



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Review

# Electronic Health Record Network Research in Infectious Diseases

Ravi Jhaveri, MD<sup>1,2</sup>; Jordan John, BS<sup>2</sup>; and Marc Rosenman, MD<sup>2,3</sup>

<sup>1</sup>Division of Pediatric Infectious Diseases, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois; <sup>2</sup>Northwestern University Feinberg School of Medicine, Chicago, Illinois; and <sup>3</sup>Mary Ann & J. Milburn Smith Child Health Outcomes, Research, and Evaluation Center, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois

## ABSTRACT

With the marked increases in electronic health record (EHR) use for providing clinical care, there have been parallel efforts to leverage EHR data for research. EHR repositories offer the promise of vast amounts of clinical data not easily captured with traditional research methods and facilitate clinical epidemiology and comparative effectiveness research, including analyses to identify patients at higher risk for complications or who are better candidates for treatment. These types of studies have been relatively slow to penetrate the field of infectious diseases, but the need for rapid turnaround during the COVID-19 global pandemic has accelerated the uptake. This review discusses the rationale for her network projects, opportunities and challenges that such networks present, and some prior studies within the field of infectious diseases. (*Clin Ther.* 2021;43:1668–1680.) © 2021 Elsevier Inc.

**Key words:** EHR network, electronic health records, infectious diseases, PCORnet.

## INTRODUCTION

In the past decade, the field of medicine has been transformed in both positive and negative ways by widespread use of the electronic health record (EHR). With EHRs offering the promise of increased data access for clinicians and incentivized by passage of the Patient Protection and Affordable Care Act and the Centers for Medicare & Medicaid Services meaningful use program, approximately 90% of physicians and hospitals use EHR systems.<sup>1</sup> Although initially designed to improve clinical decision support,

information sharing between practitioners, scheduling, and accounting, billing, and revenue (the major driver of the ongoing development and proliferation of EHRs), researchers have steadily realized that EHR systems can facilitate many aspects of clinical research that are not possible with traditional clinical trials.<sup>2</sup> Given the historical emphasis within the field of infectious diseases on laboratory-based and traditional clinical research, its embrace of EHR-based research has been slow. However, the SARS-CoV-2 pandemic has created an environment that has accelerated the uptake of EHR-based studies. This review discusses the limitations of traditional clinical studies, highlights the capabilities, advantages, and challenges of EHR-based research, and briefly summarizes several studies in the field of infectious diseases. This review does not cover the use of EHR to implement best practices or quality improvement initiatives and limits the discussion to research uses of the EHR.

## LIMITATIONS OF THE GOLD STANDARD CLINICAL STUDIES

The current goal of most medical clinicians is to practice evidence-based medicine. However, not all evidence is equal. Several articles have discussed how the evidence pyramid should be adjusted, but randomized clinical trials (RCTs) are consistently considered the highest-quality evidence followed by large cohort studies, case-control studies, and then case series and reports at the lowest level.<sup>3</sup> To be

Accepted for publication September 2, 2021  
<https://doi.org/10.1016/j.clinthera.2021.09.002>  
0149-2918/\$ - see front matter

© 2021 Elsevier Inc.

successful, RCTs require relatively large patient sample sizes, frequently require multiple centers, and usually require significant financial resources [Figure 1A](#).<sup>4</sup> In infectious diseases, many conditions are relatively uncommon and are sporadic in nature, neither of which is conducive to conducting an RCT. As a result, the field is perpetually caught in the cycle of trying to conduct better case-control and case series studies, optimizing the lower part of the pyramid. Systematic reviews and meta-analyses are often used to try to cobble together many studies to solve the problem of small sample sizes, but these methods can only partially compensate for and cannot completely erase the fundamental methodologic flaws of these studies.

Researchers have tried to compensate for some of these limitations by creating large cohorts using regional or national administrative claims databases. Usually established by large public or private insurers, these databases allow researchers to identify patients with a common diagnostic condition (*International Classification of Diseases, Tenth Revision* [ICD-10] code) or procedure (*Current Procedural Terminology* or other code), establish frequencies across a large population, and collect all the associated health care events associated with the specified population. These types of studies can be helpful to define patterns of practice and highlight large variations in practice across hospital systems and regions. However, claims data usually have only dispensing records (and not orders) for medications, limited or no detailed laboratory results or radiographic data, and no real-time clinical information, such as vital signs or pain ratings.<sup>5,6</sup>

Given the limitations of the evidence base, clinicians in the real world are left to practice the art of medicine. They encounter patients with relatively infrequent conditions, they factor in their own personal experience treating patients with similar conditions along with evidence from similar conditions, and they come up with a custom plan. The clinicians who have more extensive experience and who are better at synthesizing this kind of customized treatment approach are often acknowledged as master clinicians, but the best they can do is summarize their experience in a case series that then gets rated as low-quality evidence. As a result, the field of infectious diseases continues with the status quo, with practitioners claiming, “There are no good data on this topic!”

## FUNDAMENTALS OF EHR-BASED RESEARCH NETWORKS

What if the collective experience of these practitioners could be combined into a large real-world cohort database with patients divided into subgroups based on the different treatment they receive? Consider the example of the invasive fungal infection mucormycosis.<sup>7</sup> Because this infection is a rare event, RCTs are virtually impossible. However, the collective number of patients seen across most of the United States may be significant, with different institutions having slightly different practices. There is ongoing debate about whether different antifungal prophylaxis strategies have a significant impact on rates of invasive mucormycosis. One could develop a real-world study on the number of cases of mucormycosis to compare the effect of using voriconazole versus posaconazole as prophylaxis in high-risk oncology patients and patients undergoing hematopoietic stem cell transplantation. The clinical information for all patients in such a database would be standardized so data could be pooled easily and analyzed across institutions [Figure 1B](#). In an established, multicenter, multipurpose (inclusive of patients with any health condition or health care encounter) EHR-based database, the costs to collect such data for mucormycosis would be far less than what is required for running a multicenter clinical trial and would offer the advantage of results obtained in actual practice. These are the potential benefits that can be realized with EHR-based research networks. Another advantage to EHR-based clinical networks is that one can take an unbiased approach to the data for discovery purposes. In the mucormycosis example, one could identify that although the number of fungal infections did not differ between the 2 antifungal agents being compared, 1 treatment led to increased laboratory abnormalities that required additional interventions. Using an administrative claims database, one needs to predefine which variables to evaluate and thus would not find an association unless specifically searching for it (a traditional biased approach).

