Horizon Scan of Electronic Health Record Databases

Phase 1 Final Report

Prepared by: Joshua J Gagne, PharmD, ScD,[1] Jennifer G Lyons, PhD, MPH,[5] Rishi Desai, MS, PhD,[1] Winnie Ho, BA,[1] Shamika More, MS,[1] Catherine Corey, MSPH,[6] Sheryl Kluberg, PhD,[5] Mayura Shinde, MPH, DrPH,[5] Adee Kennedy, MS, MPH,[5] David S. Carrell, PhD,[2] Keith Marsolo, PhD,[3] Michael E. Matheny MD, MS, MPH,[4] Sebastian Schneeweiss, MD, ScD,[1] Jeffrey S. Brown, PhD,[5] Luke Zabotka, BA[1] Catherine Lerro PhD, [6] Jie Li, PhD[6]

Author Affiliations: 1 Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2 Kaiser Permanente Washington Health Research Institute, Seattle, WA; 3 Department of Population Health Sciences, Duke University School of Medicine, Durham, NC; 4 Vanderbilt University Medical Center Department of Biomedical Informatics, Nashville, TN; 5 Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA; 6 US Food and Drug Administration, Silver Spring, MD

Version 1.0

October 8, 2021

Horizon Scan of Electronic Health Record Databases

Phase 1 Final Report

Table of Contents

History of Modifications

Version	Date	Modification	Author
1.0	10/1/2021	Original Version	Sentinel Innovation Center and Sentinel Operations Center

Background

A main focus of the US Food and Drug Administration's Sentinel System Five-Year Strategy [1], 2019-2023, is to expand access to and use of electronic health record (EHR) data to support regulatory decision-making about the safety and effectiveness of medical products. The Innovation Center, one of three Sentinel centers, seeks to incorporate emerging data science innovations, such as natural language processing (NLP) and machine learning, to address a lack of valid and robust computable phenotypes for many health outcomes of interest in Sentinel queries.[2] A key objective of the Innovation Center is to identify EHR data sources to incorporate into Sentinel to expand Sentinel's data infrastructure, to enhance its computable phenotyping capabilities, and to support signal detection and inferential activities related to regulated medical products.

The Innovation Center launched the horizon scan of EHR databases and empirical queries project to identify and assess potential partners that could contribute the necessary data to Sentinel. These potential partners could include existing Data, Expansion, and Innovation Partners as well as data sources not currently included in Sentinel. The first phase of this project will develop and implement a horizon scan and series of interviews to identify potential EHR sources and registries and conduct initial feasibility assessments of these potential partners for enhancing the Sentinel System. The second phase of this project will conduct empirical queries in a set of the most promising EHR databases identified and recommended from phase to better understand the data available from potential partners, to evaluate the potential partners' processes and readiness to perform the queries, and to identify potential barriers to integrating the potential partners into Sentinel.

Project Overview

To address the Real-World Evidence (RWE) Data Enterprise mandate,[*] FDA has tasked the Sentinel Innovation Center (IC) and Sentinel Operations Center (SOC) with establishing a query-ready, quality-checked distributed data network containing electronic health records for at least 10 million lives with reusable analysis tools. A key first step in establishing such a network is identifying and assessing potential partners that could contribute the necessary data for this system, including existing data, expansion, and innovation partners, as well as data sources not currently included in Sentinel.

The purpose of the *horizon scan of electronic health record databases* project (hereafter referred to as EHR horizon scan) is to develop and implement a horizon scan and series of interviews to: (1) identify viable EHR sources to fulfill the RWE data enterprise mandate;(2) conduct interviews with these potential partners to understand their data and capabilities for meeting the needs of this system; and (3) conduct empirical queries with a smaller subset of potential EHR data partners to more closely examine the suitability of the data for Sentinel.

Although existing Sentinel Data Partners can already provide query-ready and quality-checked EHR data for more than 10 million lives, the horizon scan will ensure a thorough search of US EHR resources with a particular focus on identifying and evaluating potential new sources of data on pediatric, cancer, and pregnancy/birth outcomes, cause of death data, and COVID-19 patients and their health outcomes and treatments.

---

[*] https://www.fda.gov/news-events/fda-voices/fda-budget-matters-cross-cutting-data-enterprise-real-world-evidence

The EHR horizon scan project involves four sequential steps conducted in two phases. As part of phase 1, Step 1 included a literature review to identify viable EHR sources appropriate for further consideration in fulfilling FDA's RWE data enterprise mandate. Step 2 compiled information on the most promising EHR data sources and involved a preliminary questionnaire (herein referred to as Step 2a), completed by the partners, and a secondary questionnaire (herein referred to as Step 2b) administered as a structured phone interview. Step 3 of the project provided further opportunity for representatives from promising EHR data sources to share additional information and understand the specific goals of the IC to ensure alignment with meeting the RWE data enterprise mandate. Finally, the last step, conducted in Phase 2 will select three of the most promising EHR databases identified and recommended from Phase 1 to plan for and execute empirical queries. These queries will be used to inform planning of the data expansion process. The queries will be designed to: (1) better understand the data available from each potential partner; (2) evaluate the potential partners' processes and readiness to perform the queries; and (3) identify potential barriers to integrating the potential partners into Sentinel.

The goal of this report is to summarize the work conducted in Phase 1 of this project by highlighting the findings from both the preliminary and secondary questionnaires that were administered in November - December 2020 and January 2021, respectively.

Objectives

The purpose of the horizon scan of electronic health record databases and empirical query project is twofold:

In the first phase of this project, we aim to develop and implement a horizon scan and series of interviews to: (1) identify viable EHR sources to fulfill the RWE data enterprise mandate; and (2) conduct interviews with these potential partners to understand their data and capabilities for meeting the needs of this system.

In the second phase of this project, we aim to conduct empirical queries in three of the most promising EHR databases identified and recommended from phase 1. These queries will be used to inform planning of the data expansion process.

This report summarizes the findings of the Phase 1. A separate report will summarize the activities of Phase 2, once those are completed.

Methods (Step 1)

Search strategy

We used PubMed to conduct a scoping review to identify EHR data sources of interest. We sought to identify databases that have been used in at least one published biomedical research article to indicate that the database may be capable of supporting scientific studies. As most medical institutions in the US now have electronic medical records, we focused on sources that comprised data from at least two sites. We developed a search strategy to select for human subject research articles published in the English language between the dates of January 1, 2018 to June 18, 2020. A recent timeframe was chosen to limit the identification of data sources that are no longer available. The search strategy was developed with Medical Subject Headings (MeSH) terms referencing electronic health records, and selected for certain study types, including Clinical Studies, Comparative Studies, Observational Studies, and Randomized Controlled Trials (see **Appendix A** for full search filter).

Review and Exclusion

Titles and abstracts were examined by one member of the study team (WH) to identify articles of interest for further review. We excluded papers that were clearly not focused on biomedical research performed with EHR data, studies utilizing only non-US data, and those performed at a single institution or site. Articles that focused solely on improving EHR systems and processes were excluded. We supplemented the PubMed search with additional searches of the world wide web to identify and exclude those databases that no longer exist and those that appeared to be EHR software vendors without an accessible EHR database. Based on the web searches, we also combined certain EHR data sources into single unique entities, where appropriate. For example, when a particular database changed names or when two or more organizations' data sources had been merged, we grouped the corresponding articles together under the current name. We supplemented the search by including additional data sources that were identified via the web search and that met eligibility criteria.

Data Extraction

The following information on each unique EHR data source was extracted from the identified articles and supplementary web searches conducted between May, 2020 to December, 2020: name of EHR data source or organization, number of sites or centers from which EHR data are sourced, geographic region of the US that the data source covers, care settings (e.g., inpatient, outpatient, both), approximate patient population size, mentions of linkages to other data sources (including health insurance claims), and years of EHR data coverage.

Results (Step 1)

The PubMed search yielded a total of 4,446 hits (**Figure 1**). After an initial title and abstract screen, 348 (8%) articles were retained for further review. These articles identified a total of 110 EHR data sources that potentially met eligibility criteria, including those identified by the web search. Of these, 36 were subsequently excluded as duplicates (n = 12), single site EHR data sources (n = 9), EHR software vendors with no evidence of an accessible research database (n = 4), and data sources that no longer exist or do not appear to have recent information available (n = 3); 8 were excluded for not meeting other eligibility criteria. Of the 74 remaining unique data sources, 21 (28%) were existing Sentinel collaborators, 2 (3%) were government entities, and 51 (69%) were non-government EHR databases that were not current Sentinel collaborators. Of these 51, 9 (17%) were EHR aggregators, EHR networks, or Health Information Exchanges; 19 (37%) were health systems; and 23 (45%) were specialty data sources.

Appendix B provides an overview of all 74 unique data sources, including the type of EHR data they contain, the general size of the database in terms of numbers of facilities or providers or numbers of patients covered, the US region covered, a brief description of each data source, and references to related articles. The unique data sources varied widely in their size and coverage. For example, one EHR database included data from two hospitals and another organization holds data from >160,000 providers.

The specialty data sources covered a wide range of clinical areas, including allergy, anesthesia, cancer, cardiovascular, dental, dermatology, kidney, neurology, ophthalmologic, pediatric, trauma, and wound care. Some data sources were specific to certain states or cities while others included data from around the US.

Discussion (Step 1)

This scoping review identified many potentially viable EHR data sources that might be able to support expanding Sentinel's capabilities. The data sources identified vary widely in their geographical coverage, inclusion of specialties, size, and coverage of various care settings.
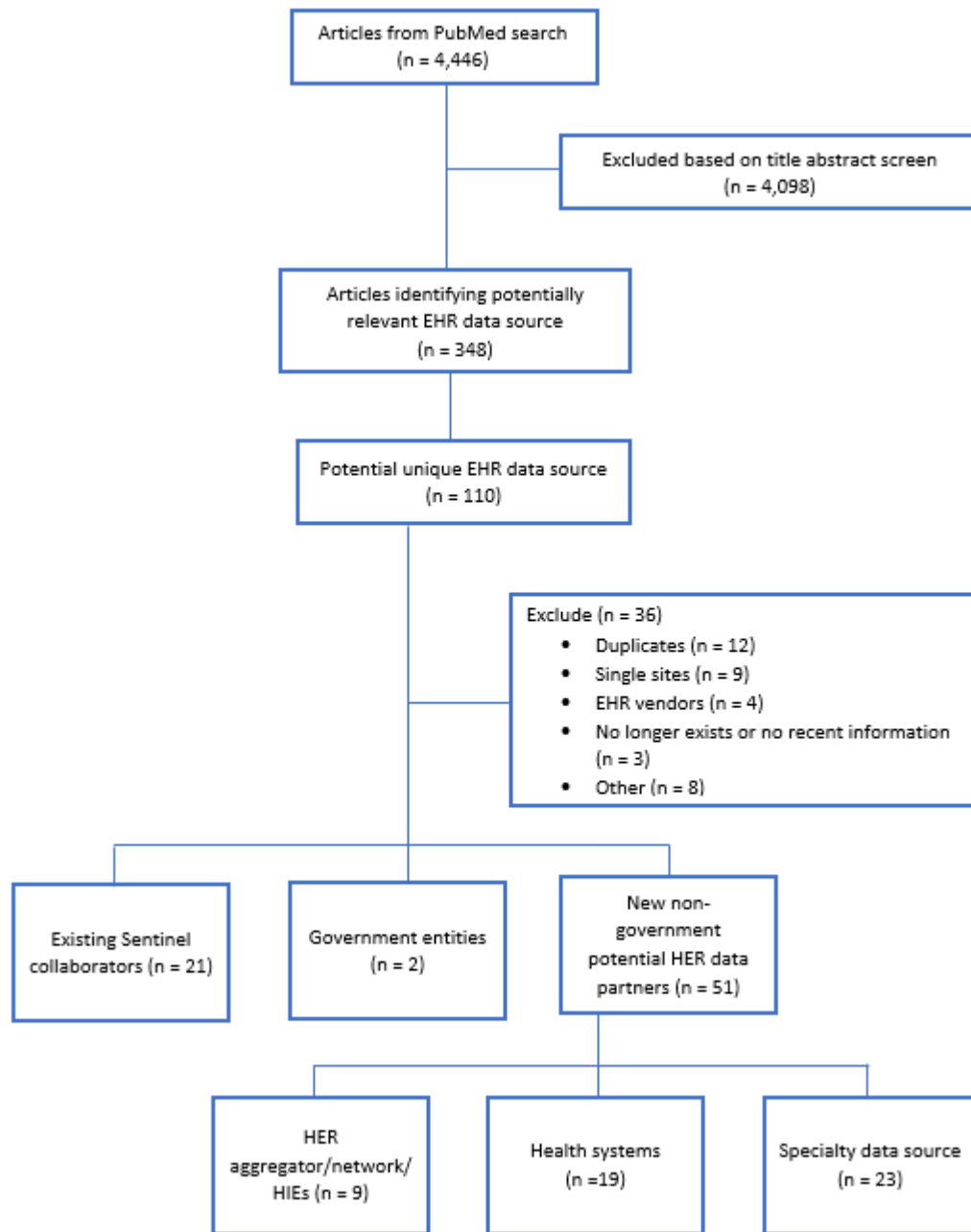
The results of this scoping review should be viewed in the context of several limitations. First, to keep the search and review process tractable, we limited the search to databases identifiable via biomedical research articles in PubMed. While this helped to ensure that identified data sources might be research-ready, this approach likely missed newer or emerging data sources from which research articles might not yet be published.

This scoping review was also limited to identifying potentially useful EHR data sources based on information available through published articles and on the world wide web. We therefore could not do a thorough review of the content, completeness, and quality of each data source nor did we assess whether the organizations holding the data would be capable of supporting Sentinel activities from operational or governance perspectives.

To gain greater insights into EHR data source characteristics and well as operational and governance considerations, Steps 2 and 3 of this project will entail interviews with select potential partners to understand their data and capabilities for meeting the needs of Sentinel. The last step, conducted as part of Phase 2,will extend that assessment further through conducting empirical queries in select EHR databases to better understand the data available from potential partners, to evaluate the potential partners' processes and readiness to perform the queries, and to identify potential barriers to integrating the potential partners into Sentinel.

In conclusion, this scoping review identified many potential data sources that could be useful for expanding Sentinel capabilities for analyses of EHR data. Subsequent assessments of these data sources – including surveys and empirical queries – are needed to further understand the viability of these data sources for supporting Sentinel activities.

