HHS Public Access

Author manuscript

J Registry Manag. Author manuscript; available in PMC 2020 July 01.

Published in final edited form as:

J Registry Manag. 2019 ; 46(2): 30–36.

Utility of Using Cancer Registry Data to Identify Patients for Tobacco Treatment Trials

Paul Krebs, PhD[1], **Erin Rogers, DrPH**[1], **Amanda Greenspan, MSc**[1], **Keith Goldfeld, DrPH**[1], **Lei Lei, MA**[1], **Jamie S Ostroff, PhD**[2], **Bridgette E. Garrett, PhD**[3], **Behnoosh Momin, DrPH**[4], **S. Jane Henley, MSPH**[4]

[1]NYU School of Medicine, New York, NY;

[2]Memorial Sloan Kettering Cancer Center, New York, NY;

[3]Office on Smoking and Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA;

[4]Division of Cancer Prevention and Control, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA

Abstract

Background: Many tobacco dependent cancer survivors continue to smoke after diagnosis and treatment. This study investigated the extent to which hospital-based cancer registries could be used to identify smokers in order to offer them assistance in quitting. The concordance of tobacco use coded in the registry was compared with tobacco use as coded in the accompanying Electronic Health Records (EHRs).

Methods: We gathered data from three hospital-based cancer registries in New York City during June 2014 to December 2016. For each patient identified as a current combustible tobacco user in the cancer registries, we abstracted tobacco use data from their EHR to independently code and corroborate smoking status. We calculated the proportion of current smokers, former smokers, and never smokers as indicated in the EHR for the hospitals, cancer site, cancer stage, and sex. We used a logistic regression model to estimate the log odds of the registry-based smoking status correctly predicting the EHR-based smoking status.

Results: Agreement in current smoking status between the registry-based smoking status and the EHR-based smoking status was 65%, 71%, and 90% at the three participating hospitals. Logistic regression results indicated that agreement in smoking status between the registry and the EHRs varied by hospital, cancer type, and stage, but not by age and sex.

Conclusions: The utility of using tobacco use data in cancer registries for population-based tobacco treatment interventions is dependent on multiple factors including accurate entry into EHR systems, updated data, and consistent smoking status definitions and registry coding protocols. Our study found that accuracy varied across the three hospitals and may not be able to inform interventions at these hospitals at this time. Several changes may be needed to improve the coding of tobacco use status in EHRs and registries.

Introduction

Tobacco use is the leading preventable cause of illness and death in the United States.[1,2] Smoking cessation even after cancer diagnosis has many benefits including improving efficacy of cancer treatments, reducing treatment- and cancer-related symptoms, and reducing the impact of other diseases.[3–7] Cancer survivors who quit smoking experience lower risks of dying from their cancer or of developing subsequent cancers compared with those who continue to smoke.[2] However, many cancer survivors continue to smoke, and promoting smoking cessation among cancer survivors is well-established as an indicator of quality cancer care.[8]

Beyond medical-legal requirements, electronic health records (EHRs) provide opportunities to enhance patient care, embed performance measures in clinical practice and facilitate clinical research through identification of patients who might benefit from and be eligible for behavioral and biomedical intervention trials. Identifying cohorts of at-risk patients and creating patient registries with linkage to patient contact information readily enables real-time targeting of evidence-based cancer prevention and control interventions.[9] Databases for identifying at-risk populations and delivering population-based interventions include electronic patient registries (e.g., cancer registries) and EHRs, which allow for real-time access to health data and patient contact information.[9] The use of EHRs to guide and deliver population-based interventions, however, can be challenging because of the variety of EHR platforms as well as lack of interoperability, accuracy, and harmonization of patient data. For example, proactive tobacco use treatment approaches that identify current tobacco users and directly offer treatment for tobacco dependence have been shown to increase the use of tobacco use treatment and long-term abstinence in primary care and mental health populations.[10,11] The use of registries and EHRs to guide and deliver population-based interventions can be challenging because of the variety of electronic platforms available, as well as inconsistency in registry or EHR implementation, even when using the same platform.

