Published in final edited form as:
J Registry Manag. 2021 ; 48(1): 4–11.

A Case Study of Early-Onset Colorectal Cancer: Using Electronic Health Records to Support Public Health Surveillance on an Emerging Cancer Control Topic

Julie S. Townsend, MS[a],
Mary Catherine Jones, MPH[b],
Mildred N. Jones, BA, CTR[c],
Amy W. Waits, BS, CTR[c],
Kamilah Konrad, LMSW[c],
Natasha M. McCoy, MPH[b]

[a]Division of Cancer Prevention and Control, Centers for Disease Control and Prevention, Atlanta, Georgia.

[b]National Association of Chronic Disease Directors, Decatur, Georgia.

[c]Northside Hospital Cancer Institute, Atlanta, Georgia.

Abstract

Electronic health records (EHRs) are increasingly being used to support public health surveillance, including in cancer, where many population-based registries can now accept electronic case reporting. Using EHRs to supplement cancer registry data provides the opportunity to examine in more detail emerging issues in cancer control, such as the increasing incidence rates of early onset colorectal cancer (CRC). The purpose of this study was to evaluate the feasibility of a public health organization partnering with a health system to examine risk factors for early-onset CRC in a community cancer setting, and to further understand challenges with using EHRs to address emerging topics in cancer control. We conducted a mixed-methods evaluation using key informant interviews with public health practitioners, researchers, and registry staff to generate insights on how using EHRs and partnering with health systems can improve chronic disease surveillance and cancer control. A data quality assessment of variables representing risk factors for CRC and other clinical characteristics was conducted on all CRC patients diagnosed in 2016 at the participating cancer center. The quantitative assessment of the EHR data revealed that, while most chronic health conditions were well documented, around 25% of CRC patients were missing information on body mass index, alcohol, and tobacco use. Key informants offered ideas and ways to overcome challenges with using EHR data to support chronic disease surveillance. Their recommendations included the following activities: engaging EHR vendors in the development of standards, taking leadership roles on workgroups to address emerging technological issues, participating in pilot studies and task forces, and negotiating with EHR vendors so that clinical decision support tools

---

Address correspondence to Julie S. Townsend, MS, Centers for Disease Control and Prevention, 4770 Buford Hwy, NE, MS S107-4, Chamblee, GA 30341. jtownsend@cdc.gov. Telephone: (770) 488-3012.

This project was considered routine public health practice and did not require Institutional Review Board approval.

built to support public health initiatives are freely available to all users of those EHRs. Although using EHR data to support public health efforts is not without its challenges, it soon could be an important part of chronic disease surveillance and cancer control.

Keywords

colorectal neoplasms; electronic health records; public health surveillance; registries

Introduction

Electronic health records (EHRs) are increasingly being used to support public health surveillance for a variety of health topics, including syndromic surveillance, immunizations, cancer, childhood obesity, diabetes, hypertension, asthma, as well as other conditions.[1–5] While electronic laboratory records (ELRs) have supported public health surveillance for a number of years, EHRs are also being tapped to enhance reporting of immunizations and notifiable conditions to health departments and registries, including infectious diseases, and incident cancer cases (https://www.naaccr.org/meaningful-use/).[1,6–8]

While incidence rates are declining for many cancer sites, there are notable increases in rates of new cases of early-onset colorectal cancer (CRC); uterine, pancreatic, and liver cancers; as well as others in which there is a need to better understand contributing factors to rising incidence rates and implement evidence-based interventions to address them.[9,10] Early-onset CRC was also chosen as a topic because of the importance of genomic testing in this disease and the opportunities available for identifying at-risk family members through cascade testing for Lynch syndrome and other inherited disorders. Additionally, understanding the prevalence of obesity as well as alcohol and tobacco use and other chronic health conditions may be important to cancer control planners, given that early-onset CRC survivors may benefit from survivorship care interventions to mitigate adverse health outcomes and improve their quality of life.[11]

Population-based central cancer registries are the backbone of cancer surveillance in the United States, and provide valuable information on demographic and clinical characteristics of new cancer cases (https://www.cdc.gov/uscs). However, there are opportunities for EHR, laboratory, and health systems data to support or enhance data reported to cancer registries and provide information not commonly found in registries (eg, tobacco use or obesity) to support special surveillance studies for emerging public health challenges.[12–15] Some cancer registries have been linked to administrative and claims data and other sources.[15–18] In many research settings, “big data” is increasingly being used to support cancer research studies, and cloud-based infrastructure has brought together hospital cancer registry, EHR, laboratory, and pharmacy data to drive improvements in cancer treatment and care.[19,20] Linking to EHRs and other electronic data sources represents a new opportunity to examine cancer risk factors, screening test use, molecular characteristics, and chronic health conditions (among others) before or at the time of a cancer diagnosis, as well as serving as another source of information on treatment, treatment adverse effects, recurrence, and other health conditions for public health researchers and practitioners.[21–25]