Figure 1. Flowchart of search results and unique electronic health record data source identification

Articles from PubMed search
(n = 4,446)

Excluded based on title abstract screen
(n = 4,098)

Articles identifying potentially
relevant EHR data source
(n = 348)

Potential unique EHR data source
(n = 110)

Exclude (n = 36)
- Duplicates (n = 12)
- Single sites (n = 9)
- EHR vendors (n = 4)
- No longer exists or no recent information
  (n = 3)
- Other (n = 8)

Existing Sentinel
collaborators (n = 21)

Government entities
(n = 2)

New non-government
potential HER data
partners (n = 51)

HER aggregator/network/
HIEs (n = 9)

Health systems
(n =19)

Specialty data source
(n = 23)

Overview of outreach to EHR data partners (Steps 2 and 3)

Informed by the literature review conducted in Step 1, the aim of Step 2 was to identify and prioritize a subset of EHR data sources that would meet the FDA's RWE data enterprise mandate for the IC. Step 1 identified 110 unique EHR data sources, of which 36 were excluded for several reasons including being single sites, no longer existing, and acting solely as EHR vendors. **Figure 1** (shown on page 9) illustrates the findings from the Step 1 literature review. The IC contacted a total of 12 EHR integrated partners that were existing Sentinel Partners; a total of 11 completed Step 2a, by completing a preliminary questionnaire. The workgroup identified 10 potential new partners, for which six participated, two did not respond, and two declined to complete Step 2a.

Preliminary questionnaire overview (Step 2a)

The primary goal of administering the preliminary questionnaire (Step 2a) was to open communication between the IC and the potential EHR data partners and gain a better understanding of technical and governance features of both new and existing data sources. The secondary goal was to identify the subset of data sources that would best align with the RWE data enterprise mandate by FDA, before proceeding onto Step 2b, completion of a second questionnaire.

The preliminary questionnaire was administered to partners via email in November and December 2021 and focused on a specific set topics related to the technical features and characteristics of the data sources. The first three questions focused on the number of patient lives represented in the EHR data source, the number of visits or encounters a patient had, and the number of patients with at least one visit in the EHR in each of the last three years. The fourth question asked about the various data elements available in the data source for analysis. Question five asked about the time period for which data was available and question six asked about the types of care settings available. Question seven asked how frequently the data are updated. Finally, the last question asked about the types of coding schemes available. For a full list of questions and sub-questions, please refer to **Appendix C.**

Secondary questionnaire overview (Step 2b)

The goal of the secondary questionnaire was to build upon the information gleaned from Step 2a and conduct in-depth interviews with the most promising EHR data sources identified, prioritizing new partners. A detailed list of questions for these interviews was provided to interviewees in advance of the interviews to ensure efficient use of time and to allow interviewees to prepare and provide complete and accurate responses. The questions were organized into ten topic areas; the majority of time was spent on the five topic areas that covered how data was captured, data models used, research capabilities, quality control and quality assessment of the data, as well as collaboration and governance. As time allowed, partners were also asked to explain any unique characteristics of the patient population or characteristics of specialty data, including pregnancy, oncology, and death data. A detailed list of questions can be found in **Appendix D.**

Follow-up discussions (Step 3)

The goal of the follow-up conversations with a select subset of potential Data Partners was to delve deeper into the findings from the preliminary and secondary questionnaires and interviews. A selection of the most promising Data Partners, as well as several that covered specific populations and care settings of interest, were invited to talk in more detail about

certain aspects of their data and processes. These conversations were guided by a few main questions of interest, but informal enough that the Data Partner representatives could expand as much as possible on the outstanding questions of interest that were specific to each Data Partner.

Question topics included:

1. What kind of access is available to clinical text?

2. What is the time required for contracting?

3. Please characterize the quality-control processes used when mapping health data including, for example, labs, from the "X" unique DPs into your data source's "standardized format"

4. Data stability over time

5. Could you provide an example of a published analysis that illustrates the strengths of your data?

6. Is it possible to access unstructured clinical notes from the EHRs from which data are obtained, and if so, what types of notes (e.g., outpatient encounter notes, pathology reports, imaging study reports)?

7. What capabilities and advantages does your data source have that differentiate the data source from other EHRs data sources?

Findings (Steps 2 and 3)

Preliminary questionnaire findings (Step 2a)

Patient lives represented in the EHR data sources, data elements, and coding schemes available

The range of patient lives represented varied among the 17 partners that responded and was not directly related to the type of data the partner had (e.g. claims only, claims and EHR, EHR aggregators). For example, existing partners with integrated delivery systems had total patient lives ranging from 400,000 to 2.9 million. For existing claims-based insurers and/or EHR aggregators, patients live totals ranged from 53 to 150 million. All new partners that returned the preliminary questionnaire, with the exception of one focused on specialty care, had slightly higher total patient lives, ranging from 62 to 280 million.

When asked how many patients had at least one visit in the EHR data in each of the last three years, the answers were split proportionally between "few", "about half," and "most". See **Table 2.** below for the breakdown of responses.

Table 1 Patient representation in EHR over each of the last three years

	None	Few	About half	Most	All
New	--	2	3	1	--
Existing	--	3	4	4	--

Data elements available (see **Appendix C .** Question 4 for the full list) were consistent across all partners with most having access to coded data using standard terminologies, prescription administration and dispensing information, lab, imaging, pathology, and radiology procedures, as well as demographic characteristics, diagnoses, and medical history. Few of the partners have procedure notes, lab, imaging, pathology and radiology notes, or free text clinical notes. See **Figure 2** for proportion of each available data element across Data Partners.

Figure 2. Number of Data Partners with inpatient and outpatient data available, by data element
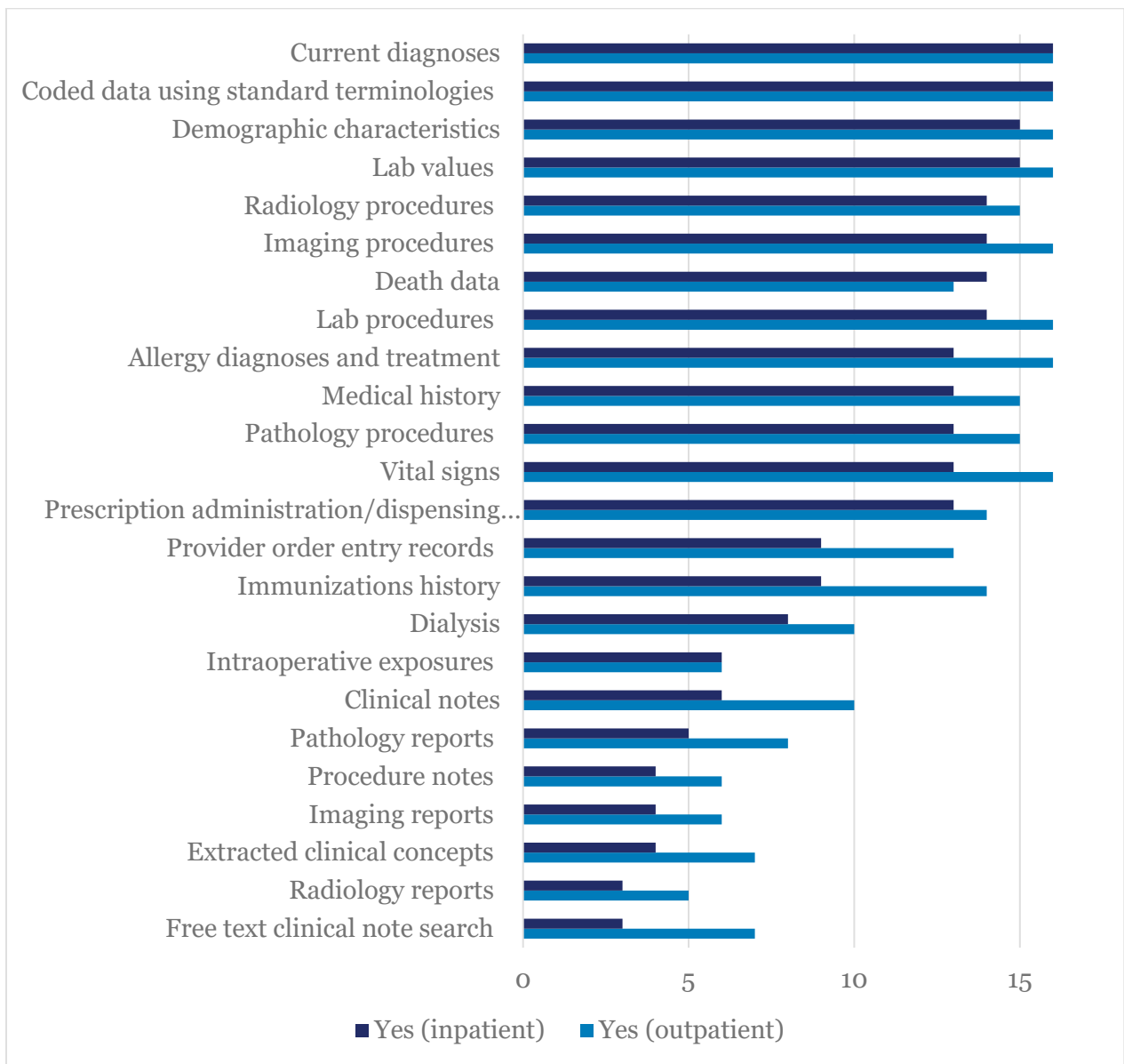
The final question of the preliminary questionnaire asks about coding schemes (e.g., NDC, ICD-9, ICD-10, LOINC, SNOMED-CT) available for use. Generally, most coding schemes listed were available at all partners. See Figure 3 for the number of Data Partners with each coding scheme available.

Figure 3. Number of Data Partners with coding schemes of interest
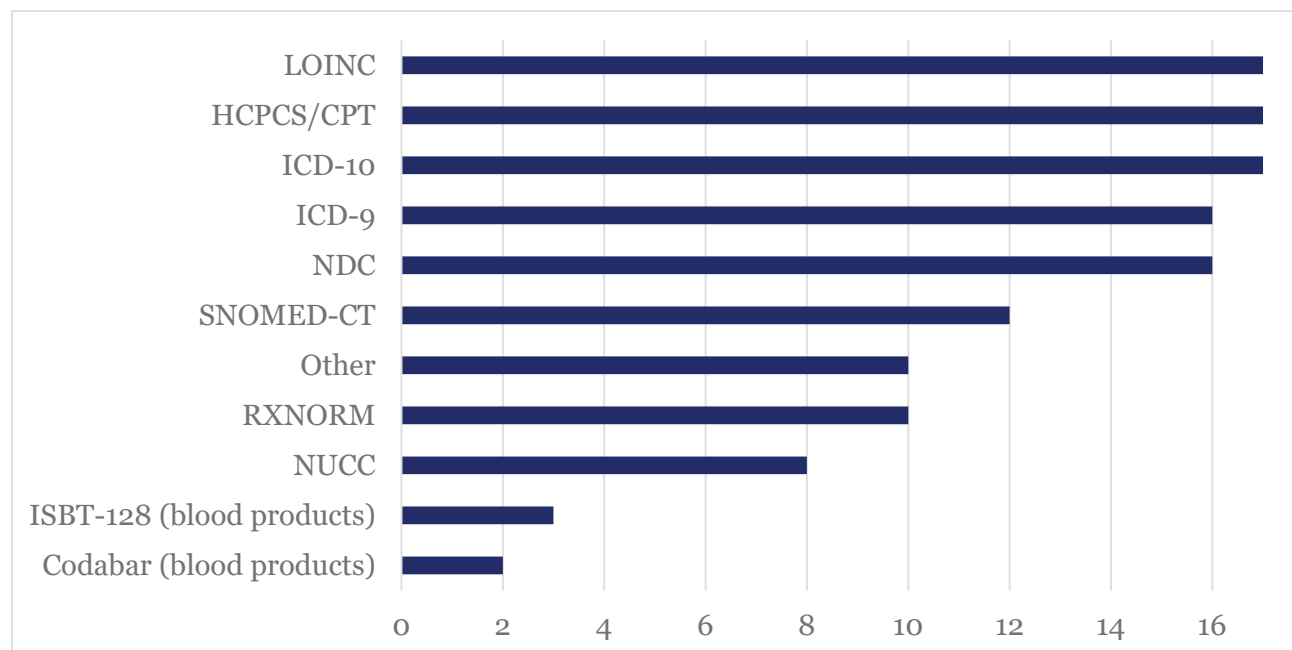
Data source time availability, care settings, and data updates

Questions 5-7 in the preliminary questionnaire cover the years of available data, care settings, and how often the data are updated. The earliest data available dates to pre-1980, though it was fairly consistent across partners that the highest volume of data was over the past five years. This is not surprising, as the mandate requiring electronic medical records for all practitioners took effect in 2014. [†]

Most data sources reported having inpatient, outpatient/ambulatory, and emergency department settings included in their data, although some are dependent on linkages or are facility dependent. A few sources only covered a single care setting (e.g., inpatient only or outpatient only). See Figure 4 for number of Data Partners with data available in each care setting of interest.

[†]American Recovery and Reinvestment Act (ARRA) of 2009
 https://www.congress.gov/bill/111th-congress/house-bill/1/text

Figure 4. Number of Data Partners with data available in specific care settings
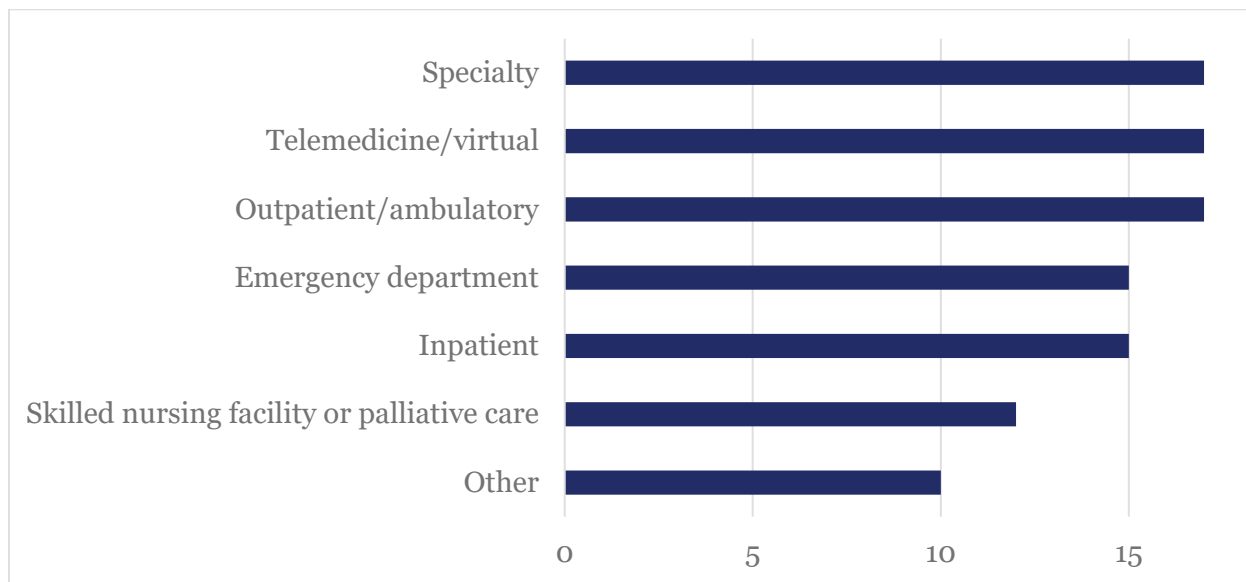
When asked about how often data is refreshed, nearly a third of partners update daily, a quarter update quarterly, and nearly one third update monthly. This breakdown did not differ between new and existing partners.

