Validation of rule-based algorithms to determine colorectal, breast, and cervical cancer screening status using electronic health record data from an urban healthcare system in New York City

Aleeza J. Leder Macek [a], Joshua D. Kirschenbaum [a], Sarah J. Ricklan [a], William Schreiber-Stainthorp [a], Britney C. Omene [b], Sarah Conderino [a,*]

[a] New York University Grossman School of Medicine, United States
[b] New York University College of Arts and Sciences, United States

ARTICLE INFO

ABSTRACT

Although cancer screening has greatly reduced colorectal cancer, breast cancer, and cervical cancer morbidity and mortality over the last few decades, adherence to cancer screening guidelines remains inconsistent, particularly among certain demographic groups. This study aims to validate a rule-based algorithm to determine adherence to cancer screening. A novel screening algorithm was applied to electronic health record (EHR) from an urban healthcare system in New York City to automatically determine adherence to national cancer screening guidelines for patients deemed eligible for screening. First, a subset of patients was randomly selected from the EHR and their data were exported in a de-identified manner for manual review of screening adherence by two teams of human reviewers. Interrater reliability for manual review was calculated using Cohen's Kappa and found to be high in all instances. The sensitivity and specificity of the algorithm was calculated by comparing the algorithm to the final manual dataset. When assessing cancer screening adherence, the algorithm performed with a high sensitivity (79%, 70%, 80%) and specificity (92%, 99%, 97%) for colorectal cancer, breast cancer, and cervical cancer screenings, respectively. This study validates an algorithm that can effectively determine patient adherence to colorectal cancer, breast cancer, and cervical cancer screening guidelines. This design improves upon previous methods of algorithm validation by using computerized extraction of essential components of patients' EHRs and by using de-identified data for manual review. Use of the described algorithm could allow for more precise and efficient allocation of public health resources to improve cancer screening rates.

1. Background

Cancer screening programs in the United States have helped to increase cancer detection and decrease cancer mortality over the last several decades (Schiffman et al., 2015; Musa et al., 2017; Siu, 2016; Curry et al., 2018; Bibbins-Domingo et al., 2016). The U.S. Preventive Service Task Force Recommendations provide guidelines for cancer screening, which vary by the procedure's invasiveness, time, and cost based on cancer screening type (Siu, 2016; Curry et al., 2018; Bibbins-Domingo et al., 2016; Vodicka et al., 2017; Sun et al., 2018; Lew et al., 2017). Colorectal cancer screening (including colonoscopy, CT colonography, flexible sigmoidoscopy, fecal occult blood test, and fecal

immunochemical test), breast cancer screening (mammography), and cervical cancer screening (Papanicolaou [Pap] smears and/or high-risk human papillomavirus [hrHPV] testing) have been extensively investigated, and their impact on mortality has been documented (Schiffman et al., 2015; Musa et al., 2017; Siu, 2016; Curry et al., 2018; Bibbins-Domingo et al., 2016). Despite these known benefits, only 50–60% of eligible adults in the U.S. receive the recommended colorectal cancer screening, while breast cancer and cervical cancer screening rates are slightly higher, at 70% and 80%, respectively (Hall et al., 2018).

Some subsets of patients have lower rates of cancer screening than others, contributing to increased mortality from cancers for which early detection can lead to better outcomes (Inadomi et al., 2012; Nelson

et al., 2009). Several studies have found that men are less likely to participate in colorectal cancer screening than women (Hall et al., 2018; Inadomi et al., 2012). Others have demonstrated that, among racial groups, African Americans show the lowest rates of colorectal cancer screening while whites are most likely to undergo colonoscopies (Hall et al., 2018; Inadomi et al., 2012). For females, smoking, high BMI, reduced healthcare access, and lack of insurance are associated with decreased breast cancer and cervical cancer screening adherence (Nelson et al., 2009; Wagholikar et al., 2012; Qureshi et al., 2000). Significantly, more than half of women diagnosed with cervical cancer have inadequate screening (Wagholikar et al., 2012). Efforts to increase screening rates, particularly in underserved populations, are effective and prudent public health measures that can and should be taken to reduce unnecessary cancer morbidity and mortality (Kietzman et al., 2019).

