bathelt. **An ETL-process design for data harmonization to participate in international research with German real-world data based on FHIR and OMOP CDM.** Int J Med Inform 2023 Vol. 169 Pages 104925

CRI Methods: SDOH
Other notable papers

- J. Phuong, E. Zampino, N. Dobbins, J. Espinoza, D. Meeker, H. Spratt, et al. **Extracting Patient-level Social Determinants of Health into the OMOP Common Data Model.** AMIA Annu Symp Proc 2021 Vol. 2021 Pages 989-998
- A. Teng and A. Wilcox. **Simplified data science approach to extract social and behavioural determinants: a retrospective chart review.** BMJ Open 2022 Vol. 12 Issue 1 Pages e048397
- J. R. Vest, J. Adler-Milstein, L. M. Gottlieb, J. Bian, T. R. Campion, Jr., G. R. Cohen, et al. **Assessment of structured data elements for social risk factors.** Am J Manag Care 2022 Vol. 28 Issue 1 Pages e14-e23
- W. E. Trick, J. C. Hill, P. Toepfer, F. Rachman, B. Horwitz and A. Kho. **Joining Health Care and Homeless Data Systems Using Privacy-Preserving Record-Linkage Software.** Am J Public Health 2021 Vol. 111 Issue 8 Pages 1400-1403

CRI Methods: Phenotyping
Other notable papers

- H. Estiri, Z. H. Strasser and S. N. Murphy. **High-throughput phenotyping with temporal sequences.** J Am Med Inform Assoc 2021 Vol. 28 Issue 4 Pages 772-781
- P. S. Brandt, A. Kho, Y. Luo, J. A. Pacheco, T. L. Walunas, H. Hakonarson, et al. **Characterizing variability of electronic health record-driven phenotype definitions.** J Am Med Inform Assoc 2023 Vol. 30 Issue 3 Pages 427-437
- M. P. Maurits, I. Korsunsky, S. Raychaudhuri, S. N. Murphy, J. W. Smoller, S. T. Weiss, et al. **A framework for employing longitudinally collected multicenter electronic health records to stratify heterogeneous patient populations on disease history.** J Am Med Inform Assoc 2022 Vol. 29 Issue 5 Pages 761-769
- J. R. Robinson, R. J. Carroll, L. Bastarache, Q. Chen, J. Pirruccello, Z. Mou, et al. **Quantifying the phenome-wide disease burden of obesity using electronic health records and genomics. (eMERGE)** Obesity (Silver Spring) 2022 Vol. 30 Issue 12 Pages 2477-2488
- W. Song, M. J. Kang, L. Zhang, W. Jung, J. Song, D. W. Bates, et al. **Predicting pressure injury using nursing assessment phenotypes and machine learning methods.** J Am Med Inform Assoc 2021 Vol. 28 Issue 4 Pages 759-765

CRI Methods: Recruitment
Other notable papers

- R. W. Grout, D. Hood, S. J. Nelson, P. A. Harris and P. J. Embi. **Selecting EHR-driven recruitment strategies: An evidence-based decision guide.** J Clin Transl Sci 2022 Vol. 6 Issue 1 Pages e108
- Y. Fang, B. Idnay, Y. Sun, H. Liu, Z. Chen, K. Marder, et al. **Combining human and machine intelligence for clinical trial eligibility querying.** J Am Med Inform Assoc 2022 Vol. 29 Issue 7 Pages 1161-1171
- T. T. Helmer, A. A. Lewis, M. McEver, F. Delacqua, C. L. Pastern, N. Kennedy, et al. **Creating and implementing a COVID-19 recruitment Data Mart.** J Biomed Inform 2021 Vol. 117 Pages 103765
- S. J. Nelson, B. Drury, D. Hood, J. Harper, T. Bernard, C. Weng, et al. **EHR-based cohort assessment for multicenter RCTs: a fast and flexible model for identifying potential study sites.** J Am Med Inform Assoc 2022 Vol. 29 Issue 4 Pages 652-659
- J. R. Rogers, J. Pavisic, C. N. Ta, C. Liu, A. Soroush, Y. Kuen Cheung, et al. **Leveraging electronic health record data for clinical trial planning by assessing eligibility criteria's impact on patient count and safety.** J Biomed Inform 2022 Vol. 127 Pages 104032

CRI Methods: Platforms
Other notable papers

- A. C. Cheng, S. N. Duda, R. Taylor, F. Delacqua, A. A. Lewis, T. Bosler, et al. **REDCap on FHIR: Clinical Data Interoperability Services.** J Biomed Inform 2021 Vol. 121 Pages 103871
- P. A. Harris, G. Delacqua, R. Taylor, S. Pearson, M. Fernandez and S. N. Duda. **The REDCap Mobile Application: a data collection platform for research in regions or situations with internet scarcity.** JAMIA Open 2021 Vol. 4 Issue 3 Pages ooab078
- P. A. Harris, J. Swafford, E. S. Serdoz, J. Eidenmuller, G. Delacqua, V. Jagtap, et al. **MyCap: a flexible and configurable platform for mobilizing the participant voice.** JAMIA Open 2022 Vol. 5 Issue 2 Pages ooac047
- W. He, K. G. Kirchoff, R. R. Sampson, K. K. McGhee, A. M. Cates, J. S. Obeid, et al. **Research Integrated Network of Systems (RINS): a virtual data warehouse for the acceleration of translational research.** J Am Med Inform Assoc 2021 Vol. 28 Issue 7 Pages 1440-1450
- J. R. Vest, J. Blackburn, S. Cash-Goldwasser, E. Peters Bergquist and P. J. Embi. **Mask-Wearing Behavior at the 2021 NCAA Men's Basketball Tournament.** JAMA 2021 Vol. 326 Issue 13 Pages 1324-1325

AI/ML and CRI

Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis
Adams, R… et al. Saria, S. Nature Medicine.

- Early recognition and treatment of sepsis are linked to improved patient outcomes. Machine learning-based early warning systems may reduce the time to recognition, but few systems have undergone clinical evaluation.
- In this prospective, multi-site cohort study, we examined the association between patient outcomes and provider interaction with a deployed sepsis alert system called the **Targeted Real-time Early Warning System (TREWS**).
- During the study, **590,736 patients** were monitored by TREWS across five hospitals.
- Focused analysis on **6,877 patients with sepsis** identified by the alert before initiation of antibiotic therapy.

Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis

Adams, R… et al. Saria, S. Nature Medicine.

- Patients whose alerts were confirmed by a provider within 3h of the alert had a reduced in-hospital mortality rate of 3.3% adjusted absolute reduction, and 18.7% adjusted relative reduction in organ failure and length of stay compared with patients whose alert was not confirmed by a provider within 3h.
- Improvements in mortality rate (4.5%) and organ failure were larger among those patients who were additionally flagged as high risk.

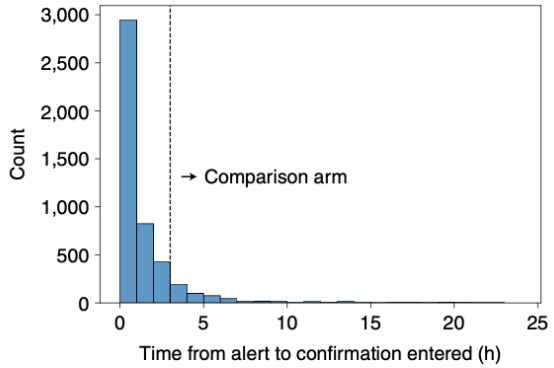


Fig. 2 | Distribution of the time from which the TREWS alert was given to the time confirmation was entered for the target population. Each bar has a width of 1 h, and the bar height represents the number of subjects who had their alert confirmed in that 1-h time bin. The dashed line is placed at 3 h, such that bars to the left and right of the line correspond to subjects in the study and comparison arms, respectively. Note that 1,840 (27%) patients in the target population either did not have an evaluation entered or had their alert dismissed by a provider and thus do not appear in this plot.