Prioritization for Step 2b
Based on findings from Step 2a, the workgroup prioritized all new partners for Step 2b.

Structured questionnaire (Step 2b)

As described above, Step 2b focused on 10 topic areas of interest, with a focus on understanding if each data source aligns with the needs of the RWE Data Enterprise, the mechanism for accessing the data for safety assessments, the degree of overlapping patient lives with the current Sentinel Common Data Model (SCDM), quality control and assessment, and governance issues that could deter a partner from engaging in future queries.

Cross-cutting considerations
Though partners completed the preliminary questionnaire in Step 2a, outlining the data elements available, Step 2b was intended to provide more details on some of those elements of interest (including pregnancy, lab, and death data).

Of the seven DPs who completed secondary questionnaire:

- Patient lives in EHRs with claims linkage: four had all or a portion of EHR data linked to claims data
- Extracted clinical concepts (a subdomain of natural language processing): four had extracted clinical concepts, one of which had this for both the inpatient and outpatient caresetting
- Pregnancy: four had information on pregnancy, one of which also had the ability to link the mother and infant
- Death: four had information on death, with the caveats including ability to identify only in the inpatient setting or if captured in the HER. Additionally, one data partner mentioned a lengthy lag for death data.
- Specialty care settings (e.g. oncology): six had access to specialty care settings and two of those also had access to registries

- Lab procedures and values: six had lab data available, two of which had both inpatient and outpatient labs

In addition to considering the types of data elements at each partner, a goal of this phase was to assess the degree of ease in accessing the data for safety assessments, the type of data model being used, quality control and assessment, and governance issues that could deter a partner from engaging in future queries. Most partners followed their own data model, and all had their own quality control and quality assessment measurements; some partners provided additional supporting documentation outlining this information.

Data partners covered from 2.7 million patient lives to 280 million patient lives, most with 1-10 encounters per patient per year. Most used their own, OMOP or FHIR® standards as their technical data environment. Four data partners currently link to claims data, with one in planning to link their EHR data to claims data in the future. All have quality control and quality assessment processes that vary by Data Partner and include FHIR® validator technical standards to partner-specific processes and checks. Few Data Partners allow direct access to data, and would need Data Use Agreements or other contracts to allow collaboration.

Follow-up discussions (Step 3)

Six Data Partners participated in follow-up discussions including one specialty care setting Data Partner that had not previously participated in the questionnaires and structured interview. These additional conversations provided the workgroup with information to further refine the list of potential participants in Phase 2 of the project, the empirical querying. After these discussions, the workgroup determined three Data Partners were the most relevant and promising based on the type and amount of data available, their experience working with investigators, and their ability to contract with external parties.

Synthesis and considerations

There are a range of issues to consider when investigating partnering with new EHR-based data sources to meet FDA's RWE Data Enterprise mandate and the analytic needs associated with supporting that mandate. Potential options range from population-based sources with patients and information across disease areas to disease-specific sources (similar to registries) that have depth within the relevant disease area. It is important to note that certain EHR-based data sources which could facilitate greater access to granular patient-level information, such as clinical notes, may be more appropriate to meet methodological research needs, to fill gaps in special populations and outcomes, and to expand the capacity of Sentinel for post-marketing surveillance and safety issues. However, additional long-term considerations are important to ensure a smooth functioning operational partnership, including the mechanisms available for querying, governance of data access and use, the population covered, the breadth and depth of the data, access to clinical notes, operational complexity, and cost.

Challenges to data use

- **Query Approach**: Data sources vary in how queries are executed and the level of interactive querying, including the need and availability of data source engagement and support. It is still unclear whether the IC would need to run queries on a platform and how much involvement would be required of a team at each partner. For example, one data source is largely accessible only via their own front-end interface and enables real-time interactive querying with user support from a dedicated team. Others enable use of project-specific datasets but have limited support.
- **Responsiveness**: The workgroup will need to explore turn-around times from each partner in Step 3.
- **Governance**: Governance issues span from contracting, to data access, to dissemination, to cost; all are critical in understanding how well a new data source can be used to support FDA needs. Access to source data and the ability to engage directly with patients, providers, and health systems that generate the data are also critical. Finally, the ability to enhance data sources via project-specific chart review, NLP, or other mechanisms will be assessed.

Unique opportunities

- **Emerging methods**: New developments in NLP and chart review are emerging and there is a need to understand the capabilities that exist at a given partner. If capabilities to process unstructured texts exists, the workgroup needs to consider the level of effort that could be offloaded from the IC and the cost/benefit analysis. The workgroup will examine this further in Phase 2.
- **Specialty data**: Although some partners have a smaller subset of the patient population in comparison to other partners, the workgroup should examine the benefit that specialty data could offer in answering safety questions for FDA.

Conclusion

To address the Real-World Evidence (RWE) Data Enterprise mandate, FDA has tasked Sentinel with establishing a query-ready, quality-checked distributed data network containing electronic health records. This report describes the development and findings of a horizon scan to identify viable EHR sources to fulfill the RWE data enterprise mandate as well as interviews with these potential partners to understand their data and analytic capabilities for meeting the needs of

this system. The horizon scan identified many potential data sources that could be useful for expanding Sentinel capabilities for analyses of EHR data. The subsequent discussions with data partners highlighted unique opportunities with respect to methodologic capabilities and data granularity, as well as challenges and considerations in technical, operational and governance domains. In the second phase of this activity, we will work with selected data sources to conduct empirical queries to gain further insights into the considerations and challenges in establishing a query-ready electronic health record network.

Appendix A. PubMed Search Strategy

PubMed Search Strategy (June 18, 2020)

("electronic health records"[MeSH Terms] OR "data aggregation" OR "electronic health records data" OR "electronic health record data" OR "EHR Data" OR "EHR" or "Electronic Health Data") AND (("2018/01/01"[PDAT] : "2020/6/18"[PDAT]) AND English[lang]) NOT "Address" [Publication type] NOT "Autobiography" [Publication type] NOT "Bibliography" [Publication type] NOT "Biography" [Publication type] NOT "Case Reports" [Publication type] NOT "Clinical Conference" [Publication type] NOT "Clinical Trial, Veterinary" [Publication type] NOT "Comment" [Publication type] NOT "Congress" [Publication type] NOT "Consensus Development Conference" [Publication type] NOT "Consensus Development Conference, NIH" [Publication type] NOT "Corrected and Republished Article" [Publication type] NOT "Dataset" [Publication type] NOT "Dictionary" [Publication type] NOT "Directory" [Publication type] NOT "Duplicate Publication" [Publication type] NOT "Editorial" [Publication type] NOT "Festschrift" [Publication type] NOT "Government Publication" [Publication type] NOT "Guideline" [Publication type] NOT "Historical Article" [Publication type] NOT "Interactive Tutorial" [Publication type] NOT "Interview" [Publication type] NOT "Lecture" [Publication type] NOT "Legal Case" [Publication type] NOT "Legislation" [Publication type] NOT "Letter" [Publication type] NOT "News" [Publication type] NOT "Newspaper Article" [Publication type] NOT "Observational Study, Veterinary" [Publication type] NOT "Patient Education Handout" [Publication type] NOT "Periodical Index" [Publication type] NOT "Personal Narrative" [Publication type] NOT "Portrait" [Publication type] NOT "Practice Guideline" [Publication type] NOT "Published Erratum" [Publication type] NOT "Retracted Publication" [Publication type] NOT "Retraction of Publication" [Publication type] NOT "Scientific Integrity Review" [Publication type] NOT "Scientific Integrity Review" [Publication type] NOT "Technical Report" [Publication type] NOT "Twin Study" [Publication type] NOT "Video-Audio Media" [Publication type] NOT "Webcast" [Publication type]

Filters:

- Clinical Study
- Clinical Trial Protocol
- Clinical Trial, Phase I (and Phase II, Phase II, Phase IV)
- Comparative Study
- Controlled Clinical Trial
- Journal Article
- Multicenter Study
- Observational Study
- Pragmatic Clinical Trial
- Randomized Controlled Trial

Appendix B. Overview of unique EHR databases identified by scoping review

Appendix A. Overview of unique electronic health record [3] databases identified by scoping review					
No.	Name	Type of EHR data	Number of facilities or providers or patients covered	Region	Comments	Citation(s)
1	Allergy Partners	New: Specialty (Allergy)	>110 Locations	Spans over 20 states	Largest US single-specialty network of allergists and immunologists. Available data reportedly includes patient demographics, diagnoses, prescription medications, laboratory test results, and other allergy-specific information.	4
2	Allina Health System	New: Health System	>100 Locations	Minnesota, North Dakota, Wisconsin	Single, comprehensive EHR system across Allina Health System.	5
3	Amgen Oncology Services Comprehensive Electronic Records (OSCER)	New: Specialty (Cancer)	>50 Locations	National	In operation since 2004, powered by Flatiron Health since 2015. Covers all payer types.	6
4	Armed Forces Health Longitudinal Technology Application (AHLTA)	Government Entity	>800 Locations	International	US Military EHR for Uniformed Services members, retirees, and their families. Provides inpatient and outpatient information.	7
5	AthenaHealth	New: Specialty (Ambulatory)	>160,000 Providers	National	AthenaHealth provides EHR services to other organizations and practices.	8
6	Atrius Health	New: Specialty (Ambulatory)	Approx. 30 Locations	Greater Boston Area	In use since 1999. Includes Harvard Vanguard.	9
7	axiUm Ascend	New: Specialty (Dental)	90% of North American Dental Schools	National	Complete dental institution management software solution serving many dental schools and institutions.	10
8	Axon Registry	New: Specialty (Neurology)		National	Hosted by the American Academy of Neurology. Collects quality improvement data.	11
9	BioMe Bank (Mount Sinai Health System)	New: Health System	~24 Locations	New York City	Electronic Medical Record (EMR)-Linked BioBank established in 2007, focusing on genetic research. Patients consent to DNA sequencing and allow contact from researchers. [12]	13
10	Cancer Research Network (CRN)	Existing Sentinel Collaborator	15 Health Plans	National	The CRN comprises of 15 participating health groups that are also part of the Health Care Systems Research Network (HCSRN). Their Virtual Data Warehouse extracts data from the participating health care systems' EHR for a common data model. This network was funded by a cooperative agreement grant from 1999-2018 and had published over 1,500 articles to date	14
11	CAPriCORN (Chicago Area Patient-	Existing Sentinel Collaborator: PCORnet	>340 Locations	Chicago area	Serves more than 10 million patients and 74,000 providers. Network sites include:	15

	Centered Outcomes Research Network)*				AllianceChicago, Cook County Health, Edward Hines Jr. VA hospital and Jesse Brown VA Medical Center, Loyola Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, NorthShore University Health System, Northwestern Medicine, Rush University Medical Center, University of Chicago, University of Illinois Hospital & Health Sciences System, Medical Research Analytics and Informatics Alliance	
12	Center for Kidney Disease Research, Education, and Hope (CURE-CKD)	New: Specialty (Kidney)	Providence St. Joseph Health (WA, MT, OR, AK, CA), and UCLA Health	West Coast	Focuses specifically on Kidney Disease research. Was developed to bridge the two health systems' distinct data bases to identify patients with chronic kidney disease (CKD) or were at risk for CKD from the EHR data.	16,17
13	Centura Health	New: Health System	>20 Locations	Colorado, western Kansas	Implemented in 2007. Centura Health is Colorado's largest health care provider. Connected with CORHIO health information exchange. [18]	19
14	Cerner HealthFacts	New: EHR Aggregator	>90 Healthcare Systems	National	It contains '100% of patients in Orchid, Keck Care and KIDS'. [20]	21, 22, 23, 24
15	Children's Hospital of Philadelphia [25] + Pediatric Research Consortium (PeRC)	New: Specialty (Pediatrics)	>30 Locations	Greater Philadelphia area (PA, NJ, DE)	EHR system for a multi-state pediatric integrated delivery system.	26, 27, 28, 29
16	Colorado Health Observational Regional Data Service (CHORDS)	New: EHR Network	~14 Locations	Colorado	Distributed Data Network (DDN) - EHR-based public health surveillance for health and mental health care providers. [30]	31, 32
17	Columbia University Medical Center/New York Presbyterian Hospital	New: Health System	>150 Locations	New York City	Jointly managed EHR for one of the largest physician practices in the Northeast.	33
18	Community Health Applied Research Network	New: Specialty (Community Health Centers)	17 Community Health Centers	9 States	CHARN consists of four research node centers (Alliance of Chicago Community Health Services, Association of Asian Pacific Community Health Organizations, Fenway Health, and Oregon Community Health Information Network) and a data coordinating center (Kaiser Permanente Center for Health Research) and was formed to study underserved populations.	34, 35
19	Department of Veterans Affairs (VA)	Government Entity	>1700 Locations	National	The Veterans Health Administration is the largest integrated national healthcare delivery system, caring for nearly 9 million veterans by 180,000 providers. Formerly, VITSA (Veterans Health Information Systems and Technology Architecture), the VA is in the	36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64,

					process of transitioning to Cerner Millennium as of 2018, with projected completion in 2028.	
20	Duke University Health System	Existing Sentinel Collaborator (PCORnet)	>200 Locations	North Carolina, primarily Durham County	Switched to an integrated system in August 2013, with some pre-2014 data available. Provides inpatient, outpatient encounters.	65-67
21	eMERGE (Electronic Medical Records and Genomics) Network	New: EHR Network	11 Locations	National	Biobank with funding from National Institutes of Health. Announced in Sept. 2007 with a goal of combining biorepositories with EMR systems. Has at least 12 centers across the country from which it draws data, with the coordinating center housed at Vanderbilt University. [68]	69, 70, 71,72, 73
22	Epic Cosmos	New: EHR Aggregator	70 Facilities	National	Multispecialty EHR with data from over 250 million patients drawn from hundreds of participating organizations. [74]	25,27,75-77, 78, 79, 80, 81, 82, 83, 84, 85
23	Flatiron Health (OncoEMR)	New: Specialty (Cancer)	>280 Clinics	National	Part of the OncoCloud suite, providing additional EMR support. Founded in 2012 . Includes partnerships with 7 major academic research centers and multiple top therapeutic oncology companies. [86]	87, 88, 89-91, 92, 93,94, 95, 96, 97
24	Geisinger Health System (part of DiscovEHR)	New: Health System	3 Hospitals and Multiple Clinics	Rural northeastern and central Pennsylvania	Integrated with lab and pharmacy services, and including "Proven Care" clinical decision support tools. Geisinger Health System serves has more than 3 million patients and contains both inpatient and outpatient information. Established in 1993. [98]	99, 100, 101, 102,103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119
25	GPC (Greater Plains Collaborative)*	Existing Sentinel Collaborator: (PCORnet)	12 Locations	9 states - Midwest	Includes University of Kansas, Allina, Indiana University, Intermountain Healthcare, Marshfield Clinic Research Institute, Medical College of Wisconsin, University of Iowa, University of Missouri, University of Nebraska, University of Texas Health Science Center at San Antonio, University of Texas Southwestern, and University of Utah.	120
26	Harris Health System	New: Health System	>30 Locations	Harris County, Texas	Harris County is the third most populous county in the US. An agreement with the Houston Dept. of Public Health and Human Services was reached in 2013 to create a unified EHR system with Harris Health System. [121] There are more than 250 million patients in this system. [122]	123
27	Health Data Compass (UCHealth System)	New: Health System	~4 Locations	Colorado	This is an enterprise health data warehouse integrating patient clinical data from UCHealth and Children's Hospital of Colorado. This also includes CU Medicine. Compass can also be linked to Colorado APCD and Center for Improving Value in Health Care.	19

