

## **Title:** Obtaining the Data from Electronic Health Records.

### **Theory:**

There is growing interest in using data captured in electronic health records (EHRs) for patient registries. Both EHRs and patient registries capture and use patient-level clinical information, but conceptually, they are designed for different purposes. A patient registry is defined as “an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure and that serves one or more predetermined scientific, clinical, or policy purposes.”

An EHR is an electronic system used and maintained by healthcare systems to collect and store patients’ medical information. EHRs are used across clinical care and healthcare administration to capture a variety of medical information from individual patients over time, as well as to manage clinical workflows. EHRs contain different types of patient-level variables, such as demographics, diagnoses, problem lists, medications, vital signs, and laboratory data. According to the National Academies of Medicine, an EHR has multiple core functionalities, including the capture of health information, orders and results management, clinical decision support, health information exchange, electronic communication, patient support, administrative processes, and population health reporting.

In summary, registries are patient-centered, purpose-driven, and designed to derive information on defined exposures and health outcome. In contrast, EHRs are visit-centered and transactional. Despite these differences, EHRs capture a wealth of data that is relevant to patient registries. EHRs also may assist in certain functions that a patient registry requires (e.g., data collection, data cleaning, data storage), and a registry may augment the value of the information collected in an EHR (e.g., comparative safety, effectiveness and value, population management, quality reporting).

Registries that capture EHR data from multiple health systems typically interface with EHRs to receive data on an interval basis (i.e., EHR-linked or EHR-reported registries), although automating such efforts and creating a bidirectional exchange of information are still challenging. The Meaningful Use program has propelled the development of both EHR-linked and EHR-integrated registries.

Driven in large part by Meaningful Use, EHR vendors and clinical providers are incentivized to develop processes that would facilitate the design and launch of EHR-based registries. Yet, despite these incentives, the practice of using EHR-based registries is still relatively immature and, like all evolving research programs, faces many challenges.

EHRs provide various types of data that can be linked, integrated, or merged directly into a registry. The Meaningful Use program has led to the collection of a Common Clinical Data Set (CCDS) across most providers. Also, EHRs capture a considerable amount of unstructured data (e.g., clinical notes) that can be further processed to extract specific data of importance to a registry (e.g., specific information extracted from radiology reports to determine eligibility). Data types commonly extracted from EHRs and imported into registries are patient identifiers, demographics, diagnoses, medications, procedures, laboratory results, vital signs, and utilization events.

Registries that incorporate EHR data may use a variety of IT system architectures. Registry architects must consider the number of participating sites (single-site or multi-site), variety of underlying EHRs (one enterprise-level EHR, multiple EHR installations of the same vendor, multiple EHRs from different vendors), existence and connectivity to Health Information Exchanges (HIEs) (centralized, federated or distributed), and other factors that affect interoperability. Following are examples of three “hypothetical” EHR-based registry types, each with a different combination of stakeholders and IT infrastructures. Registries designed to support clinical care are often based on single enterprise-level EHRs, while registries designed for research are often hosted external to EHRs but may receive EHR extracts from multiple sources. Public health registries, similar to registries designed for research, are often hosted by health departments outside of a single EHR environment but receive EHR reports on a regular basis.

## IT infrastructure and other features of sample registry types using EHR data

<b>TYPE/SPECS</b>	<b>REGISTRY TO SUPPORT CLINICAL CARE</b>	<b>REGISTRY DESIGNED FOR RESEARCH</b>	<b>PUBLIC HEALTH REGISTRY</b>
Scope	<p>Depending on the provider network's size:</p> <ul style="list-style-type: none"> <li>Local</li> <li>City</li> <li>State</li> <li>Regional</li> <li>National (e.g., VHA)</li> </ul>	<p>Depending on the research aim/goal:</p> <ul style="list-style-type: none"> <li>State</li> <li>Regional</li> <li>National</li> <li>International</li> </ul>	<p>Depending on the public health authority:</p> <ul style="list-style-type: none"> <li>City</li> <li>County</li> <li>State</li> <li>National/Federal</li> </ul>
Stakeholders	<ul style="list-style-type: none"> <li>providers (usually within the network)</li> <li>biopharmaceutical and medical device companies [optional]</li> <li>employers [optional]</li> <li>payers [optional]</li> </ul>	<ul style="list-style-type: none"> <li>research institutes</li> <li>government (local, state or federal)</li> <li>non-profit organization</li> <li>disease associations</li> <li>biopharmaceutical and medical device companies</li> <li>employers (e.g., professional sports)</li> </ul>	government (local, city, county, state or federal)
Sources of Data	<ul style="list-style-type: none"> <li>mainly EHRs</li> <li>sometimes EHR-based patient portals</li> <li>sometimes merged insurance claims</li> <li>infrequently EHR-embedded surveys</li> </ul>	<ul style="list-style-type: none"> <li>surveys</li> <li>custom EHR extracts</li> <li>other data sources (e.g., biobanks, administrative health insurance claims, eCRFs, other registries)</li> </ul>	<ul style="list-style-type: none"> <li>surveys</li> <li>EHRs</li> <li>other data sources (e.g., HIEs, environmental, health department, surveillance systems)</li> </ul>
Number of EHRs	Usually one enterprise-level EHR	Multiple/various EHRs	Multiple/various EHRs
EHR Interoperability Requirement	low	high	medium