Table 2 | Associations between alert confirmation and patient outcomes

	Treatment	Comparison	ARD or ARR	*P* value[a]
All included	*n* = 4,220	*n* = 2,657		
In-hospital mortality, no. (rate)	617 (14.6%)	509 (19.2%)	ARD −3.34% (−5.10, −1.67%)	<0.001
			ARR −18.18% (−26.31, −9.65%)	<0.001
SOFA progression at 72 h[b]	−0.8 ± 2.7	−0.4 ± 2.9	ARD −0.26 (−0.42, −0.11)	0.001
Median length of stay (h)[c]	156 (99–260)	190 (118–323)	ARD −11.58 (−18.13, −5.03)	0.001
High-risk cohort	*n* = 1,430	*n* = 935		
In-hospital mortality, no. (rate)	422 (29.5%)	320 (34.2%)	ARD −4.50% (−8.31, −0.78%)	0.012
			ARR −13.19% (−22.81, −2.45%)	0.012
SOFA progression at 72 h[b]	−1.5 ± 3.5	−1.2 ± 3.6	ARD −0.38 (−0.71, −0.05)	0.025
Median length of stay (h)[c]	210 (129–351)	246 (155–434)	ARD −14.21 (−32.47, −4.04)	0.127

Mortality is reported as 'count (percentage)', SOFA progression is reported as 'mean ± s.d.' and length of stay is reported as 'median (IQR)'. Associations are reported as either an ARD or ARR and are presented as 'ARD/ARR (95% CI)'. [a]*P* values for in-hospital mortality were based on nonparametric bootstrap resampling using 5,000 bootstrap samples. *P* values for SOFA progression and median length of stay were based on Student's *t*-tests. All tests were two sided and not adjusted for multiple comparisons. [b]SOFA progression at 72 h excludes patients who were discharged to hospice, left against medical advice or transferred to another acute care facility within 72 h of the alert. [c]Median length of stay was calculated only on patients who did not die in hospital.

Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis

Adams, R… et al. Saria, S. Nature Medicine.

- Conclusion: Early warning systems have the potential to identify sepsis patients early and improve patient outcomes and that sepsis patients who would benefit the most from early treatment can be identified and prioritized at the time of the alert
- One of four pubs related to this team's TREWS work
  - In combination, these papers address 1) the development process for ML in the hospital, 2) HCI factors involved in their use, 3) guidelines for facilitating adoption, and 4) the evaluation after the ML was in use for a time.
- Also, accompanying editorial:

CRITICAL CARE

Harnessing AI in sepsis care

A real-time early warning system for sepsis detection shows promising adoption by healthcare providers and important improvements in patient outcomes.

David W. Bates and Ania Syrowatka

- **CRI Implications:**
  - **Adds to the literature on how such AI-enabled systems can work to improve outcomes**
  - **Studies of this kind are enabled by CRI efforts, and will become increasingly important to care and research**
  - **How these are implemented, deployed, governed is key…**

A framework for the oversight and local deployment of safe and high-quality prediction models

Bedoya, AD, et al. JAMIA. 2022

- AI/ML models are being rapidly developed and used in clinical practice.
- Many models are deployed without a clear understanding of clinical or operational impact and frequently lack monitoring plans that can detect potential safety signals.
- There is a lack of consensus in establishing governance to deploy, pilot, and monitor algorithms within operational healthcare delivery workflows.
- This group describes a governance framework that combines current regulatory best practices and lifecycle management of predictive models being used for clinical care.
- Since January 2021, they've added models to their governance portfolio and are currently managing 52 models.

A framework for the oversight and local deployment of safe and high-quality prediction models

Bedoya, AD, et al. JAMIA. 2022

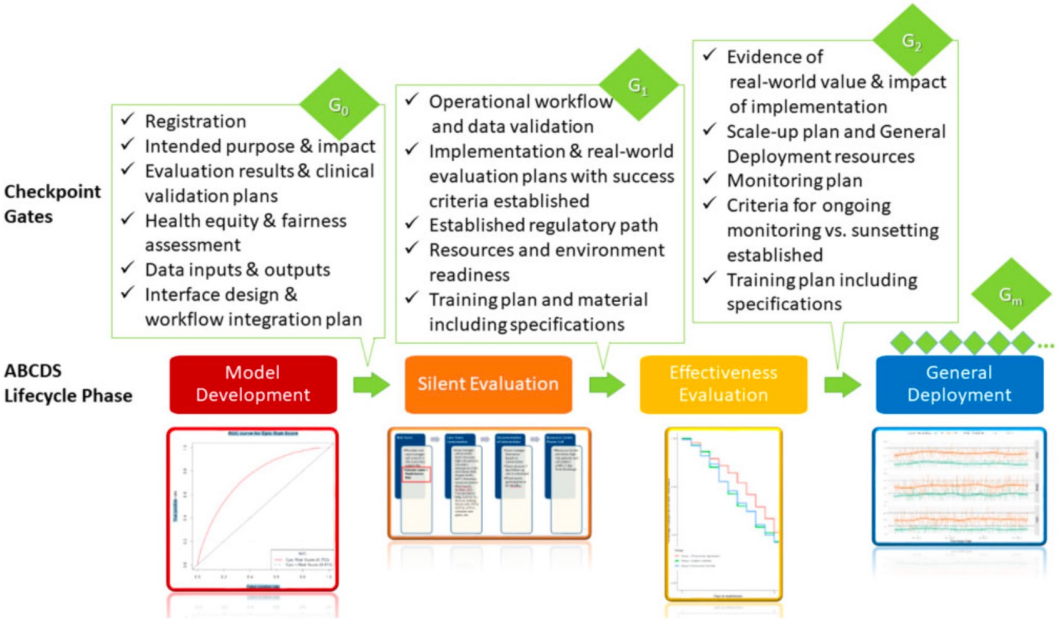


Figure 1. “Just-in-time” checkpoint gates (Gx) along the ABCDS lifecycle that help development teams prepare for approval to proceed to the next ABCDS life- cycle phase. Top: Examples of key information requested from development teams for Gx approval. Bottom: Example information provided by development teams to proceed for a third-party readmission model (see ABCDS examples in Supplementary Materials for further details).

Algorithm-based clinical decision support (**ABCDS**); model development checkpoint (G0) silent evaluation checkpoint (G1); effectiveness evaluation checkpoint (G2); general deployment check- point (Gm).