In 2017–2019, we undertook a pilot project to assess the feasibility of partnering with a large health system (Northside Hospital) in Atlanta, Georgia to assess risk factors for early-onset CRC and opportunities for prevention and control among the patient population at its community cancer center (Northside Hospital Cancer Institute), using both EHR and cancer registry data. As part of the project's evaluation, we completed a data quality assessment of EHR variables and conducted key informant interviews with public health professionals and cancer registry staff to identify successes, challenges, and barriers to using EHR and other health systems data to support chronic disease surveillance and special investigations, with the goal of identifying recommendations for public health departments that are interested in partnering with health systems on cancer-related projects at the local level to inform comprehensive cancer control efforts.

Methods

Project Description

This project was a joint collaboration between the National Association of Chronic Disease Directors, the Centers for Disease Control and Prevention (CDC), and the Northside Hospital Cancer Institute, an American College of Surgeons Commission on Cancer (CoC)-accredited comprehensive community cancer center with 3 acute-care hospitals serving the metropolitan Atlanta area at the time of the study. Briefly, one of the primary activities of our pilot project was conducting a descriptive, retrospective study of all 2016 CRC cancer cases at Northside Hospital Cancer Institute. Data elements were selected based on known CRC risk factors from the scientific literature, available registry data, and emerging conditions of interest. We obtained permission so that our certified tumor registrars (CTRs) could access and abstract data from 2 EHR systems: the hospital's and an affiliated gastroenterology practice. CTR team members went through an initial 10-case quality assurance (QA) review for each abstractor, with feedback and education provided along with an ongoing 10% QA review by a senior CTR team member to ensure data completeness and accuracy.

Mixed-Methods Evaluation Approach

Given that the project was a feasibility study, an evaluation plan using a mixed-methods approach to collect both quantitative and qualitative data was developed to address the specific study questions of the project; namely:

1. Can medical data from an integrated health care delivery system be rapidly assessed and used to determine accurate and high-quality information on early onset CRC without the need to contact the patient?
2. How can we use what we learn to build capacity among other integrated health care delivery systems and their public health partners, particularly those in the community cancer setting?

Quantitative Data Quality Assessment

We addressed the first question through an analysis conducted during a data-quality assessment of key variables needed to assess potential risk factors for early-onset CRC not typically collected as part of the cancer registry abstract. These included body mass index (BMI), tobacco use, alcohol use, CRC screening history, tumor screening for Lynch syndrome, and family history of CRC and related Lynch syndrome cancers (Figure 1). Other variables collected and assessed included demographic characteristics (eg, driving distance from the patient's residence to the cancer center, patient's preferred language, and patient's status as a caregiver). Clinical characteristics included a history of chronic health conditions, such as inflammatory bowel disease and diabetes. In total, an additional 114 data elements were abstracted from 2 EHR systems. During the data-quality assessment, we analyzed the number of patients with missing or unknown information using SAS statistical software.

Qualitative Key Informant Interviews

We completed qualitative key informant interviews with subject matter experts in the areas of laboratory reporting, state cancer registries, hospital cancer registries, state-level chronic disease epidemiology, and syndromic surveillance to better understand how to build capacity among other integrated health care delivery systems and their public health partners. Interviewees were selected based on project team recommendations with the goal of including different professional experiences with EHRs from the public health field. With the exception of 1 expert who did not respond to our inquiry, all experts invited to an interview completed one.

A semistructured discussion guide was developed that addressed the following:

1. The current landscape of using health systems data and EHRs to support public health surveillance
2. The facilitators (ie, keys to success) and barriers to health care systems partnering with public health organizations on surveillance efforts
3. The processes, policies, or practices that can help to overcome the barriers and capitalize on facilitating factors

Interviews were conducted with 9 subject matter experts who had experience with EHRs and health systems data to support public health surveillance or research. Subject matter experts included cancer center registry staff, state cancer registry staff, a gastroenterologist, and employees at state and federal government agencies. All interviews were conducted via telephone, except for 1 in-person team interview with cancer center staff involved in the project. Discussions were conducted in segments of 30 to 60 minutes. The team evaluator led key informant interviews and involved team members in contributing to the discussion with subject matter experts, including providing contextual information about the project, encouraging authentic discussion, and asking follow-up questions to prompt for additional insights and observations. The team evaluator took notes during the sessions and synthesized themes that emerged through the discussions with the subject matter experts. The themes were organized according to the potential audience (public health professionals, health systems, and industry/professional organizations) and recommendations/actions that could be done by the audience to advance the use of EHRs for public health surveillance.