Alongside the potential benefits, EHR-based observational studies themselves have noteworthy limitations and potential for bias. Without randomization, statistical techniques, such as propensity scores, must be applied in an attempt to handle baseline differences between groups. Such techniques have limitations,<sup>8-10</sup> and EHR-derived data are at risk for bias for other reasons as well. The challenges involve those inherent

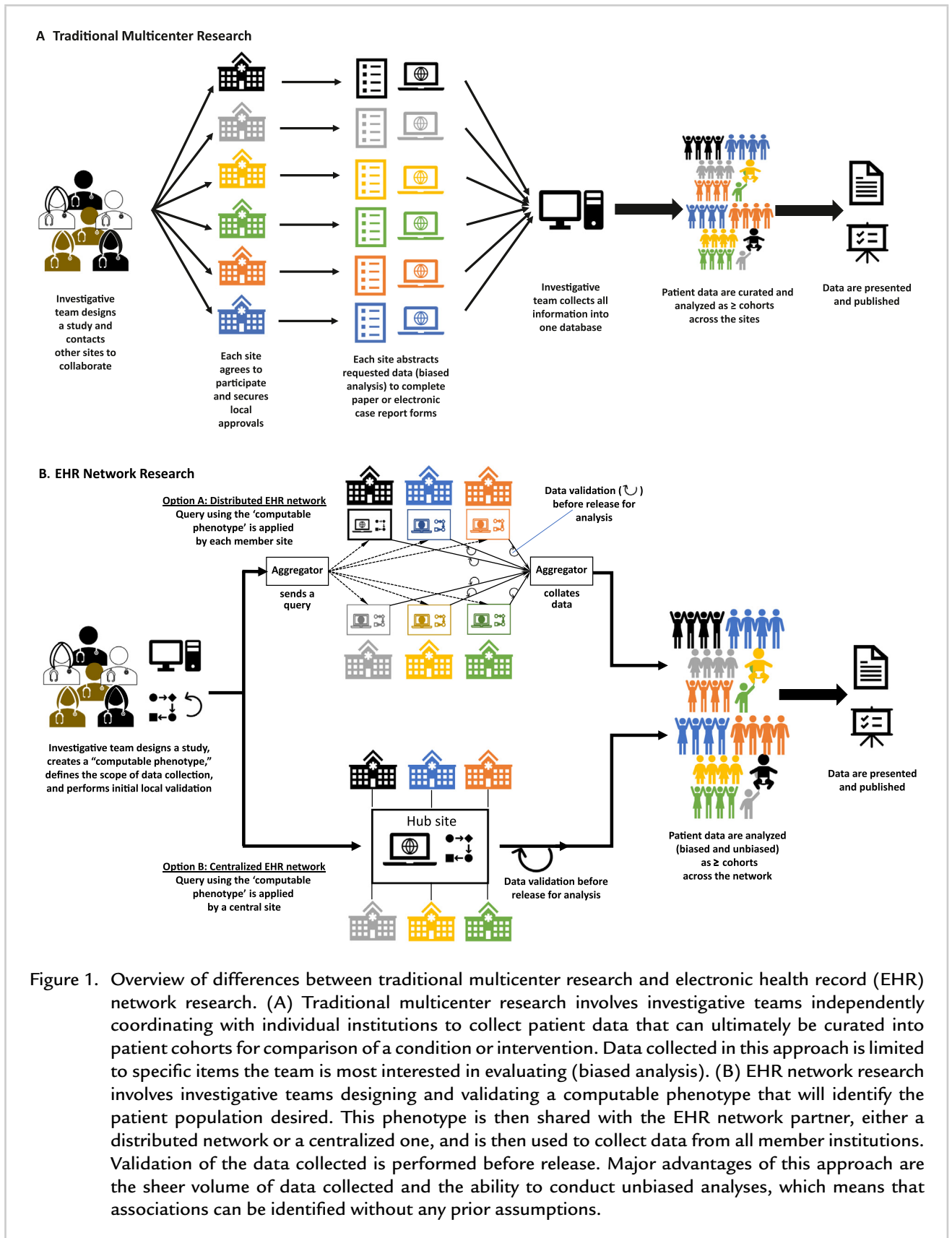


Figure 1. Overview of differences between traditional multicenter research and electronic health record (EHR) network research. (A) Traditional multicenter research involves investigative teams independently coordinating with individual institutions to collect patient data that can ultimately be curated into patient cohorts for comparison of a condition or intervention. Data collected in this approach is limited to specific items the team is most interested in evaluating (biased analysis). (B) EHR network research involves investigative teams designing and validating a computable phenotype that will identify the patient population desired. This phenotype is then shared with the EHR network partner, either a distributed network or a centralized one, and is then used to collect data from all member institutions. Validation of the data collected is performed before release. Major advantages of this approach are the sheer volume of data collected and the ability to conduct unbiased analyses, which means that associations can be identified without any prior assumptions.

in any observational study plus complexities specific to EHRs, clinical workflows, and patterns of care.<sup>11-13</sup> Two oft-cited (>6000-fold taken together) articles published 21 years ago put forth the promise of observational studies, finding that they had results and effect sizes similar to those of RCTs of the same research questions,<sup>14,15</sup> but, as Banerjee and Prasad<sup>16</sup> and others<sup>17,18</sup> have summarized, considerable evidence has emerged in recent years to the contrary. In many cases, the answer to the question, “Can real-world data really replace RCTs?” is no.<sup>19-21</sup> Nevertheless, observational data have an important role to play in studies that develop the evidence base and in fostering learning health systems.<sup>16,22,23</sup> EHR-based research networks can help build the necessarily nuanced methods for applying and for understanding when and how to apply observational EHR data on their own or as complementary data in the conceptualization, implementation, and analyses of pragmatic RCTs.<sup>24-28</sup> A key prerequisite for the success of EHR research networks is to set up the system so that a common data language and format is used by all member institutions within the network; the institutions must also extract, transform, and load their local EHR repository data into the network database with good quality and completeness. Data relevant to the particular studies that use data from the network should also be easily shareable. Another key limitation of EHR-based research is the challenge of applying criteria or clinical severity scores. The Sequential Organ Failure Assessment (SOFA) score is an example of a measure that was developed primarily in adults that could be evaluated across many institutions. However, the SOFA score may not be applicable to many other patient subpopulations (infants, children, and pregnant women), so this limits EHR-based research in these groups to the clinical variables directly assessed during care.<sup>29</sup>

### WHAT DOES AN OPTIMIZED EHR NETWORK LOOK LIKE?

Administrative claims databases have long been used as an effective tool for outcomes research in medicine. Many investigators have been reluctant to let go of using these databases and ask, “What do EHR data networks offer that claims databases don’t?” DeShazo and Hoffman<sup>30</sup> conducted a study designed to answer this question. They compared a frequently used national administrative claims database (Agency for Healthcare Research and Quality Healthcare

Cost and Utilization Project Nationwide Inpatient Sample) with inpatient records in Health Facts, an EHR network maintained by Cerner Corporation. For general demographic characteristics and discharges, the 2 sources offered similar information. While most diagnoses were similar in frequency between the 2 sources, a few notable categories were more often represented in the Nationwide Inpatient Sample versus Health Facts: Psychiatry and Behavioral Disorders, Obstetrics & Gynecology, Pregnancy and Neonatal/Perinatal Disorders, and (we noted) Infectious Diseases. These resources note that hospitals in the Nationwide Inpatient Sample skewed smaller (<99 beds) and were less likely to have fully converted their systems to EHRs (as of 2010, when this comparison study was conducted). This observation highlights one of the limitations of EHR network research, which is that larger systems and institutions are generally the ones with the resources to adopt and maintain EHR systems and to participate in the new EHR data-sharing networks discussed below. This limitation creates a bias in representation that needs to be accounted for when conducting population-based studies. The National Inpatient Sample is weighted to help the studies that analyze its data be nationally representative.