A framework for the oversight and local deployment of safe and high-quality prediction models

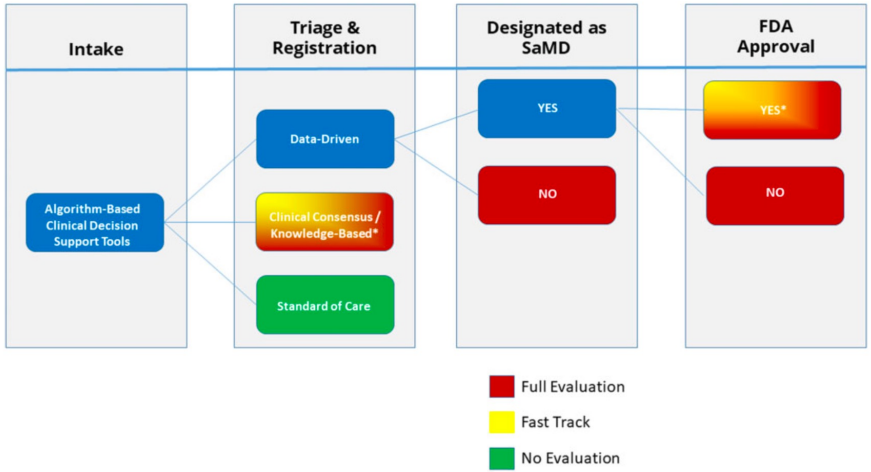
Bedoya, AD, et al. JAMIA. 2022

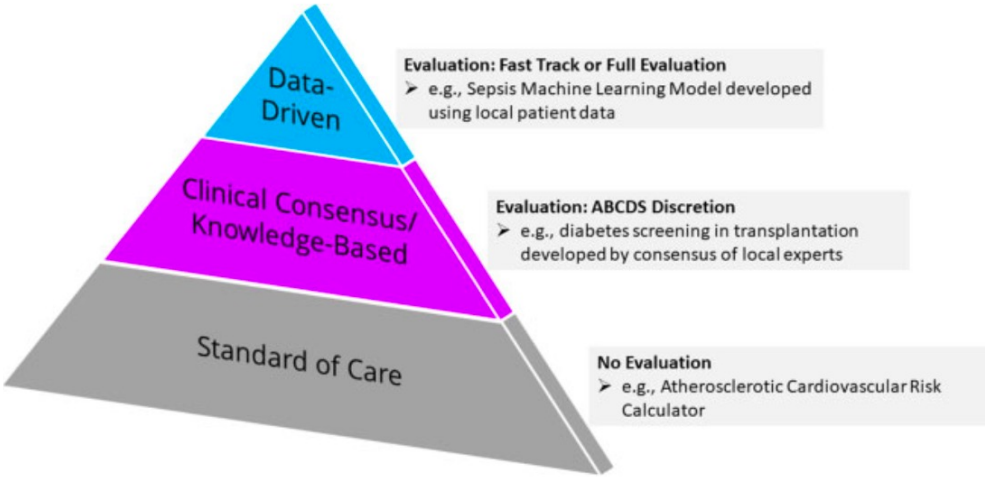
Table 1. ABCDS oversight structure

Subcommittee	Expertise	Responsibilities
Evaluation	Data scientists & physician-scientists with deep expertise on model development & evaluation	• Independently review model performance characteristics • Provide recommendations on model readiness for application • Provide guidance on designing clinical effectiveness study to development teams
Implementation & Monitoring	Individuals from IT with expertise in workflow design & implementation; clinical operational leads with experience in project deployment & change management	• Assist in integrating models into clinical workflow • Guide adoption of best practices in technical & change management • Assist in identifying operational resources to ensure adoption success • Ensures central monitoring of all deployed models
Regulatory	Regulatory & legal backgrounds plus expertise from IRB	• Review FDA regulations, guidances, and policies, and assess applicability to each ABCDS tool • Provide guidance to Oversight Committee on the overall ABCDS framework

A framework for the oversight and local deployment of safe and high-quality prediction models

Bedoya, AD, et al. JAMIA. 2022





AI/ML and CRI:
Other notable papers

- P. J. Embi. **Algorithmovigilance-Advancing Methods to Analyze and Monitor Artificial Intelligence-Driven Health Care for Effectiveness and Equity.** JAMA Netw Open 2021 Vol. 4 Issue 4 Pages e214622
- Y. Park, J. Hu, M. Singh, I. Sylla, I. Dankwa-Mullan, E. Koski, et al. **Comparison of Methods to Reduce Bias From Clinical Prediction Models of Postpartum Depression.** JAMA Netw Open 2021 Vol. 4 Issue 4 Pages e213909
- H. Estiri, Z. H. Strasser, S. Rashidian, J. G. Klann, K. B. Wagholikar, T. H. McCoy, et al. **An objective framework for evaluating unrecognized bias in medical AI models predicting COVID-19 outcomes.** J Am Med Inform Assoc 2022 Vol. 29 Issue 8 Pages 1334-1341

AI/ML and CRI:
Other notable papers

- A. Foryciarz, S. R. Pfohl, B. Patel and N. Shah. **Evaluating algorithmic fairness in the presence of clinical guidelines: the case of atherosclerotic cardiovascular disease risk estimation.** BMJ Health Care Inform 2022 Vol. 29 Issue 1
- L. Oakden-Rayner, W. Gale, T. A. Bonham, M. P. Lungren, G. Carneiro, A. P. Bradley, et al. **Validation and algorithmic audit of a deep learning system for the detection of proximal femoral fractures in patients in the emergency department: a diagnostic accuracy study.** Lancet Digit Health 2022 Vol. 4 Issue 5 Pages e351-e358
- F. Li, P. Wu, H. H. Ong, J. F. Peterson, W. Q. Wei and J. Zhao. **Evaluating and mitigating bias in machine learning models for cardiovascular disease prediction.** J Biomed Inform 2023 Vol. 138 Pages 104294
- M. Yan, M. J. Pencina, L. E. Boulware and B. A. Goldstein. **Observability and its impact on differential bias for clinical prediction models.** J Am Med Inform Assoc 2022 Vol. 29 Issue 5 Pages 937-943

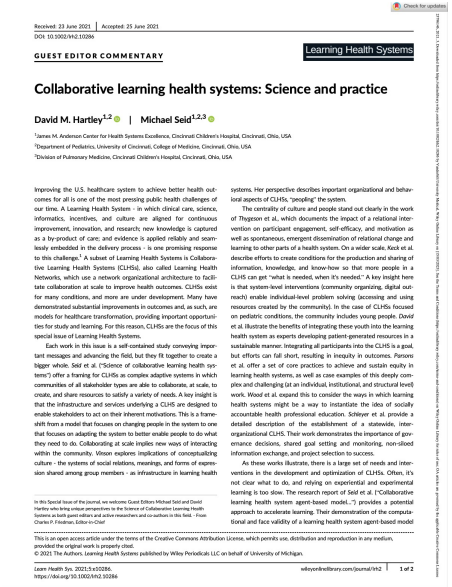
AI/ML and CRI:
Other notable papers

- X. Yang, A. Chen, N. PourNejatian, H. C. Shin, K. E. Smith, C. Parisien, et al. **A large language model for electronic health records.** NPJ Digit Med 2022 Vol. 5 Issue 1 Pages 194
- Y. Yan, T. Schaffter, T. Bergquist, T. Yu, J. Prosser, Z. Aydin, et al. **A Continuously Benchmarked and Crowdsourced Challenge for Rapid Development and Evaluation of Models to Predict COVID-19 Diagnosis and Hospitalization.** JAMA Netw Open 2021 Vol. 4 Issue 10 Pages e2124946
- P. Rajpurkar, E. Chen, O. Banerjee and E. J. Topol. **AI in health and medicine.** Nat Med 2022 Vol. 28 Issue 1 Pages 31-38
- B. Vasey, M. Nagendran, B. Campbell, D. A. Clifton, G. S. Collins, S. Denaxas, et al. **Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI.** Nat Med 2022 Vol. 28 Issue 5 Pages 924-933
    - Adds to previous like: **Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension**

Learning Health Systems and Delivery Science

Collaborative Learning Health Systems

- Journal of Learning Health Systems
- Special issue in July 2021
- Science of collaborative LHS, including methods
- Studies concerned with translational research, e.g. experiments to better understand process or effect.
- Studies concerned with implementation research, testing hypotheses in real-world settings, adjusting for context and setting, applying quality improvement methods, and assess/improve outcomes.