<b>TYPE/SPECS</b>	<b>REGISTRY TO SUPPORT CLINICAL CARE</b>	<b>REGISTRY DESIGNED FOR RESEARCH</b>	<b>PUBLIC HEALTH REGISTRY</b>
EHR Integration Level	usually automated	sometimes manual frequently semi-automated rarely fully automated	automated for some (e.g., immunization records)  manual for others (e.g., non-MU registries)
EHR Integration Tools	EHR built-in tools  sometimes custom-built APIs / extracts	custom-built APIs / extracts	EHR build-in tools for select reporting (MU program)  Custom-built tools/APIs for non-MU registries
System Dependency	EHR-based	May not be EHR-based	not EHR-based
Common Architecture	Often Centralized	Centralized  Distributed  Federated	Centralized  Distributed  Federated
Dominant Hosting DB	EHR-embedded (e.g., EHR registry data warehouse)	Could be distributed and likely to include non-EHR DB (i.e., research-specific data collection)	non-EHR DB (i.e., public health database)
Alternative Names	clinical quality registry  improvement/measure registry  chronic disease management registry  high risk [population] registry	product or disease registry  clinical research registry  research network registry	outbreak registries  vaccination registries  disease surveillance (e.g., cancer)
Typical Functions	clinical workflow management  disease/cohort management (e.g., care coordination)	evidence for effectiveness, comparative effectiveness, safety and/or value for clinicians, patients and payers	public health services  outbreak surveillance  syndromic surveillance  epidemiological research

<b>TYPE/SPECS</b>	<b>REGISTRY TO SUPPORT CLINICAL CARE</b>	<b>REGISTRY DESIGNED FOR RESEARCH</b>	<b>PUBLIC HEALTH REGISTRY</b>
	population health management (e.g., case management)	natural history of disease studies	biopharmaceutical research, e.g., vaccine effectiveness and safety
Timeliness	usually real-time sometimes periodic (daily extracts)	mostly periodic (daily, weekly or monthly extracts) sometimes real-time	sometimes real-time sometimes periodic (daily, weekly or monthly extracts)
Scalability	limited to individual EHR vendors	depends on registry architectures adopting EHR interoperability standards	depends on adopting interoperability standard and future stages of MU for public health reporting

API = Application Programming Interface; DB = Database; eCRF = electronic Case Report Forms; HER = Electronic Health Record; HIE = Health Information Exchange; MU = Meaningful Use; VHA = Veteran Health Affairs.

In a fully interoperable ecosystem, registry-specific functionality could be presented in a software-as-a-service or middleware model, interacting with the EHR as the presentation layer on one end and the registry database on the other. In this ideal model, the EHR is a gateway to multiple registries and clinical research activities through an open architecture that leverages best-in-class functionality and connectivity. Full interoperability would enable registries to interact across multiple EHRs, and EHRs to interact with multiple registries.

- **EHR-Integrated Registries To Support Clinical Care**

Healthcare providers often develop and manage EHR-based registries that are used to support clinical care and meet operational goals. To develop clinical registries, providers typically use EHR-based tools that are developed by EHR vendors. These EHR-based registries can facilitate clinical workflow, monitor quality metrics, enable disease/cohort management, and offer population health management features. In particular, the Triple Aim of care, health and cost has provided a framework to achieve value-based care while reducing cost. This framework promotes ‘population health’ while enhancing the individual’s experience of care and lowering cost. Effective population health management is essential to ensuring that resources are directed towards improving health outcomes of patients at the highest risk for developing undesired outcomes.

A major challenge with EHR-integrated clinical registries is the lack of out-of-network data in a health network’s EHR. In other words, data generated during patient encounters with out-of-network providers, who may not be using the same EHR, will be missed in the registry resulting in incomplete and sometimes outdated data. Individual health networks often complement their EHR data with insurance claims to generate a more complete picture of a patient’s health status; however, use of insurance claims is not always practical given that a large patient population of a health delivery network may use dozens, if not hundreds, of different insurers. Many challenges of EHR-based

population health registries are derived from the overarching challenges within the broader domain of population health informatics.