The Health Information Technology for Economic and Clinical Health (HITECH) Act,[12] the Patient Protection and Affordable Care Act (ACA) and the U.S. Public Health Service have provided guidelines to systematically collect data on key clinical outcomes, with the ACA enabling financial incentives for health care providers to ask patients about tobacco use during clinical encounters and document patient responses in the EHR. For instance, HITECH[7] has a recommended EHR outline for capturing tobacco use organized into five categories: history, assessment,[13] plan (quit date/counseling), pharmacotherapy, and a follow up plan. Since these categories are not mandatory fields in the recommended EHR outline and may be overlooked by providers, the frequency and quality of the data gathering are unknown.[9] Some investigations have begun to examine the effect of variation in EHR platform use and patient outcomes.[9,14] For example, Bae and colleagues reviewed the effects of using basic and advanced EHRs to document smoking status, delivery of smoking cessation counseling, and smoking cessation medication recommendations.[9] They concluded that “more sophisticated EHRs are associated with better smoking cessation support by physicians.”[9] Additionally, they recognized the need for an analysis of EHR platforms

and functions and that future meaningful use incentive programs need explicit coding for documentation of tobacco use and follow ups.[9]

While all states and the District of Columbia have central cancer registries (to which hospital-based cancer registries and other health care facilities send data on cancer diagnoses and related variables), tobacco use variables are not currently required by the North American Association of Central Cancer Registries to be collected.[15] Some central cancer registries have continued to collect tobacco use variables, but not in a standardized format. A recent assessment of tobacco use data collected from 10 National Program of Cancer Registries (NPCR) registries funded for an enhanced data collection project concluded that "studies to evaluate the validity of specific tobacco-related variables and the ability of cancer registries to capture this information from the medical record are needed."[16] Addressing barriers for this identification and accurate documentation of tobacco use in the EHR is essential because proactive tobacco dependence treatment approaches that identify smokers and directly offer them cessation treatment for tobacco use have been shown to increase its use and long-term abstinence in primary care and mental health populations.[16]

This study draws from the preliminary work done in preparation for a tobacco dependence treatment clinical trial, the aim of which was to assess the feasibility and effectiveness of using hospital cancer registry-based tobacco use status to identify and offer tobacco cessation treatment to smokers with a recent diagnosis of cancer. The sampling and recruitment plan called for use of hospital cancer registry data to identify a cohort of eligible tobacco dependent cancer survivors. The current study has two primary aims: 1) to describe how the EHRs used at three large cancer care settings and corresponding cancer registries coded for tobacco use status, and 2) to compare and examine extent of agreement of tobacco use data coded in the cancer registries to that coded in the patient's EHR.

Methods

Study population.

We worked with cancer registry programmers at three hospital-based cancer treatment settings in New York City (described below) to identify our study sample using cancer staging, administrative, and treatment date codes contained in the cancer registry. These study locations were chosen as they have affiliations with the primary project location (New York University Langone Health). The locations encompassed a National Cancer Institute designated comprehensive cancer center (New York University Perlmutter Cancer Center), Bellevue Hospital, the flagship public hospital in New York City, and the VA New York Harbor, the Veterans Administration hospital encompassing Manhattan and Brooklyn. Each hospital system used different EHR systems and maintains their own independent cancer registry. Eligibility requirements were: a) currently smoked a tobacco product in the past 30 days. b) any cancer diagnosis of less than stage 3B in the previous two years (June 2014 to December 2016) and c) no diagnosis of dementia. We used pathological staging codes as the primary source of staging data and used clinical staging codes when pathological staging was not available.[17]

Tobacco use status from cancer registry.

We identified tobacco users using an administrative code for tobacco status available in the registry. Each case in the cancer registry was categorized using New York state-mandated codes: 1) never used, 2) current cigarette smoker, 3) current cigar/pipe smoker, 4) current snuff/chew/smokeless tobacco user, 5) mixed use, 6) former smoker, or 7) unknown. We selected all patients with codes 2, 3 or 5 as representing patients currently using combustible forms of tobacco. Non-combustible tobacco use (chewing tobacco, snuff or other smokeless tobacco) and former smoker were not a target of the cessation intervention and therefore were not included in the sampling plan. Use of electronic cigarettes also was not included.

Tobacco use abstraction from EHR.