Study Approval

CDC review determined this project to be public health practice. Office of Management and Budget approval was not required for data collection because fewer than 9 nonfederal key informants were interviewed, and information was collected through secondary data sources for the data assessment. The data assessment was approved by the Northside Hospital Research Oversight Committee.

Results

What are Factors that Influence the Quality and Accuracy of, and Ease of Access to, Information on Early-Onset Colorectal Cancer Risk Factors in the Cancer Center's Records?

Findings from the data quality assessment revealed that data completeness (percentage of unknown or missing information) varied, depending on the variable collected (Table 1). Unknown/missing values ranged from 5% for common chronic health conditions to around 25% for health behaviors like alcohol and tobacco use. Around 25% of CRC patients had missing information on their BMI, and similar proportions had missing information for a family history of CRC or endometrial or ovarian cancer. Variables with the highest percentage of missing data or unknown information included the patient's preferred language (27%), the patient's caregiving status (36%), history of polyps (31%), history of a previous cancer (31%), and time from onset of symptoms to diagnosis (33%).

Key informants from the cancer center who participated in this project worked with multiple practices that use different EHRs, which did not necessarily "communicate" with each other. The use of multiple EHRs complicated data analysis for various reasons, including variation among EHRs in the headings, fields, and ways that risk factor data are documented, which challenged analysis across EHRs. There were also different security requirements for each EHR. Within EHRs, there was inconsistency in how and where information is documented. Some data were found in multiple locations within the EHR, such as patient history and the intake form, and sometimes the information conflicted. A physician's office may collect the information differently from a surgical preadmission form about the same topic (eg, do you smoke vs history of smoking/ever smoked). Key informants noted that there needs to be a protocol for determining which data to consider for risk factors. Additionally, key informants noted inconsistencies among providers in how often risk factor data were updated. For example, family history may be collected at intake but never updated over the course of the patient's care.

Some medical information continues to be collected on paper, outside of EHRs. We collected some data elements needed for the analysis from documents scanned into the EHR rather than entered into electronic fields. This had to be retrieved manually, which slowed the assessment and added cost in staff time. We learned that extending the use of the EHR from patient care to surveillance requires a shift in how the data are collected and analyzed. Creating user-defined fields in the cancer registry software to capture information not readily available in existing fields was resource-intensive.

What are Systemic Factors that Influence the Availability and Quality of Electronic Health Data That Can be Used for Public Health Surveillance, Especially for Emerging Issues in Cancer Control?

Subject matter experts noted that the implementation of the American Reinvestment and Recovery Act—Health Information Technology for Economic and Clinical Health (HITECH), which spurred adoption of EHRs by hospitals and physician practices, and Meaningful Use, which laid additional groundwork for health systems' data to be used for public health, have substantially increased the availability and quality of EHR data. Various national-level mechanisms to promote quality improvement and value-based care, including reporting on quality standards, have been instrumental. The CDC has worked to engage health systems, EHR vendors, and other stakeholders in surveillance of cancer, immunizations, asthma, diabetes, and syndromes. Key informants noted that a competitive marketplace, including competition among EHR vendors for market share and competition among health care providers for patients may have influenced the availability and quality of EHR data, the increased consumption of health care quality information by patients, and the increased use of technology for patients to monitor and report health care data to providers. Additionally, professional organizations have developed support to health systems staff in improving quality and accuracy of patient information in cancer registries.

Despite the convergence of these factors in increasing the availability and use of EHR data, multiple key informants noted the workload challenges that the technology presents to health care providers. One key informant noted that, despite advancements in EHR technology, they have not necessarily made data collection and entry more efficient or translated into more time for patient care:

> "Finding a way to make up for the increased workflow required by data entry is a challenge. The way EHR systems are constructed is very old school, the electronic version of someone taking notes or writing them in a paper record."

What are Some Opportunities for Health Systems, Public Health, and Allied Agencies to Increase the Value and Use of Electronic Health Record Data for Chronic Disease Surveillance, Especially the Identification of Risk Factors for Early-Onset Colorectal Cancer?

Table 2 summarizes recommendations for what health systems, public health agencies, and professional and industry associations can do to improve the use of EHR data for chronic disease surveillance, organized along with the themes that emerged from the qualitative data analysis of the interviews. These themes included engaging stakeholders/building partnerships, task force/work group participation, administrative/systems change, improving data quality and data use, and communication/dissemination. Notably, health systems and public health agencies can collaborate to improve population health, and one starting point is participating in state or national level pilot studies, task forces, and work groups. There are opportunities for health systems to engage providers in developing clinical decision support tools that are implemented with population health management platforms with existing interfaces with EHRs. Health systems can also develop and enforce protocols and standards for documentation among providers.

