Electronic Health Record–Based Registries: Clinical Research Using Registries in Colon and Rectal Surgery

Jacob Carlson, MD¹ Jonathan Laryea, MD, FACS, FASCRS, FWACS¹

¹Department of Surgery, University of Arkansas for Medical Sciences, Little Rock, Arkansas

Clin Colon Rectal Surg 2019;32:82–90.

Address for correspondence Jonathan Laryea, MD, FACS, FASCRS, FWACS, Department of Surgery, University of Arkansas for Medical Sciences, 4301 W Markham Street, Slot 520, Little Rock, AR 72205 (e-mail: jalaryea@uams.edu).

Abstract

Keywords

- Electronic Health Records
- registries
- interoperability
- clinical research
- outcomes
- colorectal surgery

Electronic health records (EHRs) or electronic medical records (EMRs) contain a vast amount of clinical data that can be useful for multiple purposes including research. Disease registries are collections of data in predefined formats for population management, research, and other purposes. There are differences between EHRs and registries in the data structure, data standards, and protocols. Proprietary EHR systems use different coding systems and data standards, which are usually kept secret. For EHR data to flow seamlessly into registries, there is the need for interoperability between EHR systems and between EHRs and registries. The levels of interoperability required include functional, structural, and semantic interoperability. EHR data can be manually mapped to registry data, but that is a tedious, resource-intensive endeavor. The development of data standards that can be used as building blocks for both EHRs and registries will help overcome the problem of interoperability.

Electronic health records (EHRs) or electronic medical records (EMRs) have become widespread throughout U.S. healthcare facilities and private practices and represent a dramatic departure from the dossier-style medical record invented by Dr. Henry Plummer at Mayo Clinic in 1907. EMRs and EHRs are sometimes used interchangeably, but there are differences. EMR refers to an electronic version of a patient's paper chart and it is used only by the practice or hospital for care of the patient. EHR refers to a much more encompassing entity which includes the EMR and other information. EHR allows for exchange of data between healthcare facilities and allows the patient's information to follow them wherever they go. It allows for patient interaction with care providers through the system, including ability to amend their health record. The Office of National Coordinator for Health Information Technology mandated the meaningful use of EHR exclusively. Most health informatics groups also use EHR exclusively. As of 2015, 84% of all nonfederal acute care hospitals had adopted an EHR with at least basic functionality, such as documenting clinician notes.¹ EHRs are an enormous source of data. The data generated in EHRs are primarily for patient care. However, it can be used for other purposes such as administrative, billing, quality improve-

ment/assurance, legal purposes, outcomes research, and policy functions. Administrative data such as insurance claims data have been used for outcomes and comparative effectiveness research for decades. However, the information gathered relates only to claims data, which is only a fraction of the data generated in EHRs. Using EHR data for research can be challenging, since EHRs contain both structured and unstructured data, and it is very difficult to data mine unstructured data (i.e., free-text of operative reports) for meaningful research. Institutions that have databases and registries have to employ individuals to manually re-enter the data from EHRs into the databases or registries by data managers or research assistants. This is very laborious and resource intensive. Some EHRs have functionality for building registries that can be used for population management and research. However, these may not be comprehensive enough as they may not contain all the data related to patient care. Some of the data may be in other legacy systems, which would not be available in these registries.

A registry is an organized system that uses observational study methods to collect uniform data to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure and serves one or more

predetermined scientific, clinical, or policy purposes.² Registries can serve several purposes including evaluating the natural history of a disease, evaluating patient outcomes, assessing cost-effectiveness, assessing comparative effectiveness of treatments, assessing safety and quality, evaluating rare diseases, population management, public health, and help with resource allocation.

Differences between Electronic Health Records and Registries

Both EHRs and registries use clinical information, but registries are mainly focused on population management and are designed to obtain information on predefined health outcomes data. EHRs, however, use health-related information pertinent to an individual patient. Although all the information required for registries may be present within the EHR, there is usually a lack of interoperability between the two. EHRs include data on all patients, whereas patients in a registry are selected based on predefined criteria (e.g. ICD-10 code 555.x, Crohn's disease) or protocols. Different providers enter information differently into EHRs based on how they customize the EHR, but data entry into registries is standardized, and hence there is limited, if any, variability. As a result, data in registries are structured and based on controlled vocabularies. EHRs contain both structured and unstructured data. EHR data entry is usually triggered by patient encounters, but registry data entry is based on protocol. There is limited data quality assurance in EHRs, while registries have research-specific validation rules and audits. *The ideal situation is for EHR data to flow seamlessly into registries.*

Registries

Registries can be made for any disease or condition and are useful for research, population management, as well as clinical and policy functions. A registry is a data information source that collects information about patients diagnosed with a particular medical condition. The information collected includes demographic information (age, gender, race/ethnicity, birthplace, residence), medical history (physical exam findings, screening information), diagnostic findings, therapies used in treatment, and follow-up details. Cancer registries have been used for decades, and various databases have been used to collect data on specific cancers. In addition to the information collected for other registries, a cancer registry also records information regarding previous cancer diagnoses and the results of procedures such as imaging and biopsies that were used to diagnose cancer. They also collect information regarding the specific type of cancer, such as the primary site of cancer, cell type, and extent of disease. Treatment modalities such as surgery, radiation therapy, chemotherapy, hormone therapy, and immunotherapy are also documented. Of particular importance for research, registries also contain information regarding the patient's surveillance, survival, and recurrence after diagnosis and treatment.