Interventions using electronic health records (EHR) are effective at identifying individual patients in need of cancer screening, ultimately leading to increased screening adherence among eligible adults (Wagholikar et al., 2012; Hsiang et al., 2019; Green et al., 2013). Computerized clinical decision support systems (CDSS) are programs that can be trained using medical records and lab test results to offer clinical suggestions (Wagholikar et al., 2012). Such programs can have a great impact on cancer screening by providing physicians with individualized cancer screening reminders and recommendations within the EHR (Wagholikar et al., 2012). Indeed, several studies have demonstrated that EHR-based interventions help prompt and guide physicians in implementing evidence-based care, ultimately increasing patient adherence with cancer screening guidelines (Hsiang et al., 2019; Green et al., 2013; Coronado et al., 2014). However, many studies that evaluate the effectiveness of CDSS cite key limitations, including lack of data standardization across EHRs and difficulty acquiring patient data from outside hospital systems (Hsiang et al., 2019; Green et al., 2013; Coronado et al., 2014; Petrik et al., 2016). Given these limitations, the medical community would likely benefit from research validating the use of standardized EHR algorithms with the potential to be applied across institutional and geographic boundaries for the purposes of disease prevention, screening, and treatment. Of note, such research can be applied to meaningful public health work.

Extraction of data from within an EHR for purposes of validating an algorithm typically requires extensive manual chart review. These reviews are often time consuming and require certified reviewers who have been trained to use software that, in many instances, is not user friendly (Panacek, 2007; Vassar and Holzmann, 2013). Separately, bias in medicine and clinical decision-making—particularly in regard to race and socioeconomic status—has been well documented (Williams et al., 2015; Chapman et al., 2013). Full EHR reviews, which contain patient demographic information including but not limited to race, employment status, and insurance status may introduce unnecessary bias and resultant reviewer error. Extraction of relevant EHR data in an automated, de-identified manner, however, can avoid both bias and the laborious full chart review process. Our study aims to use de-identified manual review to validate a rule-based algorithm that determines adherence to cancer screening guidelines using EHR patient data.

2. Materials and methods

2.1. Study population and data collection

We utilized EHR data extracted from the Epic platform from New York University Langone Health (NYULH). A subset of NYULH patients were selected by conducting a simple random sample among patients who qualified for relevant cancer screening by age. In total, 305, 298, and 300 patient charts were selected for each cancer screening type (colorectal cancer, breast cancer, and cervical cancer respectively) to be reviewed. This sample size has been previously validated for these types of studies; power analysis additionally found that these sample sizes give

a 95% credibility interval (Kadhim-Saleh et al., 2013). These selected patients were managed using REDCap (Research Electronic Data Capture) electronic data capture tools hosted at NYULH (Chapman et al., 2013; Harris et al., 2009). REDCap is a secure, web-based software platform designed to support data capture for research studies, providing 1) an intuitive interface for validated data capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for data integration and interoperability with external sources (Harris et al., 2009; Harris et al., 2019). Our REDCap project contained only the information on patients included in the study; all information was extracted from the EHR, de-identified to remove protected health information, and subsequently imported into REDCap (Appendix 2).

Within these selected charts, some patients were excluded because of their medical history. Patients who had a total colectomy or who were previously diagnosed with colorectal cancer were considered ineligible for colorectal cancer screening (Baker et al., 2015). For breast cancer, patients who had a double mastectomy, two unilateral mastectomies, or a prior diagnosis of breast cancer were considered ineligible (Kern et al., 2013). Patients who had received a hysterectomy, were HIV positive, or had been previously diagnosed with cervical cancer were considered ineligible for cervical cancer screening (Kern et al., 2013). Patients over age 65 were also excluded from cervical cancer analysis. Eligible patients were then determined to be either adherent or non-adherent to the cancer screening guidelines recommended by the U.S. Preventive Services Task Force (Table 1) using both a rule-based algorithm approach and manual review (Fig. 1).

A rule-based algorithm was developed to automatically classify patients' screening eligibility and adherence based on standardized codes found within the Epic Clarity database (Appendix 1). Patients' historical medical records were queried using Structured Query Language (SQL) to identify any relevant International Classification of Disease (ICD)-10-CM and ICD-9-CM diagnostic or procedure codes, Current Procedural Terminology (CPT) codes, Healthcare Common Procedure Coding System (HCPCS) codes, or Logical Observation Identifiers Names and Codes (LOINC) laboratory codes (Appendix 1). Among the individuals who were classified as eligible for screening, the rule-based algorithm classified patients as adherent if any cancer screening occurrences were identified that adhered to the criteria and timeframes of the U.S. Preventive Task Force recommendations (Table 1).