Public health entities may consider assisting health systems in identifying where quality improvement and financial incentives align with public health surveillance priorities as a way to build support and generate interest in using EHR data for chronic disease surveillance. At the local level, public health agencies could collaborate with health systems to identify priority health topics or disease areas where the health system and community would derive significant benefit from EHR improvement (eg, diabetes management). Public health agencies could assist health systems in identifying shared data needs related to surveillance and with vetting variables to be monitored and reported. Within public health agencies, coordination across disease areas and administrative units to work with health systems, including building on existing efforts related to syndromic surveillance and immunization, may synergize efforts and increase efficiencies for chronic disease surveillance.

Allied organizations, including professional and industry associations, could engage EHR vendors in collaborations, especially in developing standards and discussions of how to increase interoperability. They could also provide incentives to pilot novel approaches to EHR vendor and health systems' collaborations to support chronic disease surveillance and obtain agreements from vendors that any products resulting from the collaboration are available freely to all providers using that EHR system.

Discussion

The interviews with key informants on using EHRs and health systems data to support the investigation of emerging topics in cancer control and other chronic disease surveillance activities revealed several key domains: forming partnerships/engaging stakeholders, participating in task forces/work groups, providing education on systems and administrative changes, improving EHR data quality, and communicating/disseminating findings. Addressing each of these domains may improve the use of EHRs to support cancer control and other public health efforts at the local level. Public health, health systems, and professional/industry associations can all play a role across these domains.

We specifically addressed one key theme, improving EHR data quality, through our own data quality assessment. Our health system partner's EHR and those of its affiliated gastroenterology practices could readily provide information on the prevalence of chronic health conditions, but around one-fourth of CRC patients each had incomplete data for family history of cancer, health behaviors, and other clinical factors (duration of symptoms, history of polyps, tumor testing, etc) that could be of interest to public health partners who need relevant data to inform community interventions. Although some patients with missing data may have come from nonaffiliated gastroenterology practices outside of the health system, our findings on data completeness may be typical with EHR data.[26,27] Even well-established health care research networks using virtual data warehouses have noted the challenges with electronic capture of molecular data, particularly data elements that may only be found in scanned imaging reports and are not captured in standardized EHR data fields or as site-specific factors in tumor registries.[25] Even with these limitations, analysis of available data may be helpful for public health surveillance purposes and generating new hypotheses, which can be tested further in prospective studies. Additionally, it provides a snapshot of care that goes beyond analyzing traditional cancer registry data elements, which can provide helpful local data that public health partners could potentially use to improve access to care and train providers on use of clinical guidelines, such as tumor screening for Lynch syndrome and genetic counseling referral.[28]

Although EHRs are increasingly being used to support chronic disease surveillance, their use so far has been limited to a few topic areas (eg, diabetes, obesity, asthma, hypertension, cancer electronic reporting) and geographic areas of the United States. However, partnerships between public health organizations and health systems are increasingly becoming common to address a variety of chronic health conditions and implement interventions to improve health. For example, CDC-funded cancer programs at health departments and universities are partnering with health systems on projects to increase cancer screening and generate survivorship care plans.[11,29] Given that these efforts rely on accurate data, projects such as these may contribute to overall improvements in useful and quality data.

Health care technology is constantly evolving, and it may be challenging for public health organizations to keep up with new technologies, like HL7 Fast Healthcare Interoperability Resources (https://www.hl7.org/fhir/overview.html), that can streamline data exchange and make it possible to get regular data feeds so that the most current patient data are available.[30] Other technologies are increasingly becoming available that may help standardize data across different EHR platforms and capitalize on natural language processing techniques to make data more accessible for public health needs.[31,32] As distributed data networks become more commonplace, it may be increasingly important for public health organizations to be engaged with health systems around data so that emerging topics in cancer control can be quickly assessed and appropriate interventions and timely access to clinical care applied.[33–35] These efforts hinge on having public health and health system partners with adequate skills in data science and the information technology infrastructure for big data.[36] CTRs may continue to play a key role in ensuring data quality, along with their expertise in the types of information to capture and consolidation of data across multiple information streams.[37]

There are some limitations to our mixed-methods evaluation. During the EHR data quality assessment, we did not evaluate internal validity, whether certain patient or provider characteristics and referral patterns played a role in data completeness, or if the sample with complete information was representative of the overall patient population. Therefore, we did not evaluate all potential domains of data quality proposed for assessing EHR data for research use.[38] We only examined 1 year of data for one cancer site (CRC) at a single health system, limiting generalizability to other patient populations. Although our key informants represented federal and state health departments and health systems perspectives, we did not recruit key informants employed by EHR vendors, who may have lent a different perspective on EHR use.

Despite these limitations, there are some strengths to our study. We evaluated a practical use case scenario using cancer registry data supplemented by EHR data elements to better understand risk factors among early- and late-onset CRC patients. We were able to leverage trained CTRs using a data dictionary that we developed to capture standardized information from divergent EHRs. Our key informants represented many different user experiences and lent valuable insights into using EHRs to support investigating emerging topics in cancer control.