Cancer Registries

Cancer registries collect a great deal of information about the patient and thus are an invaluable tool that can be utilized for research. The information collected in a registry is used to help identify groups of people who are more prone to develop cancer, identify risk factors for the development of cancer, and evaluate patient outcomes and their quality of life. It is also used to calculate incidence and survival rates, analyze referral patterns, determine how resources are allocated, and evaluate the efficacy of different treatment modalities.

Several different types of cancer registries exist. Hospital-based registries collect information on patients diagnosed with or treated for cancer at particular clinic or hospital. Facilities then provide their information to the state cancer registry. Most state central cancer registries then report their information to the National Program of Cancer Registries (NPCR) and the Surveillance, Epidemiology, and End Results (SEER) program, which then use this information to publish an annual report on the incidence and mortality of cancer.³ Population-based registries collect information on cancers diagnosed or treated within a particular geographic distribution. Special purpose registries collect information on particular types of cancer. The development of these cancer registries is the product of a keen interest to better understand cancer that was met with an abundance of political support.

Yale-New Haven Hospital in New Haven, CT, established the first hospital-based cancer registry in 1926.⁴ The first state-based central cancer registry was then established in Connecticut in 1935. In efforts to improve the quality of care provided to cancer patients, the American College of Surgeons, through its Commission on Cancer, developed a policy that required all approved cancer programs to establish a hospital-based cancer registry.^{4,5} The National Cancer Act, signed by President Richard Nixon in December 1971, allocated funds to the National Cancer Institute (NCI) with the intent to promote cancer research.⁶ The first national cancer registry was established in 1973 with the creation of the SEER Program of the NCI. The National Program of Cancer Registries was established by Congress throughout the enactment of the Cancer Registries Amendment Act Public Law 102-515, which authorized the Centers for Disease Control and Prevention (CDC) to provide funds to states to "improve existing cancer registries; plan and implement registries where they do not exist; develop model legislation and regulations for states to enhance the viability of registry operations; set standards for data completeness, timeliness, and quality; provide training for registry personnel; and help establish a computerized reporting and data processing system."⁷

Electronic Health Records

The Institute of Medicine defines four major functions of an EHR: health information and data, computerized physician order entry (CPOE), results management, and decision support.⁸ Not all EHRs have the full compendium of functionality. Some institutions use different vendors and systems to achieve the different functions of the EHR. Beginning in the 1960s, over

70 individual hospitals had established projects focused on the development of an EMR.⁹ In 1968, Dr. Lawrence L. Weed wrote an article entitled “Medical Records That Guide and Teach,” which championed the documentation of a problem-oriented medical record. In contrast to Dr. Plummer’s method, this style of documentation would detail the patient’s medical issues in a more organized fashion and could be used to optimize patient care.¹⁰ In 1968, Massachusetts General Hospital in collaboration with Harvard University developed the Computer Stored Ambulatory Record (COSTAR), which was noteworthy for its compartmentalization of various tasks that could be performed within the system.⁹

In 1971, the Lockheed Corporation created an EMR that was notable for its ability to allow multiple users to simultaneously access the system and capacity for CPOE.¹¹ Dr. Weed’s work and coordination with Jan R. Schultz led to the development of the Problem-Oriented Medical Information System (PROMIS) in 1976. PROMIS was an interactive, text-based system with touch screen capabilities, designed with the purpose of accessing and maintaining a large volume of health care records.¹² In the 1970s, the federal government also began using EHRs when the Department of Veteran Affairs adopted the Decentralized Hospital Computer Program (DHCP), which is now known as Vista.¹³

Support for National Implementation of the Electronic Health Record

Numerous initiatives have been undertaken since that time to facilitate the development and adoption of EHRs. In 1997, the Institute of Medicine released a revised edition of *The Computer-Based Patient Record*, which promotes the implementation of EHRs as a cost-effective way to provide higher quality care to patients.⁹ Despite the perceived benefits of EHRs, many institutions remain reluctant to invest in this technology. Commonly cited barriers to the adoption of EHRs included initial capital cost, lack of funds and staff for maintenance of the system, not being able to find system relevant to their respective practice, being unsure about the return on their investment, resistance from physicians to change from the systems currently in place,¹⁴ and lack of interoperability between institutional departments and computer operating systems.¹⁵