Received: 23 June 2021 | Accepted: 25 June 2021
DOI: 10.1002/lrh2.10286

GUEST EDITOR COMMENTARY

Learning Health Systems

Collaborative learning health systems: Science and practice

David M. Hartley[1,2] | Michael Seid[1,2,3]

[1]James M. Anderson Center for Health Systems Excellence, Cincinnati Children's Hospital, Cincinnati, Ohio, USA
[2]Department of Pediatrics, University of Cincinnati, College of Medicine, Cincinnati, Ohio, USA
[3]Division of Pulmonary Medicine, Cincinnati Children's Hospital, Cincinnati, Ohio, USA

Improving the U.S. healthcare system to achieve better health outcomes for all is one of the most pressing public health challenges of our time. A Learning Health System - in which clinical care, science, informatics, incentives, and culture are aligned for continuous improvement, innovation, and research; new knowledge is captured as a by-product of care; and evidence is applied reliably and seamlessly embedded in the delivery process - is one promising response to this challenge.[1] A subset of Learning Health Systems is Collaborative Learning Health Systems (CLHSs), also called Learning Health Networks, which use a network organizational architecture to facilitate collaboration at scale to improve health outcomes. CLHSs exist for many reasons, and most are under development. Many have demonstrated substantial improvements in outcomes and, as such, are models for healthcare transformation, providing important opportunities for study and learning. For this reason, CLHSs are the focus of this special issue of Learning Health Systems.

Each work in this issue is a self-contained study conveying important messages and advancing the field, but they fit together to create a bigger whole. Seid et al. ("Science of collaborative learning health systems") offer a framing for CLHSs as complex adaptive systems in which communities of all stakeholder types are able to collaborate, at scale, to create, and share resources to satisfy a variety of needs. A key insight is that the infrastructure and services underlying a CLHS are designed to enable stakeholders to act on their inherent motivations. This is a frame-shift from a model that focuses on changing people in the system to one that focuses on adapting the system to better enable people to do what they need to do. Collaborating at scale implies new ways of interacting within the community. *Vinson* explores implications of conceptualizing culture - the systems of social relations, meanings, and forms of expression shared among group members - as infrastructure in learning health systems. Her perspective describes important organizational and behavioral aspects of CLHSs, "peopling" the system.

The centrality of culture and people stand out clearly in the work of Thygeson et al., which documents the impact of a relational intervention on participant engagement, self-efficacy, and motivation as well as spontaneous, emergent dissemination of relational change and learning to other parts of a health system. On a wider scale, Keck et al. describe efforts to create conditions for the production and sharing of information, knowledge, and know-how so that more people in a CLHS can get "what is needed, when it's needed." A key insight here is that system-level interventions (community organizing, digital outreach) enable individual-level problem solving (accessing and using resources created by the community). In the case of CLHSs focused on pediatric conditions, the community includes young people. *David* et al. illustrate the benefits of integrating these youth into the learning health system as experts developing patient-generated resources in a sustainable manner. Integrating all participants into the CLHS is a goal, but efforts can fall short, resulting in inequity in outcomes. *Parsons* et al. offer a set of core practices to achieve and sustain equity in learning health systems, as well as case examples of this deeply complex and challenging (at an individual, institutional, and structural level) work. *Wood* et al. expand this to consider the ways in which learning health systems might be a way to instantiate the idea of socially accountable health professional education. *Schleyer* et al. provide a detailed description of the establishment of a statewide, inter-organizational CLHS. Their work demonstrates the importance of governance decisions, shared goal setting and monitoring, non-siloed information exchange, and project selection to success.

As these works illustrate, there is a large set of needs and interventions in the development and optimization of CLHSs. Often, it's not clear what to do, and relying on experiential and experimental learning is too slow. The research report of Seid et al. ("Collaborative learning health system agent-based model…") provides a potential approach to accelerate learning. Their demonstration of the computational and face validity of a learning health system agent-based model

In this Special Issue of the journal, we welcome Guest Editors Michael Seid and David Hartley who bring unique perspectives to the Science of Collaborative Learning Health Systems as both guest editors and active researchers and co-authors in this field. - Frizan Charles P. Friedman, Editor-in-Chief

Collaborative learning health systems: Science and practice
D. M. Hartley; M. Seid
Learn Health Syst 2021
Vol. 5 Issue 3 Pages e10286

Building a Learning Health System: Creating an Analytical Workflow for Evidence Generation to Inform Institutional Clinical Care Guidelines

D. Dash, A. Gokhale, B. S. Patel, A. Callahan, J. Posada, G. Krishnan, et al.

Appl Clin Inform 2022 Vol. 13 Issue 1 Pages 315-321

- **Objectives:** This article provides a background on the current state of observational data generation in institutional guideline creation and details our institution's experience in creating a novel workflow to
  - (1) demonstrate the value of such a workflow,
  - (2) demonstrate a real-world example, and
  - (3) discuss difficulties encountered and future directions.
- **Methods:** Multidisciplinary team of database specialists, clinicians, and informaticists, created a workflow for identifying and translating a clinical questions into a queriable format in their clinical data warehouse, creating data summaries and feeding this information back into clinical guideline creation.
- **Results**: Clinical questions posed by the hospital medicine division were answered in a rapid time frame and informed creation of institutional guidelines for the care of patients with COVID-19.
- The cost of setting up a workflow, answering the questions, and producing data summaries required around 300 hours of effort and $300,000 USD.
- **Conclusion**: Example of “learning from data” generated during care delivery.
  Good case example with practical lessons learned.

How Dissemination and Implementation Science Can Contribute to the Advancement of Learning Health Systems

K. E. Trinkley, et al. Acad Med 2022

- Challenges to becoming mature learning health systems.
- Some elements have been described (i.e., building system-level supporting infrastructure and the accessibility of inclusive, integrated, and actionable data), others are underrecognized:
  - Balancing evaluation rapidity with rigor, applying principles of health equity and classic ethics, focusing on external validity and reproducibility (generalizability), and designing for sustainability.
- Dissemination and implementation (D&I) science offers solutions.
- This survey of the current state of LHSs, variability, and challenges many face, propose solutions to current challenges, focusing on the contributions of D&I science with other methods, and propose key components and characteristics of a mature LHS model that others can use to plan and develop their LHSs.

How Dissemination and Implementation Science Can Contribute to the Advancement of Learning Health Systems
K. E. Trinkley, et al. Acad Med 2022