28	HealthCore Integrated Research Database (HIRD)*	Existing Sentinel Collaborator (PCORnet)		14 States	Alongside PRACnet, HealthCore is one of two Health Plan Research Networks within PCORnet. A subsidiary of Anthem insurance, one of the largest health benefits company in terms of medical enrollment in the US. Established in 1996. Specializes in health economics and outcomes research. Has claims for more than 48 million individuals and lab result data for more than 17 million. [124]
					124
29	HealthPartners	Existing Sentinel Collaborator	>100 Locations	Minnesota, western Wisconsin	Part of the Patient Outcomes Research to Advance Learning (PORTAL) Clinical Data Research Network supported by PCORI.
					125, 126, 127
30	Henry Ford Health System	New: Health System	>35 Locations	Detroit, Michigan	Exists in partnership with Epic. Has more than 1,700 employed physicians. [128]
					126
31	IBM Explorys EHR Database (Watson Health) - includes Truven MarketScan	Existing Sentinel Collaborator	>63 Million Patients	National	Explorys has an aggregate of EHR from 26 US healthcare systems. Includes ambulatory, to inpatient, to specialty care from integrated delivery networks, clinically integrated networks, and accountable care organizations. Updated weekly, integration with claims data is available.
					129-131, 132, 133
32	Indiana Network for Patient Care	New: Health Information Exchange/Local Health Information Infrastructure	>120 Locations	Indiana	Established and collected data for over 30 years. INPC is the nation's largest clinical data repository with over 12 billion data elements.
					134, 135
33	INSIGHT-NYC*	Existing Sentinel Collaborator (PCORnet)	5 Locations	New York City	Largest urban clinical network in the US. Includes Weill Cornell, Albert Einstein COM, Hospital for Special Surgery, Icahn SOM, NYPH, NYU SOM, NYU SOM-Medicaid, and Columbia University.
					136
34	IQVIA Real-World Data Electronic Medical Records	New: EHR Aggregator	>16 Million Providers	National	Formerly GE Centricity. One of the world's largest accessible real world data platforms with connections to 20 regularly updated datasets. Includes PharMetrics Plus, a health plan database of adjudicated medical and pharmacy claims established since 2006.
					137-139, 140, 141
35	IRIS (Intelligent Research in Sight) Registry	New: Specialty (Ophthalmology)	>2300 Locations	National	Operated by the American Academy of Ophthalmology, as the nation's first comprehensive clinical registry of eye disease.
					142, 143
36	Johns Hopkins Medicine – EpicCare Ambulatory	New: Health System	>4 Hospitals	Maryland and District of Columbia	Includes records for the entire Johns Hopkins Medicine enterprise with the exception of the Johns Hopkins All Children's Hospital.
					144
37	Kaiser Permanente	Existing Sentinel Collaborator	>650 Locations	National	Kaiser Permanente has multiple regions: Northern California, Southern California, Colorado, Northwest, Washington state, Hawaii, Maryland, Virginia, and District of Columbia.
					145,146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156,157, 158, 159, 126, 160, 161, 162, 163, 164, 165, 166, 167, 19, 168, 169
38	Keck Care EHR*	New: Health System	2 Hospitals	Southern California	Based in University of Southern California School of Medicine. Uses Cerner technology.
					170, 171
39	Maine Health Information Exchange	New: Health Information Exchange/Local	>460 Locations	Maine	Statewide Health Information Exchange since 2010.
					172,173

	Network (HealthInfo Net)	Health Information Infrastructure				
40	Marshfield Clinic Health System EHR - 'Cattails Software Suite'*	Existing Sentinel Collaborator	~50 Locations	Northern, Central, and Western Wisconsin	Was among one of the first developers of a working EMR. [174]	[175]
41	Mayo Clinic	New: Health System	~90 Locations	National	Links to a Mayo Clinic Bio bank with at least 50,000 patients. [176]	[177-181]
42	MedStar Health System	New: Health System	~195 Locations	DC, Maryland, Virginia (DMV) Area	MedStar is the largest healthcare provider in the DMV area, and is one of the first health systems to join the Cerner Learning Health Network.	[182]
43	MetroHealth	New: Health System	>30 Locations	Ohio (Cleveland Metro area), Midwestern US	The first safety-net health care system in the US to implement EHR with its ambulatory clinics in 1999. Has more than 1.5 million patients. [183]	[26]
44	Michigan Emergency Department Improvement Collaborative (MEDIC)	New: Specialty (Emergency Departments)	15 EDs	Michigan	Member sites contribute EHR, trained abstractors add supplementary data. Physician-led partnership.	[184]
45	Modernizing Medicine Data Services	New: EHR Aggregator		National	Provides EHR sources for the following specialties: dermatology, gastroenterology, ophthalmology, orthopedics, otolaryngology, pain management, plastic surgery, and urology.	[185]
46	National Basketball Association EHR	New: Specialty (NBA Players)	30 NBA Teams	National	Information on injuries, illnesses, and NBA game participation.	[186]
47	National Cardiovascular Data Registry	New: Specialty (Cardiovascular)	10 Registries	National	Developed by American College of Cardiology. Includes the following hospital registries: Chest Pain – MI, AFib Ablation, CathPCI, ICD, iMPACT, LAAO, PVI, STS/ACC TVT; in addition to outpatient registries: diabetes collaborative, PINNACLE, and CathPCI. COVID-19 data available for Chest Pain – MI and CathPCI.	[187]
48	National Football League (NFL) EHR	New: Specialty (NFL Players)	32 NFL Teams	National	Initiative began in 2014 to capture injury and treatment information for players through an EHR system.	[188]
49	National Institute of Child Health and Human Development Pediatric Trials Network	New: Specialty (Pediatrics)	9 Centers	National	Created by Pediatric Trials Network (PTN) with nearly 265,000 pediatric patients with inpatient encounters between January 6, 2013 – June 30, 2017. Had 147 mandatory and 99 optional data elements.	[189]
50	National Trauma Databank	New: Specialty (Trauma Registries)	Approx. 900 Trauma Centers	National	Largest aggregation of US trauma registry data .	[190]
51	Nebraska Medicine EHR System	New: Health System	Multiple hospitals and clinics	Nebraska	The Nebraska Medicine Community Connect EPIC Network spans over 13 counties.	[191]
52	Net Health Wound Care EHR	New: Specialty (Wound Care Centers)	Approx. 58 Locations	National	Formerly known as WoundExpert, with >20,000 clinicians. Can connect with any hospital EHR system.	[192]
53	New York City Dept. of Health	New: Specialty (Ambulatory)	>500 Locations	New York City	Aggregate data from Primary Care Providers participating in the NYC	[32]

	and Mental Hygiene EHR – Macroscope				Primary Care Information Project (PCIP) launched in 2005. Serves more than 700,000 patients .	
54	New York University Langone Family Health Centers EHR	New: Specialty (Family Health Centers)	>45 Locations	New York City	Formerly known as Lutheran Family Health Centers. A community-based health network with more than 100,000 patients. [193]	194
55	Observational Health Data Sciences and Informatics (OHDSI)	New: EHR Network	Numerous Research Locations Internationally	International	Consists of federated datasets. Multi-stakeholder, interdisciplinary collaborative. Houses at least half a billion unique patient records. International but led out of Columbia University.	195
56	OCHIN Network	New: Specialty (Safety Net Clinics)	>140 Locations	National	OCHIN (not an acronym) is a nonprofit health information technology organization that provides a single, linked (each patient has a single identification number and medical record shared across every clinic in the network) instance of the Epic EHR.	196
57	OneFlorida	Existing Sentinel Collaborator (PCORnet)	>1260 Locations	Florida	Includes UF Health, Bond Community Health Center, CommunityHealth IT, Advent Health, Florida State University, Tallahassee Memorial, Nicklaus Children's Hospital, Orlando Health, and University of Miami.	197
58	Optum (Humedica)	Existing Sentinel Collaborator	>7100 Locations	National	Integrated Healthcare Delivery Networks. EMR agnostic, normalizing and integrating provider data from different platforms. Includes the following specialties: oncology, infectious disease, central nervous system disorders, immunology, metabolic disorders, cardiovascular, gastrointestinal, and respiratory.	198, 199, 200, 201, 202, 203, 204, 205
59	PaTH Towards a Learning Health System*	Existing Sentinel Collaborator (PCORnet)	8 Health Systems	8 states - New England, Mid Atlantic, midwestern regions	University of Pittsburgh Medical Center, Geisinger Health System, Johns Hopkins University, The Ohio State University, Penn State Hershey Medical Center, Temple University, and University of Michigan.	206
60	Patient Outcomes to Advance Learning (PORTAL) Network	Existing Sentinel Collaborator (PCORnet)	10 Healthcare Delivery Systems	National	Includes data from Kaiser Permanente (Southern California, Colorado, Northwest, Washington state, Hawaii, Maryland, Virginia, District of Columbia), Denver Health, and HealthPartners. Data elements include membership status, demographics, vital signs, health care use, laboratory values, and pharmacy dispensing.	125,207
61	Partners Research Patient Data Registry (RPDR)	Existing Sentinel Collaborator	Approx. 75 Locations	Massachusetts	Also contains the Partners' Enterprise-wide Allergy Repository (PEAR) EHR. RPDR is a centralized clinical data registry, pulling data from multiple sources including Mass General Brigham (formerly known as Partners Healthcare), Clinical Data Repository, Epic, and the Enterprise Patient Master Index. [208]	209, 210, 211, 212, 213, 214
62	PEDSnet*	Existing Sentinel Collaborator	8 Locations	12 States	CHOP, Cincinnati Children's, Nemour Children's, Research	215

#	Name	Type	Partners/Locations	Region	Description	Ref
		(PCORnet)/Specialty: Pediatrics			Institute at Nationwide children's Hospital, Seattle Children's Research Institute, U Colorado Denver, Washington University at St. Louis .	
63	PRACnet (Humana)*	Existing Sentinel Collaborator (PCORnet)		Midwest	Listed as the research subsidiary of Humana insurance company. HealthCore and PRACnet make up the two Health Plan Research Networks in PCORnet.	216
64	REACHnet (Research Action for Health Network)*	Existing Sentinel Collaborator (PCORnet)	6 Partners	Louisiana	Based on a partnership with Greater New Orleans Health Information Exchange (GNOHIE) run by Louisiana Public Health Institute, Pennington Biomedical Research Center's (PBRC) HarmonIQ data warehouse, Baylor Scott & White, Ochsner Health, Tulane University, and University Medical Center. Supports research into obesity, diabetes, sickle cell disease, and some rare types of cancer.	217
65	San Francisco Bay Collaborative Research Network	New: Specialty (Primary Care)	At least 3 health systems	Northern California	Hosted by University of California San Francisco. This seems to be a study-specific collaboration that created a data repository. The health systems are de-identified in the study.	218
66	Sight Outcomes Research Collaborative Ophthalmology Data Repository (SOURCE)	New: Specialty (Ophthalmology)	Approx. 11 Clinics		Based out of University of Michigan Kellogg Eye Center, Ann Arbor. Contains more than 500,000 patients with ocular diseases, 1.2 million visits, 36,000 eye surgeries, 8 million lab test results, 17.8 medication orders, and 530,000 images of the retina. [219]	3
67	STAR (Stakeholders, Technology, and Research Clinical Research Network)*	Existing Sentinel Collaborator (PCORnet)	Academic Health Centers, Community Hospitals, clinics, etc.	Southeast US	Includes Vanderbilt, Duke, Health Sciences South Carolina, Mayo Clinic, Meharry Medical College, UNC Chapel Hill, Vanderbilt Health Affiliated Network, and Wake Forest Baptist Health.	220
68	Sutter Community Connect	New: Health System	>170 Locations	Northern California	Sutter Health System's EHR is Epic-integrated EHR, includes integration with clinical data repository, lab, radiology, and pharmacy data.	25
69	University of Pennsylvania Health System (Penn Medicine)	New: Health System	Approx. 15 Locations	Southeast Pennsylvania, Central and South New Jersey, and Delaware.	Penn Medicine has a Clinical Data Warehouse known as Penn Data Store. This collects patient data from 12 core Penn Medicine information systems including both outpatient and inpatient data. [221]	222, 223, 224
70	University of Utah Health System	New: Health System	>120 Locations	Utah	Has records dating from 1990 and more than 1.4 million patient records.	225
71	University of Washington Medicine Healthcare System	New: Health System	Approx. 8 Locations	Seattle and King County, WA State	Replaced its Cerner and Epic EHRs starting 2018. It appears to have ORCA [20] and EPIC Care. UW Health System is comprised of at least 8 different entities and as of 2021 is currently undergoing transformation to an enterprise-wide single EHR. [226]	227

72	University of Wisconsin Health System	New: Health System	Approx. 27 Locations	Wisconsin	Implemented in the University of Wisconsin's emergency departments in 2008. [228]	229
73	Vanderbilt University Medical Center (VUMC) – BioVU, Synthetic Derivative, and eStar EHR	Existing Sentinel Collaborator (PCORnet)	Medical Complex System	Central Tennessee	Biorepository of DNA, linked to de-identified medical records in the Synthetic Derivative. Synthetic Derivative is derived from Vanderbilt's EMR, linked to the BioVU BioBank.	230, 178, 231, 232, 233, 234, 235-237, 238
74	Veradigm	Existing Sentinel Collaborator		National	Veradigm houses a few EHR data sources including: Practice Fusion (150,000 medical professionals), AllScripts (Professional EHR, Touchworks EHR), Paragon, Sunrise.	76,239,240
*Identified via web search					

Appendix C. Preliminary Questionnaire

Sentinel Innovation Center Electronic Health Record (EHR) Horizon Scan

The following questions are designed to help us understand more about your EHR data source and the data elements that are available. This is intended to be a brief overview and should not take longer than 15 minutes to complete. For the purposes of this initial questionnaire, if an exact response is not readily available, please provide your best estimate. For all questions, consider the most relevant answer for the past five years. We understand that answering questions such as these often have substantial subtlety which we plan to address in our follow-up discussions. Please feel free to add context to the answers as appropriate.

