4-13-2016

Developing an Algorithm to Identify History of Cancer Using Electronic Medical Records

Christina L. Clarke
Kaiser Permanente Institute for Health Research, christina.l.clarke@kp.org

Heather S. Feigelson
Kaiser Permanente Institute for Health Research, heather.s.feigelson@kp.org

The Electronic Data Methods (EDM) Forum is supported by the Agency for Healthcare Research and Quality (AHRQ), Grant 1U18HS022789-01. eGEMs publications do not reflect the official views of AHRQ or the United States Department of Health and Human Services.

Developing an Algorithm to Identify History of Cancer Using Electronic Medical Records

Abstract

Introduction/Objective: The objective of this study was to develop an algorithm to identify Kaiser Permanente Colorado (KPCO) members with a history of cancer.

Background: Tumor registries are used with high precision to identify incident cancer, but are not designed to capture prevalent cancer within a population. We sought to identify a cohort of adults with no history of cancer, and thus, we could not rely solely on the tumor registry.

Methods: We included all KPCO members between the ages of 40-75 years who were continuously enrolled during 2013 (N=201,787). Data from the tumor registry, chemotherapy files, inpatient and outpatient claims were used to create an algorithm to identify members with a high likelihood of cancer. We validated the algorithm using chart review and calculated sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for occurrence of cancer.

Findings: The final version of the algorithm achieved a sensitivity of 100% and specificity of 84.6% for identifying cancer. If we relied on the tumor registry alone, 47% of those with a history of cancer would have been missed.

Discussion: Using the tumor registry alone to identify a cohort of patients with prior cancer is not sufficient. In the final version of the algorithm, the sensitivity and PPV were improved when a diagnosis code for cancer was required to accompany oncology visits or chemotherapy administration.

Conclusion: EMR data can be used effectively in combination with data from the tumor registry to identify health plan members with a history of cancer.

Acknowledgements

This work was funded by the National Cancer Institute (Contracts HHSN261201400644P and HHSN261201300460P). We thank LeeAnn Rohm, MSW and Kate Burniece, BS for completing the chart abstraction and Erica Blum-Barnett, MS for manuscript preparation.

Keywords

algorithm, electronic health record, cancer

Disciplines

Databases and Information Systems | Health Information Technology | Oncology | Theory and Algorithms

eGEMs
Generating Evidence & Methods to improve patient outcomes

Developing an Algorithm to Identify History of Cancer Using Electronic Medical Records

Christina L. Clarke, MS; Heather S. Feigelson, PhD, MPH[i]

ABSTRACT

Introduction/Objective: The objective of this study was to develop an algorithm to identify Kaiser Permanente Colorado (KPCO) members with a history of cancer.

Background: Tumor registries are used with high precision to identify incident cancer, but are not designed to capture prevalent cancer within a population. We sought to identify a cohort of adults with no history of cancer; thus, we could not rely solely on the tumor registry.

Methods: We included all KPCO members between the ages of 40–75 years who were continuously enrolled during 2013 (N=201,787). Data from the tumor registry, chemotherapy files, and inpatient and outpatient claims were used to create an algorithm to identify members with a high likelihood of cancer. We validated the algorithm using chart review and calculated sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for occurrence of cancer.

Findings: The final version of the algorithm achieved a sensitivity of 100 percent and specificity of 84.6 percent for identifying cancer. If we relied on the tumor registry alone, 47 percent of those with a history of cancer would have been missed.

Discussion: Using the tumor registry alone to identify a cohort of patients with prior cancer is not sufficient. In the final version of the algorithm, the sensitivity and PPV were improved when a diagnosis code for cancer was required to accompany oncology visits or chemotherapy administration.

Conclusion: Electronic medical record (EMR) data can be used effectively in combination with data from the tumor registry to identify health plan members with a history of cancer.