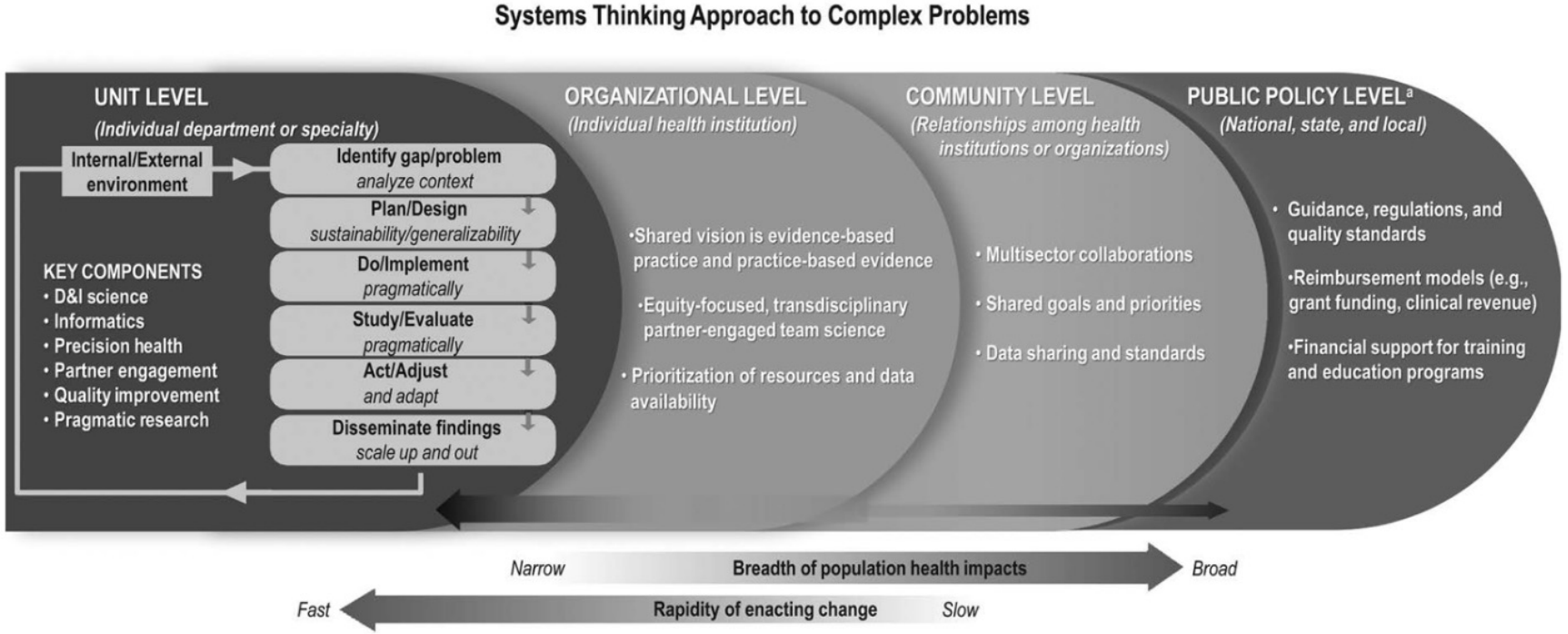


Figure 1 An aspirational model for a mature LHS. Positive outcomes of a fully mature LHS include high value of health care (high-quality, cost-effective, timely, patient-centered), health equity, generalizability, sustainability, and scalability of learning cycle findings. Application of this model addresses the following challenges facing existing LHSs: building system-level supporting infrastructure; the accessibility of inclusive, integrated, and actionable data; balancing evaluation rapidity with rigor; applying principles of health equity and classic ethics; focusing on external validity and reproducibility (generalizability); and designing for sustainability. Abbreviations: LHS, learning health system; D&I, dissemination and implementation. [a]Few examples of LHSs at this level.

Learning Health Systems: Other notable papers

- L. Bastarache, J. S. Brown, J. J. Cimino, D. A. Dorr, et al. **Developing real-world evidence from real-world data: Transforming raw data into analytical datasets.** Learn Health Syst 2022 Vol. 6 Issue 1 Pages e10293
- P. R. O. Payne, A. B. Wilcox, et al. **Better together: Integrating biomedical informatics and healthcare IT operations to create a learning health system during the COVID-19 pandemic.** Learn Health Syst 2022 Vol. 6 Issue 2 Pages e10309
- A. Ostropolets, P. Zachariah, P. Ryan, R. Chen and G. Hripcsak. **Data Consult Service: Can we use observational data to address immediate clinical needs?** J Am Med Inform Assoc 2021 Vol. 28 Issue 10 Pages 2139-2146
- T. Schleyer, L. Williams, J. Gottlieb, C. Weaver, M. Saysana, J. Azar, et al. **The Indiana Learning Health System Initiative: Early experience developing a collaborative, regional learning health system.** Learn Health Syst 2021 Vol. 5 Issue 3 Pages e10281

Learning Health Systems: Other notable papers

- L. Woods, R. Eden, R. Duncan, Z. Kodiyattu, S. Macklin and C. Sullivan. **Which one? A suggested approach for evaluating digital health maturity models.** Front Digit Health 2022 Vol. 4 Pages 1045685
- M. Williams, B. E. Bray, R. A. Greenes, J. McCusker, B. Middleton, G. Perry, et al. **Summary of fourth annual MCBK public meeting: Mobilizing computable biomedical knowledge-metadata and trust.** Learn Health Syst 2022 Vol. 6 Issue 1 Pages e10301
- S. M. Greene and K. L. Holmes. **Learn to fly: Training and competencies to support the multidisciplinary workforce needs of learning health systems.** Learn Health Syst 2022 Vol. 6 Issue 4 Pages e10347
  - Introduction to a special issue focused on training and competencies for LHS
- M. W. Semler, J. D. Casey, B. D. Lloyd, P. G. Hastings, M. A. Hays, J. L. Stollings, et al. **Oxygen-Saturation Targets for Critically Ill Adults Receiving Mechanical Ventilation.** N Engl J Med 2022 Vol. 387 Issue 19 Pages 1759-1769
  - Data-enabled Pragmatic Clinical Trial…

NIH Pragmatic Trials Collaboratory

Initiated through the NIH Common Fund in 2012

Goal: Strengthen the national capacity to implement cost-effective, large-scale research studies that engage healthcare delivery organizations as research partners

Vision: Support the design and execution of innovative pragmatic clinical trial Demonstration Projects to establish best practices and proof of concept

Demonstration Projects

- Pragmatic trials embedded in healthcare
  systems to address questions of major public health importance
- Projects span multiple NIH Institutes, Centers, and Offices
- Projects have 1-year planning phase followed by implementation phase
- Coordinating Center supports methods-focused cores

The Living Textbook *of Pragmatic Clinical Trials*

- www.rethinkingclinicaltrials.org

NIH PRAGMATIC TRIALS COLLABORATORY

Rethinking Clinical Trials®


CRI Ethics & Policy

Defining AMIA's artificial intelligence principles

A. E. Solomonides, et al. JAMIA 2022

- Defines and provides a rationale for principles that should guide the commission, creation, implementation, maintenance, and retirement of AI systems as a foundation for governance throughout the lifecycle.

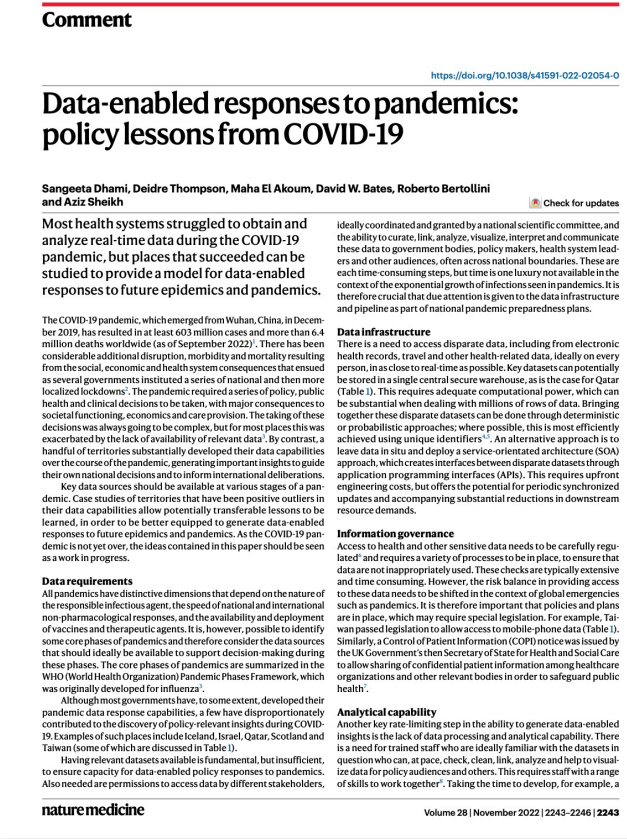
Table 1. Summary of principles governing AI