For each patient identified as a current combustible tobacco smoker in the cancer registries, we abstracted tobacco use data from their EHR to independently code smoking status. Each hospital EHR documented tobacco use differently (see Table 1). The hospital that used the EPIC EHR system[18] documented smoking status as a variable in the substance use history. This EHR platform enabled more granular fields to categorize patients as current every day smoker, current some day smoker, former smoker, heavy smoker, light smoker, never assessed, never smoker, passive smoke exposure- never smoker, smoker-current status unknown, and unknown if ever smoked. The VA and Bellevue clinics documented smoking status and history under social history in provider notes. While the VA creates tobacco use reminders that providers must complete every 6 months in the Computerized Patient Record System (CPRS) EHR, these data are not able to be seen by provider-users of the EHR and only available via a group data query. The Bellevue system (Misys) has neither clinical reminders nor tobacco use history available as a provider-facing variable. For all clinics, we reviewed provider notes within each patient's EHR to confirm that the listed smoking status was accurate and updated. The most recent smoking status was used to code patients as never, former, or current smokers. When no smoking history information could be found, patients were coded as never smokers, consistent with registry practice. For current and former smokers, we extracted years of smoking and the most recent 'cigarettes per day' information when available. When 'pack per day' was used in a patient's EHR instead of 'cigarettes per day', we calculated 'cigarettes per day' based on a standard pack with 20 cigarettes. We also included date of quitting for former smokers.

Data Analysis.

Descriptive statistics were calculated for each hospital to assess the number of current combustible tobacco users identified using the registry for all cancer patients diagnosed during 2014–2016, to determine the availability of detailed tobacco use data in the EHR for these patients, and to compare the coding of registry-based smoking status to that from the EHR. We calculated the proportion of current smokers, former smokers, and never smokers as indicated in the EHR for the hospitals, cancer site, cancer stage, and sex. We used a logistic regression model to estimate the log odds of the registry-based cigarette smoking status correctly predicting the EHR-based cigarette smoking status. The dependent variable in the model was an indicator of the prediction status; this indicator had a value of 1 if the smoking status in the EHR was "current smoker" and a value of 0 if the smoking status

was either “former smoker” or “never smoker”. The model adjusted for hospital, cancer site, cancer stage, sex, and age. We did not conduct sensitivity analyses as it was not feasible to check registry patients who did not have an indication of tobacco use against the EHRs. Analyses were conducted using R 3.4.3.[19]

Results

Current tobacco use (inclusive of cigarettes and cigar/pipe), at each of the three study locations using cancer registry data is reported in Table 2, and ranged from 6.3% at NYU, 7.5% at Bellevue, and 19.1% at the VA.

As shown in Table 3, agreement in current cigarette smoking status between the registry-based smoking status and the EHR-based cigarette smoking status ranged from 65% at the VA to 71% at NYU and 90% at Bellevue. About 25% of patients identified by the NYU and VA cancer registries as current smokers were classified by their EHRs as former smokers. A small fraction (4% to 9%) of patients identified as current smokers in the cancer registries were classified as never smokers in the EHRs. The overall positive predictive value aggregating across clinics was 72.2% (95% CI: 70.9%, 73.4%) (data not shown in table).

Logistic regression results indicated that agreement in cigarette smoking status between the registry and the EHR varied by hospital, cancer type, and stage, but not by age and sex (Table 4). Agreement in smoking status between the registry and the EHR was more likely at Bellevue than at NYU or the VA and among persons with endocrine (OR = 2.79, 95% CI: 1.15–6.23), genitourinary (OR = 1.99, 95% CI: 1.20–3.29), and hematologic malignancies (OR = 2.10, 95% CI: 1.04–4.25) compared to those with lung cancer, as well as for those with stage I cancer compared to stage 0 (OR = 1.96, 95% CI: 1.07–3.58). Age was not associated with agreement in current cigarette smoking status (OR = 1.01, 95% CI: 1.00–1.03). Classification accuracy did not differ by sex.

The Bellevue EHR had the least amount of data about cigarettes per day (32.1%) and years smoked (22.8%). NYU had data available for about half of their patients, while VA had cigarettes per day for 55% of patients, but we could not find years smoked for any of the registry-identified smokers.

Discussion

This is the first study to compare smoking status recorded in hospital cancer registries with that recorded in EHRs. Using tobacco use data recorded in cancer registries could be a potential way to identify cancer survivors who smoked at the time of diagnosis. However, because information about tobacco use in cancer registries has not been routinely collected, questions remain about the quality, validity, and usefulness of these data.