Butame et al<sup>31</sup> analyzed the barriers and facilitators to EHR data collection in HIV trials in adolescents. They performed qualitative interviews with national experts and key stakeholders in the field of adolescent HIV. They found that respondents used 17 different EHR systems (among 29 participants), and several used >1 at their site to achieve different tasks (billing vs patient management vs reporting requirements from federal funders, such as the US Health Resources and Services Administration). Other barriers were identified: specific variable collection, general data collection, and HIV-specific variable data collection. The authors summarize their findings as describing a “fractured EHR landscape,” a phrase that summarizes the situation for most practitioners and researchers well beyond the field of infectious diseases. Many of the barriers highlighted in this study are addressed when using the PCORnet Common Data Model (CDM), but some of the barriers highlighted still exist.

To avoid some of the barriers mentioned, an EHR network would need to have developed a system that facilitates data sharing, formatting, and harmonization across institutions, with a robust process for improving and maintaining the quality and completeness of the

data. Fortunately for clinical investigators based in the United States, such an EHR-based data collection system already exists: PCORnet, the National Patient-Centered Clinical Research Network.

Guided by the Patient-Centered Outcomes Research Institute, Dr Francis Collins, and various stakeholders, PCORnet was created in 2014 and has been expanding and evolving in the ensuing years. The impetus for PCORnet was, in part, what is discussed above: to address some of the key challenges in building large, multi-institution study cohorts. Local EHR data repositories are idiosyncratic and not easy to combine across institutions. The overarching rationale for PCORnet was to be able to strengthen the clinical evidence base by making large pragmatic trials much more feasible.<sup>32</sup> PCORnet is a distributed network in which each of the institutions extracts, transforms, and loads (ETLs) data from its local EHR repository into a common data model data mart, which the institution also maintains locally. The PCORnet CDM data mart specifications are issued by the PCORnet coordinating center, which is run (2014 to the present) jointly by Duke University and Harvard Pilgrim. The specifications address not only the data mart format (tables and fields) but also the content by requiring data element values and units of measure to conform to standards (RxNorm, LOINC [Logical Observation Identifiers Names and Codes], and others, by providing guidance on mapping local values to the common terminologies, and through quarterly refreshes that are accompanied by data quality and completeness checks that have increased in stringency.<sup>33</sup> The common data model format, along with a query-and-data transmission software that also has been adopted by every participating institution, allows the coordinating center to issue a query that can be run at all participating institutions. The query outputs (aggregate results or individual-level data sets) then are collated by the coordinating center and/or by a project's investigator team.

The harmonization of formats, content, and processes creates substantial efficiencies. In addition, each data mart includes *all* of an institution's patients who had a health care record from 2011 forward; therefore, studies of clinical research questions in any specialty or subspecialty can be pursued. An early example from PCORnet in pediatrics is a study of whether there is an association between early antibiotic exposure and weight gain.<sup>34</sup> A recent example is the Centers for Disease Control and Prevention (CDC)-

sponsored COVID-19 project that began in May 2020; this project provides the CDC with regular reports on data related to the pandemic.<sup>35</sup> Forty-three PCORnet institutions are participating. Every specific project affords an opportunity to improve the PCORnet data marts with regard to that project's particular data elements or required collaborative processes; the CDC COVID-19 project prompted each site to improve its ETL procedures for SARS-CoV-2 testing and vaccination data and to establish more frequent (approximately biweekly) data refreshes.

As of the middle of 2021, PCORnet contained 9 clinical research networks (with EHR-derived data for a total of approximately 80 million patients) and 2 health care payer data marts (with claims-derived data for a total of approximately 60 million patients).<sup>36</sup> The 9 clinical research networks include a total of 70 EHR data-contributing institutions, within which there are 337 hospitals, approximately 170,000 physicians, and more >3500 primary care practices.<sup>36</sup> Each of the 9 clinical research networks has a hub site and may perform some studies on its own (without the PCORnet coordinating center). Many of the clinical research networks are in a particular geographic area (eg, the Greater Plains Collaborative includes institutions from the Midwest to Texas); by contrast, the PEDSnet clinical research network is composed of an increasing number of children's hospitals from across the country.

In PCORnet (and in any other EHR data network), a common data model and harmonization are helpful, up to a point. With inclusion criteria that include all patients since 2011 and with the large number of data tables and fields, not everything can be audited (although the quarterly data checks have continued to expand), not all scenarios can have been envisioned, and not all of the underlying idiosyncrasies can have been resolved.<sup>33,37</sup> Even if application of the PCORnet data mart specifications were perfect at all institutions, there would still be some variation because of the data provenance upstream of the PCORnet data marts (eg, the electronic systems and data flows and repositories that a hospital has in place for its surgery data or for its physician billing data may affect how much data and what types make it into the PCORnet data mart). In addition, many PCORnet data-contributing institutions are complex institutions themselves. Many are composed of >1 hospital. Most have multiple affiliated outpatient clinic locations. Understandably, there tends to be intrainstitutional variation, upstream

of one's PCORnet data mart, that affects how various data types do or do not flow in from the various reaches of the institution. These situations are related to the fundamental challenges that PCORnet and other similarly motivated multi-institutional networks, such as Sentinel, arose to address.<sup>38</sup>

Institutions also have seen the need to craft (1) research data repositories that are more expansive than a PCORnet data mart and (2) research collaborations that cut across institutions and data models to foster learning health systems science in various specialty and subspecialty domains. Duke University built the Duke Clinical Research Datamart by incorporating the PCORnet data model plus other data types and data tables.<sup>39</sup> PEDSnet is a PCORnet clinical research network and maintains a PCORnet data mart to participate in PCORnet projects but also has designed and implemented a PEDSnet Common Data Model.<sup>23</sup> The PEDSnet Common Data Model is based largely on the robust Observational Medical Outcomes Partnership (OMOP) model.<sup>40,41</sup> Using an OMOP-like model allows PEDSnet to expand its pediatric research capabilities and its data quality monitoring infrastructure,<sup>42-44</sup> PEDSnet is an exemplar for networks that seek to be (and to build) learning health systems<sup>22,23</sup>; it has facilitated collaboration among subspecialists, generalists, and informaticians in the fields of gastroenterology, nephrology, and several others.<sup>42,45</sup>