1. **How many patient lives are represented in your EHR data source?**	N=
A. **How many patients contribute both EHR data and claims data to your data source?**	None Few About half Most All
	Additional comments:
B. **For patients with both EHR and claims data, how much of a patient's observation time is overlapping between both EHR and claims data, on average?**	None Few About half Most All
	Additional comments:
2. **On average, how many visits or encounters (e.g., hospitalization, office visit, lab draw) does a patient have in your EHR data source per year?**	<1 1-10 >10 NA
3. **How many patients have at least one visit in the EHR data in each of the last three years?**	None Few About half Most All

4.	Are the following data elements available from your EHR data for analysis?			
A.	Coded data using standard terminologies (see question 8 for examples)	Yes (inpatient)	Yes (outpatient)	No	NA
B.	Extracted clinical concepts (a subdomain of Natural Language Processing)	Yes (inpatient)	Yes (outpatient)	No	NA
C.	Clinical notes	Yes (inpatient)	Yes (outpatient)	No	NA
D.	Free text clinical note search (ability to search chart notes for a large cohort of patients simultaneously)	Yes (inpatient)	Yes (outpatient)	No	NA
E.	Prescription administration/dispensing information	Yes (inpatient)	Yes (outpatient)	No	NA
F.	Lab procedures (test occurred and/or administrative codes)	Yes (inpatient)	Yes (outpatient)	No	NA
G.	Lab values (microbiology, chemistries, hematology, inflammatory markers, pregnancy tests, viral respiratory panels)	Yes (inpatient)	Yes (outpatient)	No	NA
H.	Vital signs	Yes (inpatient)	Yes (outpatient)	No	NA
I.	Death data	Yes (inpatient)	Yes (outpatient)	No	NA
J.	Imaging procedures (test occurred and/or administrative codes)	Yes (inpatient)	Yes (outpatient)	No	NA
K.	Imaging reports	Yes (inpatient)	Yes (outpatient)	No	NA
L.	Pathology procedures (test occurred and/or administrative codes)	Yes (inpatient)	Yes (outpatient)	No	NA
M.	Pathology reports (biopsies, tumor grade)	Yes (inpatient)	Yes (outpatient)	No	NA
N.	Radiology procedures (test occurred and/or administrative codes)	Yes (inpatient)	Yes (outpatient)	No	NA

	Yes (inpatient)	Yes (outpatient)	No	NA
O. Radiology reports (pregnancy related ultrasounds, head MRI)	Yes (inpatient)	Yes (outpatient)	No	NA
P. Demographic characteristics	Yes (inpatient)	Yes (outpatient)	No	NA
Q. Current diagnoses	Yes (inpatient)	Yes (outpatient)	No	NA
R. Medical history	Yes (inpatient)	Yes (outpatient)	No	NA
S. Allergy diagnoses and treatment	Yes (inpatient)	Yes (outpatient)	No	NA
T. Immunizations history	Yes (inpatient)	Yes (outpatient)	No	NA
U. Provider order entry records (orders)	Yes (inpatient)	Yes (outpatient)		NA
V. Procedure notes (surgical procedure notes)	Yes (inpatient)	Yes (outpatient)		NA
W. Intraoperative exposures (anesthesia)	Yes (inpatient)	Yes (outpatient)		NA
X. Dialysis	Yes (inpatient)	Yes (outpatient)		NA

5. **What is the time period of available EHR data?**

    _____

A. **Is the time period available different for different locations or providers?**	Yes	No	NA

6. **Care settings that are covered:**

A. **Inpatient**	Yes	No	NA
B. **Outpatient/ambulatory**	Yes	No	NA
C. **Telemedicine/virtual**	Yes	No	NA
D. **Emergency department**	Yes	No	NA
E. **Skilled nursing facility or palliative care**	Yes	No	NA
F. **Specialty (e.g., oncology)**	Yes	No	NA
G. **Other**	Yes	No	NA

	Daily	Monthly	Quarterly	Yearly	Other
7. Thinking about EHR data that are available for secondary use, how frequently are the research data updated?					
A. Do refreshes vary by the specific field?	Yes	No			NA
B. How old are the freshest data in your data source?	Updated daily	Week(s) old	Month(s) old	Year old	Longer than a year old
8. What coding schemes are available in your EHR data source?					
A. NDC	Yes	No			NA
B. RXNORM	Yes	No			NA
C. ICD-9	Yes	No			NA
D. ICD-10	Yes	No			NA
E. HCPCS/CPT	Yes	No			NA
F. ISBT-128 (blood products)	Yes	No			NA
G. Codabar (blood products)	Yes	No			NA
H. LOINC	Yes	No			NA
I. NUCC	Yes	No			NA
J. SNOMED-CT	Yes	No			NA
K. Other	Yes	No			NA

Appendix D. Secondary Questionnaire

Question	Notes
INTRODUCTION AND GOALS	
OVERVIEW	
DATA CAPTURE AND DATA ELEMENTS	
1. What types of data elements are collected?	
2. Do you use any natural language processing or extract standardized information for research purposes?	
3. How are providers referenced?	
3A. Provider number for clinician	
3B. Provider number for hospital/clinic	
3C. Specific department/clinic in a larger hospital/facility	
4. Is it possible to separate history of a condition from a new diagnosis?	
4A. If so, how?	
5. Are metadata available?	
DATA MODEL, RESEARCH CAPABILITIES AND TECHNICAL ENVIRONMENT	
1. Do you use any specific data model(s)? If so, describe the existing data model	
1A. In what format are the data elements?	
1B. How often are your data used for research updated?	
ABILITY TO LINK TO OTHER SOURCES OF DATA	
1. Patients frequently seek care at multiple clinical locations. How complete are the data with respect to patients receiving care at other health care systems?	
2. Is your data source able to link to other data sources?	
2A. If so, do you have any existing linkages?	
3. How many patients are linked to health insurance claims data (defined as medical, hospital and pharmacy claims)?	
3A. Are these open or closed claims?	
3B. How many linked patients have simultaneous "person-time" overlap?	

3BI. What is the average duration of overlapping "person-time"?	
4. Any other issues to discuss about linkages?	
5. Are the claims data in your data source currently available in sentinel's distributed database?	
QUALITY CONTROL AND QUALITY ASSESSMENT	
1. Describe your source's procedures for quality checking and assurance	
2. Are there rules for acceptable values, particularly for medication dose and strength	
COLLABORATION, EXPERIENCE, AND GOVERNANCE	
1. Has data or information from your data source been used in research?	
1A. Has this research been published in a peer-reviewed journal or presented at a national or international meeting?	
1B. Has the research been presented or used in your institutional setting?	
2. Is an IRB review required to conduct research with your data source?	
3. Are there existing data sharing agreements that regulate other entities' use of your data?	
3A. What contracts and/or data sharing agreements will need to be in place in order to utilize your data?	
4. Are there any additional governance barriers that need to be considered prior to using your data for research?	
5. Does your data source have existing relationships with other collaborators?	
6. Can the data be shared externally for data analysis or must analyses be conducted in house?	
7. If analyses are done at the data source, in what format are results shared (i.e., pdf, excel)	
8. Is the data source a distributed data network?	

POPULATION CHARACTERISTICS	
1. Please describe the population characteristics	
1A. Age distribution	
1B. Sex distribution	
1C. Number of unique patients over time	
2. What is the geographic distribution of the patient population?	
2A. How is the state or region identified? For example, is it the location of the clinic/provider or the home residence of the patient?	
2A1. If applicable, how is the location of the home residence captured?	
2A2. How are changes in patient residence captured?	
2B. How timely is geographic information? For example, is the zip code of the patient captured a single time or at every visit?	
3. Is your patient population restricted in any way (e.g. oncology patients, pediatrics, etc.)?	
4. Does the data source capture pediatric patient and outcomes (e.g., BMI, blood pressure)?	
DEATH DATA	
1. Are cause of death data available? If so, for what data source?	
1A. NDI	
1B. State-based death registry	
1C. Other	
ONCOLOGY DATA	
1. Does the data source capture cancer outcomes? If so, what about the following characteristics:	
1A. Disease-free survival	
1B. Mortality	
1C. Treatment regimens	
1D. Severity (e.g., histology, staging)	
LAB DATA	
1. Are lab values available?	
1A. If so, are units normalized across patients/visits/labs?	

PREGNANCY DATA	
1. Does the data source capture pregnancies? If so, what about the following characteristics:	
1A. LMP (last menstrual period)	
1B. EDC/EDD (estimated due date)	
1C. Pregnancy test results	
1D. Use of prenatal vitamins	
1E. Mother-infant linkage	
1F. Multiple births	
1G. Infant outcomes	
1H. Pregnancy outcomes	
1I. Maternal outcomes	
1J. Is the entire pregnancy captured and quantified (i.e., estimated pregnancy start through delivery date)?	

References

1. FDA. Sentinel System Five-Year Strategy 2019-2023. (FDA, 2019).
2. Brown, J.S., Maro, J.C., Nguyen, M. & Ball, R. Using and improving distributed data networks to generate actionable evidence: the case of real-world outcomes in the Food and Drug Administration's Sentinel system. *J Am Med Inform Assoc* **27**, 793-797 (2020).
3. Stein, J.D.*, et al.* Evaluation of an Algorithm for Identifying Ocular Conditions in Electronic Health Record Data. *JAMA Ophthalmol* **137**, 491-497 (2019).
4. Pilon, D.*, et al.* Asthma control, lung function, symptoms, and corticosteroid sparing after omalizumab initiation in patients with allergic asthma. *Allergy Asthma Proc* **39**, 127-135 (2018).
5. Sidebottom, A.C.*, et al.* Assessing the impact of the heart of New Ulm Project on cardiovascular disease risk factors: A population-based program to reduce cardiovascular disease. *Prev Med* **112**, 216-221 (2018).
6. Hernandez, R.K.*, et al.* Incidence of bone metastases in patients with solid tumors: analysis of oncology electronic medical records in the United States. *BMC Cancer* **18**, 44 (2018).
7. Cahill, M.S.*, et al.* Implications of New Cholesterol Treatment Guidelines on Statin Utilization, Intensity, and Costs in Active Duty Service Members. *Mil Med* **183**, e66-e70 (2018).
8. García, M.C.*, et al.* Opioid Prescribing Rates in Nonmetropolitan and Metropolitan Counties Among Primary Care Providers Using an Electronic Health Record System - United States, 2014-2017. *MMWR Morb Mortal Wkly Rep* **68**, 25-30 (2019).
9. Krakower, D.S.*, et al.* Development and validation of an automated HIV prediction algorithm to identify candidates for pre-exposure prophylaxis: a modelling study. *Lancet HIV* **6**, e696-e704 (2019).
10. Yansane, A.*, et al.* Utilization and Validity of the Dental Diagnostic System over Time in Academic and Private Practice. *JDR Clin Trans Res* **4**, 143-150 (2019).
11. Baca, C.M.*, et al.* Axon Registry® data validation: Accuracy assessment of data extraction and measure specification. *Neurology* **92**, 847-858 (2019).
12. BioMe™ BioBank Program. (Icahn School of Medicine at Mount Sinai, 2021).
13. Abul-Husn, N.S.*, et al.* Exome sequencing reveals a high prevalence of BRCA1 and BRCA2 founder variants in a diverse population-based biobank. *Genome Med* **12**, 2 (2019).
14. Uno, H.*, et al.* Determining the Time of Cancer Recurrence Using Claims or Electronic Medical Record Data. *JCO Clin Cancer Inform* **2**, 1-10 (2018).
15. CAPriCORN. (2021).
16. Norris, K.C.*, et al.* Rationale and design of a multicenter Chronic Kidney Disease (CKD) and at-risk for CKD electronic health records-based registry: CURE-CKD. *BMC Nephrol* **20**, 416 (2019).
17. Tuttle, K.R.*, et al.* Clinical Characteristics of and Risk Factors for Chronic Kidney Disease Among Adults and Children: An Analysis of the CURE-CKD Registry. *JAMA Netw Open* **2**, e1918169 (2019).
18. Centura Health Named 2017 Most Wired. (Centura Health, 2017).
19. Duca, L.M.*, et al.* Estimating the prevalence of congenital heart disease among adolescents and adults in Colorado adjusted for incomplete case ascertainment. *Am Heart J* **221**, 95-105 (2020).
20. Cerner. Cerner Health Facts: Multi-Site Electronic Health Records for Data Mining and Health Economics.

21. Rhee, C.*, et al.* Sepsis Surveillance Using Adult Sepsis Events Simplified eSOFA Criteria Versus Sepsis-3 Sequential Organ Failure Assessment Criteria. *Crit Care Med* **47**, 307-314 (2019).

22. Jamil, K.*, et al.* The burden of illness of hepatorenal syndrome (HRS) in the United States: a retrospective analysis of electronic health records. *J Med Econ* **22**, 421-429 (2019).

23. Rasmy, L.*, et al.* A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set. *J Biomed Inform* **84**, 11-16 (2018).

24. Bodagh, N.*, et al.* Feasibility of real-time capture of routine clinical data in the electronic health record: a hospital-based, observational service-evaluation study. *BMJ Open* **8**, e019790 (2018).

25. Romanelli, R.J.*, et al.* Longitudinal Weight Outcomes From a Behavioral Lifestyle Intervention in Clinical Practice. *Diabetes Educ* **45**, 529-543 (2019).