Conclusion

Major efforts are underway at the federal, state, academic, and local health care levels to tap into EHRs, laboratory data, biobanks, and genomics data to integrate information for a more complete picture of population health.[6,26] The key domains we identified through our key informant interviews may be able to guide public health practitioners, health systems, and professional associations/vendors on how to navigate this unchartered territory by providing concrete actions that can be undertaken through this journey. Our data quality findings may be used to identify problem areas in EHRs that need attention, such as improving the documentation of health behaviors and cancer family history that may impact the cancer patient's prognosis through the treatment and survivorship period, demographic characteristics related to the social determinants of health, and other clinical characteristics that can inform community-level interventions with health system partners. Big data analytics using integrated, cloud-based data may one day allow public health professionals, researchers, and cancer control planners to better understand emerging topics in cancer control, including early-onset CRC.

Acknowledgements

We would like to thank our project team members: Maurshanda Renee Matthews, Sheema Ahmed, Dr. Marc Sonenshine, Kate Canterbury, Dr. Xu Xhang, Dr. C. Brooke Steele, Frank Bright, and Dr. Paulette Valliere for their contributions to this project. We would like to acknowledge the key informants from the CDC, National Cancer Institute, University of Kentucky, and Hawaii Department of Health who kindly agreed to provide their thoughts and opinions about using electronic health records to support chronic disease surveillance and cancer control.

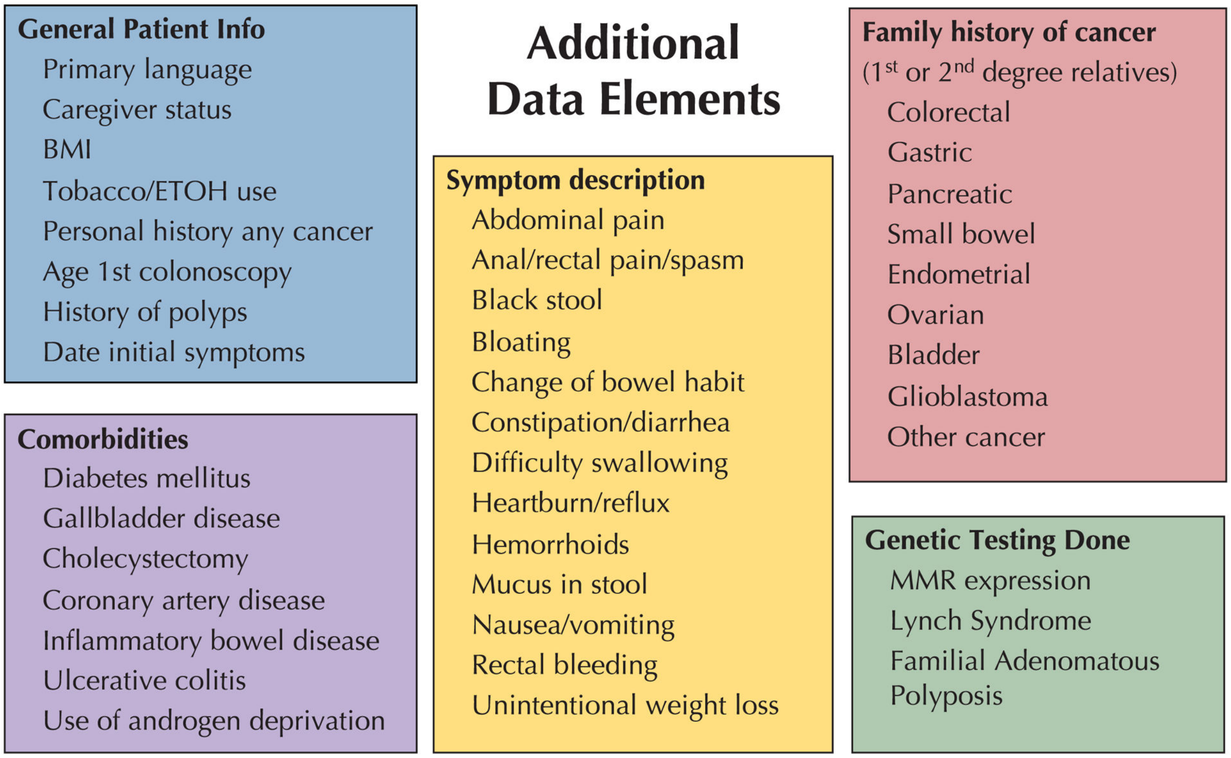
The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or the National Association of Chronic Disease Directors. Funding for the Health Behaviors and Other Risk Factors in Early-Onset Colorectal Cancer project was provided by Cooperative Agreement Number 5NU38OT000225-05, sponsored by the Centers for Disease Control and Prevention, Division of Cancer Prevention and Control (CDC), and the National Association of Chronic Disease Directors (NACDD).