We found that while most patients classified as current smokers in the cancer registry were also classified as current smokers in the EHRs, about a quarter were classified as former smokers and a small fraction as never smokers. The largest discrepancies occurred when the registry categorizes patients as current smokers while the EHRs categorizes them as former smokers. Smoking status was abstracted from the EHR 6 to 24 months after it was

recorded in the cancer registry so it is possible that patients could have quit smoking during that interval. The discrepancies also could be due to differences in the way that the tobacco use questions were worded, differences in the way the patient responded to tobacco use questions, or differences in the way that responses were recorded.

Our study found that there was inconsistency in how tobacco use was recorded across three EHR systems. Only one EHR system (EPIC at NYU) included tobacco use as a user-facing variable; it also included the number of cigarettes smoked per day and the number of years smoked as variables. Thus, rates were highest for this information compared to Bellevue and the VA, which did not have variables for these details. The VA system prompts providers to assess tobacco use yearly, but does not allow users to see previous answers, and the Misys system at Bellevue had no tobacco prompts or standardized fields. In terms of agreement with registry data, agreement with current EHR information ranged from 65% to 90% for current cigarette use. Missing data rates for cigarettes per day and years smoked were high as systems did not have standardized or required fields for these data points.

Examining the EHR systems suggests reasons for such variability. Even though Bellevue Mysis did not have tobacco use prompts, the agreement between EHR and registry data was highest at 90%. This could be attributable to the fact that tobacco use data is recorded only in clinical notes, which are able to be viewed by the registry. The lowest rate of agreement was at the VA, which prompts for and records tobacco use, but does not allow this information to be seen by EHR users, including registry abstractors. Thus registry can only record it if providers describe tobacco use in a clinical note. NYU's EPIC build makes tobacco use data available, but abstractors prioritize oncology notes over the Social History/ Smoking Status section, which is updated more frequently.

While this is the first study to compare EHR and cancer registry data, rates of recording smoking status in EHRs have been examined in previous studies, particularly in the lung cancer screening literature. A study of 4 lung cancer screening programs found that pack history and quit date could only be determined for 44% of those screened at one clinic and 44% at another, leading to considerable uncertainty about screening eligibility.[20] Another study found that EHRs underreported pack years in 85% of cases,[21] and the VA lung cancer screening demonstration project found inaccurate pack years in 39% of patients.[22] Another recent study surveyed 200 patients and compared survey data to EHR data for determining eligibility for lung cancer screening found that only 70% had complete data for smoking status, years smoked, and pack years, rates which were on par or higher than our best results, found in NYU's EPIC system.[23]

In terms of limitations, we were not able to conduct a full sensitivity analysis. Our screening was for the purposes of enrollment for a smoking cessation clinical trial and we did not have resources to examine false negatives in the registry, that is, patients who were listed as non-tobacco users in the registry but who were categorized as current smokers in the EHRs. Such data would help to elucidate the quality of registry coding for tobacco use. Additionally, sample sizes, and thus reliability of estimates differed across clinics. Nevertheless, our data provides valuable insights for evaluating the utility of registry data and for suggesting improvements related to tobacco use.

A number of challenges exist in terms of ensuring accurate tobacco use coding and documentation in the EHR. First, no one health care provider or other staff may have sole workflow responsibility for entering tobacco use history, the format of questioning patients is not standardized across health systems and variations in categorizing tobacco use are not well-defined, which may lead to inaccurate data collection. For instance, asking someone “are you a smoker” rather than the recommended “have you used any tobacco product in the past 30 days” often resulted in a significantly different answer.[24] Defining tobacco use behaviorally is essential as it prevents misinterpretation about non-frequent and light smoking and avoids using stigmatizing terms (i.e. “smoker”) which people are apt to deny. Moreover, misreporting of current smoking status may be attributable to the stigma associated with current smoking, particularly for cancer patients making it even more important to ask smoking status questions in an empathic and non-ambiguous manner. Second, the tobacco use categories and data entry fields are not standardized across EHR systems (e.g. “smoker” vs. “used tobacco product in past 30 days”), or do not exist as structured fields at all resulting in interoperability challenges. Third, delays from diagnosis to data coding into a registry has been noted as a barrier to using a registry for actionable decision making as patient smoking status can change during this time.[25] Finally, without a defined field in an EHR for tobacco use, registry abstraction is difficult and left to each registrar to develop their own policy regarding where to look for tobacco use history data.