AI systems		
Rule	Principle	Definitions
I.	Autonomy	AI systems must protect the autonomy of all people and treat them with courtesy and respect including facilitating informed consent.
II.	Beneficence	AI systems must be helpful to people modeled after compassionate, kind, and considerate human behavior.
III.	Nonmaleficence	AI systems shall "do no harm" by avoiding, preventing, and minimizing harm or damage to any stakeholder.
IV.	Justice	AI systems must include equity for people in representation and access to AI, its data, and its benefits. AI must support social justice.
V.	Explainability	AI developers must describe AI systems in context-appropriate language so that their scope, proper application, and limitations are understandable.
VI.	Interpretability	AI developers must endow their systems with the functionality to provide plausible reasoning for decisions or advice in accessible language.
VII.	Fairness	AI systems must be free of bias and must be nondiscriminatory.
VIII.	Dependability	AI systems must be robust, safe, secure, and resilient. Failure must not leave any system in an unsafe or insecure state.
IX.	Auditability	AI systems must provide and preserve a performance "audit trail" including internal changes, model state, input variables, and output for any system decision or recommendation.
X.	Knowledge management	AI systems must be maintained including retraining of algorithms. AI models need listed creation, revalidation, and expiration dates.
Organizations deploying or developing AI		
XI.	Benevolence	Organizations deploying or developing AI must be committed to use AI systems for positive purposes.
XII.	Transparency	AI must be recognizable as such or must announce its nature. AI systems do not incorporate or conceal any special interests and deal even-handedly and fairly with all good faith actors.
XIII.	Accountability	AI systems must be the subject of active oversight by the organization, and any risk attributed to AI must be reported, assessed, monitored, measured, and mitigated as needed. Complaints and redress must be guaranteed.
Special considerations		
XIV.	Vulnerable populations	AI applied to vulnerable populations requires increased scrutiny to avoid worsening the power differential among groups.
XV.	AI research	Academic and industrial research organizations must continue to research AI to address inherent dangers as well as benefits.
XVI.	User education	AI developers have a responsibility to educate healthcare providers and consumers on machine learning and AI systems.

Data-enabled responses to pandemics: policy lessons from COVID-19

S. Dhami, D. Thompson, M. El Akoum, D. W. Bates, R. Bertollini and A. Sheikh
Nat Med 2022

- Key sections:
  - Data requirements
  - Information governance
  - Analytical capabilities
  - Transparency
  - International co-operation
- Ready access to high quality multi-dimensional data is fundamental to generating effective evidence and informed policy responses to pandemics, but most places have struggled with this.
- Many analyses need to extend across international boundaries, which is most likely to be achieved through federated analytical approaches, but will require coordination between governments.
- A few territories have excelled in health data science during the pandemic, which offers a framework that might be developed and deployed in future epidemics and pandemics.



Comment

https://doi.org/10.1038/s41591-022-02054-0

Data-enabled responses to pandemics: policy lessons from COVID-19

Sangeeta Dhami, Deidre Thompson, Maha El Akoum, David W. Bates, Roberto Bertollini and Aziz Sheikh

Check for updates

Most health systems struggled to obtain and analyze real-time data during the COVID-19 pandemic, but places that succeeded can be studied to provide a model for data-enabled responses to future epidemics and pandemics.

The COVID-19 pandemic, which emerged from Wuhan, China, in December 2019, has resulted in at least 603 million cases and more than 6.4 million deaths worldwide (as of September 2022)[1]. There has been considerable additional disruption, morbidity and mortality resulting from the social, economic and health system consequences that ensued as several governments instituted a series of national and then more localized lockdowns[2]. The pandemic required a series of policy, public health and clinical decisions to be taken, with major consequences to societal functioning, economics and care provision. The taking of these decisions was always going to be complex, but for most places this was exacerbated by the lack of availability of relevant data[3]. By contrast, a handful of territories substantially developed their data capabilities over the course of the pandemic, generating important insights to guide their own national decisions and to inform international deliberations.

Key data sources should be available at various stages of a pandemic. Case studies of territories that have been positive outliers in their data capabilities allow potentially transferable lessons to be learned, in order to be better equipped to generate data-enabled responses to future epidemics and pandemics. As the COVID-19 pandemic is not yet over, the ideas contained in this paper should be seen as a work in progress.

Data requirements

All pandemics have distinctive dimensions that depend on the nature of the responsible infectious agent, the speed of national and international non-pharmacological responses, and the availability and deployment of vaccines and therapeutic agents. It is, however, possible to identify some core phases of pandemics and therefore consider the data sources that should ideally be available to support decision-making during these phases. The core phases of pandemics are summarized in the WHO (World Health Organization) Pandemic Phases Framework, which was originally developed for influenza[3].

Although most governments have, to some extent, developed their pandemic data response capabilities, a few have disproportionately contributed to the discovery of policy-relevant insights during COVID-19. Examples of such places include Iceland, Israel, Qatar, Scotland and Taiwan (some of which are discussed in Table 1).

Having relevant datasets available is fundamental, but insufficient, to ensure capacity for data-enabled policy responses to pandemics. Also needed are permissions to access data by different stakeholders, ideally coordinated and granted by a national scientific committee, and the ability to curate, link, analyze, visualize, interpret and communicate these data to government bodies, policy makers, health system leaders and other audiences, often across national boundaries. These are each time-consuming steps, but time is one luxury not available in the context of the exponential growth of infections seen in pandemics. It is therefore crucial that due attention is given to the data infrastructure and pipeline as part of national pandemic preparedness plans.

Data infrastructure

There is a need to access disparate data, including from electronic health records, travel and other health-related data, ideally on every person, in as close to real-time as possible. Key datasets can potentially be stored in a single central secure warehouse, as is the case for Qatar (Table 1). This requires adequate computational power, which can be substantial when dealing with millions of rows of data. Bringing together these disparate datasets can be done through deterministic or probabilistic approaches; where possible, this is most efficiently achieved using unique identifiers[4,5]. An alternative approach is to leave data in situ and deploy a service-orientated architecture (SOA) approach, which creates interfaces between disparate datasets through application programming interfaces (APIs). This requires upfront engineering costs, but offers the potential for periodic synchronized updates and accompanying substantial reductions in downstream resource demands.

Information governance

Access to health and other sensitive data needs to be carefully regulated[6] and requires a variety of processes to be in place, to ensure that data are not inappropriately used. These checks are typically extensive and time consuming. However, the risk balance in providing access to these data needs to be shifted in the context of global emergencies such as pandemics. It is therefore important that policies and plans are in place, which may require special legislation. For example, Taiwan passed legislation to allow access to mobile-phone data (Table 1). Similarly, a Control of Patient Information (COPI) notice was issued by the UK Government's then Secretary of State for Health and Social Care to allow sharing of confidential patient information among healthcare organizations and other relevant bodies in order to safeguard public health[7].