Table 1
Cancer screening guidelines for colorectal, breast, and cervical cancer screening according to the U.S. Preventive Services Task Force (3–5).

| Cancer Screening Type	Adherence Criteria
Colorectal Cancer	Patient 50–75 years old meeting any of the following criteria: Had at least 1 colonoscopy within 10 years of most recent encounter Had a ct colonography within 5 years of most recent encounter Had a flexible sigmoidoscopy within 5 years of most recent encounter Had a fecal occult blood test (fobt) within 1 year of most recent encounter Had a fecal immunochemical test (fit)-dna within 3 years of most recent encounter
Breast Cancer	Patient 50–74 years old that had a mammogram within 2 years of most recent encounter
Cervical Cancer	Patient 21–29 years old that had a cytology/pap smear within 3 years of most recent encounter Patient 30–65 years old meeting any of the following criteria: Had a cytology/Pap smear within 3 years of most recent encounter Had a high-risk HPV test within 5 years of most recent encounter Had a high-risk HPV test with a cytology/Pap smear within 5 years of most recent encounter
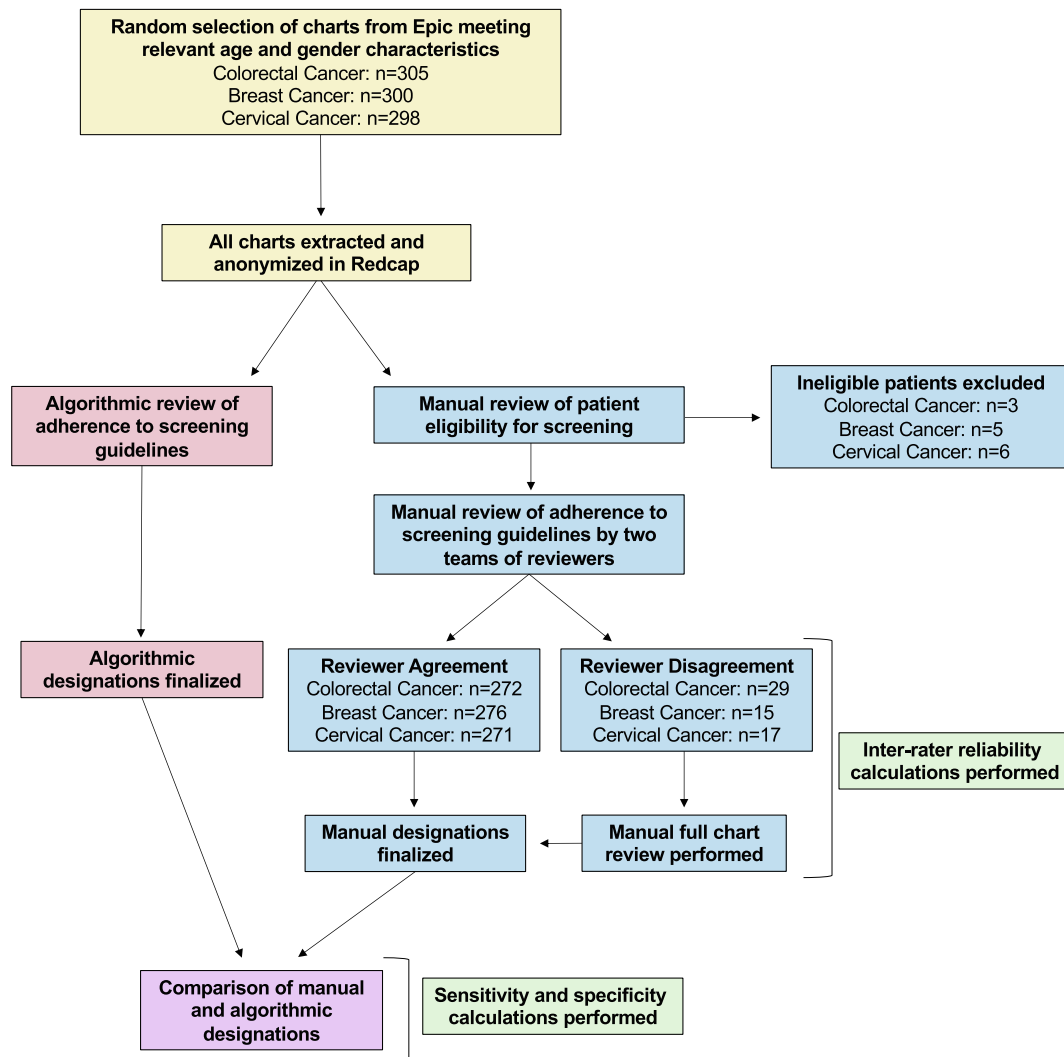
Fig. 1. Algorithm validation workflow. Diagram demonstrating the parallel workflows of the manual review teams and algorithm.

Two teams of trained manual reviewers each evaluated the de-identified dataset in REDCap to determine if the patients were eligible for cancer screening (Fig. 1). If they were eligible, the manual reviewers determined their adherence status. If manual reviewers were unable to make a designation or were unsure of their designation based on the de-identified data available in REDCap, the chart was designated for full review in Epic, after which a final determination about eligibility and adherence was made. Once both teams finished reviewing their datasets, all disagreements (i.e. discrepancies in eligibility and/or adherence designations) between the teams concerning both eligibility and adherence were identified. As necessary, a subsequent round of full chart reviews was performed to make final eligibility and adherence determinations for each conflict and to create a final manual review dataset to be compared directly to results produced by the rule-based algorithm. Individual adherence/non-adherence categories (e.g., colonoscopy vs. FIT testing) were not examined in detail because the algorithm did not distinguish adherence at this level. The datasets used to evaluate screening adherence included only patients who were deemed eligible by both manual review and the algorithm. This study was approved by the New York University School of Medicine Institutional Review Board and waived informed consent.