Improvements are needed at multiple levels from provider to EHR system design to increase data accuracy, perhaps the least burdensome for providers would be the patient updating social/medical history via an EHR-linked tablet prior to a visit. Such a system could readily adopt the recommendations of the National Cancer Institute-American Association for Cancer Research (NCI-AACR) Cancer Patient Tobacco Use Assessment Task Force, which has developed a systematic tobacco screening protocol, the Cancer Patient Tobacco Use Questionnaire (C-TUQ).[24] While intended for clinical trials, the structure of the questions provides essential data to inform a full range of tobacco-related trials and clinical services including brief cessation counseling (e.g., Ask-Advise-Refer), population health interventions, and automated eligible case finding and clinical reminders for lung cancer screening.

EHRs offer the functionality for systematically assessing and documenting tobacco use, yet, as demonstrated by this study, problems remain in implementing reminders within the systems, quality of how tobacco use is assessed by the healthcare team, availability of tobacco use data, and comprehensiveness of the tobacco use history. These limitations may hamper the ability of health systems to support tobacco cessation efforts via direct care by healthcare providers as well as through use of proactive, population-based tobacco cessation programs.

Strategies are underway to improve screening and recording of tobacco use in EHRs that would lead to increased accuracy and usability for population health interventions, namely modifying the tobacco use screening process, and improved technologies. Raz et al found that improving the tobacco use screen was minimally burdensome and increased the quality of data.[26] Another method for improving the quality of tobacco use data could be to institute technology-enabled survey devices that link to EHRs where patients can complete patient

reported outcomes and routine health update questions, including tobacco use history, prior to their appointments, thereby standardizing the questions and ensuring accurate data entry.[27] While not widespread, some commercial systems are in place that allow patients to report data via tablet computer, and pilot studies have indicated acceptability on the part of patients and providers.[28,29] Researchers have tried machine learning to search clinical notes with moderate success.[30,31]

Because tobacco use data collected by cancer registries is dependent on the documentation of tobacco use in the medical record, the quality of these data may benefit from improving provider training in assessing and documenting tobacco use in the EHR. Providing patients with a rationale for asking about tobacco use and discussing current tobacco use in an empathic manner is likely to improve accurate patient reporting.[32] Documentation of tobacco use should be improved by educating health care providers about the importance and utility of tobacco use data and implementing organizational policies that encourage documenting tobacco use.

Tobacco use data in cancer registries is typically recorded at the time of diagnosis, and may not accurately reflect changes in tobacco status over the course of cancer treatment and survivorship. Our study results suggest registry-based tobacco use data does not appear to be sufficiently accurate at present to serve as a foundation for identifying eligible smokers for clinical trials and providing a referral to tobacco treatment services without greater attention to improving the quality of patient data capture. One has to consider available resources and weigh the benefits of offering cessation services to the cohort of current smokers versus reaching patients who may have stopped smoking (former smokers) since inclusion in the registry. For studies that need real-time tobacco status, registry-based tobacco status data may not be sufficient. However, registry-based tobacco status may be useful in examining the association between tobacco use at time of diagnosis and cancer outcomes and in making decisions about survivorship care.[16] Efforts are needed to improve the accuracy of tobacco use screening and data capture protocols in EHRs. Registrars could also increase attention to accuracy of tobacco use coding, which can better inform proactive methods for providing smokers with treatment for tobacco dependence.

Funding:

Support provided through a supplemental award to grant number U48-DP001909-05 to the New York University, City University of New York Health Promotion & Prevention Research Center (U48 DP005008-02S3) from the National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention. Findings reported here are from Special Interest Project (SIP) 15–003: Proactive Care Coordination for Cancer Survivors Who Smoke.

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.