[i]Kaiser Permanente Institute for Health Research

Introduction and Objectives

Tumor registries are considered the "gold standard" for identifying incident cancer; that is, identifying new cancer cases within a defined population within a defined period.[1] The case finding procedures used by tumor registrars do not focus on the capture of prevalent cancer, or on indications of a history of cancer. We are interested in developing a prospective cohort to study the development of incident cancer over time. For such studies, it is important to begin with a population with no prior cancer history. In order to identify all cancers, the tumor registry may not be the only data source. The objective of this study was to develop an algorithm that used data available from the electronic medical record (EMR) that could identify individuals with a history of cancer, which included data sources to supplement the tumor registry.

Background

Health systems that capture clinical data in an EMR system find many ways to use these data to improve health care and answer important scientific questions.[2-7] Kaiser Permanente Colorado (KPCO) is an integrated health plan with EMR data collected over several decades, including information such as medication use, medical conditions, laboratory test results, disease onset, and subsequent treatment. When paired with a tumor registry, the EMR can be a powerful tool for conducting studies of cancer incidence and prognosis; however, EMR data are not without limitations. In particular, events occurring outside of the health plan are not well captured in most EMRs.

KPCO maintains a tumor registry dating back to 1989, and it is used reliably to identify incident cancer diagnosed and treated within KPCO.[1,2] Cancers diagnosed and treated outside of KPCO, usually prior to KPCO membership, are not followed in the tumor registry. Recently diagnosed cases may also be missed, because the tumor registry has a lag time of about 12 months. We sought to identify a cohort of adults from our current KPCO member population who had never been diagnosed with cancer. To accomplish this, we used our electronic data systems to develop an algorithm to identify individuals with a history of cancer, and validated the algorithm using manual chart reviews.

Methods

KPCO maintains an EMR for each of its members. Data from the EMR are collected into a virtual data warehouse (VDW), which contains content areas such as pharmacy (including chemotherapy), inpatient and outpatient claims, enrollment, and patient demographics.[8-10] We included all KPCO members between the ages of 40–75 years who were continuously enrolled during 2013. We used administrative and EMR data including the tumor registry, chemotherapy files, and inpatient and outpatient claims to identify those with a high likelihood of prior cancer. This project was reviewed and approved by the KPCO institutional review board (IRB). The requirement for informed consent was waived.

Our goal was to identify individuals with any prior cancer, with the exception of nonmelanoma skin cancer. Thus, for the initial algorithm, we cast a wide net to capture any possible incidence or history of cancer. Patients who ever had a behavior code indicating an in situ or invasive cancer from the tumor registry were flagged as having a history of cancer. We flagged anyone who ever had an inpatient or outpatient claim with an International Classification of Disease Ninth Edition (ICD-9) code indicating cancer. We also flagged any patient who had at least three visits in the oncology department on separate days, or at least two records on separate days of receiving a chemotherapeutic drug (specific codes are provided in Table 1), dating back to the beginning of our EMR in 1998.