Clinical registries usually use a centralized architecture and often have an EHR data warehouse as their backbone along with multiple data marts containing various registry data. The centralized architecture accumulates and manages data in a single and centralized repository. The advantages of a centralized model are: simplicity and efficiency; greater data consistency; and easier patient linkage if the same patient identifiers are used across the healthcare network. Potential disadvantages of a centralized model include: data capture that is limited to users of a single EHR vendor across the healthcare network (e.g., trouble with integrating a different EHR vendor if a new facility joins the network); and difficult data exchange with registries developed by other networks due to a lack of interoperability.

Healthcare networks often develop clinical registries based on their underlying enterprise-wide EHR architecture as shown in the below fig.1. Data collected at different facilities of a healthcare delivery network (e.g., hospitals and outpatient clinics) are aggregated in a common data repository such as an EHR's data warehouse. Facilities not using the same EHR platform face extra work to harmonize and standardize their data before feeding it into the data warehouse. Data warehouses can be used to develop multiple data marts feeding into various registries for different purposes such as quality measures, disease management, population health management, and public health reporting. Internal clinical registries are sometimes linked to external registries for reporting purposes (e.g., PQRS reporting), although interoperability challenges may limit such exchanges.

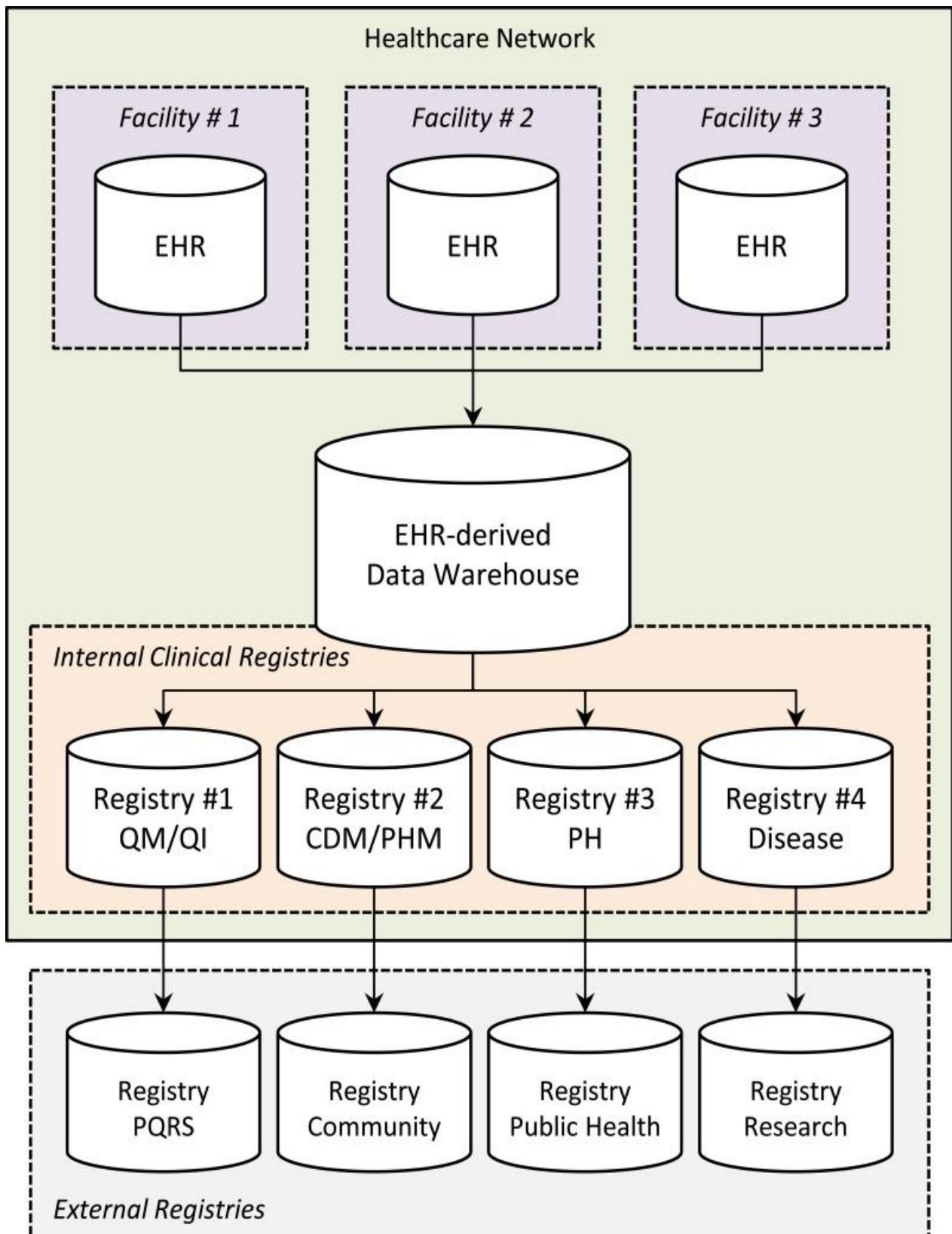


Fig.1: Common architecture of EHR-integrated registries to support clinical care.

## EHR-Linked Registries Designed for Research

Registries designed for research purposes (referred to here as ‘research registries’) may use EHR data on a variety of levels. At the low end, research registries may use EHR data to identify and enroll eligible patients into studies that use supplementary registry-specific data collection. In this scenario, EHR data are used to identify eligible patients (based on the registry’s inclusion and exclusion criteria), and minimal EHR data (e.g., family history of breast cancer) are imported into the registry. The remaining registry-specific data are captured through another means, usually a dedicated data repository that allows for entry of eCRFs and web-based survey forms. On the other end of the spectrum, some research registries have been built entirely using EHR data (e.g. California Cancer registry). Many other research registries use a combination of self-reported and EHR data (e.g. Autism treatment network). Registries in which EHR-based extracts are merged with registry data on a periodic basis are referred to here as EHR-linked registries.

EHR-linked research registries collect EHR data using a variety of mechanisms, ranging from automated EHR-embedded push protocols to manual ad-hoc EHR-database pulls. Triggers for EHR data extraction include standardized protocols that follow the inclusion and exclusion criteria of the research registry (i.e., phenotyping queries; retrieve protocols). After receiving the EHR data, research registries use a multi-phase process to import incoming EHR data as shown in figure2). Extract, transform, and load functions may include data curation activities such as data preparation, data standardization, secure data transfer, data mapping, data redaction, data integration/merging, and data reconciliation.