Other networks also have been building unprecedented opportunities for observational research. The OMOP data model was initiated in 2007 by a public-private partnership among the US Food and Drug Administration, the Foundation for the National Institutes of Health, and Pharmaceutical Research and Manufacturers of America.<sup>46</sup> Seven years later, a new organization, Observational Health Data Sciences and Informatics (OHDSI), was formed to continue to develop this model and to encourage its use in research. OHDSI is a federated network with a central office at Columbia University; each OHDSI institution uses ETL procedures to create its own OMOP data mart. OHDSI and OMOP also offer analytical tools to help sites understand the quality and completeness of their data and to facilitate research. OHDSI now includes sites in approximately 30 countries and 600 million patients.<sup>47</sup> Among many computable phenotypes that OHDSI has developed, those related to infectious diseases include HIV infection, human papillomavirus infection, influenza, methicillin-resistant *Staphylococcus aureus*,

otitis media, pneumonia, respiratory syncytial virus infection, and tuberculosis.<sup>47</sup> OHDSI facilitates the efficient analysis of large cohorts, as seen in an international study of 34,128 patients hospitalized with COVID-19.<sup>48</sup>

Big EHR vendors also have been developing capabilities for multi-institution observational research. The Cerner Corporation (North Kansas City, Missouri) has Cerner Real World Data (formerly Cerner Health Facts), a centralized database that uses data from a subset of the institutions that use a Cerner EHR system<sup>49</sup>; this database includes data from >60 institutions and >65 million patients.<sup>50-52</sup> Some data are available back to the year 2000.<sup>51</sup> Several recently published studies in infectious diseases have used this Cerner resource.<sup>51-55</sup> Epic Systems Corporation (Verona, Wisconsin) has the Epic Health Research Network, which is composed of a subset of the institutions that use an Epic EHR system<sup>56</sup>; this research network includes >50 million patients and is envisioned as a component of Epic's proposed Cosmos system.<sup>57,58</sup> Epic Systems Corporation personnel have used multi-institution Epic EHR data in a study of COVID-19.<sup>59</sup>

Other noteworthy examples of multi-institutional EHR networks have arisen in university and medical school settings and in collaboration with professional organizations. The Indiana Network for Patient Care (INPC) was established in 1995 to build on the Regenstrief Medical Record System (which began in 1972); INPC includes a large number of health care institutions across Indiana and has facilitated many observational EHR studies and the development of EHR research methods and standards.<sup>60,61</sup> Wiehe et al<sup>62,63</sup> studied sexually transmitted infections in youth and young adults based on INPC data and have built productive collaborations and observational studies with public health and justice system partners.<sup>62,63</sup> The electronic Pediatric Research in Office Systems network and its Comparative Effectiveness Research Through Collaborative Electronic Reporting Consortium was developed by Fiks et al<sup>64</sup> and the American Academy of Pediatrics<sup>65</sup>; it contains EHR data from 222 clinical sites in 27 states for >1.2 million children. This network was developed to study pediatric medication tolerability and effectiveness and has facilitated observational research on many other questions, including in infectious disease.<sup>65-68</sup> Via the National Center for Advancing Translational Sciences, a large group of investigators and institutions in the

Clinical and Translational Science Award program formed the National COVID Cohort Collaborative (N3C); N3C has addressed many infrastructure challenges in data governance, extraction, quality, harmonization, and analytics to create a centralized database for COVID-19 studies.<sup>69,70</sup> The N3C now includes >7 billion rows of EHR data and >2.1 million patients with COVID-19 (and >6.3 million patients overall) and has more than 200 projects under way [Table Supplemental 1](#).<sup>70</sup> Supplemental table 1 offers a summary of the EHR networks discussed.

A custom, cross-cutting network in infectious disease would build on this track record. It would facilitate collaboration by experts across the large number of institutions from which data would be needed to address rare infections and related conditions. It would help PEDSnet, PCORnet, and other networks enhance the ways that microbiology and virology laboratory results (including antibiotic susceptibility results) and antimicrobial medication data records are stored and made accessible for research projects. It also would support the development of learning health systems.

#### REVIEW OF EXISTING LITERATURE ON EHR NETWORK STUDIES IN INFECTIOUS DISEASE

There are several examples of EHR-network research outside infectious diseases that illustrate the potential advantages of using EHR-based systems. Hornik et al<sup>71</sup> within the Pediatric Trials Network developed their own EHR-based data repository with 9 of the highest-enrolling centers within their existing research network. They created a data model with 147 mandatory elements and 99 optional ones for all children discharged between January 2013 and June 2017. The authors used the PCORnet CDM as a starting point for their project because it already established many of the elements their team needed to evaluate drug exposure information in children. Their final data repository consisted of institutions using both Epic- and Cerner-based systems. They designed an implementation guide to help each site contribute its data accurately and ultimately collected data from >380,000 encounters from >260,000 children. This project is a good example of the steps required in creating a custom EHR network using a predefined group of investigators and/or institutions.

PEDSnet is one of the PCORnet member networks that has helped advance EHR-based research for a number of pediatric conditions outside infectious

diseases. Khare et al<sup>42</sup> used PEDSnet to create a computable phenotype for children with Crohn disease. They created several versions of this phenotype, ranging from less stringent (1 diagnosis code or medication associated with Crohn disease) to more stringent ( $\geq 3$  encounters and/or medications associated with Crohn disease). Using less stringent criteria, they were able to identify approximately 12,000 children with Crohn disease, and using more stringent criteria, they identified approximately 8000. They validated their data using a national registry as well as manual record review at participating institutions. They also found few false-negative results, which means that they were effectively capturing patients with the desired diagnoses. This study is a great example of the importance of the computable phenotype and how the rigor with which one defines the phenotype dictates the nature of the cohort one is left to study.

Within the field of infectious diseases, studies performed in the pre-COVID-19 era using EHR networks offer a glimpse of the potential benefits of EHR-based research. One study from the United Kingdom by Esan et al<sup>72</sup> evaluated the total burden of *Campylobacter* and nontyphoidal *Salmonella* infections within the National Health Service (NHS) by evaluating data from the Clinical Practice Research Datalink (CRPD).<sup>72</sup> The CRPD is an EHR data warehouse that consists of all outpatient primary care encounters within the NHS. CRPD data contain demographic information and specific encounter-level data associated with *International Classification of Diseases, Ninth Revision* (ICD-9) and ICD-10 diagnosis codes and are linked to inpatient hospital encounter data, mortality records, and socioeconomic status. They identified >20,000 patients (approximately 18,000 with *Campylobacter* and >2000 with nontyphoidal *Salmonella*) within the database. They described the rate of secondary complications, such as reactive arthritis, and found the peak onset of this complication within the first month after infection. Using the EHR data, they were also able to evaluate risk factors. They identified proton pump inhibitors as being associated with an elevated risk of *Campylobacter* infections (adjusted odds ratio = 2.1; 95% CI, 1.5–2.9). They were also able to define the annual costs to NHS of >£1.5 million and found increased health care use after infections and their associated complications. This study is a good example of using an EHR network to evaluate the impact of infections across a large population and bring attention



to the public health burden that may not otherwise be appreciated at any single center or practice.