Volume 4 (2016) Issue Number 1

Table 1. ICD-9 Codes Used to Flag Patients History or Incidence of Cancer

SOURCE	VERSION OF ALGORITHM	CODE TYPE	CODE/LOGIC
Tumor registry	1 and 2	Behavior	In situ or invasive
Inpatient and outpatient claims	1 and 2	ICD-9	Any code between 140 and 239 or V10.x, **Excluding:** 173.x, 199.1, 209.4x, 209.5x, 209.6x, 210.x-229.x, 232.x, 233.1, 238.2, 238.4, 238.7x, 238.9, 239, 239.1-239.5, 239.8-239.9 or V10.83
Inpatient and outpatient claims to oncology	1 Only	Encounters	≥3 visits on separate days to Oncology
Inpatient and outpatient claims indicating chemotherapy was given at least twice	1 Only	ICD-9	17.70, 99.25, 99.28, V58.11, V07.3, V07.39
		DRG	410, 492 if between years 1998 and 2007, or 837-839, 846-848 if between years 2008 and 2013
		HCPCS	A9600, A9604, C1086, C1166, C1167, C1178, C8953, C8954, C8955, C9004, C9012, C9110, C9205, C9207, C9213, C9214, C9215, C9235, C9257, C9262, C9265, C9414, C9415, C9417, C9418, C9419, C9420, C9421, C9422, C9423, C9424, C9425, C9426, C9427, C9429, C9431, C9432, C9433, C9434, C9437, C9438, C9440, G0355, G0357, G0358, G0359, G0360, G0361, G0362, G8372, G8373, G8374, G9021, G9022, G9023, G9024, G9025, G9026, G9027, G9028, G9029, G9030, G9031, G9032, J0490, J0594, J0894, J1094, J1100, J1190, J1457, J2323, J3262, J7150, J7527, J8510, J8520, J8521, J8530, J8540, J8560, J8561, J8562, J8565, J8600, J8610, J8700, J8705, J8999, J9000, J9001, J9002, J9010, J9015, J9017, J9019, J9020, J9025, J9027, J9033, J9035, J9040, J9041, J9042, J9045, J9050, J9055, J9060, J9062, J9065, J9070, J9080, J9090, J9091, J9092, J9093, J9094, J9095, J9096, J9097, J9098, J9100, J9110, J9120, J9130, J9140, J9150, J9151, J9165, J9170, J9171, J9178, J9180, J9181, J9182, J9185, J9190, J9200, J9201, J9206, J9207, J9208, J9211, J9230, J9245, J9250, J9260, J9261, J9263, J9264, J9265, J9266, J9268, J9270, J9280, J9290, J9291, J9293, J9300, J9302, J9303, J9305, J9307, J9310, J9315, J9320, J9328, J9330, J9340, J9350, J9351, J9355, J9357, J9360, J9370, J9375, J9380, J9390, J9600, J9999, Q0083, Q0084, Q0085, Q2017, Q2024, Q2049, S0087, S0088, S0115, S0116, S0172, S0176, S0178, S0182, S5019, S5020, S9329, S9330, S9331
		CPT-4	0519F, 36640, 4180F, 61517, 96400, 96401, 96402, 96405, 96406, 96408, 96409, 96410, 96411, 96412, 96413, 96414, 96415, 96416, 96417, 96420, 96422, 96423, 96425, 96440, 96445, 96446, 96450, 96542, 96545, 96549, C9287, J9043, J9179
		Revenue Codes	331, 332, 335

To test the algorithm, we conducted manual chart reviews, powered on specificity.[11-12] We randomly selected 297 charts from KPCO members who had any utilization in 2013. Of these, 69 patients were identified by the algorithm as having a history of cancer, and 228 patients were classified by the algorithm as cancer free. Cancer risk increases with age, so for the chart review we oversampled those older than the median age using a ratio of 2:1. We used a ratio of approximately 4:1 of individuals flagged as cancer free to those with cancer. Sampling 228 cancer-free individuals could detect an 80 percent specificity with a 95 percent confidence interval (CI) of 75 percent to 85 percent. Finally, we excluded cases from chart review that were flagged by the tumor registry—as we consider the tumor registry to be a validated source for identifying incident cancer, and our aim was to develop a method to identify cancers not included in the tumor registry. We used PROC SURVEYSELECT (SAS 9.2) to obtain a weighted random sample fitting the above criteria.

Each chart was fully reviewed to find any mention of cancer, or to confirm there was no history of cancer—using all notes in the chart dating back to the beginning of a patient's enrollment, or to the beginning of the EMR in 1998 (up to 15 years of medical utilization and history). The chart reviewer first looked for the exact date of diagnosis for patients who had been flagged with a history of cancer. If a diagnosis for cancer was not found, the reviewer then examined the chart from the administrative diagnosis date in the EMR going forward in time for a mention of cancer. Patients who were not identified as having cancer were fully reviewed from the start of the EMR forward. We did not review medical record data prior to 1998 (available in paper charts), as it is unlikely that information about cancer history would only be recorded prior to 1998 and not captured in EMRs spanning 1998–2013.