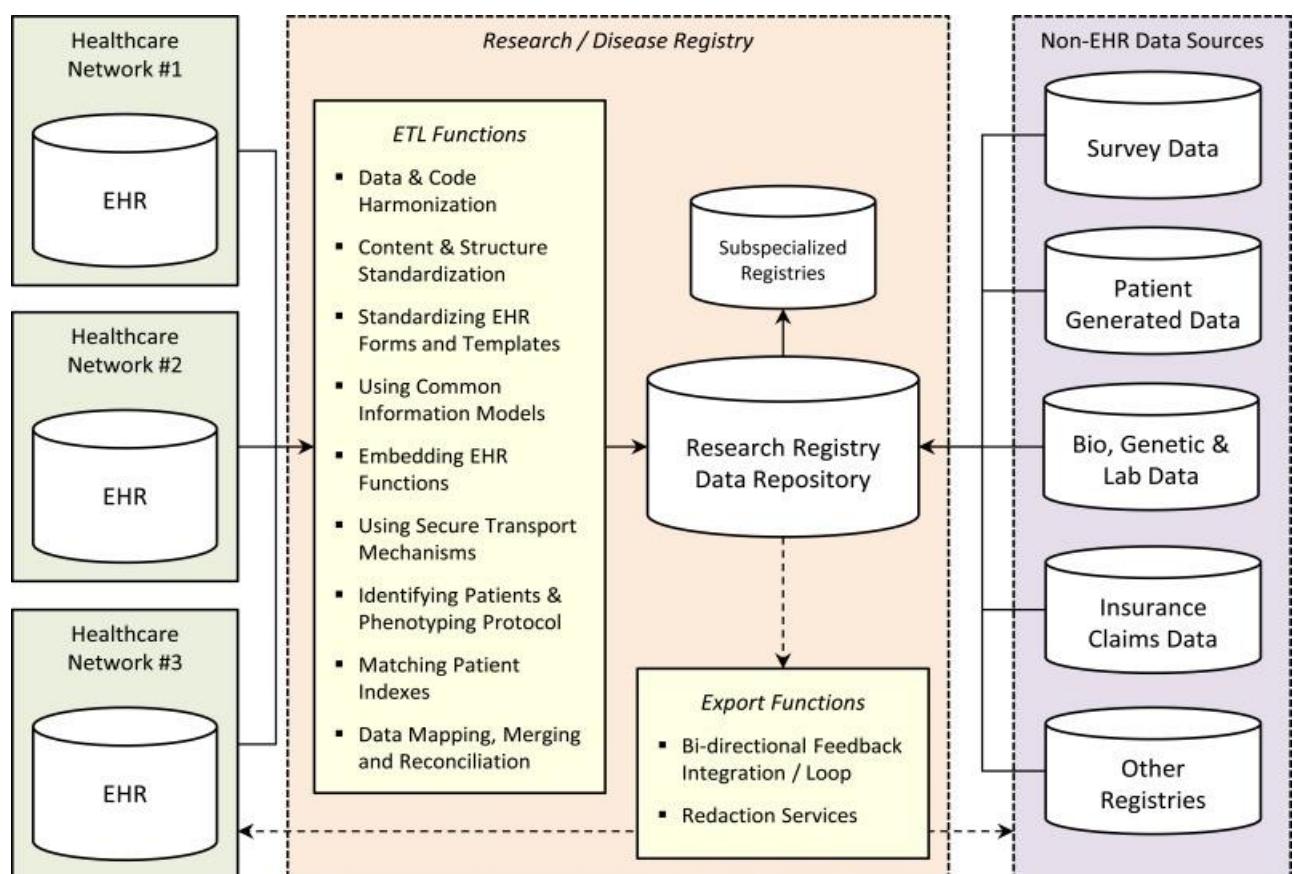


Fig.2: Common architecture of EHR-linked research registries.

Importing and merging data from EHRs into research registries is challenging. Automating the data imports requires high degrees of interoperability, data curation, and post-hoc harmonization as well as attention to data quality.

## EHR-Linked Public Health Registries

Public health agencies have long used registries for surveillance and tracking purposes. For example, local and state public health departments usually maintain immunization registries that receive information from clinicians and other entities such as schools and pharmacies. Other common public health registries include syndromic surveillance and specialized registries such as birth defects, chronic diseases, and traumatic injury registries. In recent years, coincident with the rising EHR adoption among providers, public health entities began to link various registries with EHRs. A significant driver of increased EHR integration has been the Meaningful Use program, which incentivized clinicians to share EHR immunization and syndromic surveillance data with public health agencies.

EHR-linked public health registries follow a similar architecture to that of EHR-linked research registries as shown in figure 4.2; however, the methods used to collect data from EHRs may vary as not all public health registries require patient-level data (e.g., counts are sufficient for some purposes). Methods used include but are not limited to: (1) semi-automated forms/templates to collect public health specific information about patients that fit a certain criteria (e.g., S&I Framework SDC) (2) data exchange protocols for receiving case reports from certified EHRs (e.g., MU public health reporting objectives) (3) tools to mine EHR and HIE data for signs and symptoms relevant to public health emergencies and outbreaks (e.g., ESSENCE Syndromic Surveillance System) and, (4) distributed data network queries to collect aggregated data from multiple providers when the identity of patients is not relevant (e.g., PopMedNet).