Vihta et al<sup>73</sup> evaluated trends in *Escherichia coli* infections during an 18-year period in the region of Oxfordshire, United Kingdom.<sup>73</sup> They used the Infections in Oxfordshire Research Database, which has records of all admissions to the Oxford University Hospital NHS trust (a collection of 4 hospitals in the region) since 1997. This database includes inpatient data, including microbiology, biochemistry, and hematology results, along with demographic information and out-of-hospital mortality data. The authors evaluated all episodes of bloodstream infections and urinary tract infections (UTIs), evaluated episodes of relapse or recurrent infections, catalogued antibiotic prescriptions for these infections, and defined the outcomes of these infections. Their results indicated >5000 patients with *E coli* bloodstream infection and a relapse or recurrence rate of 9%. The authors found that almost half of these cases were nosocomial or in patients with chronic medical conditions who were recently admitted for treatment. They found year over year increases in both nosocomial and community cases of *E coli* bacteremia. Their analysis identified >137,000 patients with *E coli* UTI and a recurrence rate of 40% and found that 70% of all UTIs were de novo from the community. They found significant increases in community UTIs during the study period but a significant decrease in nosocomial UTIs during the study period. They identified that *E coli* with resistance to amoxicillin-clavulanate increased in bloodstream and UTI isolates during the study, reaching >40% in 2016. They were able to define increased antibiotic use as a major risk factor for subsequent infections with antibiotic-resistant isolates. This study is an example of evaluating a more selected population in detail for outcomes of infection as well as resistance trends over time.

With the onset of the COVID-19 pandemic, researchers and clinicians needed to collect data from large numbers of patients across domestic and international networks to rapidly predict which patients would develop severe disease and to determine which treatments were most effective at reducing mortality. Klann et al<sup>74</sup> organized many institutions across the United States and Europe to attempt to answer many of these questions. They used data from the early reports of COVID-19 cases to build a computable phenotype for COVID-19 cases and for severe cases.

Across 12 participating sites, they were able to evaluate >10,000 patients with COVID-19 infection as well as >3000 with severe disease. Importantly, they also performed validation via medical record review at each participating site, as well as cross-referencing EHR case data with locally acquired data on COVID-19 cases. They documented variation in rates of severe COVID-19 across sites and evaluated the different measures of disease severity (eg, acute respiratory disease, need for intubation, and vasopressor use) They constructed a Venn diagram that evaluated those patients who had markers of severe infection, those who had laboratory values indicating severe disease, and those who had the presence of a diagnosis or procedure code consistent with severe disease. There was not complete overlap among these 3 categories. Lastly, they simultaneously attempted to build a machine learning algorithm to help identify those patients with severe COVID-19 infection, but the first iteration did not perform as well as the expert-derived computable phenotype at identifying patients with severe COVID-19. Klann also played a key role in the N3C initiative detailed above.<sup>69</sup>

During the early phase of the epidemic, most of the focus was on adults given the severity and mortality observed in many settings. Data on children with COVID-19 were much harder to come by. PEDSnet was able to generate a national picture of what the COVID-19 pandemic looked like from the pediatric perspective. Bailey et al<sup>44</sup> reported the outcomes for >135,000 children tested for SARS-CoV-2 in the United States. They found that >4000 children had confirmed COVID-19 infection (4%) and confirmed many of the racial and ethnic disparities that were first reported in adults: a higher proportion of Black and Latine persons with positive SARS-CoV-2 test results and a higher proportion of children with underlying disorders who tested positive. They also confirmed a very low overall hospitalization rate (359 hospitalizations [7% of total cases identified]), a low rate of severe infection and mechanical ventilatory support (99 and 33 cases, respectively), and an extremely low death rate (8 total [0.2%]). Interestingly, in their unbiased data analysis, they identified a 40% decrease in cases of Kawasaki disease across all sites. Kawasaki disease has long been suspected to have an infectious origin; a similar observation had been reported in a single-center study and was hypothesized to be attributable to masking and social distancing having reduced transmission of an infectious agent.<sup>75</sup>

## WHY HAS THE UPTAKE FOR EHR-BASED RESEARCH BEEN SLOW IN INFECTIOUS DISEASE?

Given all the purported benefits reviewed above, one obvious question that could be asked is, “Why has there not been more progress and/or more studies in infectious disease using EHR-based methods?” There are likely several reasons for this, some of which were mentioned earlier. Infectious disease has traditionally been a bench research specialty, so there is not a long tradition of prior EHR research to draw on. Infectious disease is also an older specialty, so there may be significant discomfort with embracing EHR-based research when everyday EHR use can be a challenge. It is also likely that many infectious disease-related projects are performed by people doing more informatics-related research within hospital-based medicine or generalist, health services research, or epidemiology divisions. Informatics expertise is still a limited commodity so there are likely many institutions that lack a core group of personnel or sufficient bandwidth among their personnel to support the desired studies. These are all speculations, and the truth likely encompasses elements of all the above. In addition, because the clinical epidemiology questions that arise in infectious disease are diverse and often complex and nuanced, limitations remain in the EHR research networks (as of 2021) that are particularly relevant to infectious disease. Some types of data (eg, ICD-10 codes) cannot simply be retrieved from these networks. Microbiology culture and antibiotic susceptibility results are among the most difficult EHR elements for institutions to store in a way that makes the data amenable to use in subsequent multi-institutional cohort studies and in networks.<sup>76,77</sup> EHR results data for sexually transmitted infection tests, especially syphilis (and any test involving titers), are also complex to examine in an EHR network. Challenges in neonatal and pediatric intensive care unit research arise when one wants to analyze the details of mechanical ventilatory support or of medication dose changes and their timing. There may be ample (and unwieldy) data for these elements in an institution’s own EHR, but the ETL processes that build the data marts in the EHR networks do not bring in all these details. The ETL processes create simplifications and structure but cannot at the same time be all-encompassing.

## WHAT ARE THE STEPS REQUIRED FOR CONDUCTING EHR-BASED RESEARCH?

For those interested in beginning to conduct EHR-based research, there are several key steps that are critical in determining the success of any project. The previous example of mucormycosis serves as a good framework for considering what is required in each step.

The first step is to define the question. In general, a precise focus a priori will improve the preliminary feasibility assessments (step 2), the clarity of the data collected, and the statistical analyses. For mucormycosis, the question would be something such as, “For institutions that use voriconazole versus posaconazole for antifungal prophylaxis for high-risk oncology and stem cell transplants, how do the ultimate rates of invasive mucormycosis compare between the 2 agents?”

The second step is to define the size of the network that would be needed and explore whether the data elements available are suitable/ The size of the network will be dictated in part by the prevalence of the condition to be studied and in part by what networks and sites may be available to use. For a common condition such as *E coli* UTI, the Oxfordshire regional EHR network mentioned above identified >100,000 patients who could be included in a cohort study. For a rare condition such as mucormycosis, one would require a large national or international network capable of contributing several years of data. A PCORnet clinical research network may be a place to start, with the thought of expanding to other networks after generating preliminary data and additional resources. Preparatory-to-research queries in PCORnet or other networks may find robust voriconazole and posaconazole data but also may find that fungal culture data require additional work on ETL procedures before an observational study would be feasible.