References

1. Carter BD, Abnet CC, Feskanich D, et al. Smoking and mortality — beyond established causes. N Engl J Med 2015;372(7):631–640. doi:10.1056/NEJMsa1407211. [PubMed: 25671255]
2. U.S. Department of Health and Human Services. The health consequences of smoking: 50 years of progress. A report of the Surgeon General. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease

Prevention and Health Promotion, Office on Smoking and Health; 2014: https://www.cdc.gov/tobacco/data_statistics/sgr/50th-anniversary/index.htm

3. Browman GP, Wong G, Hodson I, et al. Influence of cigarette smoking on the efficacy of radiation therapy in head and neck cancer. N Engl J Med 1993;328(3):159–163. 10.1056/nejm199301213280302. [PubMed: 8417381]
4. Zhang Z, Xu F, Wang S, Li N, Wang C. Influence of smoking on histologic type and the efficacy of adjuvant chemotherapy in resected non-small cell lung cancer. Lung Cancer 2008;60(3):434–440. 10.1016/j.lungcan.2007.10.013. [PubMed: 18036699]
5. Fleshner N, Garland J, Moadel A, et al. Influence of smoking status on the disease-related outcomes of patients with tobacco-associated superficial transitional cell carcinoma of the bladder. Cancer 1999;86(11):2337–2345. [PubMed: 10590376]
6. Waller LL, Weaver KE, Petty WJ, Miller AA. Effects of continued tobacco use during treatment of lung cancer. Expert Rev Anticancer Ther 2010;10(10):1569–1575. 10.1586/era.10.140. [PubMed: 20942628]
7. Thomsen T, Tonnesen H, Moller AM. Effect of preoperative smoking cessation interventions on postoperative complications and smoking cessation. Br J Surg 2009;96(5):451–461. 10.1002/bjs.6591. [PubMed: 19358172]
8. Shields PG, Herbst RS, Arenberg D, et al. Smoking cessation, version 1.2016, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 2016;14(11):1430–1468. [PubMed: 27799513]
9. Bae J, Ford EW, Huerta TR. The electronic medical record's role in support of smoking cessation activities. Nicotine Tob Res 2016;18(5):1019–1024. 10.1093/ntr/ntv270. [PubMed: 26659415]
10. Fu SS, van Ryn M, Sherman SE, et al. Proactive tobacco treatment and population-level cessation: A pragmatic randomized clinical trial. JAMA Intern Med 2014;174(5):671–677. 10.1001/jamainternmed.2014.177. [PubMed: 24615217]
11. Rogers ES, Fu SS, Krebs P, et al. Proactive tobacco treatment for smokers using Veterans Administration mental health clinics. Am J Prev Med 2018;54(5):620–629. 10.1016/j.amepre.2018.02.011. [PubMed: 29551324]
12. Lawson PJ, Flocke SA, Casucci B. Development of an instrument to document the 5A's for smoking cessation. Am J Prev Med 2009;37(3):248–254. 10.1016/j.amepre.2009.04.027. [PubMed: 19666161]
13. U.S. Department of Health and Human Services. Treating tobacco use and dependence: 2008 update. Rockville (MD): U.S. Department of Health and Human Services, U.S. Public Health Service; 2008: https://www.ahrq.gov/professionals/clinicians-providers/guidelines-recommendations/tobacco/index.html.
14. Alkureishi MA, Lee WW, Webb S, Arora V. Integrating patient-centered electronic health record communication training into resident onboarding: curriculum development and post-implementation survey among housestaff. JMIR Med Educ 2018;4(1):e1. 10.2196/mededu.8976. [PubMed: 29301735]
15. Thornton M, ed NAACCR data dictionary. standards for cancer registries volume ii: data standards and data dictionary, twenty first edition. Springfield, Ill.: North American Association of Central Cancer Registries; 2018.
16. Siegel DA, Henley SJ, Wike JM, et al. Capture of tobacco use among population-based registries: Findings from 10 National Program of Cancer Registries states. Cancer 2018;124(11):2381–2389. doi:10.1002/cncr.31326. [PubMed: 29579317]
17. Edge S, Byrd D, Compton C, Fritz A, Greene F, Trotti A, eds. AJCC cancer staging maual 7th ed. New York, NY: Springer; 2010.
18. EPIC [computer program]. Verona, WI2018.
19. R Core Team. Vienna, Austria: R foundation for statistical computing, 2017.
20. Gould MK, Sakoda LC, Ritzwoller DP, et al. Monitoring lung cancer screening use and outcomes at four cancer research network sites. Ann Am Thorac Soc 2017;14(12):1827–1835. 10.1513/AnnalsATS.201703-237OC. [PubMed: 28683215]
21. Modin HE, Fathi JT, Gilbert CR, et al. Pack-year cigarette smoking history for determination of lung cancer screening eligibility. comparison of the electronic medical record versus a

shared decision-making conversation. Ann Am Thorac Soc 2017;14(8):1320–1325. 10.1513/AnnalsATS.201612-984OC. [PubMed: 28406708]