In 2009, President Barack Obama signed into law the Health Information Technology for Economic and Clinical Health (HITECH) Act, as part of the American Recovery and Reinvestment Act (ARRA), which provided \$19.2 billion toward the advancement of health information technologies, \$17.2 billion of which was to be used “for financial incentives to physicians and hospitals through Medicare and Medicaid to promote the use of EHRs and other health information technology” for meaningful use.^{16,17} “Meaningful use” is defined in three main categories as (1) the use of certified EHR in a meaningful manner such as e-prescribing (CPOE); (2) use of certified EHR technology for electronic exchange of health information to improve quality of healthcare, such as promoting care coordination; and (3) use of certified EHR technology to submit clinical quality and other measures.¹⁸ The “meaningful use” incentives under the ARRA have led to

a drastic increase in the implementation of EHRs.^{19,20} The Department of Health and Human Services reported that the percentage of primary care physicians using a basic EMR system increased from 20% in 2009 to 39% in 2011, and that the proportion of physicians using at least a basic EMR system has increased from 22 to 34% over that same period.²¹ In 2003, only 17% of office-based physicians used EHR. This rose to 74% in 2012.²²

Electronic Health Records for Patient Care

While there are multiple barriers to the adoption of EHRs, proponents suggest that they also offer a host of potential benefits in terms of providing higher quality patient care. These benefits impact not only the patient but also the treating physician, policymakers, and those involved in research and pharmaceuticals.²³ Potential benefits of EHRs include documenting records in a more organized fashion, reduced costs of health care overall, and producing fewer medical errors in documentation.^{23–25} Use of health information technologies and EHRs can improve adherence to evidence-based guidelines for the management of chronic disease, cancers, and preventive care.¹⁹ They also have the potential to reduce gaps in the quality of care provided to underserved populations.²⁶

From the perspective of a health care provider, EHRs can improve data tracking, provide data for the analysis of practice patterns and assessment of patient outcomes, and provide information that can be used for quality-improvement interventions.²⁴ EHRs can also integrate “triggers” or suggestions within order sets to promote utilization of predefined clinical care pathways (e.g., best practice alerts) and alert providers of potential deviations from the recommended guidelines, thereby improving adherence to evidence-based care through clinical decision support.^{26,27} EHRs can facilitate identification of a particular population and issue reminders when a screening test needs to be performed. This can be accomplished by notifying the physician and/or the patient of the need for a test to be performed. The result of these screening tests can then be documented in a system that is widely accessible to the patient’s other providers, which can improve the coordination of care received by the patient.^{26,28} EHRs also offer a unique, rich source of information that can be employed for clinical research.

Prior to the advent of EHRs, administrative/payer claims data were used to evaluate treatment and patient outcomes. Claims data contain information only about the patient’s demographics, diagnoses, and procedures documented for billing purposes.²⁷ Compared with claims data, EHRs offer a more robust collection of data that can be used in research. EHRs contain more detailed clinical information such as the patient’s personal and family history, physical exam findings, laboratory results, and outcomes and can be queried using diagnosis and procedure codes to extract data on the patient population.^{23,29} However, one of the main challenges to this approach is determining which query will extract data on the cohort under investigation.²⁴ EHRs offer a bountiful collection of information, but this does not guarantee that they are complete or free from error. Using EHRs in this fashion

requires providers to document information accurately, completely, and expeditiously.³⁰

Drawbacks of Using the Electronic Health Record

EHRs are a rich source of data, but there are limitations and obstacles to implementing their use for research. The accuracy of information within an EHR relies on careful documentation by health care professionals. Issues with regard to using EHRs for research arise when there is missing information, erroneous data, data that cannot be accessed or interpreted, and inconsistencies in documentation.^{20,31}

Missing Information

It is imperative that data used for outcomes research are complete and accurate so that the results generated are applicable to the study population being examined. Providers record information within an EHR for the purpose of documenting patient care and billing, and therefore may not document information that is relevant for research purposes.^{23,32} Lack of physician documentation in the EHR has been studied. Pascoe et al compared the records of a regional cancer registry (Northern and Yorkshire Cancer Registry and Information Service) to those maintained by five primary care practices in the United Kingdom. They found that the EHRs managed by the primary care practices contained electronic codes for cancer with only a 29.4% level of completeness and 65.6% level of correctness when compared with the regional cancer registry.³³ They also found that 19.3% of patients with a diagnosis of malignancy in their paper charts had no electronic code for malignancy documented in their EHR.³³ In a comparison of data recorded in the Oncology Services Comprehensive Electronic Records (OSCER) and the SEER cancer registry, Lau et al found that data regarding race were missing for 40% of patients and tumor stage was missing for approximately 70% of patients.³²