Some public health agencies have been able to directly integrate their registries with the EHRs of clinicians who provide care in their jurisdiction. The prime example of such a fully-integrated EHR-linked public health registry is the New York City (NYC) Population Health Registry. This registry collects information from NYC's eligible healthcare professionals across several domains (e.g., Influenza-like-Illnesses). The NYC's Population Health Registry has been successful as most eligible professionals in NYC use the same EHR system, one which is capable of reporting data in real-time to local public health agencies. The Population Health Registry is part of NYC MacroScope Hub, a surveillance system for tracking conditions managed by primary care practices (e.g., obesity, diabetes, hypertension, and smoking).

EHR-based registries fulfill different purposes and use different IT system architectures, but many technical issues and operational challenges are common across the range of registries. Several common challenges, such as identification of eligible patients; data quality; unstructured data; interoperability; data sharing and patient privacy; data access and patient privacy; and human resources. Data sharing is an additional concern in the context of EHR-based registries. Decisions must be made about whether a single institutional review board (IRB) will suffice or whether all sites will require local IRB approval. Governance is also challenging as the rules around sharing of data (identifiable or de-identified) vary depending on the organizations involved and the purpose of the research.

**Conclusion:** EHRs and patient registries play crucial roles in the healthcare ecosystem. EHRs provide a comprehensive and ongoing record of individual patient health data, while registries serve specific research and quality improvement purposes. Leveraging the strengths of both systems can lead to enhanced healthcare outcomes, better population health management, and valuable insights for research and healthcare decision-making in the field of data science. EHRs are used to support clinical care and meet operational goals. To develop clinical registries, providers typically use EHR-based tools that are developed by EHR vendors. These EHR-based registries can facilitate clinical workflow, monitor quality metrics, enable disease/cohort management, and offer population health management features.