The third step is to create the computable phenotype. A precise set of criteria should be created for the conditions of interest (denominator and numerator populations, exposure[s], covariate[s], and outcome[s]) that maximizes the total number of patients identified and minimizes the number of patients who have related (or unrelated) conditions who should not be included in the final analysis. For mucormycosis, creation of an overlapping (Venn diagram) set of criteria would help to enrich the patient population. Patients could be identified by searching for the appropriate ICD-9/ICD-10 codes for various invasive mucormycosis

diagnoses and independently identifying all fungal culture results that yield a species within the *Mucor* group of fungi. At the same time, one would limit the patient population to those with a primary diagnosis of an oncologic condition and/or a procedure code that corresponds to hematopoietic stem cell or cord blood transplant, while excluding other nononcologic populations who also are susceptible to mucormycosis infections (patients with diabetes, burns, or SARS-CoV-2 and corticosteroid therapy).

The fourth step is to perform initial local validation. The easiest way to evaluate the computable phenotype initially is to evaluate how it performs within one's own EHR. Do the criteria designed above capture the known patients recently treated for mucormycosis? If yes, one could proceed to broader data collection. If not, the phenotype needs to be refined and tested again with local validation. However, as noted in step 2, even if it works well locally, a computable phenotype may not work as well across the other data-contributing sites in an EHR network; such networks are standardized and harmonized only up to a point. Most projects conduct data quality and completeness assessments at the data-contributing sites before or during the execution of the main queries.

The fifth step is to determine the scope of data to be collected. EHR networks are good at gathering structured data. These data may include vital signs, laboratory test results, medication dispensing and dosing information, primary and secondary diagnoses, procedures performed, intensive care admissions, and deaths. EHR networks, for the most part, are not yet collecting unstructured data: practitioner progress notes, radiology reports, pathology reports, and any other test results with extensive free text or narrative discussion. Efforts are constantly under way to improve the ability to capture and analyze unstructured data, but at the moment it is not feasible on a large scale. For the mucormycosis study, the structured data mentioned should allow for analysis of the rates of mucormycoses identified, the number of patients given voriconazole and posaconazole as daily prophylaxis before the diagnosis being made, and laboratory and outcomes data to determine if there are adverse events associated with the 2 different prophylaxis strategies that could negate any benefit in preventing a few cases of mucormycosis.

The sixth step is to share the search strategy and data requirements with partners. By sharing the refined

computable phenotype and data requirements with network partners or a centralized data coordinating center, one can gather the desired raw data at this step. For the mucormycosis project, this step would identify the numerator of all mucormycosis cases in oncology patients and patients undergoing stem cell transplantation within the national network and the denominator of all oncology patients and patients undergoing stem cell transplantation who received voriconazole and posaconazole prophylaxis during their treatment. One would also obtain the structured data elements mentioned previously (to the extent available) across all patients. Microbiology culture results, because of their complex hierarchical structure, are the most challenging type of laboratory result to store in a data mart in an analyzable way; fungal culture data in this study may require more work.

The seventh step is to perform additional validation via medical record review at each site. Assuring data integrity at each partner site is a critical step in maintaining a robust network. The best studies ensure that each site performs an audit of a sample of all cases identified to make sure that the patients have the study condition, received the particular drug, and had the electronically ascribed outcome. For the mucormycosis study, each site would confirm the diagnosis of fungal infection, confirm the underlying diagnosis of oncology or stem cell transplantation, and confirm the final outcomes (survived and discharged or died in the hospital or subsequently).

The eighth step is data and statistical analysis. Acquisition of the data described would be useless unless an investigator has the ability for proper data analysis. Having collaborators with experience analyzing large EHR datasets is crucial to the success of any EHR network study. The experience entails not only an analysis of the expected effects but also an unbiased search for differences that were not expected but that could generate new hypotheses for future research. In the mucormycosis study, an unbiased analysis might reveal that widespread use of posaconazole prophylaxis led to small but significant increases in *Aspergillus ustus* infections, a multidrug-resistant species that can be seen with extensive antifungal exposure in highly susceptible patients.

Although the steps outlined above do not capture every aspect of EHR-based research, they highlight key aspects of a successful project.

## CONCLUSIONS

With the widespread adoption of EHRs across the United States and many parts of the world, there is a tremendous opportunity to interrogate many aspects of clinical care. It should no longer be acceptable to claim that data are not available on any topic. EHR networks may allow researchers to evaluate the clinical experience of a large pool of treating physicians, to describe trends across a regional or national population, and to evaluate seemingly rare conditions that are otherwise difficult to study. As the field of infectious diseases moves to more widely embrace EHR-based studies, data on outcomes and comparative effectiveness will be more readily available.

## DISCLOSURE

None.

## REFERENCES

- Moriarty A. Does Hospital EHR Adoption Actually Improve Data Sharing. *Definitive Healthcare*. <https://www.definitivehc.com/blog/hospital-ehr-adoption>.
- Evans RS. Electronic health records: then, now, and in the future. *Yearb Med Inform*. 2016;Suppl 1:S48–S61.
- Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. *Evid Based Med*. 2016;21:125–127.
- Mc Cord KA, Hemkens LG. Using electronic health records for clinical trials: where do we stand and where can we go? *CMAJ*. 2019;191:E128–E133.
- Rosenman MB, Tao G, Szucs KA, Wang J, Ambuehl R, Mahon BE. Prenatal syphilis screening rates measured using medicaid claims and electronic medical records. *Sex Transm Dis*. 2008;35:387–392.
- Zeltzer D, Balicer RD, Shir T, Flaks-Manov N, Einav L, Shadmi E. Prediction accuracy with electronic medical records versus administrative claims. *Med Care*. 2019;57:551–559.
- Cornely OA, Alastruey-Izquierdo A, Arenz D, et al. Global guideline for the diagnosis and management of mucormycosis: an initiative of the European Confederation of Medical Mycology in cooperation with the Mycoses Study Group Education and Research Consortium. *Lancet Infect Dis*. 2019;19:e405–e421.
- Osborne V, Lane S, Shakir SAW. The role of the contextual cohort to resolve some challenges and limitations of comparisons in Pharmacoepidemiology. *Drug Saf*. 2021;44:835–841.
- Sainani KL. Propensity scores: uses and limitations. *PM R*. 2012;4:693–697.
- Streiner DL, Norman GR. The pros and cons of propensity scores. *Chest*. 2012;142:1380–1382.
- Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care*. 2013;51:S30–S37.
- Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. *BMJ*. 2018;361:k1479.
- Callahan A, Shah NH, Chen JH. Research and reporting considerations for observational studies using electronic health record data. *Ann Intern Med*. 2020;172:S79–S84.
- Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med*. 2000;342:1878–1886.
- Concato J, Shah N, Horwitz RJ. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000;342:1887–1892.
- Banerjee R, Prasad V. Are observational, real-world studies suitable to make cancer treatment recommendations? *JAMA Netw Open*. 2020;3.
- Kumar A, Guss ZD, Courtney PT, et al. Evaluation of the use of cancer registry data for comparative effectiveness research. *JAMA Netw Open*. 2020;3.
- Soni PD, Hartman HE, Dess RT, et al. Comparison of population-based observational studies with randomized trials in Oncology. *J Clin Oncol*. 2019;37:1209–1216.
- Ramagopalan SV, Simpson A, Sammon C. Can real-world data really replace randomised clinical trials? *BMC Med*. 2020;18:13.
- Collins R, Bowman L, Landray M. Randomization versus real-world evidence. reply. *N Engl J Med*. 2020;383:e21.
- Collins R, Bowman L, Landray M, Peto R. The magic of randomization versus the myth of real-world evidence. *N Engl J Med*. 2020;382:674–678.
- Deans KJ, Sabihi S, Forrest CB. Learning health systems. *Semin Pediatr Surg*. 2018;27:375–378.
- Forrest CB, Margolis PA, Bailey LC, et al. PEDSnet: a national pediatric learning health system. *J Am Med Inform Assoc*. 2014;21:602–606.
- Tseng HF, Sy LS. Use of real-world evidence to evaluate the effectiveness of Herpes Zoster vaccine. *J Infect Dis*. 2018;218(suppl 2):S63–S67.
- Simon GE, Bindman AB, Dreyer NA, et al. When can we trust real-world data to evaluate new medical treatments? *Clin Pharmacol Ther*. 2021 May 1 [Epub ahead of print].
- Watanabe JH, Simon GE, Horberg M, Platt R, Hernandez A, Califf RM. When are treatment blinding and treatment standardization necessary in real-world clinical trials? *Clin Pharmacol Ther*. 2021 Apr 8 [Epub ahead of print].