References

1. Birkhead GS, Klompas M, Shah NR. Uses of electronic health records for public health surveillance to advance public health. Annu Rev Public Health. 2015;36:345–359. [PubMed: 25581157]
2. Perlman SE, McVeigh KH, Thorpe LE, Jacobson L, Greene CM, Gwynn RC. Innovations in population health surveillance: using electronic health records for chronic disease surveillance. Am J Public Health. 2017;107(6):853–857. [PubMed: 28426302]
3. Gutilla MJ, Davidson AJ, Daley MF, Anderson GB, Marshall JA, Magzamen S. Data for community health assessment in rural Colorado: a comparison of electronic health records to public health surveys to describe childhood obesity. J Public Health Manag Pract. 2017;23(suppl 4):S53–S62.
4. Klompas M, Cocoros NM, Menchaca JT, et al. State and local chronic disease surveillance using electronic health record systems. Am J Public Health. 2017;107(9):1406–1412. [PubMed: 28727539]

5. Thorpe LE, McVeigh KH, Perlman S, et al. Monitoring prevalence, treatment, and control of metabolic conditions in New York City adults using 2013 primary care electronic health records: a surveillance validation study. EGEMS (Wash DC). 2016;4(1):1266. [PubMed: 28154836]
6. Cooney MA, Iademarco MF, Huang M, MacKenzie WR, Davidson AJ. The public health community platform, electronic case reporting, and the digital bridge. J Public Health Manag Pract. 2018;24(2):185–189. [PubMed: 29360697]
7. Blumenthal W, Alimi TO, Jones SF, et al. Using informatics to improve cancer surveillance. J Am Med Inform Assoc. 2020;27(9):1488–1495. [PubMed: 32941600]
8. Pollack LA, Jones SF, Blumenthal W, et al. Population heath informatics can advance interoperability: National Program of Cancer Registries Electronic Pathology Reporting Project. 2020;4:985–992.
9. Siegel RL, Fedewa SA, Anderson WF, et al. Colorectal cancer incidence patterns in the United States, 1974–2013. J Natl Cancer Inst. 2017;109(8):djw32.
10. Henley SJ, Ward EM, Scott S, et al. Annual report to the nation on the status of cancer, part I: national cancer statistics. Cancer. 2020;126(10):2225–2249. [PubMed: 32162336]
11. Rohan EA, Miller N, Bonner F III, et al. Comprehensive cancer control: promoting survivor health and wellness. Cancer Causes Control. 2018;29(12):1277–1285. [PubMed: 30506490]
12. Hernandez MN, Voti L, Feldman JD, et al. Cancer registry enrichment via linkage with hospital-based electronic medical records: a pilot investigation. J Registry Manag. 2013;40(1):40–47. [PubMed: 23778697]
13. Ruppert LP, He J, Martin J, et al. Linkage of Indiana State Cancer Registry and Indiana Network for Patient Care Data. J Registry Manag. 2016;43(4):174–178. [PubMed: 29595920]
14. Reams C, Powell M, Edwards R. State synergies and disease surveillance: creating an electronic health data communication model for cancer reporting and comparative effectiveness research in Kentucky. EGEMS (Wash DC). 2014;2(2):1064. [PubMed: 25848604]
15. Tucker TC, Durbin EB, McDowell JK, Huang B. Unlocking the potential of population-based cancer registries. Cancer. 2019;125(21):3729–3737. [PubMed: 31381143]
16. Gallaway MS, Huang B, Chen Q, et al. Identifying smoking status and smoking cessation using a data linkage between the Kentucky Cancer Registry and health claims data. JCO Clin Cancer Inform. 2019;3:1–8.
17. Blanchette PS, Chung H, Pritchard KI, et al. Influenza vaccine effectiveness among patients with cancer: a population-based study using health administrative and laboratory testing data from Ontario, Canada. J Clin Oncol. 2019;37(30):2795–2804. [PubMed: 31465264]
18. Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. Med Care. 2002;40(8 suppl):IV-3–18.
19. Rubinstein SM, Warner JL. CancerLinQ: origins, implementation, and future directions. JCO Clin Cancer Inform. 2018;2:1–7.
20. Tran Q, Warren JL, Barrett MJ, et al. An evaluation of the utility of big data to supplement cancer treatment information: linkage between IQVIA pharmacy database and the Surveillance, Epidemiology, and End Results Program. J Natl Cancer Inst Monogr. 2020;2020(55):72–81. [PubMed: 32412073]
21. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. JMIR Med Inform. 2019;7(2):e12239. [PubMed: 31066697]
22. Banerjee I, Bozkurt S, Caswell-Jin JL, Kurian AW, Rubin DL. Natural language processing approaches to detect the timeline of metastatic recurrence of breast cancer. JCO Clin Cancer Inform. 2019;3:1–12.
23. Dewdney SB, Lachance J. Electronic records, registries, and the development of “big data”: crowd-sourcing quality toward knowledge. Front Oncol. 2016;6:268. [PubMed: 28194369]
24. Booth CM, Karim S, Mackillop WJ. Real-world data: towards achieving the achievable in cancer care. Nat Rev Clin Oncol. 2019;16(5):312–325. [PubMed: 30700859]