Though documentation varies by the individual provider, these studies suggest that registries maintain a more complete and accurate record of cancer diagnoses than EHRs. Lau et al found that nearly half of the patients' record with an electronic code for malignancy had only a benign condition documented in the paper chart,³¹ suggesting that these diagnoses had been incorrectly coded in the EHR,³³ highlighting the concern that data documented within an EHR are prone to error. As there is no audit process in place for the EHR, the task of documenting valid information lies solely on the health care provider.²³ Miriovsky et al suggested implementing EHRs in which certain variables, such as tumor stage, are required to be documented by the provider before the encounter can be closed.²⁰

Inconsistencies in Documentation

Documents within the patient's chart are often entered as free text, either typed or dictated by the health care provider. Since the free-text portions of providers' notes are often a rich, concentrated source of information, including data from this portion of the record can contribute significantly to the

completeness of the patient's record. Documentation patterns also vary based on the provider constructing the record. There are many variations on medical terminology, spelling, formatting, and descriptions and interpretations of findings or results, all of which add to the complexity of the medical record. While one provider may document the stage of disease in a "structured" field, another provider may include this in a free-text portion of the document. All of these factors lead to challenges with the interoperability and interpretation of the EMR.

Currently, there exists standardized coding for medical terminology, diagnoses, and procedures, but EHR systems lack standardized coding for recording cancer progression and between systems in general.²³ Standardization of data will improve the organization of data within the EHR overall and facilitate the interoperability of various EHR systems as mandated by meaningful use.²⁰ The present lack of standardization prohibits researchers from being able to accurately compare data across different EHRs without significant data collection efforts.²⁴ Unless data are documented in a standardized fashion, comparative effectiveness research will be limited by the breadth of information available to researchers, which may produce results that are not generalizable to populations beyond the specific cohort being studied.^{23,31}

Inaccessible Data

Free-text fields potentially contain information vital to research, such as the patient's history, physical exam, radiology reports, operative reports, pathology reports, procedural reports, and discharge summaries. This "unstructured data" that are not entered into discrete fields within the EHR are difficult to extract in an accurate, reliable, and cost-effective fashion.³⁰ The standard "chart review" to extract this information is a slow, expensive process that may lack quality control.³⁰ Not including such information in outcomes research can lead to a misrepresentation of clinical quality measures.³¹

To use the data within EHRs, researchers must be able to locate and extract the relevant information effectively and efficiently.³⁴ Barriers to accurately extracting this information include the presence of irrelevant references to terms within the record, variation in terminology and spelling, lack of distinction between different conditions with similar names, lack of distinction of the timing of conditions or procedures, and inability to distinguish between the presence and denial of a symptom or condition, such as what is recorded in a review of systems.³⁴

A novel approach to addressing the issue of extracting unstructured data is the implementation of natural language processing (NLP), which extracts particular information from a free-text document.^{23,27,35} Xu et al developed a two-step framework utilizing NLP and machine learning to extract from the Vanderbilt University EHR a list of patients with a diagnosis of colorectal cancer by scanning free-text clinical notes and coded data. A case detection F₁ score of 0.93 was achieved.^{27,35} Denny et al developed a method utilizing NLP to identify the timing and status of completion of screening colonoscopies with a precision of 0.95. The NLP used also identified 29% more completed colonoscopies within the

EHR compared with institution's billing records.^{27,36} Application of this technology could lead to identification of previously unknown associations and relationships among factors such as a patient's personal history, medications, and behaviors.^{23,30} There is potentially valuable information to be obtained by examining the interplay of cancer patients' disease and response to treatment with factors such as their personal histories, tumor characteristics, psychological factors, lifestyle choices, and adherence to treatment.³⁷

Another obstacle to obtaining and incorporating valuable information from the patient's record is that data can be "trapped" in other forms. Information that exists only in paper charts and documents that have not been scanned into the patient's EHR are not capable of having data extracted from them using NLP.²³ At present, data from such records can be accessed and retrieved only by the meticulous and time-consuming manual chart review. With greater numbers of providers employing EHRs and greater quantities of data being entered into patient records, the utilization of these records for clinical research is promising.

Advantages of Using Electronic Health Record for Research

EHRs can provide data on the efficacy, effectiveness, and safety of treatment patterns. They offer an opportunity for researchers to evaluate dosing patterns and treatment combinations.^{24,31} By accumulating longitudinal data on existing patients, there is potential to uncover relationships among factors that impact a patient's disease process and their respective response to treatment. Public health can be improved with electronic reporting of cases of a diagnosis to the appropriate condition-specific registry and prompting patients to undergo screening and/or treatment at the appropriate time.^{19,26} The actualization of this potential of the EHR hinges on the interoperability of various systems at both the institutional level and on a regional scale.^{23,27} By collecting and integrating data among EHRs from various institutions and regions, a more complete picture of the study population can be represented in the study cohorts. Combining data from multiple institutions and regions will also increase the number of cases of rare cancers or conditions to be evaluated, allowing projects to attain sufficient power for results of their studies to be statistically significant.²³

Connecting the Electronic Health Record with Registries

EHRs and registries can serve as extraordinary sources of information for outcomes research, but this potential depends on the completeness and accuracy of the EHR. Regional registries can be used as an external validation source to improve the accuracy and completeness of data within practice-based registries.^{31,37} Institutional EHRs contain a great deal of information regarding a patient's treatment but do not necessarily contain information on treatment received by that patient outside of that particular institution and the ensuing outcomes.³⁷ Ideally, the EHR

should include records from all patient encounters, both from the inpatient and outpatient settings. Implementing EHR systems that are interoperable between departments and among institutions with the ability to link to registries would provide a more complete data source, which could be accessed for condition-specific research.