We calculated sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for occurrence of cancer, and used the chart review results to identify specific codes or logic patterns that could improve the performance of the algorithm.[13] CIs were calculated using the efficient-score method corrected for continuity.[11] Based on the results of the first chart review, we refined the algorithm, then conducted a second review of 200 novel charts, using the same sampling criteria specified above; except we selected those flagged as not having cancer in a ratio of 2:1 (137 no indication of cancer, 63 flagged with a history of cancer), and recalculated the aforementioned statistics. The second chart review of 137 charts for those who were flagged as cancer free could detect an 80 percent specificity with a 95 percent CI of 73 percent to 87 percent.

Findings

A total of 201,787 members met our initial inclusion criteria. The median age of this cohort as of January 1, 2013 was 56 years, 53.8 percent were female, and the average continuous enrollment time including 2013 was 8.75 years (standard deviation= 6.35 years). The initial algorithm flagged 25,824 (12.8 percent) people as having a history of cancer. Table 2 describes the number of people with a history of cancer identified by each "flag" specified in the initial algorithm. As we would expect, a large proportion of cases (n=11,410; 44.2 percent), were included in the tumor registry. Another 8,906 (34.5 percent) were identified with only a diagnosis code of cancer; 2,348 (9.1 percent) had only a receipt of chemotherapy; and having at least three visits to oncology accounted for 1,570 (6.1 percent) cases. The remaining 1,590 (6.2 percent) cases were a combination of the aforementioned categories.

Volume 4 (2016) Issue Number 1

Table 2. Results from Initial Algorithm Indicating How Patients Were Flagged as Having Cancer

INCLUSION CRITERIA*	NUMBER OF SUBJECTS	% NOT IN TUMOR REGISTRY/IN TUMOR REGISTRY	% OF TOTAL
NOT IN TUMOR REGISTRY (N= 14414 CASES)			
Chemotherapy only	2,348	16.3%	9.1%
Diagnosis only	8,906	61.8%	34.5%
Oncology visits only	1,570	10.9%	6.1%
Diagnosis and Chemotherapy	262	1.8%	1.0%
Oncology and Chemotherapy	324	2.2%	1.3%
Oncology and Diagnosis	773	5.4%	3.0%
Oncology, Diagnosis and Chemotherapy	231	1.6%	0.9%
TUMOR REGISTRY (N= 11410 CASES)			
Chemotherapy only	5	0.0%	0.0%
Diagnosis only	5,319	46.6%	20.6%
Oncology visits only	60	0.5%	0.2%
Diagnosis and Chemotherapy	306	2.7%	1.2%
Oncology and Chemotherapy	11	0.1%	0.0%
Oncology and Diagnosis	2,587	22.7%	10.0%
Oncology, Diagnosis and Chemotherapy	2,910	25.5%	11.3%
Tumor Registry Alone	212	1.9%	0.8%

Notes: *Patients were flagged as having a history of cancer if they were in the tumor registry, had at least 1 diagnosis of cancer, >=2 records of chemotherapy, or >=3 visits to oncology.

Using manual chart review as the gold standard, the sensitivity, specificity and NPV for a history of cancer from the first iteration of the algorithm (algorithm V1) were all relatively high (92.3 percent, 87.2 percent, and 98.7 percent, respectively); however, the PPV was low (52.2 percent) (Figure 1). Of the 69 patients identified by the algorithm as having cancer, 8 (11.6 percent) were identified by oncology visits only, 10 (14.5 percent) were identified by chemotherapy visits only, and 5 (7.3 percent) had both chemotherapy and oncology visits, but did not have any diagnosis of cancer. Of those 23 patients, only 1 was confirmed to have cancer through chart review. We reviewed these 23 further to determine why these patients had encounters consistent with cancer treatment, but no evidence in the medical record indicated a cancer diagnosis. This review revealed that patients may be seen in oncology