2.2. Statistical analysis

Within the original manual review datasets (i.e., before disagreements were resolved), we first assessed interrater reliability (IRR) between the two sets of manual raters through both percent agreement and calculation of Cohen's Kappa. The Cohen's Kappa for colorectal cancer was 0.723, for breast cancer was 0.892, and 0.867 for cervical cancer. These Cohen's Kappa values indicate there was a substantial or almost perfect amount of agreement for adherence designations (Kern et al., 2013). Within the final manual review dataset, we calculated the percentage of adherent patients for each type of cancer screening. Next, sensitivity and specificity of the algorithm were calculated by comparing algorithm results to manually determined results. For our study, sensitivity refers to the ability of the algorithm to correctly identify patients who were up to date with screening while specificity refers to the ability to correctly identify patients who were *not* up to date with screening. All analyses were conducted using SPSS 25 (IBM SPSS Statistics 25, Armonk NY).

3. Results

A total of 305 charts were considered for colorectal cancer screening; 3 (0.98%) were excluded due to their medical history. A total of 24.2% (73/302) of charts were found to be up to date with screening by manual review. The algorithm performed with a sensitivity of 79.5% (58/73; 95% CI 70.2%–88.7%) and a specificity of 91.7% (210/229; 95% CI 88.1%–95.3%) when assigning adherence status (Table 2). This means that the colorectal cancer screening algorithm correctly identifies

Table 2
Sensitivity and specificity of the algorithm for cancer screening adherence.

	Sensitivity (n, %, 95% CI)	Specificity (n, %, 95% CI)
Colorectal Cancer	58/73, 79.45%, 70.2%–88.7%	210/229 91.70% 88.1%–95.3%
Breast Cancer	79/113, 69.9%, 61.5%–78.4%	176/177, 99.4%, 98.3%–100%
Cervical Cancer	94/105, 89.5%, 83.7%–95.4%	180/186, 96.8%, 94.2%–99.3%

approximately 80% of patients who are adherent with colorectal cancer screening guidelines but misclassifies around 8% of patients who are not adherent as adherent.

A total of 298 charts were considered for breast cancer screening; 5 (1.7%) were excluded due to medical history. A total of 39.6% (116/293) of charts were subsequently determined to have demonstrated adherence to the breast cancer screening guidelines. For determinations of adherence, the algorithm displayed a sensitivity of 69.9% (79/113; 95% CI 61.5%–78.4%) and a specificity of 99.4% (176/177; 95% CI 98.3%–100%) (Table 2).

A total of 300 charts were considered for cervical cancer screening; 6 (2.0%) were excluded due to age and 3 (1.0%) were excluded due to their specific medical history. A total of 36.1% (105/291) of patients were ultimately determined to be adherent to the cervical cancer screening guidelines (Table 1). The algorithm performed with a sensitivity of 89.5% (94/105; 95% CI 83.7%–95.4%) and a specificity of 96.8% (180/186; 95% CI 94.2%–99.3%) when determining adherence to the screening guidelines in comparison with the manual reference standard (Table 2).

4. Discussion

Our study demonstrates that the previously described standardized algorithm for determining patient adherence to cancer screening guidelines is valid given the high sensitivities (79%, 70%, 90%) and specificities (92%, 99%, 97%) produced in colorectal cancer, breast cancer, and cervical cancer screening, respectively (Table 2). Although cancer screening has effectively reduced cancer mortality, patient adherence with these preventative measures ranges from 50 to 80% and is lower among certain groups (Schiffman et al., 2015; Musa et al., 2017; Hall et al., 2018; Inadomi et al., 2012; Nelson et al., 2009). EHRs have been used previously to increase individual patient adherence with cancer screening recommendations; however, these studies in creating EHR-based tools are generally time-consuming, require extensive reviewer training, and are potentially vulnerable to bias (Chapman et al., 2013; Antoniou et al., 2011; Wagholikar et al., 2012; Qureshi et al., 2000; Kietzman et al., 2019; Hsiang et al., 2019; Green et al., 2013; Coronado et al., 2014). The technique we have described for automated, de-identified data extraction prior to manual review of patient charts can decrease the time required for review and help decrease bias. The construction and validation of an algorithm that can accurately determine cancer screening adherence for large patient groups can also allow targeted public health efforts to improve cancer screening rates within particular neighborhoods or among specific groups.