A cancer registry contains data on all incident cancers, patient demographics, tumor sites, histopathology, stage at diagnosis, and data on mortality. This information is certainly useful in terms of conducting clinical research, but it is by no means exhaustive. Cancer registries often do not include information regarding patient risk factors, such as their socioeconomic status, screening history, and lifestyle factors such as obesity and smoking. They contain no information on diagnostic tests or side effects of treatment, patient comorbidities, and disease recurrence.^{23,37} Tumor registries are also often 6 to 24 months behind on coding, so the information extracted at any given time may not be up to date, thus impacting the completeness and accuracy of the data being used for research.³⁰

The ideal situation is to have an EHR that seamlessly interfaces with multiple registries in an automated fashion. However, that system does not yet exist. Most EHRs are proprietary and highly customized. The codes and standards used in these systems are usually top secret. Therefore, there is the problem of interoperability between EHR systems and more importantly between EHRs and registries. Interoperability between EHRs is a core goal of the standards and certification criteria established by the Office of the National Coordinator for Health Information Technology.³⁸ Interoperability requires that two or more systems communicate with each other leading to accurate and consistent exchange of data in a form that is useable.³⁹ Three levels of interoperability are described: foundational interoperability, syntactic or structural interoperability, and semantic interoperability.⁴⁰ Foundational interoperability allows exchange of data between two systems without the requirement for interpretation of the data. Syntactic interoperability refers to the ability of different EHR systems to exchange data such that the meaning and purpose of the data are unaltered. Semantic interoperability refers to the ability of EHR systems to understand or interpret the exchanged data.⁴¹ This requires similar data models based on standard vocabularies and common data elements.^{40,42,43} From the standpoint of integrating EHRs and registries, functional interoperability requires both structural and semantic interoperability. For EHRs and registries to communicate, there needs to be uniform coding systems and data standards. Currently, EHR and registry vendors use different coding systems.

There is the need for open standards and EHRs need to be certified to meet functional interoperability. Despite the hurdles of interoperability, some EHRs have been used to populate registries. For example, the Society of Thoracic Surgeons and others use a central data repository that receives data from multiple systems through batch transfers using custom-programmed specifications and definitions for data collection elements. Another way of achieving interoperability is through the use of open-standard building

blocks that will lead incrementally to functional interoperability solutions. In this approach, the different components such as the physical network connection, web services, integration profiles, content profiles, data standards, and data security are standardized and built in incremental modular fashion to achieve interoperability and scalability.⁴⁴

Some EHRs have functionality for building registries within the system. This is mainly used for population management. An example of this is “Healthy Planet” from Epic Systems Corporation. These registries are built by having predefined criteria for enrolling patients. Patients who meet these criteria will be automatically added to the registry. Even though patients may be added to the registry automatically based on predefined criteria, there is still the need to manually manage the information that goes into the registry to ensure consistency of the data. Although this is a great approach to building registries, its robustness as a comprehensive registry for research is limited by the extent of EHR implementation. This means that if an institution does not have all the full components of the EHR, their registry will be limited in scope and some important information may exist in some other legacy system and will be inaccessible.

Another way of using EHR data for research is the use of data warehouses, where data from multiple sources including EHRs, laboratory data, billing data, claims data, and other data sources related to patient care are stored. It is a complex process where the data are kept in a staging area where it is scrubbed/cleansed and transformed and then stored in a data warehouse. The data within the warehouse is grouped and stored in Data Marts for retrieval.

Examples of Application to Colorectal Conditions

Currently, publications on EHRs and their registries in our field are few, and this represents a largely untapped resource for clinical research in colorectal surgery. Application of an EHR-based cancer registry could provide many potential benefits in advancing the screening, detection, diagnosis, and treatment of colorectal disease. Despite advancements in each of these areas, discrepancies in the incidence and survival by various ethnic groups persist. For example, African Americans have a higher incidence of CRC and a lower 5-year survival rate of CRC compared with other ethnic groups, which may be attributed to disparities in access to healthcare.⁴⁵ In a cross-sectional study of 12 community health centers within the Community Health Applied Research Network (CHARN), Baker et al found that the percentage of patients within the practices who were up to date on CRC screening via any screening modality ranged from 9.7 to 67.2% (median: 30.6%).⁴⁶ The lower rates of screening at certain centers were proposed to be due to a patient population with lower levels of education and income, language barriers, and lack of health insurance.^{19,46} Burnett-Hartman et al found that African Americans were less likely (adjusted odds ratio [OR] 0.89) and Asian/Pacific Islanders were more likely (adjusted OR: 1.16) than whites and Hispanics to undergo CRC screening within 18 months of entering the Population-based Research Optimizing Screening Through Personalized