Analytical capability

Another key rate-limiting step in the ability to generate data-enabled insights is the lack of data processing and analytical capability. There is a need for trained staff who are ideally familiar with the datasets in question who can, at pace, check, clean, link, analyze and help to visualize data for policy audiences and others. This requires staff with a range of skills to work together[8]. Taking the time to develop, for example, a

CRI Ethics & Policy: Other notable papers

- L. Cook, J. Espinoza, N. G. Weiskopf, N. Mathews, D. A. Dorr, K. L. Gonzales, et al. **Issues With Variability in Electronic Health Record Data About Race and Ethnicity: Descriptive Analysis of the National COVID Cohort Collaborative Data Enclave.** JMIR Med Inform 2022 Vol. 10 Issue 9 Pages e39235
- S. Arvisais-Anhalt, A. Ravi, B. Weia, J. Aarts, H. B. Ahmad, E. Araj, et al.. **Paging the Clinical Informatics Community: Respond STAT to Dobbs v. Jackson's Women's Health Organization.** Appl Clin Inform 2023 Vol. 14 Issue 1 Pages 164-171
- E. W. Clayton, P. J. Embi and B. A. Malin. **Dobbs and the future of health data privacy for patients and healthcare organizations.** J Am Med Inform Assoc 2022 Vol. 30 Issue 1 Pages 155-160
- D. W. Bates. **How to regulate evolving AI health algorithms.** Nat Med 2023 Vol. 29 Issue 1 Pages 26

CRI Ethics & Policy:
Other notable papers

- R. M. Califf. **Now is the time to fix the evidence generation system.** Clin Trials 2023 Vol. 20 Issue 1 Pages 3-12
- S. Emani, J. A. Rodriguez and D. W. Bates. **Racism and Electronic Health Records (EHRs): Perspectives for research and practice.** J Am Med Inform Assoc 2023
- E. A. Mendonca, R. L. Richesson, H. Hochheiser, D. M. Cooper, M. N. Bruck and E. S. Berner. **Informatics education for translational research teams: An unrealized opportunity to strengthen the national research infrastructure.** J Clin Transl Sci 2022 Vol. 6 Issue 1 Pages e130
- S. R. Morain, J. Bollinger, K. Weinfurt and J. Sugarman. **Ethics challenges in sharing data from pragmatic clinical trials.** Clin Trials 2022 Vol. 19 Issue 6 Pages 681-689

Notable CRI News/Events

COVID and CRI

- Many funding opportunities– significant research funding
- Leveraging our data and capabilities
- Monitor and track, it's not over
- Pivoting now to longer-term, lessons…
- Launch of long-covid studies initiative
  - Much work to be done
  - Our specialty at the center
  - We're all counting on each other!

Open Funding Opportunities

Explore COVID-19 funding opportunities from NIH.

Trying to Make Sense of Long COVID Syndrome

Posted on January 19th, 2021 by Dr. Francis Collins

Credit: NIH

NIH National Institutes of Health
Turning Discovery Into Health

Search NIH

NIH Employee Intranet | Staff Directory | En Español

Health Information | Grants & Funding | News & Events | Research & Training | Institutes at NIH | About NIH

COVID-19

- Get the latest public health information from CDC
- NIH staff guidance on coronavirus (NIH Only)
- Get the latest research information from NIH | Español

Home » About NIH » Who We Are » The NIH Director

THE NIH DIRECTOR

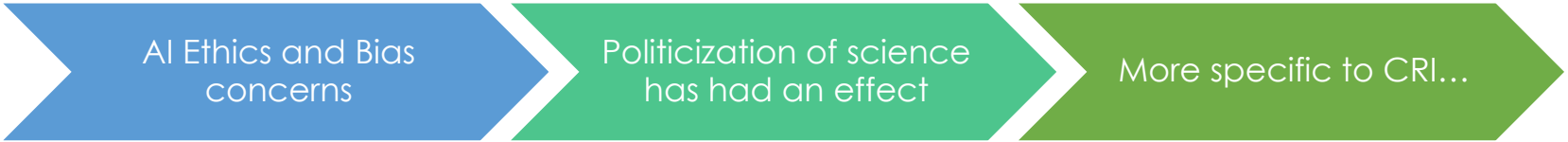
The NIH Director
Photo Gallery
Congressional Testimonies
Advisory Groups
Video & Sound Gallery
Articles
Statements

February 23, 2021

NIH launches new initiative to study "Long COVID"

I write to announce a major new NIH initiative to identify the causes and ultimately the means of prevention and treatment of individuals who have been sickened by COVID-19, but don't recover fully over a period of a few weeks. Large numbers of patients who have been infected with SARS-CoV-2 continue to experience a constellation of symptoms long past the time that they've recovered from the initial stages of COVID-19 illness. Often referred to as "Long COVID", these symptoms, which can include fatigue, shortness of breath, "brain fog", sleep disorders, fevers, gastrointestinal symptoms, anxiety, and depression, can persist for months and can range from mild to incapacitating. In some cases, new symptoms arise well after the time of infection or evolve over time. In December, NIH held a workshop to summarize what is known about these patients who do

Notable CRI-Related Events:
Societal events continue to affect health/research/informatics



- Examples in major industries, including health
- Opportunities for us to lead and address concerns we understand best

- Must continue do good work to address anti-science sentiment
- Financial issues plaguing heathcare starting to affect academy

- RWD/RWE boom
- AI/ML boom
- New capabilities to enable democratization of data
- Continued growth in CRI leadership roles (e.g. CRIO)
- Pragmatic research on the rise
- Continued focus on intersection with LHS

NIH Rule on Data Sharing

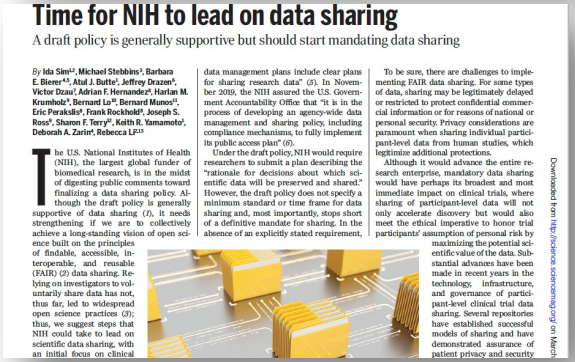
- During comment period, feedback from our community, including AMIA and thought leaders
- Final data sharing rule released in October 2020.
- Took effect on Jan 25, 2023
- https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

AMIA: NIH Misses Mark on Data Sharing Proposals

Monday, January 13, 2020

Nation's health informatics experts urge NIH to dramatically revise draft data management and sharing policy to maximize the value of scientific data

BETHESDA, MD – Last week the American Medical Informatics Association (AMIA) warned the National Institutes of Health (NIH) that its proposed data management and sharing policy would be detrimental to data-driven discovery and lead to increased compliance burdens for researchers. The organization



Time for NIH to lead on data sharing

A draft policy is generally supportive but should start mandating data sharing

By Ida Sim[1,2], Michael Stebbins[1], Barbara E. Bierer[4,5], Atul J. Butte[1], Jeffrey Drazen[6], Victor Dzau[7], Adrian F. Hernandez[8], Harlan M. Krumholz[9], Bernard Lo[10], Bernard Munos[11], Eric Perakslis[12], Frank Rockhold[13], Joseph S. Ross[9], Sharon F. Terry[14], Keith R. Yamamoto[1], Deborah A. Zarin[4], Rebecca Li[2,3]

The U.S. National Institutes of Health (NIH), the largest global funder of biomedical research, is in the midst of digesting public comments toward finalizing a data sharing policy. Although the draft policy is generally supportive of data sharing (1), it needs strengthening if we are to collectively achieve a long-standing vision of open science built on the principles of findable, accessible, interoperable, and reusable (FAIR) (2) data sharing. Relying on investigators to voluntarily share data has not, thus far, led to widespread open science practices (3); thus, we suggest steps that NIH could take to lead on scientific data sharing, with an initial focus on clinical data management plans include clear plans for sharing research data" (5). In November 2019, the NIH assured the U.S. Government Accountability Office that "it is in the process of developing an agency-wide data management and sharing policy, including compliance mechanisms, to fully implement its public access plan" (6).

Under the draft policy, NIH would require researchers to submit a plan describing the "rationale for decisions about which scientific data will be preserved and shared." However, the draft policy does not specify a minimum standard or time frame for data sharing and, most importantly, stops short of a definitive mandate for sharing. In the absence of an explicitly stated requirement, To be sure, there are challenges to implementing FAIR data sharing. For some types of data, sharing may be legitimately delayed or restricted to protect confidential commercial information or for reasons of national or personal security. Privacy considerations are paramount when sharing individual participant-level data from human studies, which legitimize additional protections.