27. Gokhale M, Sturmer T, Buse JB. Real-world evidence: the devil is in the detail. *Diabetologia*. 2020;63:1694–1705.
28. Jones WS, Mulder H, Wruck LM, et al. Comparative effectiveness of Aspirin dosing in Cardiovascular Disease. *N Engl J Med*. 2021;384:1981–1990.
29. Matics TJ, Sanchez-Pinto LN. Adaptation and validation of a pediatric sequential organ failure assessment score and evaluation of the Sepsis-3 definitions in critically ill children. *JAMA Pediatr*. 2017;171.
30. DeShazo JP, Hoffman MA. A comparison of a multistate inpatient EHR database to the HCUP Nationwide Inpatient Sample. *BMC Health Serv Res*. 2015;15:384.
31. Butame SA, De Leon JM, Lee SJ, et al. Barriers and facilitators to the collection and aggregation of electronic health record HIV data: an analysis of study recruitment venues within the adolescent medicine trials network for HIV/AIDS interventions (ATN). *Eval Health Prof*. 2021;44:168–176.
32. Collins FS, Hudson KL, Briggs JP, Lauer MS. PCORnet: turning a dream into reality. *J Am Med Inform Assoc*. 2014;21:576–577.
33. Qualls LG, Phillips TA, Hammill BG, et al. Evaluating foundational data quality in the National Patient-Centered Clinical Research Network (PCORnet(R)). *EGEMS (Wash DC)*. 2018;6:3.
34. Block JP, Bailey LC, Gillman MW, et al. Early antibiotic exposure and weight outcomes in young children. *Pediatrics*. 2018;142(6).
35. Public Health Informatics Institute. PCORnet CDC COVID-19 Healthcare Data Initiative. <https://www.phii.org/electronic-healthcare-data-initiative>. Published 2021. Accessed July 1, 2021.
36. Forrest CB, McTigue KM, Hernandez AF, et al. PCORnet(R) 2020: current state, accomplishments, and future directions. *J Clin Epidemiol*. 2021;129:60–67.
37. Lin PD, Daley MF, Boone-Heinonen J, et al. Comparing prescribing and dispensing data of the PCORnet common data model within PCORnet antibiotics and childhood growth study. *EGEMS (Wash DC)*. 2019;7:11.
38. Curtis LH, Brown J, Platt R. Four health data networks illustrate the potential for a shared national multipurpose big-data network. *Health Aff (Millwood)*. 2014;33:1178–1186.
39. Hurst JH, Liu Y, Maxson PJ, Permar SR, Boulware LE, Goldstein BA. Development of an electronic health records datamart to support clinical and population health research. *J Clin Transl Sci*. 2020;5:e13.
40. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc*. 2012;19:54–60.
41. Hripcsak G, Schuemie MJ, Madigan D, Ryan PB, Suchard MA. Drawing reproducible conclusions from observational clinical data with OHDSI. *Yearb Med Inform*. 2021;30:283–289.
42. Khare R, Kappelman MD, Samson C, et al. Development and evaluation of an EHR-based computable phenotype for identification of pediatric Crohn's disease patients in a National Pediatric Learning Health System. *Learn Health Syst*. 2020;4:e10243.
43. Khare R, Utidjian L, Ruth BJ, et al. A longitudinal analysis of data quality in a large pediatric data research network. *J Am Med Inform Assoc*. 2017;24:1072–1079.
44. Bailey LC, Razzaghi H, Burrows EK, et al. Assessment of 135794 pediatric patients tested for severe acute Respiratory Syndrome Coronavirus 2 across the United States. *JAMA Pediatr*. 2021;175:176–184.
45. Denburg MR, Razzaghi H, Bailey LC, et al. Using electronic health record data to rapidly identify children with Glomerular Disease for clinical research. *J Am Soc Nephrol*. 2019;30:2427–2435.
46. Weeks J, Pardee R. Learning to share health care data: a brief timeline of influential common data models and distributed health data networks in U.S. health care research. *EGEMS (Wash DC)*. 2019;7:4.
47. Observational Health Data Sciences and Informatics. [www.ohdsi.org](http://www.ohdsi.org). Published 2021. Accessed July 1, 2021.
48. Burn E, You SC, Sena AG, et al. Deep phenotyping of 34,128 adult patients hospitalised with COVID-19 in an international network study. *Nat Commun*. 2020;11:5009.
49. Cerner Data and Research Inc. <https://www.cerner.com/ap/en/solutions/data-research>. Published 2021. Accessed July 1, 2021.
50. Southern California Clinical and Translational Science Institute. Cerner Health Facts. <https://sc-ctsi.org/resources/cerner-health-facts>. Published 2021. Accessed July 1, 2021.
51. Ewherhemuepha L, Gasperino G, Bischoff N, Taraman S, Chang A, Feaster W. HealthDataLab - a cloud computing solution for data science and advanced analytics in healthcare with application to predicting multi-center pediatric readmissions. *BMC Med Inform Decis Mak*. 2020;20:115.
52. Stroever SJ, Ostapenko D, Scatena R, et al. Medication use among patients with COVID-19 in a large, national dataset: cerner real-world data. *Clin Ther*. 2021;43:e173–e196.
53. Kadri SS, Lai YL, Warner S, et al. Inappropriate empirical antibiotic therapy for bloodstream infections based on discordant in-vitro susceptibilities: a retrospective cohort analysis of prevalence, predictors, and mortality risk in US hospitals. *Lancet Infect Dis*. 2021;21:241–251.
54. Rhee C, Dantes R, Epstein L, et al. Incidence and trends of Sepsis in US hospitals using clinical vs claims data, 2009-2014. *JAMA*. 2017;318:1241–1249.