In this study, we developed a rule-based algorithm that efficiently identified patient adherence to colorectal cancer, breast cancer, and cervical cancer screenings. In all instances, the algorithm was able to determine the adherence status with a high specificity and relatively high sensitivity. Interestingly, the algorithm had the lowest sensitivity when determining breast cancer screening adherence (70%) when compared to colorectal cancer (79%) and cervical cancer (90%) cancer screening, despite breast cancer screening guidelines having the simplest adherence criteria. This could be because the CPT and ICD codes selected for the SQL algorithm were not necessarily a universal list of mammogram codes, a procedure which may have more codes than

other screening modalities.

When identifying the adherence status of patients, for each cancer type, the algorithm performed with a higher specificity than sensitivity. Given the importance of identifying true negatives—individuals not adherent to the guidelines—to better focus public health efforts, we are inclined to accept the high specificity at the expense of sensitivity. We must accept that if we use this algorithm on EHRs to estimate screening rates in a given population, we will slightly underestimate the proportion of patients who are adherent because of the lower sensitivity. We believe that the high specificity of the algorithm we present in this study can accordingly aid in the efficient allocation of public health resources.

Our study demonstrates an efficient model to validate rule-based algorithms. Many studies have validated such algorithms; however, the initial extraction of the data in these studies was manual (Antoniou et al., 2011; Ginde et al., 2008; Thirumurthi et al., 2010). Our design improves upon these studies' methods by employing computerized deidentification of relevant data from the EHR into REDCap, thereby reducing the time to complete the manual review process and potentially reducing bias. This does, however, make the process of validation entirely dependent upon the accuracy of the transfer. Our comprehensive workflow (Fig. 1) allowed for full EHR chart reviews when manual reviewers were unsure of the designation based on the information in REDCap. These checks allowed us to intermittently assess the accuracy of the REDCap data in an observational manner while maintaining an efficient workflow.

Many studies have also demonstrated the utility of algorithms capable of searching an HER in order to identify patients with a given condition, or having undergone a specific treatment, or procedure (Wagholikar et al., 2012; Smischney et al., 2014; Amra et al., 2017; Barnes et al., 2020). Our study improves on these by performing similar analysis using limited, de-identified data digitally extracted from Epic into REDCap. A previous study demonstrated that algorithms used on limited, manually abstracted EHR data were able to identify individuals infected with HIV with a high sensitivity and specificity (Antoniou et al., 2011). However, their data extraction was manual, while ours was computerized (Antoniou et al., 2011). Computerized EHR abstractions have been used before; however, they have been limited by the number of variables collected as well as the number of patients involved in these studies (Kurdyak et al., 2015; Hu et al., 2017). Furthermore, many of these studies determined whether the patient had a given condition by examining only ICD codes (Ginde et al., 2008; Thirumurthi et al., 2010). Our study examined billings, orders, problem lists, and labs in addition to ICD codes. Our results indicate the capacity of our algorithm to quickly and accurately provide answers to simple inquiries on sets of limited data from large patient populations.

Despite the promising results of this study, there are several limitations. The sample population for this study only included NYULH patients who were randomly selected, including those who had only one encounter at NYULH. Because of this, some patients' cancer screening status may not be represented in their NYULH chart. Additionally, since we only tested the SQL algorithm on one hospital EHR system, the generalizability of this algorithm's abilities and performance statistics may be limited. There may be differences in the completeness and accuracy of EHR coding and documentation between hospitals, which could affect the ability of the algorithm to correctly identify individuals who are not up-to-date with screening recommendations. This algorithm also fails to capture individuals who are not currently within the health system or are not regularly receiving medical care. Further work would be needed to identify these vulnerable populations for cancer screening programs. Many of these described limitations, however, did not affect our ability to validate the algorithm itself. Instead they highlight key considerations for adopting this algorithm in the future to inform public health efforts.

We suggest that the algorithm validated in this study, when applied to a large multi-institutional de-identified data set, has the potential to inform the allocation of public health resources for robust cancer

screening. Currently, many public health efforts on the national, state, and even city levels are targeted based on race, ethnicity, socioeconomic status, or insurance status because many of these variables have been linked to lower cancer screening rates (Levano et al., 2014; Itzkowitz et al., 2016; Hall et al., 2012). While this effort is well intentioned, it may be inefficient because it may spread funding too thinly across communities without accounting for the various intersecting factors that contribute to cancer screening adherence in a subgroup of the general population. Acknowledging this, New York City has implemented several public health measures aimed specifically at high-risk neighborhoods by prevalence of preventable cancers as well as responses to a CDC national survey on cancer screening (Itzkowitz et al., 2016; Screening Amenable Cancers in New York State, 2020). These efforts, however, can be expensive, and the cost-effectiveness of many commonly used methods has yet to be studied in detail (Andersen et al., 2004). Our algorithm has the potential to provide the data to allow for more efficient targeting of communities and individuals within New York City and other urban areas. Given the amount of data that could be collected when applied to health systems across a densely populated city, our algorithm could allow for a reduction of cost on a large scale by ensuring resources are directed to those communities and hospitals where they are most needed.