25. Burnett-Hartman AN, Udaltsova N, Kushi LH, et al. Clinical molecular marker testing data capture to promote precision medicine research within the cancer research network. JCO Clin Cancer Inform. 2019;3:1–10.
26. Hiatt RA, Tai CG, Blayney DW, et al. Leveraging state cancer registries to measure and improve the quality of cancer care: a potential strategy for California and beyond. J Natl Cancer Inst. 2015;107(5):djv047. [PubMed: 25766400]
27. Wright A, McCoy AB, Hickman TT, et al. Problem list completeness in electronic health records: a multi-site study and assessment of success factors. Int J Med Inform. 2015;84(10):784–790. [PubMed: 26228650]
28. Provenzale D, Gupta S, Ahnen DJ, et al. Genetic/familial high-risk assessment: colorectal version 1.2016, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2016;14(8):1010–1030. [PubMed: 27496117]
29. Tangka FKL, Subramanian S, Hoover S, et al. Identifying optimal approaches to scale up colorectal cancer screening: an overview of the centers for disease control and prevention (CDC)'s learning laboratory. Cancer Causes Control. 2019;30(2):169–175. [PubMed: 30552592]
30. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. J Am Med Inform Assoc. 2016;23(5):899–908. [PubMed: 26911829]
31. Mandl KD, Kohane IS. A 21st-Century health IT system: creating a real-world information economy. N Engl J Med. 2017;376(20):1905–1907. [PubMed: 28514623]
32. Warner JL, Levy MA, Neuss MN, Warner JL, Levy MA, Neuss MN. ReCAP: Feasibility and accuracy of extracting cancer stage information from narrative electronic health record data. J Oncol Pract. 2016;12(2):157–158; e169–157. [PubMed: 26306621]
33. Tabano DC, Cole E, Holve E, Davidson AJ. Distributed data networks that support public health information needs. J Public Health Manag Pract. 2017;23(6):674–683. [PubMed: 28628584]
34. Khan S, Shea CM, Qudsi HK. Barriers to local public health chronic disease surveillance through health information exchange: a capacity assessment of health departments in the Health Information Network of South Texas. J Public Health Manag Pract. 2017;23(3):e10–e17.
35. Bacon E, Budney G, Bondy J, 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. 2019;25(5):498–507. [PubMed: 31348165]
36. McFarlane TD, Dixon BE, Grannis SJ, Gibson PJ. Public health informatics in local and state health agencies: an update from the Public Health Workforce Interests and Needs Survey. J Public Health Manag Pract. 2019;25(2 suppl):S67–S77. [PubMed: 30720619]
37. Peterson J Encouraging health information management graduates to pursue cancer registry careers. J Registry Manag. 2016;43(1):37–41. [PubMed: 27195997]
38. Weiskopf NG, Bakken S, Hripcsak G, Weng C. A data quality assessment guideline for electronic health record data reuse. EGEMS (Wash DC). 2017;5(1):14. [PubMed: 29881734]
39. Nine essential principles of software usability. Healthcare Information and Management Systems Society website. Published April 9, 2015. Accessed December 4, 2019. https://www.himss.org/himss-emr-usability-evaluation-guide-clinicians-practices-9-essential-principles-software-usability



Additional Data Elements

General Patient Info
- Primary language
- Caregiver status
- BMI
- Tobacco/ETOH use
- Personal history any cancer
- Age 1st colonoscopy
- History of polyps
- Date initial symptoms

Comorbidities
- Diabetes mellitus
- Gallbladder disease
- Cholecystectomy
- Coronary artery disease
- Inflammatory bowel disease
- Ulcerative colitis
- Use of androgen deprivation

Symptom description
- Abdominal pain
- Anal/rectal pain/spasm
- Black stool
- Bloating
- Change of bowel habit
- Constipation/diarrhea
- Difficulty swallowing
- Heartburn/reflux
- Hemorrhoids
- Mucus in stool
- Nausea/vomiting
- Rectal bleeding
- Unintentional weight loss

Family history of cancer
(1st or 2nd degree relatives)
- Colorectal
- Gastric
- Pancreatic
- Small bowel
- Endometrial
- Ovarian
- Bladder
- Glioblastoma
- Other cancer

Genetic Testing Done
- MMR expression
- Lynch Syndrome
- Familial Adenomatous Polyposis

Figure 1.
Select Additional Data Elements Collected for the Colorectal Cancer Cohort

Table 1.