Figure 1. Performance of Each Algorithm for Identifying Individuals with a History of Cancer

A. Chart review 1, Algorithm V1

	Cancer	No Cancer	Total
Cancer Flag	36	33	69
No Cancer Flag	3	225	228
Total	39	258	297

Sensitivity	92.3%	(78%, 98%)
Specificity	87.2%	(82%, 91%)
PPV	52.2%	(40%, 64%)
NPV	98.7%	(96%, 100%)

B. Chart review 1, Algorithm V2

	Cancer	No Cancer	Total
Cancer Flag	35	11	46
No Cancer Flag	4	247	251
Total	39	258	297

Sensitivity	89.7%	(75%, 97%)
Specificity	95.7%	(92%, 98%)
PPV	76.1%	(61%, 87%)
NPV	98.4%	(96%, 99%)

C. Chart review 2, Algorithm V2

	Cancer	No Cancer	Total
Cancer Flag	38	25	63
No Cancer Flag	0	137	137
Total	38	162	200

Sensitivity	100%	(89%, 100%)
Specificity	84.6%	(78%, 90%)
PPV	60.3%	(47%, 72%)
NPV	100%	(97%, 100%)

Legend: For each table, the results from the tumor registry (the gold standard) are shown in the columns, and the results from the algorithm are shown in the rows. Sensitivity, specificity, positive predicative value (PPV), and negative predictive value (PPV) and their associated 95 percent confidence intervals are shown for each round of chart review. Confidence intervals were calculated using the efficient-score method corrected for continuity.[11]
Panel A shows results from chart review 1, algorithm version 1.
Panel B shows results from chart review 1, algorithm version 2.
Panel C shows results from chart review 2, algorithm version 2.

several times for suspected tumors, or they may undergo chemotherapy for noncancer-related conditions such as idiopathic thrombocytopenic purpura (ITP).