5. Conclusions

Our study has validated an effective algorithm as well as a novel de-identified method of evaluating cancer screening adherence as documented within an EHR. We have also presented an efficient workflow for validating rule-based algorithms, albeit with a small initial dataset. With further data and analysis, our algorithm may eventually be used to highlight areas or populations with lower cancer screening rates. In the future, we aim to apply this algorithm to a larger clinical research network containing limited, de-identified data from academic institutions across New York City (Kaushal et al., 2014). We hope that applying the algorithm to such a database will allow us to create a multi-institutional EHR-based network to improve city-wide cancer screening and EHR-based surveillance. By validating our algorithm, we have taken the first step towards this goal.

6. Declarations

Ethics approval and consent to participate: The study was approved by the New York University School of Medicine Institutional Review Board and waived informed consent.

Consent for publication: Not applicable.

Availability of data and materials: Data are housed at NYU Langone Health due to personal health information.

Funding: This project was supported by Centers for Disease Control and Prevention. Prevention Research Program Special Interest Project 19-008, National Center for Chronic Disease Prevention and Health Promotion (New York University Cooperative Agreement Application # 3U48 DP006396-01S8).

Authors' contributions: SC conceptualized and designed the project. AJLM, SJR, JDK, WSS, and BO collected the data. AJLM, SJR, JDK, and WSS cleaned the data. WSS analyzed the data. AJLM, SJR, JDK, and WSS wrote the paper and created the tables and figures. BO and SC contributed to manuscript editing.

Disclosures: This project was supported by Centers for Disease Control and Prevention. Prevention Research Program Special Interest Project 19-008, National Center for Chronic Disease Prevention and Health Promotion (New York University Cooperative Agreement Application # 3U48 DP006396-01S8). The authors declare that they have no conflict of interest to disclose.

References

Amra, S., O'Horo, J.C., Singh, T.D., Wilson, G.A., Kashyap, R., Petersen, R., et al., 2017. Derivation and validation of the automated search algorithms to identify cognitive impairment and dementia in electronic health records. J. Crit. Care 37, 202–205. https://doi.org/10.1016/j.jcrc.2016.09.026.

Andersen, M.R., Urban, N., Ramsey, S., Briss, P.A., 2004. Examining the cost-effectiveness of cancer screening promotion. Cancer 101 (S5), 1229–1238. https://doi.org/10.1002/cncr.20511.

Antoniou, T., Zagorski, B., Loutfy, M.R., Strike, C., Glazier, R.H., 2011. Validation of case-finding algorithms derived from administrative data for identifying adults living with human immunodeficiency virus infection. PLoS ONE 6 (6), e21748. https://doi.org/10.1371/journal.pone.0021748.

Baker, D.W., Liss, D.T., Alperovitz-Bichell, K., Brown, T., Carroll, J.E., Crawford, P., et al., 2015. Colorectal cancer screening rates at community health centers that use electronic health records: a cross sectional study. J. Health Care Poor Underserved 26 (2), 377–390. https://doi.org/10.1353/hpu.2015.0030.

Barnes, D.E., Zhou, J., Walker, R.L., Larson, E.B., Lee, S.J., Boscardin, W.J., et al., 2020. Development and validation of eRADAR: a tool using EHR data to detect unrecognized dementia. J. Am. Geriatr. Soc. 68 (1), 103–111. https://doi.org/10.1111/jgs.16182.

Bibbins-Domingo, K., Grossman, D.C., Curry, S.J., Davidson, K.W., Epling Jr., J.W., García, F.A.R., et al., 2016. Screening for colorectal cancer: US preventive services task force recommendation statement. JAMA 315 (23), 2564–2575.

Chapman, E.N., Kaatz, A., Carnes, M., 2013. Physicians and implicit bias: how doctors may unwittingly perpetuate health care disparities. J. Gen. Int. Med. 28 (11), 1504–1510. https://doi.org/10.1007/s11606-013-2441-1.

Coronado, G.D., Vollmer, W.M., Petrik, A., Aguirre, J., Kapka, T., DeVoe, J., et al., 2014. Strategies and opportunities to STOP colon cancer in priority populations: pragmatic pilot study design and outcomes. BMC Cancer 14 (1), 55. https://doi.org/10.1186/1471-2407-14-55.

Curry, S.J., Krist, A.H., Owens, D.K., Barry, M.J., Caughey, A.B., Davidson, K.W., et al., 2018. Screening for cervical cancer: US preventive services task force recommendation statement. JAMA 320 (7), 674–686.