Table 1 Odds of completing a CRC screening episode (during 18 mo) by race/ethnicity: PROSPR cohort 2010–2012

Characteristic	Screened, %	Adjusted odds ratio ^{a,b}	95% CI
Overall (n = 1,746,714)			
Non-Hispanic white	54.4	1.00	ref
Non-Hispanic black	50.9	0.89	0.88, 0.91
Hispanic	53.9	0.99	0.98, 1.00
Asian/Pacific Islander	59.2	1.16	1.15, 1.18
American Indian/Alaskan Native	47.7	0.76	0.72, 0.81
Multiple/other	63.6	1.31	1.28, 1.34

Abbreviations: CRC, colorectal cancer; PROSPR, Population-Based Research Optimizing Screening through Personalized Regimens.

^aAdjusted for age, sex, length of prior enrollment, and study site.

^bAdjusted for age, sex, study site, length of prior enrollment, ZIP code median income, type of residence, insurance type, and Charlson's comorbidity score.

Notes: Follow-up of 18 months after enrollment in the PROSPR cohort revealed that African Americans were less likely to undergo screening for colorectal cancer compared with whites and Hispanic in the same cohort. Reproduced with permission from Burnett-Hartman et al.⁴⁵

Regimens (PROSPR) cohort (→ **Table 1**).⁴⁵ African Americans were more likely to undergo initial screening with endoscopic screening than fecal immunochemical test compared with other ethnic groups. Asian/Pacific Islanders were less likely than other ethnic groups to follow up with colonoscopy after being found to have a positive fecal immunochemical test. The results of this study suggest that different ethnic groups may benefit from more targeted approaches.⁴⁵

EHRs can assist in the identification of patients who meet criteria for CRC screening and then issue reminders to patients or prompt physicians to reach out to patients to pursue screening or further evaluation at the appropriate time.^{19,45} In a randomized control trial evaluating delays in cancer diagnosis, Murphy et al found that patients whose physicians received EHR-based triggers underwent diagnostic evaluation for colorectal cancer significantly sooner (median: 104 vs. 200 days) than compared with controls (→ **Fig. 1**).⁴⁷ In another randomized control trial, 21,860 patients from 11 health care centers within the Harvard Vanguard Medical Associates (HVMA) group were found to be overdue for colorectal cancer screening by query of the group's common EMR. Sequist et al found that patients who received FOBT kits in the mail had a higher screening rate (44.0 vs. 38.1%) compared with control patients over a 15-month follow-up period (→ **Table 2**).⁴⁸

Conclusion

EHRs contain a wealth of information that can be useful for disease management and research in colon and rectal diseases and in particular colorectal cancer. With the widespread adoption of EHRs across individual/group practices and

Fig. 1 Effect of trigger-based intervention on time to diagnostic evaluation for colorectal cancer. Description: The time to evaluation for patients who required diagnostic evaluation for colorectal cancer was significantly reduced by the implementation of EHR-based triggers that notified patients’ respective primary care physician of this need. Reproduced with permission from Murphy et al.⁴⁷

hospitals, there is the need to combine data from all these sources into patient registries. Doing this requires the ability of all the different systems to interface seamlessly. EHRs contain both structured and unstructured data and it is difficult to search through free text to find pertinent information to map to structured data in registries. The use of NLP may help overcome some of the limitations of unstructured data. The

development of data standards will help facilitate interoperability between registries and multiple proprietary EHRs. Achieving the ideal state of interoperability between EHRs and registries is decades away, but there are efforts being made presently at various levels to develop data standards to accomplish this endeavor. Another source of data for research is the use of data warehouses, where data from multiple sources

Table 2 Receipt of colorectal cancer screening by intervention status

		Completed colorectal cancer screening ^a			
		Patient mailing intervention			
	No. of patients	Intervention, %	Control, %	Percentage point difference (95% CI)	p-Value
All patients	21,860	44.0	38.1	5.8 (4.5–7.1)	<0.001
Age group, y					
50–59	12,240	42.1	38.4	3.7 (2.0–5.5)	<0.001
60–69	5,671	45.4	38.0	7.3 (4.5–10.1)	<0.001
70–80	3,949	47.4	37.3	10.1 (7.0–13.2)	<0.001
Sex					
Female	12,439	44.3	38.6	5.7 (4.0–7.4)	< 0.001
Male	9,421	43.5	37.5	6.0 (4.1–7.9)	< 0.001
Primary care visits, no.					
0	7,643	19.6	15.6	3.9 (2.2–5.6)	<0.001
1–2	9,011	55.6	49.0	6.6 (4.7–7.4)	<0.001
≥ 3	5,206	59.5	52.3	7.1 (4.4–9.8)	<0.001