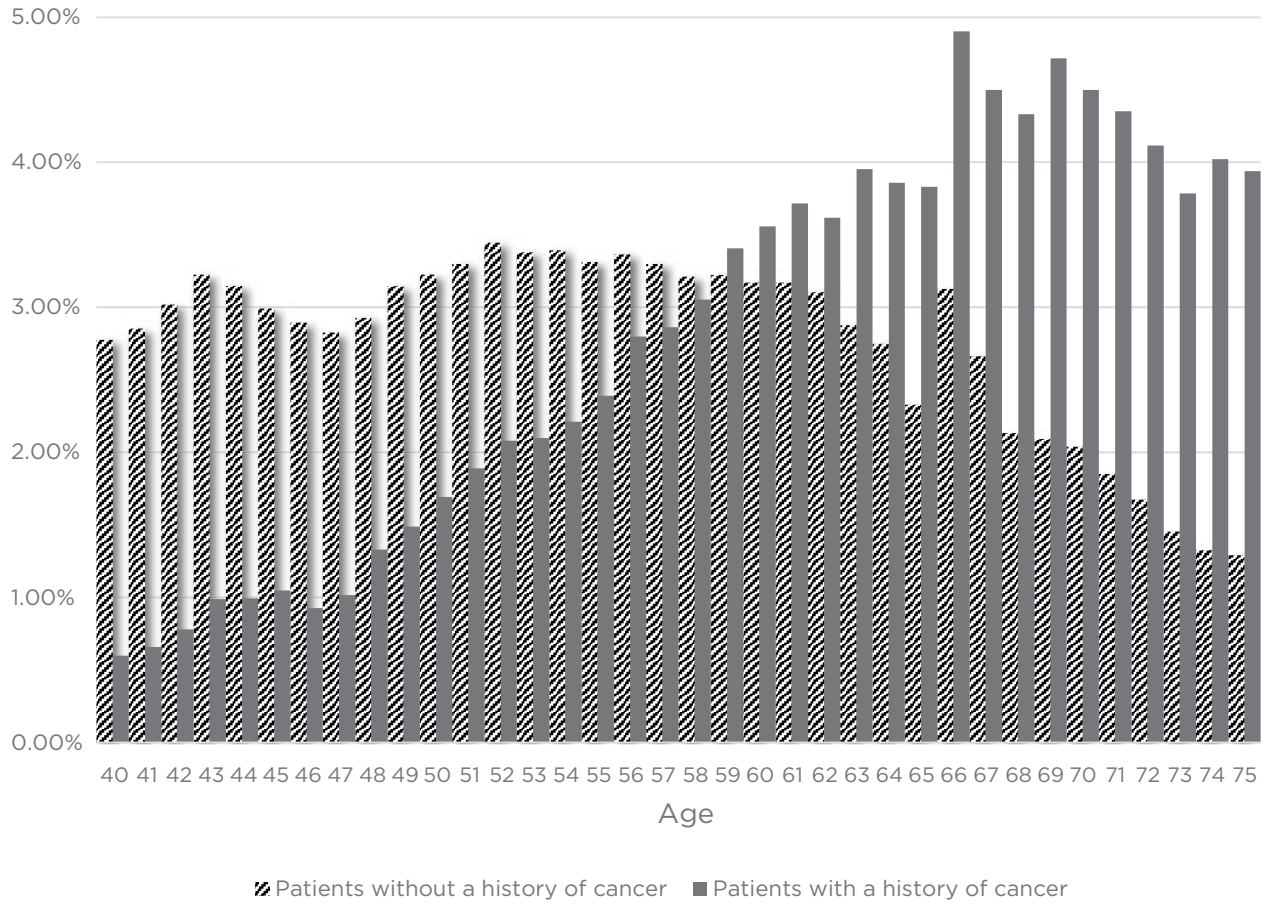
Based on this first chart review, we revised the algorithm (algorithm V2) to require either (1) diagnosis of cancer, or (2) inclusion in the tumor registry. Visits to oncology or receipt of chemotherapy alone were not sufficient to flag a record as a cancer case. This revision eliminated 4,242 (16.4 percent) of cases originally suspected as having cancer; and 46 of the 297 patients were flagged, in the revision, as having cancer. The resulting specificity and PPV improved (95.7 percent and 76.1 percent, respectively), while the sensitivity and NPV were slightly reduced to 89.7 percent and 98.4 percent, respectively (Figure 1). We then ran the revised algorithm on the full data set and conducted a second manual review on a new sample of charts. This second version of the algorithm had a sensitivity of 100 percent, specificity of 84.6 percent, NPV of 100 percent, and PPV of 60.3 percent on the second chart review, a marked improvement from version one (Figure 1).

Volume 4 (2016) Issue Number 1

Using the second version of the algorithm, most cases—19,278 (89.3 percent)—were identified as having cancers prior to 2013. Nearly half of the cancers identified were not in the tumor registry (N=10,172; 47 percent), and of those, 9,241 (91 percent) were identified prior to 2013. The top cancer indications that were not in the tumor registry prior to 2013 were "history of breast cancer," "diagnosis of stage 1 breast cancer," "history of malignant melanoma," "history of prostate cancer," and "diagnosis of prostate cancer," which accounted for 37 percent of diagnoses prior to 2013 not in the tumor registry. Of those with prior cancer, 60 percent were women. Members flagged as having cancer were significantly older than those without; the median age was 64 years and 56 years, respectively ($p < 0.0001$). Figure 2 shows the age distribution of the cancer and noncancer populations.

Figure 2. Age Distribution of Those with and Without a History of Cancer, as Defined by the Algorithm, of 201,787 KPCO Members Enrolled in 2013 and Ages 40–75 years. The distribution of individuals with cancer is skewed to older ages, which is expected.

Figure 2. Age Distribution of Those with and Without a History of Cancer, as Defined by the Algorithm, of 201,787 KPCO Members Enrolled in 2013 and Ages 40–75 years



Note: The distribution of individuals with cancer is skewed to older ages, which is expected.