Findings from Data Quality Assessment of Key Variables for CRC Cohort, N = 721

Variable	**Unknown/missing, n (%)**
Demographic characteristics	
Distance to hospital from residence	55 (7.6)
Primary language spoken	191 (26.5)
Patient caregiver status	257 (35.6)
Health behaviors	
Alcohol use history	178 (24.7)
Tobacco use history	160 (22.2)
Body mass index	186 (25.8)
Clinical factors	
Time from initial symptom to diagnosis	240 (33.3)
History of other cancer	220 (30.5)
History of polyps	222 (30.8)
Diabetes mellitus	37 (5.1)
Gallbladder disease	37 (5.1)
Coronary artery disease	37 (5.1)
Inflammatory bowel disease	37 (5.1)
Family history of cancer	
First-degree relative with CRC	180 (25.0)
First-degree relative with endometrial cancer	176 (24.4)
First-degree relative with ovarian cancer	175 (24.3)
Second-degree relative with CRC	176 (24.4)
Second-degree relative with endometrial cancer	175 (24.3)
Second-degree relative with ovarian cancer	175 (24.3)

CRC, colorectal cancer.

Table 2.

Thoughts and Ideas Provided by Key Informants for Improving Electronic Health Records to Support Chronic Disease Surveillance and Cancer Control, by Audience Type and Topic Area

Health Systems	Public Health	Professional and Industry Associations
Building partnerships/engaging stakeholders		
Have conversations with public health organizations about their common interests in chronic disease prevention, including screening and early identification, and management. Bring together providers and EHR vendors to adapt EHRs to provide the greatest benefit to practice and administration. Engage providers in developing the clinical questions that align with the public health surveillance needs. Have conversations with public health organizations about their common interests in chronic disease prevention, including screening and early identification, and management.	Coordinate across disease areas and administrative units to work with health systems, including building on existing efforts related to syndromic surveillance and immunization where appropriate. Collaborate across different levels of government to ensure that local, state, and federal opportunities are connected. Facilitate the connection of health systems with community organizations, public health coalitions, and other stakeholders that can use health systems data to inform community health and action planning. Collaborate with health systems to identify priority health topics or disease areas where the health system and community may derive significant benefit from EHR improvement. Explore opportunities to work with health plans to incentivize identification of early onset CRC risk factors for providers.	Convene diverse stakeholders at different levels of healthcare organizations, including members, with public health to facilitate problem definition and strategic planning. Learn the landscape of health information technology and its intersections with public health and assist other organizations in defining their roles. Engage EHR vendors in collaborations, especially in the development of standards and discussions of how to increase interoperability.
Taskforce/workgroup participation		
Participate in state or national level pilot studies, task forces, and workgroups to identify and implement mutually beneficial opportunities for collaboration with public health.	Lead or participate in the development of standards for EHR vendors.	Lead regional or national workgroups and task forces to address collaboration between public health and health systems around surveillance, such as the CDC/CSTE/APHL’s Electronic Laboratory Reporting Task Force, and the CDC Public Health-EHR Vendors Collaboration Initiative.
Administrative/system changes		
Encourage or incentivize their affiliated practices to use one EHR, which may eliminate inefficiencies for data collection and analysis associated with the use of multiple EHR platforms.	Assist health systems in identifying where quality improvement and financial incentives align with public health surveillance priorities. Create and promote recognition programs to highlight the efforts of health systems making strides in using their EHR for chronic disease surveillance.	Provide incentives to pilot novel approaches to EHR vendor and health systems collaborations to support chronic disease surveillance; obtain agreement from the vendor that any products resulting from the collaboration are available freely to all providers using that EHR system.
Improving data quality and data use		
Develop, implement, train on, and enforce protocols and standards for documentation among providers.	Encourage the use of existing public health data sources and healthcare data in community health needs assessments. Demonstrate the connection between public health surveillance interests and clinical quality measures. Engage EHR vendors in piloting algorithms that can improve the ease of collecting early onset CRC risk factor data.	Continue the development of tools and resources that health systems and public health can use or adapt to facilitate EHR modification to meet public health surveillance interests, such as the Healthcare Information and Management Systems Society’s (HIMSS) EMR Usability Evaluation Toolkit.[39]
Communication/dissemination		
Communicate the direct benefit of improved documentation to providers, including sharing of findings from surveillance efforts and how improved documentation links back to improved patient care.	Share successes and challenges and develop solutions with other public health organizations.	Promote the exchange of successful public health-health systems data partnerships within and across disease areas through publications, conference presentations, and webinars. Define the mutual benefit of public health-health systems collaborations on surveillance.