Abbreviation: CI, confidence interval.
^aAdjusted for clustering by physician and for physician baseline screening rate.
Notes: Patients who were due for colorectal cancer screening and were mailed screening packets were more likely to complete screening compare to control patients. Screening packets included instructions on scheduling screening colonoscopy or flexible sigmoidoscopy, a FOBT kit, and a pamphlet containing pertinent educational information. Reproduced with permission from Sequist et al.⁴⁸

including EHRs, laboratory data, billing data, claims data, and other sources related to patient care are stored. Specific data are stored in Data Marts for retrieval for research. As these systems become ubiquitous, so will opportunities to use these resources for colorectal clinical research to improve quality of care and outcomes for individual patients and populations.

References

- Non-federal Acute Care Hospital Electronic Health Record Adoption. Office of the National Coordinator for Health Information Technology website: dashboard.healthit.gov/quickstats/pages/FIG-Hospital-EHR-Adoption.php Published May 2016. Accessed September 6, 2016
- Gliklich RE, Dreyer NA, Leavy MD. Registries for Evaluating Patient Outcomes: A User's Guide. 3rd ed., Vol. 1. Rockville, MD: AHRQ Publication; 2014:17
- How Cancer Registries Work. Centers for Disease Control and Prevention website. <http://www.cdc.gov/cancer/npcr/value/registries.htm> Updated October 19, 2015. Accessed March 24, 2016
- The Cancer Registry and the Registrar. National Cancer Registrars Association website. <https://www.ncra-usa.org/files/public/Cancer-RegistryandRegistrarFactSheet2014.pdf>. Accessed March 24, 2016
- Cancer Registry. Virginia Mason Medical Center website. <https://www.virginiamason.org/cancerregistry>. Accessed March 24, 2016
- National Cancer Act of 1971. National Cancer Institute website. https://dtp.cancer.gov/timeline/flash/milestones/M4_Nixon.htm. Accessed March 24, 2016
- Cancer Registries Amendment Act. Centers for Disease Control and Prevention website. <http://www.cdc.gov/cancer/npcr/amendment.htm>. Reviewed October 13, 2015. Accessed March 22, 2016
- Institute of Medicine. Committee on Data Standards for Patient Safety. Key Capabilities of an Electronic Health Record System: Letter Report. Washington, DC: National Academies Press; 2003
- Electronic Health Records. History in the Making. Greater Than One. New York, NY: GTO Labs; 2012:4
- Weed LL. Medical records that guide and teach. *N Engl J Med* 1968; 278(11):593–600
- Tripathi M. EHR evolution: policy and legislation forces changing the EHR. *J AHIMA* 2012;83(10):24–29, quiz 30
- Schultz JR. A history of the PROMIS technology: an effective human interface. In: Goldberg A, ed., *A History of Personal Workstations*. New York, NY: ACM; 1988:439–488
- Atherton J. Development of the electronic health record. *Virtual Mentor* 2011;13(03):186–189
- DesRoches CM, Campbell EG, Rao SR, et al. Electronic health records in ambulatory care—a national survey of physicians. *N Engl J Med* 2008;359(01):50–60
- Jha AK, DesRoches CM, Campbell EG, et al. Use of electronic health records in U.S. hospitals. *N Engl J Med* 2009;360(16):1628–1638
- Electronic Health Records At A Glance. Centers for Medicare & Medicaid Services website. <https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2010-Fact-sheets-items/2010-07-132.html>. Published July 13, 2010. Accessed September 2, 2016
- Steinbrook R. Health care and the American Recovery and Reinvestment Act. *N Engl J Med* 2009;360(11):1057–1060
- American Recovery and Reinvestment Act of 2009. Section 4101. Incentives for Eligible Professionals. US Government Publishing Office website. <https://www.gpo.gov/fdsys/pkg/BILLS-111hr1enr/pdf/BILLS-111hr1enr.pdf>. Accessed April 28, 2016
- Cole AM, Tu SP, Fernandez ME, Calo WA, Hotz J, Wolver S. Reported use of electronic health records to implement evidence based approaches to colorectal cancer screening in community health centers. *J Health Care Poor Underserved* 2015;26(04):1235–1245
- Miriovsky BJ, Shulman LN, Abernethy AP. Importance of health information technology, electronic health records, and continuously aggregating data to comparative effectiveness research and learning health care. *J Clin Oncol* 2012;30(34):4243–4248
- Blumenthal D. Implementation of the federal health information technology initiative. *N Engl J Med* 2011;365(25):2426–2431
- Hsiao C-J, Hing E. Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001–2013. NCHS Data Brief, No 143. Hyattsville, MD: National Center for Health Statistics; 2014
- Kanas G, Morimoto L, Mowat F, O'Malley C, Fryzek J, Nordsyke R. Use of electronic medical records in oncology outcomes research. *Clinicoecon Outcomes Res* 2010;2:1–14
- Dean BB, Lam J, Natoli JL, Butler Q, Aguilar D, Nordsyke RJ. Review: use of electronic medical records for health outcomes research: a literature review. *Med Care Res Rev* 2009;66(06):611–638
- Hanna KE. Think Research: Using Electronic Medical Records to Bridge Patient Care and Research. *FasterCures White Paper Fall2005*:10–18
- Jha AK, DesRoches CM, Kralovec PD, Joshi MS. A progress report on electronic health records in U.S. hospitals. *Health Aff (Millwood)* 2010;29(10):1951–1957
- Turina M, Kiran RP. Electronic medical records in colorectal surgery. *Clin Colon Rectal Surg* 2013;26(01):17–22
- Fisher DA. Electronic medical records and improving the quality of the screening process. *J Gen Intern Med* 2011;26(07):683–684
- Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58(04):323–337
- Berger ML, Curtis MD, Smith G, Harnett J, Abernethy AP. Opportunities and challenges in leveraging electronic health record data in oncology. *Future Oncol* 2016;12(10):1261–1274
- Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med Care* 2013;51(08, Suppl 3):S30–S37
- Lau EC, Mowat FS, Kelsh MA, et al. Use of electronic medical records (EMR) for oncology outcomes research: assessing the comparability of EMR information to patient registry and health claims data. *Clin Epidemiol* 2011;3:259–272
- Pascoe SW, Neal RD, Heywood PL, Allgar VL, Miles JNV, Stefoski-Mikeljevic J. Identifying patients with a cancer diagnosis using general practice medical records and Cancer Registry data. *Fam Pract* 2008;25(04):215–220
- Edinger T, Cohen AM, Bedrick S, Ambert K, Hersh W. Barriers to retrieving patient information from electronic health record data: failure analysis from the TREC Medical Records Track. *AMIA Annu Symp Proc* 2012;2012:180–188
- Xu H, Fu Z, Shah A, et al. Extracting and integrating data from entire electronic health records for detecting colorectal cancer cases. *AMIA Annu Symp Proc* 2011;2011:1564–1572
- Denny JC, Peterson JF, Choma NN, et al. Extracting timing and status descriptors for colonoscopy testing from electronic medical records. *J Am Med Inform Assoc* 2010;17(04):383–388
- Kurian AW, Mitani A, Desai M, et al. Breast cancer treatment across health care systems: linking electronic medical records and state registry data to enable outcomes research. *Cancer* 2014; 120(01):103–111
- 2015 Edition Health Information Technology (Health IT) Certification Criteria, 2015 Edition Base Electronic Health Record (EHR) Definition, and ONC Health IT Certification Program Modifications. Office of the Federal Register website. <https://www.federalregister.gov/articles/2015/10/16/2015-25597/2015-edition-health-information-technology-health-it-certification-criteria-2015-edition-base>. Published October 16, 2015. Accessed September 5, 2016
- Definition of Terms. HIMSS Dictionary of Healthcare Information Technology Terms, Acronyms and Organizations. 3rd ed. Chicago, IL: HIMSS; 2013:75
- National Committee on Vital and Health Statistics (NCVHS). Report on Uniform Data Standards for Patient Medical Record Information. July 6, 2000:21–22