Discussion

We developed an algorithm with high sensitivity (100 percent) and specificity (84.6 percent) to identify patients with a history of cancer using administrative data routinely captured in an EMR. Our final algorithm identified 10.7 percent of all health plan members ages 40–75 years with a history of cancer; nearly half (47 percent) of these cases would have been missed if we relied on the tumor registry alone.

Our intent was to create an algorithm to identify anyone with a history of cancer, so that those individuals could be excluded from an analytic cohort, leaving only cancer-free individuals. Because cancer is a rare disease, the remaining cohort of cancer-free individuals would be large, and we were thus willing to except a somewhat lower specificity. In the first version of the algorithm, chart review revealed that receipt of chemotherapy or visits to oncology alone contributed to an unacceptably low PPV (52.2 percent). In particular, infused therapies are not always administered for treatment of cancer; we identified patients receiving infusions for ITP or other conditions being mistakenly flagged as cancer cases. In the final version of the algorithm, the sensitivity and PPV were improved when a diagnosis code for cancer was required to accompany oncology visits or chemotherapy administration.

The PPV of this algorithm was modest (60.3 percent), a reflection of both the specificity and the prevalence of cancer in our study population. When the prevalence of a disease is low, the PPV will not be close to 1, even if both sensitivity and specificity are high.[14] In the final version of the algorithm, our false positive rate was 15 percent; thus, in our sample of 21,582 people flagged as having cancer, we excluded 3,331 people who did not have a history of cancer. The remaining sample (n=180,205) was sufficiently large to create a cancer-free cohort. However, for other applications, this false positive rate may be unacceptably high, and one may wish to improve the specificity of this algorithm.

Our methods have other limitations that should be considered. First, we limited our cohort to adults who are ages 40–75 years; the algorithm may perform differently in other age groups. Second, it is likely that we still included people with prior cancer, in that we depend on the patients to report prior cancer to their medical providers, and for those providers to record the information using a diagnosis code. We did not review medical records prior to 1998 (available in paper charts). Although unlikely, it is possible that information about cancer history would only be recorded prior to 1998 and not captured in EMRs spanning the next 15 years. In such instances, we would have erroneously classified people with a history of cancer as "cancer free." People who have little interaction with the medical system, or who are newer members of the health plan, are more likely to have an incomplete medical history. Given that the average length of KPCO membership in this population was 8.75 years, the magnitude of this error is likely small.

We have not validated this algorithm in other data systems; however, this algorithm and these methods should be generalizable to other health plans with EMR systems. We used codes such as ICD-9 that are readily available in other systems (Table 1). KPCO has a well-developed and validated tumor registry; organizations without a tumor registry, or with a less complete registry, could find this algorithm useful to identify all cancer cases (not just prior cancers) for research purposes. This algorithm is useful for identifying a cancer-free study population that may be desirable for any number of research questions—for example, to study conditions such as heart failure[15] that may be more common among those with a history of cancer. Our algorithm may not perform as well in other EMR systems where the data are not as complete as those at KPCO. The KPCO

Volume 4 (2016) Issue Number 1

EMR system dates back to 1998, and the accuracy of our data has been validated in previous studies.[16-19] This algorithm, like other algorithms, depends on complete and accurate data; additional validation may be required when applying it to EMR data that are not as well developed as the KPCO EMR.

Conclusion

We developed an algorithm with high sensitivity (100 percent) and specificity (84.6 percent) to identify patients with a history of cancer using administrative data routinely captured in an EMR. It is not sufficient to rely on the tumor registry alone to capture those with a history of cancer. Casting a wide net will help ensure that anyone with a history of cancer, whether diagnosed recently or prior to joining the health plan, will be identified. This algorithm could be applied to other health plans with similar data coding systems.

Acknowledgments

This work was funded by the National Cancer Institute (Contracts HHSN261201400644P and HHSN261201300460P). We thank LeeAnn Rohm, MSW and Kate Burniece, BS for completing the chart abstraction and Erica Blum-Barnett, MS for manuscript preparation.