26. Czaja, A.S.*, 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* **27**, 815-822 (2018).

27. Maltenfort, M.G., Chen, Y. & Forrest, C.B. Prediction of 30-day pediatric unplanned hospitalizations using the Johns Hopkins Adjusted Clinical Groups risk adjustment system. *PLoS One* **14**, e0221233 (2019).

28. Nelson, O.*, et al.* A Visual Analytics Dashboard to Summarize Serial Anesthesia Records in Pediatric Radiation Treatment. *Appl Clin Inform* **10**, 563-569 (2019).

29. Palakshappa, D., Strane, D., Griffis, H. & Fiks, A.G. Determining Food-Insecure Families' Resource Access with Health System and Public Data. *J Health Care Poor Underserved* **30**, 265-279 (2019).

30. CHORDS (Colorado Health Observation Regional Data Service). (2020).

31. Bacon, E.*, et al.* Developing a Regional Distributed Data Network for Surveillance of Chronic Health Conditions: The Colorado Health Observation Regional Data Service. *J Public Health Manag Pract* **25**, 498-507 (2019).

32. Namulanda, G.*, et al.* Electronic health record case studies to advance environmental public health tracking. *J Biomed Inform* **79**, 98-104 (2018).

33. Levine, M.E., Albers, D.J. & Hripcsak, G. Methodological variations in lagged regression for detecting physiologic drug effects in EHR data. *J Biomed Inform* **86**, 149-159 (2018).

34. Hatch, B.*, et al.* Impacts of the Affordable Care Act on Community Health Centers: Characteristics of New Patients and Early Changes in Delivery of Care. *J Ambul Care Manage* **41**, 250-261 (2018).

35. Mayer, K.H.*, et al.* Excess Clinical Comorbidity Among HIV-Infected Patients Accessing Primary Care in US Community Health Centers. *Public Health Rep* **133**, 109-118 (2018).

36. Vair, C.L.*, et al.* Electronic Medical Record Documentation of Driving Safety for Veterans with Diagnosed Dementia. *Clin Gerontol* **41**, 66-76 (2018).

37. Liu, S.*, et al.* Outcomes associated with warfarin time in therapeutic range among US veterans with nonvalvular atrial fibrillation. *Curr Med Res Opin* **34**, 415-421 (2018).

38. Salas, J.*, et al.* Gender and the Association between Long-Term Prescription Opioid Use and New-Onset Depression. *J Pain* **19**, 88-98 (2018).

39. Feder, S.L.*, et al.* Validation of the ICD-9 Diagnostic Code for Palliative Care in Patients Hospitalized With Heart Failure Within the Veterans Health Administration. *Am J Hosp Palliat Care* **35**, 959-965 (2018).

40. Vassy, J.L.*, et al.* Yield and bias in defining a cohort study baseline from electronic health record data. *J Biomed Inform* **78**, 54-59 (2018).

41. Bensley, K.M.*, et al.* Racial/ethnic differences in the association between alcohol use and mortality among men living with HIV. *Addict Sci Clin Pract* **13**, 2 (2018).

42. Koola, J.D.*, et al.* Development of an automated phenotyping algorithm for hepatorenal syndrome. *J Biomed Inform* **80**, 87-95 (2018).

43. Rogers, E.S.*, et al.* Proactive Tobacco Treatment for Smokers Using Veterans Administration Mental Health Clinics. *Am J Prev Med* **54**, 620-629 (2018).

44. Chen, S.*, et al.* Identifying and categorizing spurious weight data in electronic medical records. *Am J Clin Nutr* **107**, 420-426 (2018).

45. Krishnamurthi, N., Francis, J., Fihn, S.D., Meyer, C.S. & Whooley, M.A. Leading causes of cardiovascular hospitalization in 8.45 million US veterans. *PLoS One* **13**, e0193996 (2018).

46. Zaman, T., Rife, T.L., Batki, S.L. & Pennington, D.L. An electronic intervention to improve safety for pain patients co-prescribed chronic opioids and benzodiazepines. *Subst Abus* **39**, 441-448 (2018).

47. Blosnich, J.R.*, et al.* Using clinician text notes in electronic medical record data to validate transgender-related diagnosis codes. *J Am Med Inform Assoc* **25**, 905-908 (2018).

48. Groppi, J.A., Ourth, H., Morreale, A.P., Hirsh, J.M. & Wright, S. Advancement of clinical pharmacy practice through intervention capture. *Am J Health Syst Pharm* **75**, 886-892 (2018).

49. Buta, E.*, et al.* Posttraumatic stress disorder diagnosis and gender are associated with accelerated weight gain trajectories in veterans during the post-deployment period. *Eat Behav* **29**, 8-13 (2018).

50. Montgomery, A.E.*, et al.* Recent Intimate Partner Violence and Housing Instability Among Women Veterans. *Am J Prev Med* **54**, 584-590 (2018).

51. Stevens, V.W.*, et al.* Characterization and survival of patients with metastatic basal cell carcinoma in the Department of Veterans Affairs: a retrospective electronic health record review. *Arch Dermatol Res* **310**, 505-513 (2018).

52. Owens, M.D.*, et al.* Receipt of alcohol-related care among patients with HCV and unhealthy alcohol use. *Drug Alcohol Depend* **188**, 79-85 (2018).

53. Williams, E.C.*, et al.* Changes in alcohol use associated with changes in HIV disease severity over time: A national longitudinal study in the Veterans Aging Cohort. *Drug Alcohol Depend* **189**, 21-29 (2018).

54. Patel, Y.R.*, et al.* Development and validation of a heart failure with preserved ejection fraction cohort using electronic medical records. *BMC Cardiovasc Disord* **18**, 128 (2018).

55. Justice, A.C.*, et al.* AUDIT-C and ICD codes as phenotypes for harmful alcohol use: association with ADH1B polymorphisms in two US populations. *Addiction* **113**, 2214-2224 (2018).

56. Morasco, B.J., Shull, S.E., Adams, M.H., Dobscha, S.K. & Lovejoy, T.I. Development of an Algorithm to Identify Cannabis Urine Drug Test Results within a Multi-Site Electronic Health Record System. *J Med Syst* **42**, 163 (2018).

57. Everett, C.M.*, et al.* Interpersonal continuity of primary care of veterans with diabetes: a cohort study using electronic health record data. *BMC Fam Pract* **19**, 132 (2018).

58. Cai, T.*, et al.* Association of Interleukin 6 Receptor Variant With Cardiovascular Disease Effects of Interleukin 6 Receptor Blocking Therapy: A Phenome-Wide Association Study. *JAMA Cardiol* **3**, 849-857 (2018).

59. DerSarkissian, M.*, et al.* Health Care Resource Utilization and Costs Associated with Transitioning to 3-month Paliperidone Palmitate Among US Veterans. *Clin Ther* **40**, 1496-1508 (2018).

60. Hebert, P.L.*, et al.* Reliance on Medicare Providers by Veterans after Becoming Age-Eligible for Medicare is Associated with the Use of More Outpatient Services. *Health Serv Res* **53 Suppl 3**, 5159-5180 (2018).
61. Alemi, F., Avramovic, S. & Schwartz, M.D. Electronic Health Record-Based Screening for Substance Abuse. *Big Data* **6**, 214-224 (2018).
62. Bensley, K.M.*, et al.* Patterns of Alcohol Use Among Patients Living With HIV in Urban, Large Rural, and Small Rural Areas. *J Rural Health* **35**, 330-340 (2019).
63. Njei, B.*, et al.* Regional and Rural-Urban Differences in the Use of Direct-acting Antiviral Agents for Hepatitis C Virus: The Veteran Birth Cohort. *Med Care* **57**, 279-285 (2019).
64. Hatef, E.*, et al.* The Impact of Social Determinants of Health on Hospitalization in the Veterans Health Administration. *Am J Prev Med* **56**, 811-818 (2019).
65. Wu, L.T.*, et al.* Substance use disorders and medical comorbidities among high-need, high-risk patients with diabetes. *Drug Alcohol Depend* **186**, 86-93 (2018).
66. Wu, L.T., Zhu, H. & Ghitza, U.E. Multicomorbidity of chronic diseases and substance use disorders and their association with hospitalization: Results from electronic health records data. *Drug Alcohol Depend* **192**, 316-323 (2018).
67. Bhavsar, N.A., Gao, A., Phelan, M., Pagidipati, N.J. & Goldstein, B.A. Value of Neighborhood Socioeconomic Status in Predicting Risk of Outcomes in Studies That Use Electronic Health Record Data. *JAMA Netw Open* **1**, e182716 (2018).
68. Electronic Medical Records and Genomics (eMERGE) Network. (Genome, 2020).
69. Stanaway, I.B.*, et al.* The eMERGE genotype set of 83,717 subjects imputed to ~40 million variants genome wide and association with the herpes zoster medical record phenotype. *Genet Epidemiol* **43**, 63-81 (2019).
70. Antommaria, A.H.M.*, et al.* Parents' attitudes toward consent and data sharing in biobanks: A multisite experimental survey. *AJOB Empir Bioeth* **9**, 128-142 (2018).
71. Aronson, S.*, et al.* Empowering genomic medicine by establishing critical sequencing result data flows: the eMERGE example. *J Am Med Inform Assoc* **25**, 1375-1381 (2018).
72. Pacheco, J.A.*, et al.* A case study evaluating the portability of an executable computable phenotype algorithm across multiple institutions and electronic health record environments. *J Am Med Inform Assoc* **25**, 1540-1546 (2018).
73. Wei, W.Q.*, et al.* LPA Variants Are Associated With Residual Cardiovascular Risk in Patients Receiving Statins. *Circulation* **138**, 1839-1849 (2018).
74. Cosmos. (Epic, 2021).
75. Dhar, J.P., Essenmacher, L., Dhar, R., Ragina, N. & Sokol, R.J. Lack of Uptake of Prophylactic Human Papilloma Virus Vaccine Among Women With Systemic Lupus Erythematosus Seen at a Regional Medical Center. *J Clin Rheumatol* **25**, 348-350 (2019).
76. Pandey, A.*, et al.* Rapid-Cycle Implementation of a Multi-Organization Registry for Heart Failure with Preserved Ejection Fraction Using Health Information Exchange Standards. *Stud Health Technol Inform* **264**, 1560-1561 (2019).
77. Schmajuk, G.*, et al.* Pneumocystis jirovecii pneumonia (PJP) prophylaxis patterns among patients with rheumatic diseases receiving high-risk immunosuppressant drugs. *Semin Arthritis Rheum* **48**, 1087-1092 (2019).
78. Ruley, M., Walker, V., Studeny, J. & Coustasse, A. The Nationwide Health Information Network: The Case of the Expansion of Health Information Exchanges in the United States. *Health Care Manag (Frederick)* **37**, 333-338 (2018).
79. Palestine, A.G., Merrill, P.T., Saleem, S.M., Jabs, D.A. & Thorne, J.E. Assessing the Precision of ICD-10 Codes for Uveitis in 2 Electronic Health Record Systems. *JAMA Ophthalmol* **136**, 1186-1190 (2018).

80. Chiu, A.S., *et al.* Association of Lowering Default Pill Counts in Electronic Medical Record Systems With Postoperative Opioid Prescribing. *JAMA Surg* **153**, 1012-1019 (2018).

81. Silvestri, M.T., *et al.* Impact of Cost Display on Ordering Patterns for Hospital Laboratory and Imaging Services. *J Gen Intern Med* **33**, 1268-1275 (2018).

82. Nayor, J., Borges, L.F., Goryachev, S., Gainer, V.S. & Saltzman, J.R. Natural Language Processing Accurately Calculates Adenoma and Sessile Serrated Polyp Detection Rates. *Dig Dis Sci* **63**, 1794-1800 (2018).

83. Nguyen, O.K., *et al.* Predicting 30-Day Hospital Readmissions in Acute Myocardial Infarction: The AMI "READMITS" (Renal Function, Elevated Brain Natriuretic Peptide, Age, Diabetes Mellitus, Nonmale Sex, Intervention with Timely Percutaneous Coronary Intervention, and Low Systolic Blood Pressure) Score. *J Am Heart Assoc* **7**(2018).

84. Turrentine, F.E., *et al.* Fewer thromboembolic events after implementation of a venous thromboembolism risk stratification tool. *J Surg Res* **225**, 148-156 (2018).

85. Westra, B.L., *et al.* Validation and Refinement of a Pain Information Model from EHR Flowsheet Data. *Appl Clin Inform* **9**, 185-198 (2018).

86. Flatiron. NICE Partners with Flatiron Health to Develop Real-World Evidence Research Methodologies. (Flatiron, 2020).

87. Velcheti, V., *et al.* Outcomes of first-line pembrolizumab monotherapy for PD-L1-positive (TPS ≥50%) metastatic NSCLC at US oncology practices. *Immunotherapy* **11**, 1541-1554 (2019).

88. Simeone, J.C., Nordstrom, B.L., Patel, K. & Klein, A.B. Treatment patterns and overall survival in metastatic non-small-cell lung cancer in a real-world, US setting. *Future Oncol* **15**, 3491-3502 (2019).

89. Carrigan, G., *et al.* An evaluation of the impact of missing deaths on overall survival analyses of advanced non-small cell lung cancer patients conducted in an electronic health records database. *Pharmacoepidemiol Drug Saf* **28**, 572-581 (2019).

90. Griffith, S.D., *et al.* Characterizing the Feasibility and Performance of Real-World Tumor Progression End Points and Their Association With Overall Survival in a Large Advanced Non-Small-Cell Lung Cancer Data Set. *JCO Clin Cancer Inform* **3**, 1-13 (2019).

91. Griffith, S.D., *et al.* Generating Real-World Tumor Burden Endpoints from Electronic Health Record Data: Comparison of RECIST, Radiology-Anchored, and Clinician-Anchored Approaches for Abstracting Real-World Progression in Non-Small Cell Lung Cancer. *Adv Ther* **36**, 2122-2136 (2019).

92. Huang Bartlett, C., *et al.* Concordance of real-world versus conventional progression-free survival from a phase 3 trial of endocrine therapy as first-line treatment for metastatic breast cancer. *PLoS One* **15**, e0227256 (2020).

93. Winfree, K.B., *et al.* Treatment patterns, duration and outcomes of pemetrexed maintenance therapy in patients with advanced NSCLC in a real-world setting. *Curr Med Res Opin* **35**, 817-827 (2019).

94. Khozin, S., *et al.* Real-World Outcomes of Patients with Metastatic Non-Small Cell Lung Cancer Treated with Programmed Cell Death Protein 1 Inhibitors in the Year Following U.S. Regulatory Approval. *Oncologist* **24**, 648-656 (2019).