Ginde, A.A., Blanc, P.G., Lieberman, R.M., Camargo, C.A., 2008. Validation of ICD-9-CM coding algorithm for improved identification of hypoglycemia visits. BMC Endocr. Disord. 8 (1), 4. https://doi.org/10.1186/1472-6823-8-4.

Green, B.B., Wang, C.Y., Anderson, M.L., Chubak, J., Meenan, R.T., Vernon, S.W., et al., 2013. An automated intervention with stepped increases in support to increase uptake of colorectal cancer screening: a randomized trial. Ann Intern Med. 158 (5 Pt 1), 301–311. https://doi.org/10.7326/0003-4819-158-5-201303050-00002.

Hall IJ, Rim SH, Johnson-Turbes CA, Vanderpool R, Kamalu NN. The African American Women and Mass Media campaign: a CDC breast cancer screening project. Journal of women's health (2002). 2012;21(11):1107-13. doi:10.1089/jwh.2012.3903.

Hall, I.J., Tangka, F.K.L., Sabatino, S.A., Thompson, T.D., Graubard, B.I., Breen, N., 2018. Patterns and trends in cancer screening in the United States. Prev Chronic Dis. 15, E97. https://doi.org/10.5888/pcd15.170465.

Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., Conde, J.G., 2009. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. J. Biomed. Inform. 42 (2), 377–381.

Harris, P.A., Taylor, R., Minor, B.L., Elliott, V., Fernandez, M., O'Neal, L., et al., 2019. The REDCap consortium: building an international community of software platform partners. J. Biomed. Inform. 95, 103208 https://doi.org/10.1016/j.jbi.2019.103208.

Hsiang, E.Y., Mehta, S.J., Small, D.S., Rareshide, C.A.L., Snider, C.K., Day, S.C., et al., 2019. Association of an active choice intervention in the electronic health record directed to medical assistants with clinician ordering and patient completion of breast and colorectal cancer screening tests. JAMA Netw Open. 2 (11), e1915619 https://doi.org/10.1001/jamanetworkopen.2019.15619.

Hu, Z., Melton, G.B., Moeller, N.D., Arsoniadis, E.G., Wang, Y., Kwaan, M.R., et al., 2017. Accelerating chart review using automated methods on electronic health record data for postoperative complications. AMIA Annu. Symp. Proc. 2016, 1822–1831.

Inadomi, J.M., Vijan, S., Janz, N.K., Fagerlin, A., Thomas, J.P., Lin, Y.V., et al., 2012. Adherence to colorectal cancer screening: a randomized clinical trial of competing

strategies. Arch. Int. Med. 172 (7), 575–582. https://doi.org/10.1001/archinternmed.2012.332.

Itzkowitz, S.H., Winawer, S.J., Krauskopf, M., Carlesimo, M., Schnoll-Sussman, F.H., Huang, K., et al., 2016. New York citywide colon cancer control coalition: a public health effort to increase colon cancer screening and address health disparities. Cancer 122 (2), 269–277. https://doi.org/10.1002/cncr.29595.

Kadhim-Saleh, A., Green, M., Williamson, T., Hunter, D., Birtwhistle, R., 2013. Validation of the diagnostic algorithms for 5 chronic conditions in the Canadian Primary Care Sentinel Surveillance Network (CPCSSN): a Kingston Practice-based Research Network (PBRN) report. J Am Board Fam Med. 26 (2), 159–167. https://doi.org/10.3122/jabfm.2013.02.120183.

Kaushal, R., Hripcsak, G., Ascheim, D.D., Bloom, T., Campion Jr., T.R., Caplan, A.L., et al., 2014. Changing the research landscape: the new york city clinical data research network. J. Am. Med. Inform. Assoc. 21 (4), 587–590. https://doi.org/10.1136/amiajnl-2014-002764.

Kern, L.M., Malhotra, S., Barron, Y., et al., 2013. Accuracy of electronically reported "meaningful use" clinical quality measures: a cross-sectional study. Ann. Int. Med. 158 (2), 77–83.

Kietzman, K.G., Toy, P., Bravo, R.L., Duru, O.K., Wallace, S.P., 2019. Multisectoral collaborations to increase the use of recommended cancer screening and other clinical preventive services by older adults. Gerontologist. 59 (Suppl 1), S57–S66. https://doi.org/10.1093/geront/gnz004.

Kurdyak, P., Lin, E., Green, D., Vigod, S., 2015. Validation of a population-based algorithm to detect chronic psychotic illness. Can. J. Psychiatry 60 (8), 362–368. https://doi.org/10.1177/070674371506000805.

Levano W, Miller JW, Leonard B, Bellick L, Crane BE, Kennedy SK, et al. Public education and targeted outreach to underserved women through the National Breast and Cervical Cancer Early Detection Program. Cancer. 2014;120 Suppl 16(0 16):2591-6. doi:10.1002/cncr.28819.