22. Kinsinger LS, Anderson C, Kim J, et al. Implementation of lung cancer screening in the Veterans Health Administration. JAMA Intern Med 2017;177(3):399–406. 10.1001/jamainternmed.2016.9022. [PubMed: 28135352]
23. Cole AM, Pflugeisen B, Schwartz MR, Miller SC. Cross sectional study to assess the accuracy of electronic health record data to identify patients in need of lung cancer screening. BMC Res Notes 2018;11(1):14. 10.1186/s13104-018-3124-0. [PubMed: 29321038]
24. Land SR, Toll BA, Moinpour CM, et al. Research priorities, measures, and recommendations for assessment of tobacco use in clinical cancer research. Clin Cancer Res 2016;22(8):1907–1913. 10.1158/1078-0432.ccr-16-0104. [PubMed: 26888828]
25. Jabour AM, Dixon BE, Jones JF, Haggstrom DA. Toward timely data for cancer research: assessment and reengineering of the cancer reporting process. JMIR Cancer 2018;4(1):e4. 10.2196/cancer.7515. [PubMed: 29496653]
26. Raz DJ, Dunham R, Tiep B, et al. Augmented meaningful use criteria to identify patients eligible for lung cancer screening. Ann Thorac Surg 2014;98(3):996–1002. 10.1016/j.athoracsur.2014.04.105. [PubMed: 25063304]
27. Zbrozek A, Hebert J, Gogates G, et al. Validation of electronic systems to collect patient-reported outcome (PRO) data-recommendations for clinical trial teams: report of the ISPOR ePRO systems validation good research practices task force. Value Health 2013;16(4):480–489. 10.1016/j.jval.2013.04.002. [PubMed: 23796281]
28. Girgis A, Durcinoska I, Levesque JV, et al. eHealth system for collecting and utilizing Patient Reported Outcome Measures for Personalized Treatment and Care (PROMPT-Care) among cancer patients: mixed methods approach to evaluate feasibility and acceptability. J Med Internet Res 2017;19(10):e330. 10.2196/jmir.8360. [PubMed: 28970188]
29. Wong D, Cao S, Ford H, et al. Exploring the use of tablet computer-based electronic data capture system to assess patient reported measures among patients with chronic kidney disease: a pilot study. BMC Nephrol 2017;18(1):356. 10.1186/s12882-017-0771-7. [PubMed: 29212466]
30. Onega T, Nutter E, Sargent J, Doherty J, Hassanpour S. Identifying patient smoking history for cessation and lung cancer screening through mining electronic health records. Cancer Epidemiol Biomarkers Prev 2017;26(3):437–437. 10.1158/1055-9965.epi-17-0032.
31. Bennett AV, Jensen RE, Basch E. Electronic patient-reported outcome systems in oncology clinical practice. CA Cancer J Clin 2012;62(5):337–347. 10.3322/caac.21150. [PubMed: 22811342]
32. Hamann HA, Ver Hoeve ES, Carter-Harris L, Studts JL, Ostroff JS. Multilevel opportunities to address lung cancer stigma across the cancer control continuum. J Thorac Oncol 2018;13(8):1062–1075. 10.1016/j.jtho.2018.05.014. [PubMed: 29800746]

Table 1.