- 41 Institute of Electrical and Electronics Engineers. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York, NY1990
- 42 Adapting a Tool To Be caBIG® Compatible. National Cancer Institute website. <https://cabig.nci.nih.gov/sharable/compatible/>. Accessed August 17, 2016
- 43 Appendix B. HIMSS Dictionary of Healthcare Information Technology Terms, Acronyms and Organizations. 2nd ed. Chicago, IL: HIMSS; 2010:190
- 44 Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. 3rd ed., Vol 2. AHRQ Publication No. 13 (14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality; April 2014. <http://www.effectivehealthcare.ahrq.gov/registries-guide-3.cfm>
- 45 Burnett-Hartman AN, Mehta SJ, Zheng Y, et al; PROSPR Consortium. Racial/Ethnic disparities in colorectal cancer screening across healthcare systems. *Am J Prev Med* 2016;51(04):e107–e115
- 46 Baker DW, Liss DT, Alperovitz-Bichell K, et al. Colorectal cancer screening rates at community health centers that use electronic health records: a cross sectional study. *J Health Care Poor Underserved* 2015;26(02):377–390
- 47 Murphy DR, Wu L, Thomas EJ, Forjuoh SN, Meyer AN, Singh H. Electronic trigger-based intervention to reduce delays in diagnostic evaluation for cancer: a cluster randomized controlled trial. *J Clin Oncol* 2015;33(31):3560–3567
- 48 Sequist TD, Zaslavsky AM, Marshall R, Fletcher RH, Ayanian JZ. Patient and physician reminders to promote colorectal cancer screening: a randomized controlled trial. *Arch Intern Med* 2009;169(04):364–371