Lew, J.-B., St John, D.J.B., Xu, X.-M., Greuter, M.J.E., Caruana, M., Cenin, D.R., et al., 2017. Long-term evaluation of benefits, harms, and cost-effectiveness of the National Bowel Cancer Screening Program in Australia: a modelling study. Lancet Public Health 2 (7), e331–e340. https://doi.org/10.1016/S2468-2667(17)30105-6.

Musa, J., Achenbach, C.J., O'Dwyer, L.C., Evans, C.T., McHugh, M., Hou, L., et al., 2017. Effect of cervical cancer education and provider recommendation for screening on screening rates: a systematic review and meta-analysis. PLoS ONE 12 (9), e0183924. https://doi.org/10.1371/journal.pone.0183924.

Nelson, W., Moser, R.P., Gaffey, A., Waldron, W., 2009. Adherence to cervical cancer screening guidelines for U.S. women aged 25–64: data from the 2005 health Information National Trends Survey (HINTS). J. Womens Health (Larchmt) 18 (11), 1759–1768. https://doi.org/10.1089/jwh.2009.1430.

Panacek, E.A., 2007. Performing chart review studies. Air Med. J. 26 (5), 206–210. https://doi.org/10.1016/j.amj.2007.06.007.

Petrik, A.F., Green, B.B., Vollmer, W.M., Le, T., Bachman, B., Keast, E., et al., 2016. The validation of electronic health records in accurately identifying patients eligible for colorectal cancer screening in safety net clinics. Fam. Pract. 33 (6), 639–643. https://doi.org/10.1093/fampra/cmw065.

Qureshi, M., Thacker, H.L., Litaker, D.G., Kippes, C., 2000. Differences in breast cancer screening rates: an issue of ethnicity or socioeconomics? J. Womens Health Gend. Based Med. 9 (9), 1025–1031.

Schiffman, J.D., Fisher, P.G., Gibbs, P., 2015. Early detection of cancer: past, present, and future. Am. Soc. Clin. Oncol. Educ. Book 35, 57–65. https://doi.org/10.14694/EdBook_AM.2015.35.57.

Screening Amenable Cancers in New York State, 2011-2015 New York Department of Health. 2020. https://www.health.ny.gov/statistics/cancer/docs/2019_screening_amenable_cancers.pdf.

Siu, A.L., 2016. Screening for breast cancer: U.S. preventive services task force recommendation statement. Ann Intern Med. 164 (4), 279–296.

Smischney, N.J., Velagapudi, V.M., Onigkeit, J.A., Pickering, B.W., Herasevich, V., Kashyap, R., 2014. Derivation and validation of a search algorithm to retrospectively identify mechanical ventilation initiation in the intensive care unit. BMC Med. Inform. Decis. Mak. 14, 55. https://doi.org/10.1186/1472-6947-14-55.

Sun, L., Legood, R., Sadique, Z., Dos-Santos-Silva, I., Yang, L., 2018. Cost-effectiveness of risk-based breast cancer screening programme, China. Bull. World Health Organ. 96 (8), 568–577. https://doi.org/10.2471/BLT.18.207944.

Thirumurthi, S., Chowdhury, R., Richardson, P., Abraham, N.S., 2010. Validation of ICD-9-CM diagnostic codes for inflammatory bowel disease among veterans. Dig. Dis. Sci. 55 (9), 2592–2598. https://doi.org/10.1007/s10620-009-1074-z.

Vassar, M., Holzmann, M., 2013. The retrospective chart review: important methodological considerations. J. Educ. Eval. Health Prof. 10, 12. https://doi.org/10.3352/jeehp.2013.10.12.

Vodicka, E.L., Babigumira, J.B., Mann, M.R., Kosgei, R.J., Lee, F., Mugo, N.R., et al., 2017. Costs of integrating cervical cancer screening at an HIV clinic in Kenya. Int. J. Gynecol. Obstetr. 136 (2), 220–228. https://doi.org/10.1002/ijgo.12025.

Wagholikar, K.B., MacLaughlin, K.L., Henry, M.R., Greenes, R.A., Hankey, R.A., Liu, H., et al., 2012. Clinical decision support with automated text processing for cervical cancer screening. J. Am. Med. Inform. Assoc. 19 (5), 833–839. https://doi.org/10.1136/amiajnl-2012-000820.

Williams, R.L., Romney, C., Kano, M., Wright, R., Skipper, B., Getrich, C.M., et al., 2015. Racial, gender, and socioeconomic status bias in senior medical student clinical decision-making: a national survey. J. Gen. Int. Med. 30 (6), 758–767. https://doi.org/10.1007/s11606-014-3168-3.