References

1. Thoburn KK, German RR, Lewis M, et al. Case completeness and data accuracy in the Centers for Disease Control and Prevention's National Program of Cancer Registries. Cancer. 2007;109(8):1607-1616.
2. Bowles EJA, Feigelson HS, Barney T, et al. Improving quality of breast cancer surgery through development of a national breast cancer surgical outcomes (BRCASO) research database. BMC Cancer. 2012;12:136.
3. Wagner EH, Greene SM, Hart G, et al. Building a research consortium of large health systems: the Cancer Research Network. J Natl Cancer Inst Monogr. 2005;(35):3-11.
4. Kahn MG, Raebel, MA, Glanz JM, et al. A Pragmatic Framework for Single-site and Multisite Data Quality Assessment in Electronic Health Record-based Clinical Research. Med Care. 2012;50(Suppl):S21–S29.
5. Platt R, Davis R, Finkelstein J, et al. Multicenter epidemiologic and health services research on therapeutics in the HMO Research Network Center for Education and Research on Therapeutics. Pharmacoepidemiol Drug Saf. 2001;10:373–377.
6. Go AS, Magid DJ, Wells B, et al. The Cardiovascular Research Network: a new paradigm for cardiovascular quality and outcomes research. Circ Cardiovasc Qual Outcomes. 2008;1:138–147.
7. Baggs J, Gee J, Lewis E, et al. The Vaccine Safety Datalink: a model for monitoring immunization safety. Pediatrics. 2011;127 Suppl 1:S45-S53.
8. Ross TR, Ng D, Brown JS et al. The HMO Research Network Virtual Data Warehouse: A public data model to support collaboration. EGEMS (Wash DC). 2014 Mar 24;2(1):1049. doi:10.13063/2327-9214.1049. eCollection 2014.
9. Ritzwoller DP, Carroll N, Delate T et al. Validation of electronic data on chemotherapy and hormone therapy use in HMOs. Med Care. 2013;51:e67-e73.
10. Hornbrook MC, Hart G, Ellis JL et al. Building a virtual cancer research organization. J Natl Cancer Inst Monogr. 2005:12-25.
11. Fleiss JL, Levin B, Paik MC. Statistical Methods for Rates and Proportions. Third Edition. New York: John Wiley & Sons 2003.
12. Newcombe, R. G. Two-sided confidence intervals for the single proportion: comparison of seven methods: comparison of seven methods. Statistics in Medicine. 1998;17.8:857-872.
13. Gordis L. Epidemiology. Fourth Edition. Philadelphia: Elsevier Saunders, 2009.
14. Altman DG and Bland JM. Diagnostic tests 2: predictive values. BMJ. 1994;09:102.
15. Bowles EJA, Wellman R, Feigelson HS, et al. Risk of heart failure in breast cancer patients after anthracycline and trastuzumab treatment: a retrospective cohort study. Journal of the National Cancer Institute. 2012;104.17:1293-1305.
16. Delate T, Bowles EJA, Pardee R, et al. Validity of eight integrated healthcare delivery organizations' administrative clinical data to capture breast cancer chemotherapy exposure. Cancer Epidemiol Biomarkers Prev. 2012;21(4):673-80.
17. Ritzwoller DP, Carroll N, Delate T, et al. Validation of Electronic Data on Chemotherapy and Hormone Therapy Use in HMOs. Med Care. 2013;51.10:e67-73.
18. Andrade SE, Moore Simas TA, Boudreau D, et al. Validation of algorithms to ascertain clinical conditions and medical procedures used during pregnancy. Pharmacoepidemiol Drug Saf. 2011 Nov;20(11):1168-76.
19. Bowles EJA, Tuzzio L, Ritzwoller DP, et al. Accuracy and complexities of using automated clinical data for capturing chemotherapy administrations: implications for future research. Med Care. 2009;47:1091–1097.