95. Davies, J., Martinec, M., Coudert, M., Delmar, P. & Crane, G. Real-world anaplastic lymphoma kinase (ALK) rearrangement testing patterns, treatment sequences, and survival of ALK inhibitor-treated patients. *Curr Med Res Opin* **35**, 535-542 (2019).

96. Khozin, S., *et al.* Characteristics of Real-World Metastatic Non-Small Cell Lung Cancer Patients Treated with Nivolumab and Pembrolizumab During the Year Following Approval. *Oncologist* **23**, 328-336 (2018).

97. Butler, A.M., *et al.* Treatment dynamics of bone-targeting agents among men with bone metastases from prostate cancer in the United States. *Pharmacoepidemiol Drug Saf* **27**, 229-238 (2018).

98. Research at Geisinger. (Geisinger, 2021).

99. Verma, A., *et al.* Human-Disease Phenotype Map Derived from PheWAS across 38,682 Individuals. *Am J Hum Genet* **104**, 55-64 (2019).

100. Staples, J., *et al.* Profiling and Leveraging Relatedness in a Precision Medicine Cohort of 92,455 Exomes. *Am J Hum Genet* **102**, 874-889 (2018).

101. Park, J., *et al.* A genome-first approach to aggregating rare genetic variants in LMNA for association with electronic health record phenotypes. *Genet Med* **22**, 102-111 (2020).

102. Moon, K.A., *et al.* Epidemiology of Lyme disease in Pennsylvania 2006-2014 using electronic health records. *Ticks Tick Borne Dis* **10**, 241-250 (2019).

103. Miller, J.E., *et al.* Systematic characterization of germline variants from the DiscovEHR study endometrial carcinoma population. *BMC Med Genomics* **12**, 59 (2019).

104. Casey, J.A., Wilcox, H.C., Hirsch, A.G., Pollak, J. & Schwartz, B.S. Associations of unconventional natural gas development with depression symptoms and disordered sleep in Pennsylvania. *Sci Rep* **8**, 11375 (2018).

105. Samad, M.D., *et al.* Predicting Survival From Large Echocardiography and Electronic Health Record Datasets: Optimization With Machine Learning. *JACC Cardiovasc Imaging* **12**, 681-689 (2019).

106. Williams, B.A., *et al.* External validation of the TIMI risk score for secondary cardiovascular events among patients with recent myocardial infarction. *Atherosclerosis* **272**, 80-86 (2018).

107. Harrington, K.M., *et al.* Validation of an Electronic Medical Record-Based Algorithm for Identifying Posttraumatic Stress Disorder in U.S. Veterans. *J Trauma Stress* **32**, 226-237 (2019).

108. Wang, X.Q., *et al.* Veterans Affairs patient database (VAPD 2014-2017): building nationwide granular data for clinical discovery. *BMC Med Res Methodol* **19**, 94 (2019).

109. Vincent, B.M., *et al.* Template matching for benchmarking hospital performance in the veterans affairs healthcare system. *Medicine (Baltimore)* **98**, e15644 (2019).

110. Kroll-Desrosiers, A.R., *et al.* Rates and Correlates of Depression Symptoms in a Sample of Pregnant Veterans Receiving Veterans Health Administration Care. *Womens Health Issues* **29**, 333-340 (2019).

111. Kramer, J.R., *et al.* Validation of HIV-infected cohort identification using automated clinical data in the Department of Veterans Affairs. *HIV Med* **20**, 567-570 (2019).

112. Rubinsky, A.D., *et al.* Utility of routine alcohol screening for monitoring changes in alcohol consumption. *Drug Alcohol Depend* **201**, 155-160 (2019).

113. Dixon, B.E., *et al.* Regional data exchange to improve care for veterans after non-VA hospitalization: a randomized controlled trial. *BMC Med Inform Decis Mak* **19**, 125 (2019).

114. Russo, E.M., *et al.* Incidence of Clostridioides difficile infections among young and middle-aged adults: Veterans Health Administration. *Infect Control Hosp Epidemiol* **40**, 997-1005 (2019).

115. Bick, A.G., *et al.* Association of APOL1 Risk Alleles With Cardiovascular Disease in Blacks in the Million Veteran Program. *Circulation* **140**, 1031-1040 (2019).

116. Le, B., *et al.* Laboratory evaluation of secondary causes of bone loss in Veterans with spinal cord injury and disorders. *Osteoporos Int* **30**, 2241-2248 (2019).

117. Fillmore, N., Do, N., Brophy, M. & Zimolzak, A. Interactive Machine Learning for Laboratory Data Integration. *Stud Health Technol Inform* **264**, 133-137 (2019).

118. Mahajan, S.M. & Ghani, R. Combining Structured and Unstructured Data for Predicting Risk of Readmission for Heart Failure Patients. *Stud Health Technol Inform* **264**, 238-242 (2019).
119. Edwards, S.T.*, et al.* Association of a Multisite Interprofessional Education Initiative With Quality of Primary Care. *JAMA Netw Open* **2**, e1915943 (2019).
120. Greater Plains Collaborative. (GPC Network, 2021).
121. Bresnick, J. Houston switches on EHRs, HIE for population health management. (EHRIntelligence, 2013).
122. New Electronic Health Record System promises to improve health and care coordination in Harris County. (The Harris Center for Mental Health and IDD, 2020).
123. Chung, C.*, et al.* Implementation of an integrated computerized prescriber order-entry system for chemotherapy in a multisite safety-net health system. *Am J Health Syst Pharm* **75**, 398-406 (2018).
124. HealthCore Announces First-Time Direct Access to its Research Data for Improved Insights and Decision Making. (BusinessWire, 2018).
125. Lewis, K.H.*, et al.* Safety and Effectiveness of Longer-Term Phentermine Use: Clinical Outcomes from an Electronic Health Record Cohort. *Obesity (Silver Spring)* **27**, 591-602 (2019).
126. Simon, G.E.*, et al.* Self-Reported Suicidal Ideation as a Predictor of Suicidal Behavior Among Outpatients With Diagnoses of Psychotic Disorders. *Psychiatr Serv* **70**, 176-183 (2019).
127. Waitzfelder, B.*, et al.* Treatment Initiation for New Episodes of Depression in Primary Care Settings. *J Gen Intern Med* **33**, 1283-1291 (2018).
128. Hospital Spotlight: Henry Ford Health System Uses Electronic Medical Records to Improve Care. (Michigan Health & Hospital Association, 2018).
129. Garg, A., Hundal, J. & Strunk, A. Overall and Subgroup Prevalence of Crohn Disease Among Patients With Hidradenitis Suppurativa: A Population-Based Analysis in the United States. *JAMA Dermatol* **154**, 814-818 (2018).
130. Reddy, S., Strunk, A. & Garg, A. Comparative Overall Comorbidity Burden Among Patients With Hidradenitis Suppurativa. *JAMA Dermatol* **155**, 797-802 (2019).
131. Abou Saleh, M., Mansoor, E., Anindo, M. & Isenberg, G. Prevalence of Small Intestine Carcinoid Tumors: A US Population-Based Study 2012-2017. *Dig Dis Sci* **64**, 1328-1334 (2019).
132. Sheffield, K.M.*, et al.* Development and validation of a claims-based approach to proxy ECOG performance status across ten tumor groups. *J Comp Eff Res* **7**, 193-208 (2018).
133. Huo, J., Yang, M. & Tina Shih, Y.C. Sensitivity of Claims-Based Algorithms to Ascertain Smoking Status More Than Doubled with Meaningful Use. *Value Health* **21**, 334-340 (2018).
134. Schleyer, T.*, et al.* Quantifying Unmet Need in Statin-Treated Hyperlipidemia Patients and the Potential Benefit of Further LDL-C Reduction Through an EHR-Based Retrospective Cohort Study. *J Manag Care Spec Pharm* **25**, 544-554 (2019).
135. Hood, S.R.*, et al.* Association Between Medication Adherence and the Outcomes of Heart Failure. *Pharmacotherapy* **38**, 539-545 (2018).
136. INSIGHT - Our Data. (INSIGHT Clinical Research Network, 2021).
137. Wysham, C.H.*, et al.* An investigation into the durability of glycemic control in patients with type II diabetes initiated on canagliflozin or sitagliptin: A real-world analysis of electronic medical records. *J Diabetes Complications* **33**, 140-147 (2019).
138. Wysham, C.H.*, et al.* Quality goal attainment and maintenance in patients with type II diabetes mellitus initiated on canagliflozin or a glucagon-like peptide-1 receptor agonist in an actual practice setting. *Curr Med Res Opin* **34**, 1125-1133 (2018).

139. Wysham, C.H., *et al.* HBA1C CONTROL AND COST-EFFECTIVENESS IN PATIENTS WITH TYPE 2 DIABETES MELLITUS INITIATED ON CANAGLIFLOZIN OR A GLUCAGON-LIKE PEPTIDE 1 RECEPTOR AGONIST IN A REAL-WORLD SETTING. *Endocr Pract* **24**, 273-287 (2018).

140. Lang, K., Nguyen, H., Huang, H., Bauer, E. & Levin, P. Real-world clinical responses in patients with type 2 diabetes mellitus adding exenatide BID (EBID) or mealtime insulin to basal insulin: a retrospective study using electronic medical record data. *Curr Med Res Opin* **34**, 1045-1051 (2018).

141. Shafrin, J., Tebeka, M.G., Price, K., Patel, C. & Michaud, K. The Economic Burden of ACPA-Positive Status Among Patients with Rheumatoid Arthritis. *J Manag Care Spec Pharm* **24**, 4-11 (2018).

142. Repka, M.X., Lum, F. & Burugapalli, B. Strabismus, Strabismus Surgery, and Reoperation Rate in the United States: Analysis from the IRIS Registry. *Ophthalmology* **125**, 1646-1653 (2018).

143. Chiang, M.F., Sommer, A., Rich, W.L., Lum, F. & Parke, D.W., 2nd. The 2016 American Academy of Ophthalmology IRIS(®) Registry (Intelligent Research in Sight) Database: Characteristics and Methods. *Ophthalmology* **125**, 1143-1148 (2018).

144. Swenor, B.K., Guo, X., Boland, M.V. & Goldstein, J.E. Leveraging Electronic Health Records to Identify and Characterize Patients with Low Vision. *Ophthalmic Epidemiol* **26**, 132-139 (2019).

145. Bahorik, A.L., *et al.* Alcohol, marijuana, and opioid use disorders: 5-Year patterns and characteristics of emergency department encounters. *Subst Abus* **39**, 59-68 (2018).

146. Landy, R., *et al.* Challenges in risk estimation using routinely collected clinical data: The example of estimating cervical cancer risks from electronic health-records. *Prev Med* **111**, 429-435 (2018).

147. Lott, J.P., *et al.* Population-Based Analysis of Histologically Confirmed Melanocytic Proliferations Using Natural Language Processing. *JAMA Dermatol* **154**, 24-29 (2018).

148. Chen, H.Y., *et al.* Association of LPA Variants With Aortic Stenosis: A Large-Scale Study Using Diagnostic and Procedural Codes From Electronic Health Records. *JAMA Cardiol* **3**, 18-23 (2018).

149. Simon, G.E., *et al.* Care Pathways Before First Diagnosis of a Psychotic Disorder in Adolescents and Young Adults. *Am J Psychiatry* **175**, 434-442 (2018).

150. Glanz, J.M., *et al.* Prediction Model for Two-Year Risk of Opioid Overdose Among Patients Prescribed Chronic Opioid Therapy. *J Gen Intern Med* **33**, 1646-1653 (2018).

151. Campbell, C.I., *et al.* Use of a prescription opioid registry to examine opioid misuse and overdose in an integrated health system. *Prev Med* **110**, 31-37 (2018).

152. Hu, W., Chan, C.W., Zubizarreta, J.R. & Escobar, G.J. Incorporating Longitudinal Comorbidity and Acute Physiology Data in Template Matching for Assessing Hospital Quality: An Exploratory Study in an Integrated Health Care Delivery System. *Med Care* **56**, 448-454 (2018).

153. Hoffmann, T.J., *et al.* A large electronic-health-record-based genome-wide study of serum lipids. *Nat Genet* **50**, 401-413 (2018).

154. Getahun, D., *et al.* Cross-sex Hormones and Acute Cardiovascular Events in Transgender Persons: A Cohort Study. *Ann Intern Med* **169**, 205-213 (2018).

155. Pocobelli, G., *et al.* One-Step Approach to Identifying Gestational Diabetes Mellitus: Association With Perinatal Outcomes. *Obstet Gynecol* **132**, 859-867 (2018).

156. Hechter, R.C., Bider-Canfield, Z. & Towner, W. Effect of an Electronic Alert on Targeted HIV Testing Among High-Risk Populations. *Perm J* **22**, 18-015 (2018).

157. Young-Wolff, K.C., *et al.* Evaluating the Impact of Eliminating Copayments for Tobacco Cessation Pharmacotherapy. *Med Care* **56**, 912-918 (2018).

158. Oni-Orisan, A.*, et al.* Characterization of Statin Low-Density Lipoprotein Cholesterol Dose-Response Using Electronic Health Records in a Large Population-Based Cohort. *Circ Genom Precis Med* **11**, e002043 (2018).

159. Richards, J.E.*, et al.* Understanding Why Patients May Not Report Suicidal Ideation at a Health Care Visit Prior to a Suicide Attempt: A Qualitative Study. *Psychiatr Serv* **70**, 40-45 (2019).

160. Haas, C.B., Phipps, A.I., Hajat, A., Chubak, J. & Wernli, K.J. Time to fecal immunochemical test completion for colorectal cancer screening. *Am J Manag Care* **25**, 174-180 (2019).

161. Green, C.A.*, et al.* Development of an algorithm to identify inpatient opioid-related overdoses and oversedation using electronic data. *Pharmacoepidemiol Drug Saf* **28**, 1138-1142 (2019).

162. Green, C.A.*, et al.* Identifying and classifying opioid-related overdoses: A validation study. *Pharmacoepidemiol Drug Saf* **28**, 1127-1137 (2019).

163. Danforth, K.N.*, et al.* Follow-up of Abnormal Estimated GFR Results Within a Large Integrated Health Care Delivery System: A Mixed-Methods Study. *Am J Kidney Dis* **74**, 589-600 (2019).

164. Hazlehurst, B.*, et al.* Using natural language processing of clinical text to enhance identification of opioid-related overdoses in electronic health records data. *Pharmacoepidemiol Drug Saf* **28**, 1143-1151 (2019).

165. Marcus, J.L.*, et al.* Use of electronic health record data and machine learning to identify candidates for HIV pre-exposure prophylaxis: a modelling study. *Lancet HIV* **6**, e688-e695 (2019).