Comparison of Tobacco Use Coding in 3 EHR Systems

Setting	EHR System	Tobacco Use Coding	Availability	Clinical reminder?
New York University Langone Health	EPIC	• current every day smoker • current some day smoker • former smoker • heavy smoker • light smoker • never assessed, never smoker • passive smoke exposure- never smoker • smoker-current status unknown • unknown if ever smoked • Cigarettes per day	Variable. in Social History	Yes. On intake
Bellevue	Mysis	None	Provider notes only	No
VA New York Harbor	CPRS	• Never • Former • Current	Variable, not user-facing; Provider notes	Yes. Every 6 months

Abbreviations: EHR: electronic health record

Table 2:

Current tobacco use rates (determined from cancer registry) among cancer patients who met study eligibility criteria from 2014–2016 in three New York City hospitals and availability of detailed data in EHR for current tobacco users

Setting	Patients N	Current tobacco users* N (%)	Data available for cigarettes/day N (%)	Data available for years smoked N (%)
New York University Langone Health	11,679	739 (6.3%)	422 (57.1%)	358 (48.4%)
Bellevue	1,461	109 (7.5%)	35 (32.1%)	27 (22.8%)
VA New York Harbor	839	160 (19.1%)	88 (55.0%)	Not available

Abbreviations: EHR: electronic health record

*Includes cigarettes and other tobacco use

Table 3:

Smoking status abstracted from the EHR among patients identified as current cigarette users using cancer registry data from three New York City hospitals 2014–2016 by select characteristics.

	Smoking Status in EHR		
	Smoker N=687	**Former Smoker N=215**	**Never Smoker N=50**
Hospital			
NYU (n=696)	494 (71%) *	170 (24%)	32 (5%)
Bellevue (n=108)	97 (90%) *	7 (6%)	4 (4%)
VA (n= 148)	96 (65%) *	38 (26%)	14 (9%)
Cancer Site			
Respiratory/Thoracic	84 (66%)	40 (31%)	4 (3%)
Head and Neck	21 (51%)	18 (44%)	2 (5%)
Breast	98 (77%)	27 (21%)	2 (2%)
Endocrine	49 (80%)	10 (16%)	2 (3%)
Gastrointestinal	102 (73%)	28 (20%)	9 (6%)
Genitourinary	159 (75%)	45 (21%)	9 (4%)
Hematologic	68 (74%)	17 (18%)	7 (8%)
Neurologic	43 (68%)	12 (19%)	8 (13%)
Skin	52 (73%)	13 (18%)	6 (8%)
Other	11 (65%)	5 (29%)	1 (6%)
Cancer Stage **			
0	43 (66%)	16 (25%)	6 (9%)
I	237 (76%)	62 (20%)	13 (4%)
II	140 (70%)	50 (25%)	9 (5%)
III	74 (73%)	25 (25%)	3 (3%)
IV	41 (63%)	19 (29%)	5 (8%)
unstaged	152 (73%)	43 (21%)	14 (7%)
Sex			
Female	279 (76%)	78 (21%)	12 (3%)
Male	408 (70%)	137 (23%)	38 (7%)

Abbreviations: EHR: electronic health record; NYU: New York University Langone Health; VA: VA New York Harbor

* The percentage of patients identified as current smokers using cancer registry data and EHR (ie, agreement in current smoking status between the registry-based smoking status and the EHR-based smoking status).

** Based on stage recorded in the EHR; cancers may have progressed from time of initial registry coding.

Table 4:

Logistic regression model of the cancer registry-based smoking status correctly predicting the EHR-based smoking status among cancer patients in 3 New York City Hospitals, 2014-2016

	OR	95% Lower Limit	95% Upper Limit
Hospital (vs. Bellevue)			
NYU	0.25	0.13	0.49 ***
VA	0.19	0.09	0.41 ***
Cancer Site (vs. lung)			
Head and Neck	0.54	0.26	1.13
Breast	1.88	0.99	3.55
Endocrine	2.79	1.25	6.23 *
Gastrointestinal	1.58	0.91	2.73
Genitourinary	1.99	1.20	3.29 **
Hematologic/Blood	2.10	1.04	4.25 *
Neurologic	1.48	0.70	3.12
Skin	1.84	0.93	3.62
Other	1.39	0.45	4.32
Cancer Stage (vs. 0)			
I	1.96	1.07	3.58 *
II	1.41	0.75	2.64
III	1.76	0.85	3.63
IV	0.91	0.40	2.06
unstaged	1.61	0.83	3.14
Male (vs. Female)	0.77	0.53	1.11
Age (per year)	1.01	1.00	1.03

Abbreviations: EHR: electronic health record; NYU: New York University Langone Health; OR: Odds Ratio; VA: VA New York Harbor

*** = $p < 0.001$,

** = $p < 0.01$,

* = $p > 0.05$