Although it would advance the entire research enterprise, mandatory data sharing would have perhaps its broadest and most immediate impact on clinical trials, where sharing of participant-level data will not only accelerate discovery but would also meet the ethical imperative to honor trial participants' assumption of personal risk by maximizing the potential scientific value of the data. Substantial advances have been made in recent years in the technology, infrastructure, and governance of participant-level clinical trial data sharing. Several repositories have established successful models of sharing and have demonstrated assurance of patient privacy and security

National Institutes of Health
Office of Science Policy

NIH Data Management and Sharing Activities Related to Public Access and Open Science

Validation and progress in biomedical research – the cornerstone of developing new prevention strategies, treatments, and cures – is dependent on access to scientific data. Sharing scientific data helps validate research results, enables researchers to combine data types to strengthen analyses, facilitates reuse of hard to generate data or data from limited sources, and accelerates ideas for future research inquiries. Central to sharing scientific data is the recognized need to make data as available as possible while ensuring that the privacy and autonomy of research participants are respected, and that confidential/proprietary data are appropriately protected.

Scientific Data Sharing

› Genomics and Health

› Scientific Data Management

Related to Public Access and Open Science

NEW Final NIH Policy for Data Management and Sharing and Supplemental Information (October 2020)

- Federal Register Notice
- NIH Guide Notices:
  - NOT-OD-21-013 – Final NIH Policy for Data Management and Sharing
  - NOT-OD-21-014– Supplemental Information to the NIH Policy for Data Management and Sharing: Elements of an NIH Data Management and Sharing Plan

More Notable CRI-Related Events

- Request for Information on the NIH Plan to Enhance Public Access to the Results of NIH-Supported Research
  - https://grants.nih.gov/grants/guide/notice-files/NOT-OD-23-091.html
- White House's Office of Science and Technology Policy: Blueprint for an AI Bill of Rights
  - https://www.whitehouse.gov/ostp/ai-bill-of-rights/
- OSTP, in Partnership with ONC, Seeks Input on Optimizing Data Capture for Clinical Trials
  - https://www.whitehouse.gov/ostp/news-updates/2022/10/31/ostp-in-partnership-with-onc-seeks-input-on-optimizing-data-capture-for-clinical-trials/
- Growing influence of Informatics across
- health and biomedicine
- 3rd edition of CRI textbook will be out this summer!
- New NEJM AI journal – adding to others like PLOS, etc.
- And so much more…

OCTOBER 31, 2022

OSTP, in Partnership with ONC, Seeks Input on Optimizing Data Capture for Clinical Trials

OSTP › BRIEFING ROOM › OSTP BLOG

By Micky Tripathi, National Coordinator for Health Information Technology, HHS

Jennifer Roberts, Assistant Director for Health Technologies, OSTP

Grail Sipes, Assistant Director for Biomedical Regulatory Policy, OSTP

In Summary…

- Informatics and CRI/Implementation Science has never been more important
  - **Much** more activity than in years past
- CRI continues to *mature* and is clearly driving science and… translation into practice!
- COVID changed a LOT – including CRI
- Multiple federal, state and local initiatives continue to advance field
- Initiatives and investments beginning to realize the vision of the "*learning health system*"
- No question CRI is and will remain relevant
- A **very** exciting time to be in CRI!

Thanks!

Special thanks to those who suggested articles/events to highlight, particularly

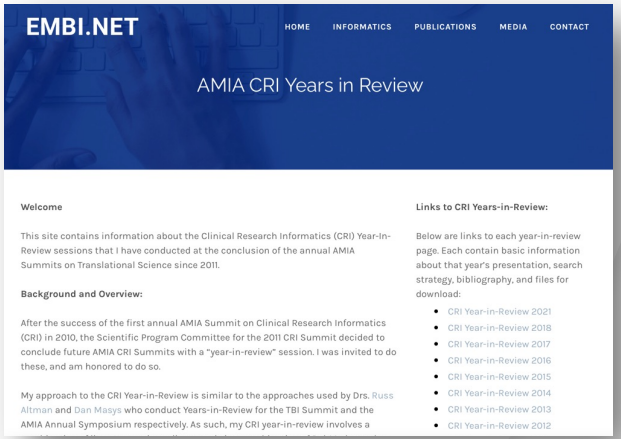
- Rachel Richesson
- Sean Mooney
- Erin Holve
- Shawn Murphy
- Paul Harris
- Adam Wilcox
- Dave Dorr
- Melissa Haendel
- Lucy Savitz
- Boyd Knosp
- Matt Semler
- Cheryl Gatto
- Brad Malin
- Laurie Novak
- Michael Matheny
- Luke Rassmusen
- Nigam Shah
- Peter Elkin
- George Hripcsak
- Nick Anderson

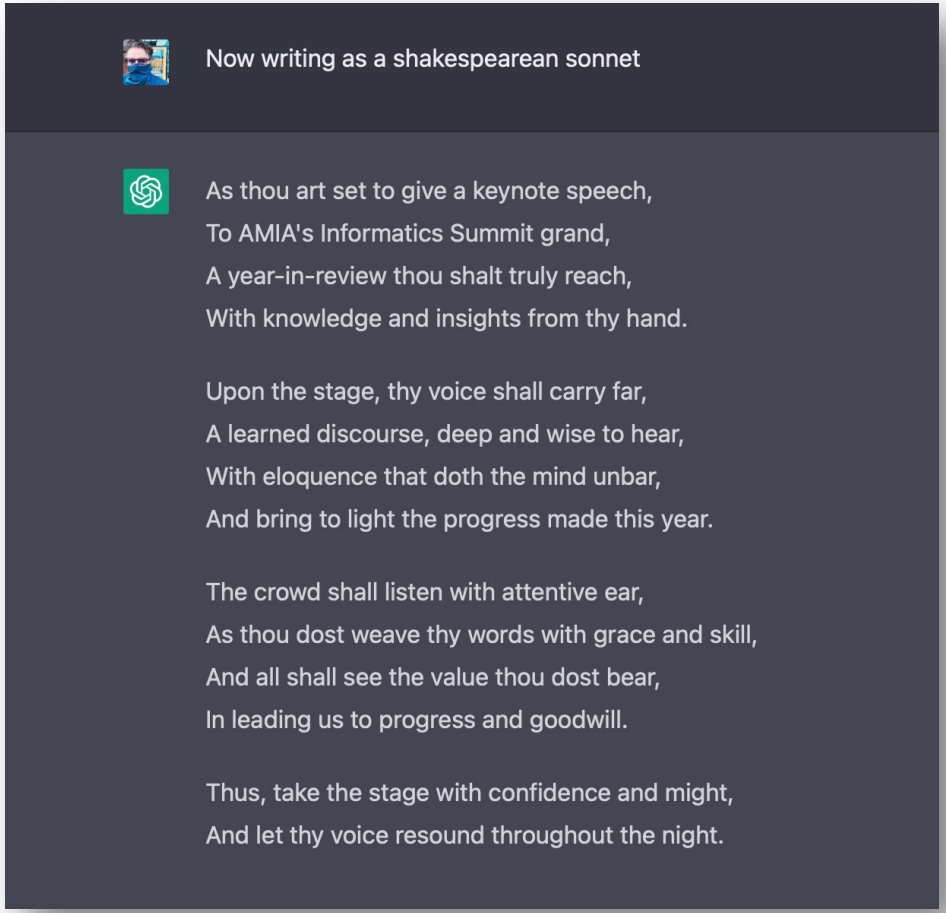
Thanks!

peter.embi@vumc.org

Twitter: @embimd

Slideshow, files:
http://www.embi.net/





Q&A