166. Wartko, P.D.*, et al.* Antidepressant continuation in pregnancy and risk of gestational diabetes. *Pharmacoepidemiol Drug Saf* **28**, 1194-1203 (2019).

167. Danforth, K.N.*, et al.* Care Quality and Variability in the Use of Intravesical Therapy for Initial Treatment of Nonmuscle Invasive Bladder Cancer Within a Large, Diverse Integrated Delivery System. *Urology* **131**, 93-103 (2019).

168. Gordon, N.P., Lin, T.Y., Rau, J. & Lo, J.C. Aggregation of Asian-American subgroups masks meaningful differences in health and health risks among Asian ethnicities: an electronic health record based cohort study. *BMC Public Health* **19**, 1551 (2019).

169. Gold, R.*, et al.* Does increased implementation support improve community clinics' guideline-concordant care? Results of a mixed methods, pragmatic comparative effectiveness trial. *Implement Sci* **14**, 100 (2019).

170. Drees, J. Keck Medicine of USC implements Cerner PDMP access for EHR. (Beckers Hospital Review, 2019).

171. Connecting patient data through interoperability helps improve care across Los Angeles area health systems. (Cerner, 2019).

172. Wang, X.*, et al.* Prediction of the 1-Year Risk of Incident Lung Cancer: Prospective Study Using Electronic Health Records from the State of Maine. *J Med Internet Res* **21**, e13260 (2019).

173. Ye, C.*, et al.* Prediction of Incident Hypertension Within the Next Year: Prospective Study Using Statewide Electronic Health Records and Machine Learning. *J Med Internet Res* **20**, e22 (2018).

174. Cattails Software Suite: A service of Marshfield Clinic. (ed. Clinic, M.) (2007).

175. Geng, S.*, et al.* Parathyroid hormone independently predicts fracture, vascular events, and death in patients with stage 3 and 4 chronic kidney disease. *Osteoporos Int* **30**, 2019-2025 (2019).

176. MAYO CLINIC BIOBANK. (Mayo Clinic, 2021).

177. Moon, S.*, et al.* Automated extraction of sudden cardiac death risk factors in hypertrophic cardiomyopathy patients by natural language processing. *Int J Med Inform* **128**, 32-38 (2019).

178. Wu, Y.*, et al.* Discovery of Noncancer Drug Effects on Survival in Electronic Health Records of Patients With Cancer: A New Paradigm for Drug Repurposing. *JCO Clin Cancer Inform* **3**, 1-9 (2019).

179. Keller, C.A.*, et al.* Utilization and Cost Analysis of Lung Transplantation and Survival After 10 Years of Adapting the Lung Allocation Score. *Transplantation* **103**, 638-646 (2019).

180. Sandefur, B.J.*, et al.* Clinical Features and Outcomes Associated with Angioedema in the Emergency Department. *West J Emerg Med* **20**, 760-769 (2019).

181. Hertzog, R.*, et al.* Diagnostic Accuracy in Primary Care E-Visits: Evaluation of a Large Integrated Health Care Delivery System's Experience. *Mayo Clin Proc* **94**, 976-984 (2019).

182. Tominaga, N., Fernandez, S.J., Mete, M., Shara, N.M. & Verbalis, J.G. Hyponatremia and the risk of kidney stones: A matched case-control study in a large U.S. health system. *PLoS One* **13**, e0203942 (2018).

183. Center for Clinical Informatics Research and Education: A Note from the Director. (MetroHealth, 2021).

184. Kocher, K.E.*, et al.* Baseline Performance of Real-World Clinical Practice Within a Statewide Emergency Medicine Quality Network: The Michigan Emergency Department Improvement Collaborative (MEDIC). *Ann Emerg Med* **75**, 192-205 (2020).

185. Armstrong, A.W.*, et al.* Real-world health outcomes in adults with moderate-to-severe psoriasis in the United States: a population study using electronic health records to examine patient-perceived treatment effectiveness, medication use, and healthcare resource utilization. *BMC Dermatol* **18**, 4 (2018).

186. Mack, C.D.*, et al.* The Establishment and Refinement of the National Basketball Association Player Injury and Illness Database. *J Athl Train* **54**, 466-471 (2019).

187. Wentworth, L.J.*, et al.* Decreased bed rest post-percutaneous coronary intervention with a 7-French arterial sheath and its effects on vascular complications. *J Clin Nurs* **27**, e109-e115 (2018).

188. Dreyer, N.A.*, et al.* Lessons on Data Collection and Curation From the NFL Injury Surveillance Program. *Sports Health* **11**, 440-445 (2019).

189. Hornik, C.P.*, et al.* Creation of a Multicenter Pediatric Inpatient Data Repository Derived from Electronic Health Records. *Appl Clin Inform* **10**, 307-315 (2019).

190. Rosenfeld, E.H.*, et al.* The value of failure to rescue in determining hospital quality for pediatric trauma. *J Trauma Acute Care Surg* **87**, 794-799 (2019).

191. Gardner, B.J., Pedersen, J.G., Campbell, M.E. & McClay, J.C. Incorporating a location-based socioeconomic index into a de-identified i2b2 clinical data warehouse. *J Am Med Inform Assoc* **26**, 286-293 (2019).

192. Raspovic, K.M.*, et al.* Effectiveness of viable cryopreserved placental membranes for management of diabetic foot ulcers in a real world setting. *Wound Repair Regen* **26**, 213-220 (2018).

193. Family Health Centers at NYU Langone. (NYU Langone, 2021).

194. Hammond, R.*, et al.* Predicting childhood obesity using electronic health records and publicly available data. *PLoS One* **14**, e0215571 (2019).

195. Wang, Q.*, et al.* Development and validation of a prognostic model predicting symptomatic hemorrhagic transformation in acute ischemic stroke at scale in the OHDSI network. *PLoS One* **15**, e0226718 (2020).

196. Bailey, S.R., Heintzman, J., Jacob, R.L., Puro, J. & Marino, M. Disparities in Smoking Cessation Assistance in US Primary Care Clinics. *Am J Public Health* **108**, 1082-1090 (2018).

197. He, Z.*, et al.* Prevalence of Multiple Chronic Conditions Among Older Adults in Florida and the United States: Comparative Analysis of the OneFlorida Data Trust and National Inpatient Sample. *J Med Internet Res* **20**, e137 (2018).

198. Kaplan, R.M.*, et al.* Stroke Risk as a Function of Atrial Fibrillation Duration and CHA(2)DS(2)-VASc Score. *Circulation* **140**, 1639-1646 (2019).

199. Noe, M.H., Shin, D.B., Doshi, J.A., Margolis, D.J. & Gelfand, J.M. Prescribing Patterns Associated With Biologic Therapies for Psoriasis from a United States Medical Records Database. *J Drugs Dermatol* **18**, 745-750 (2019).

200. Mittal, S., Rogers, J., Sarkar, S., Koehler, J. & Passman, R.S. Real-World Incidence of Pacemaker and Defibrillator Implantation Following Diagnostic Monitoring With an Insertable Cardiac Monitor. *Am J Cardiol* **123**, 1967-1971 (2019).

201. Vonesh, E.*, et al.* Cardiovascular risk profile in individuals initiating treatment for overactive bladder - Challenges and learnings for comparative analysis using linked claims and electronic medical record databases. *PLoS One* **13**, e0205640 (2018).

202. Young, J.B.*, et al.* Development of predictive risk models for major adverse cardiovascular events among patients with type 2 diabetes mellitus using health insurance claims data. *Cardiovasc Diabetol* **17**, 118 (2018).

203. Kalilani, L., Halpern, R., Seare, J. & Dedeken, P. The challenges of assessing effectiveness of lacosamide using electronic medical record databases. *Epilepsy Behav* **85**, 195-199 (2018).

204. Ammann, E.M., Kalsekar, I., Yoo, A. & Johnston, S.S. Validation of body mass index (BMI)-related ICD-9-CM and ICD-10-CM administrative diagnosis codes recorded in US claims data. *Pharmacoepidemiol Drug Saf* **27**, 1092-1100 (2018).

205. Morrison, V.A.*, et al.* Treatment Patterns and Survival Outcomes in Patients With Follicular Lymphoma: A 2007 to 2015 Humedica Database Study. *Clin Lymphoma Myeloma Leuk* **19**, e172-e183 (2019).

206. PaTH Towards a Learning Health System (PaTH). (PCORI, 2019).

207. Burnett-Hartman, A.N.*, et al.* Treatment patterns and survival differ between early-onset and late-onset colorectal cancer patients: the patient outcomes to advance learning network. *Cancer Causes Control* **30**, 747-755 (2019).

208. About RPDR Tool & Website. (Mass General Brigham, 2021).

209. Moura, L.*, et al.* Epilepsy Among Elderly Medicare Beneficiaries: A Validated Approach to Identify Prevalent and Incident Epilepsy. *Med Care* **57**, 318-324 (2019).

210. Dhopeshwarkar, N.*, et al.* Drug-Induced Anaphylaxis Documented in Electronic Health Records. *J Allergy Clin Immunol Pract* **7**, 103-111 (2019).

211. Golas, S.B.*, et al.* A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC Med Inform Decis Mak* **18**, 44 (2018).

212. Orlova, Y., Rizzoli, P. & Loder, E. Association of Coprescription of Triptan Antimigraine Drugs and Selective Serotonin Reuptake Inhibitor or Selective Norepinephrine Reuptake Inhibitor Antidepressants With Serotonin Syndrome. *JAMA Neurol* **75**, 566-572 (2018).

213. Cornu, P.*, et al.* High-priority and low-priority drug-drug interactions in different international electronic health record systems: A comparative study. *Int J Med Inform* **111**, 165-171 (2018).

214. Navathe, A.S.*, et al.* Hospital Readmission and Social Risk Factors Identified from Physician Notes. *Health Serv Res* **53**, 1110-1136 (2018).

215. PEDSnet. (PEDSnet, 2021).

216. Practice Research Network (PRACNET). (PRACnet, 2021).

217.   Research Action for Health Network. (REACHnet, 2021).

218.   Selby, K.*, et al.* Disparities in Hypertension Control Across and Within Three Health Systems Participating in a Data-Sharing Collaborative. *J Am Board Fam Med* **31**, 897-904 (2018).

219.   Machine-learning Technology Used to Provide Personalized Care. (Michigan Medicine Kellogg Eye Center, 2018).

220.   Stakeholders, Technology, and Research CRN. (STAR CRN, 2021).

221.   Data Analytics Center: Penn Data Store. (Penn Medicine, 2021).

222.   Kim, R.H.*, et al.* Variations in Influenza Vaccination by Clinic Appointment Time and an Active Choice Intervention in the Electronic Health Record to Increase Influenza Vaccination. *JAMA Netw Open* **1**, e181770 (2018).

223.   Lei, V.J.*, et al.* Risk Stratification for Postoperative Acute Kidney Injury in Major Noncardiac Surgery Using Preoperative and Intraoperative Data. *JAMA Netw Open* **2**, e1916921 (2019).

224.   Huang, J., Chen, Y., Landis, J.R. & Mahoney, K.B. Difference Between Users and Nonusers of a Patient Portal in Health Behaviors and Outcomes: Retrospective Cohort Study. *J Med Internet Res* **21**, e13146 (2019).

225.   McAdam-Marx, C.*, et al.* Impact of a guideline-based best practice alert on pneumococcal vaccination rates in adults in a primary care setting. *BMC Health Serv Res* **19**, 474 (2019).

226.   UW Medicine Clinical Transformation: Approve Project, Budget, and Internal Lending Program (ILP) Funding. (University of Washington, 2018).

227.   Hicks, K.*, et al.* Predictors of Death in the Hospital for Patients with Chronic Serious Illness. *J Palliat Med* **21**, 307-314 (2018).

228.   Electronic Health Record Helps Transform Patient Care. (University of Wisconsin Health, 2010).

229.   Liu, N.*, et al.* Weight Loss for Patients With Obesity: An Analysis of Long-Term Electronic Health Record Data. *Med Care* **58**, 265-272 (2020).

230.   Bejan, C.A., Lee, D.J., Xu, Y. & Hsi, R.S. Performance of a Natural Language Processing Method to Extract Stone Composition From the Electronic Health Record. *Urology* **132**, 56-62 (2019).

231.   Wells, Q.S.*, et al.* Accelerating Biomarker Discovery Through Electronic Health Records, Automated Biobanking, and Proteomics. *J Am Coll Cardiol* **73**, 2195-2205 (2019).

232.   Unlu, G.*, et al.* GRIK5 Genetically Regulated Expression Associated with Eye and Vascular Phenomes: Discovery through Iteration among Biobanks, Electronic Health Records, and Zebrafish. *Am J Hum Genet* **104**, 503-519 (2019).

233.   Oshikoya, K.A.*, et al.* CYP2D6 genotype and adverse events to risperidone in children and adolescents. *Pediatr Res* **85**, 602-606 (2019).

234.   Bejan, C.A.*, et al.* Mining 100 million notes to find homelessness and adverse childhood experiences: 2 case studies of rare and severe social determinants of health in electronic health records. *J Am Med Inform Assoc* **25**, 61-71 (2018).

235.   Bray, M.J., Torstenson, E.S., Jones, S.H., Edwards, T.L. & Velez Edwards, D.R. Evaluating risk factors for differences in fibroid size and number using a large electronic health record population. *Maturitas* **114**, 9-13 (2018).

236.   Crawford, D.C., Restrepo, N.A., Diggins, K.E., Farber-Eger, E. & Wells, Q.S. Frequency and phenotype consequence of APOC3 rare variants in patients with very low triglyceride levels. *BMC Med Genomics* **11**, 66 (2018).

237.   Restrepo, N.A., Laper, S.M., Farber-Eger, E. & Crawford, D.C. Local genetic ancestry in CDKN2B-AS1 is associated with primary open-angle glaucoma in an African American cohort extracted from de-identified electronic health records. *BMC Med Genomics* **11**, 70 (2018).

238. Barnado, A.*, et al.* Phenome-Wide Association Studies Uncover a Novel Association of Increased Atrial Fibrillation in Male Patients With Systemic Lupus Erythematosus. *Arthritis Care Res (Hoboken)* **70**, 1630-1636 (2018).
239. Horth, R.Z.*, et al.* Use of electronic health records from a statewide health information exchange to support public health surveillance of diabetes and hypertension. *BMC Public Health* **19**, 1106 (2019).
240. Kallenbach, L.*, et al.* Predictors and Clinical Outcomes of Treatment Intensification in Patients With Type 2 Diabetes Uncontrolled on Basal Insulin in a Real-World Setting. *Endocr Pract* **24**, 805-814 (2018).