55. Haeusler GM, Thursky KA. Electronic health record data for antimicrobial prescribing. *Lancet Infect Dis*. 2021;21:155–157.
56. Epic Systems Corporation. Epic Health Research Network. <https://ehrn.org/>. Published 2021. Accessed July 1, 2021.
57. Wahlberg D. With a nod to disco era, Epic Systems Corp. looks to Cosmos, voice-activated software. *Wisconsin State Journal*. August 28, 2019.
58. Milinovich A, Kattan MW. Extracting and utilizing electronic health data from Epic for research. *Ann Transl Med*. 2018;6:42.
59. Rubin-Miller L, Alban C, Artiga S, Sullivan S. COVID-19 Racial Disparities in Testing, Infection, Hospitalization, and Death: Analysis of Epic Patient Data. 16 Sept 2020. [Epub ahead of print]
60. McDonald CJ, Overhage JM, Barnes M, et al. The Indiana network for patient care: a working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries. *Health Aff (Millwood)*. 2005;24:1214–1220.
61. Vreeman DJ, Chiaravallotti MT, Hook J, McDonald CJ. Enabling international adoption of LOINC through translation. *J Biomed Inform*. 2012;45:667–673.
62. Wiehe SE, Rosenman MB, Aalsma MC, Scanlon ML, Fortenberry JD. Epidemiology of Sexually Transmitted Infections Among Offenders Following Arrest or Incarceration. *Am J Public Health*. 2015;105:e26–e32.
63. Wiehe SE, Rosenman MB, Chartash D, et al. A solutions-based approach to building data-sharing partnerships. *EGEMS (Wash DC)*. 2018;6:20.
64. Fiks AG, Grundmeier RW, Margolis B, et al. Comparative effectiveness research using the electronic medical record: an emerging area of investigation in pediatric primary care. *J Pediatr*. 2012;160:719–724.
65. Fiks AG, Scheindlin B, Shone L. 30th anniversary of pediatric research in office settings (PROS): an invitation to become engaged. *Pediatrics*. 2016;138(3).
66. Daymont C, Ross ME, Russell Localio A, Fiks AG, Wasserman RC, Grundmeier RW. Automated identification of implausible values in growth data from pediatric electronic health records. *J Am Med Inform Assoc*. 2017;24:1080–1087.
67. Czaja AS, Ross ME, Liu W, et al. Electronic health record (EHR) based postmarketing surveillance of adverse events associated with pediatric off-label medication use: a case study of short-acting beta-2 agonists and arrhythmias. *Pharmacoepidemiol Drug Saf*. 2018;27:815–822.
68. Gerber JS, Prasad PA, Russell Localio A, et al. Variation in antibiotic prescribing across a pediatric primary care network. *J Pediatric Infect Dis Soc*. 2015;4:297–304.
69. Haendel MA, Chute CG, Bennett TD, et al. The National COVID Cohort Collaborative (N3C): rationale, design, infrastructure, and deployment. *J Am Med Inform Assoc*. 2021;28:427–443.
70. US Department of Health and Human Services. National COVID Cohort Collaborative (N3C). <https://ncats.nih.gov/n3c>. Published 2021. Accessed July 5, 2021.
71. Hornik CP, Atz AM, Bendel C, et al. Creation of a multicenter pediatric inpatient data repository derived from electronic health records. *Appl Clin Inform*. 2019;10:307–315.
72. Esan OB, Perera R, McCarthy N, Violato M, Fanshawe TR. Incidence, risk factors, and health service burden of sequelae of campylobacter and non-typhoidal salmonella infections in England, 2000-2015: A retrospective cohort study using linked electronic health records. *J Infect*. 2020;81:221–230.
73. Vihta KD, Stoesser N, Llewelyn MJ, et al. Trends over time in Escherichia coli bloodstream infections, urinary tract infections, and antibiotic susceptibilities in Oxfordshire, UK, 1998-2016: a study of electronic health records. *Lancet Infect Dis*. 2018;18:1138–1149.
74. Klann JG, Weber GM, Estiri H, et al. Validation of an internationally derived patient severity phenotype to support COVID-19 Analytics from electronic health record data. *J Am Med Inform Assoc*. 2021;28:1411–1420.
75. Shulman S, Geevarghese B, Kim KY, Rowley A. The impact of social distancing for COVID-19 upon diagnosis of Kawasaki Disease. *J Pediatric Infect Dis Soc*. 2021;10:742–744.
76. Rosenman MB, Szucs KA, Finnell SM, et al. Nascent regional system for alerting infection preventionists about patients with multidrug-resistant gram-negative bacteria: implementation and initial results. *Infect Control Hosp Epidemiol*. 2014;35:S40–S47.
77. Wulff A, Baier C, Ballout S, et al. Transformation of microbiology data into a standardised data representation using OpenEHR. *Sci Rep*. 2021;11:10556.

---

**Address correspondence to:** Ravi Jhaveri, MD, Division of Pediatric Infectious Diseases, Ann & Robert H. Lurie Children's Hospital of Chicago, 225 E Chicago Ave, Box 20, Chicago, IL 60611-2991. E-mail: [ravi.jhaveri@northwestern.edu](mailto:ravi.jhaveri@northwestern.edu).

## SUPPLEMENTARY MATERIALS

Supplemental Table 1. Summary of EHR Networks.

Name	Geographic Coverage	# of Patients	Website
<b>PCORnet</b>	Regional or National	Up to 66 million	<a href="https://pcornet.org/">https://pcornet.org/</a>
<b>OHDSI(-OMOP)</b>	International	Up to 600 million	<a href="https://ohdsi.org/">https://ohdsi.org/</a>
<b>Cerner Real World Data</b>	National	Up to 65 million	<a href="https://www.cerner.com/solutions/real-world-data#main">https://www.cerner.com/solutions/real-world-data#main</a>
<b>Epic Health Research Network</b>	National	Up to 50 million	<a href="https://ehrn.org/">https://ehrn.org/</a>
<b>Indiana Network for Patient Care</b>	Regional	Up to 18 million	<a href="https://www.regenstrief.org/rds/data/">https://www.regenstrief.org/rds/data/</a>
<b>ePROS network (Pediatric)</b>	National	Up to 1.2 million	<a href="https://www.aap.org/en/research/pediatric-research-in-office-settings-pros/">https://www.aap.org/en/research/pediatric-research-in-office-settings-pros/</a>
<b>National COVID Cohort Collaborative (N3C)</b>	National	Up to 6.3 million	<a href="https://ncats.nih.gov/n3c">https://ncats.nih.gov/n3c</a>