

Algorithm to Identify Systemic Cancer Therapy Treatment Using Structured Electronic Data

original report

abstract

Purpose With the shift in the majority of oncology clinical care in the United States from paper records to electronic health records, researchers need efficient and validated processes to obtain accurate data about the entire treatment history of patients diagnosed with cancer. The objective of this study was to develop and validate an algorithm that is agnostic to the source of data but that can identify specific regimens in the entire course of systemic therapy treatment for patients diagnosed with breast, colorectal, or lung cancer.

Methods A cohort of patients with incident breast, colorectal, and lung cancer were randomly distributed into six groups. The algorithm was iteratively modified, and the performance was assessed until no additional modifications could be identified in the first three groups. The performance of the algorithm was confirmed in the three groups that remained.

Results The final model produced ranges of sensitivity between 97.2% and 100% for first-course systemic therapy across all cancers, with a false-positive rate of 0%. The algorithm matched the exact number of courses and the exact regimens of systemic therapy agents as captured by infusion, pharmacy, and procedure electronic medical record data for all courses of therapy 88% to 100% of the time.

Conclusion Use of our validated algorithm that characterizes entire courses of systemic therapy treatment in patients diagnosed with breast, colorectal, and lung cancer will allow researchers in a variety of settings to conduct comparative effectiveness studies related to the uptake, safety, outcomes, and costs associated with the use of both novel and standard regimens.

Clin Cancer Inform. © 2017 by American Society of Clinical Oncology

Nikki M. Carroll
Kate M. Burniece
Jeff Holzman
Deanna B. McQuillan
Angela Plata
Debra P. Ritzwoller

All authors: Institute for Health Research, Kaiser Permanente Colorado, Denver, CO.

Supported by the Strategic Allocation of Resources Committee at Kaiser Permanente Colorado, with initial infrastructure support provided by National Cancer Institute Grant No. RC2 CA148185 (Building CER Capacity: Aligning CRN, CMS, and State Resources to Map Cancer Care; co-primary investigators: Jane C. Weeks, MD, and Debra P. Ritzwoller, PhD).

Corresponding author: Nikki M. Carroll, MS, Kaiser Permanente Colorado, Institute for Health Research, 10065 E Harvard Ave, Suite 300, Denver, Colorado 80237; e-mail: nikki.m.carroll@kp.org.

INTRODUCTION

To conduct comparative effectiveness research on treatment options commonly used in community-based oncology practices, researchers need generalizable and accurate data about the entire treatment history of patients diagnosed with cancer.^{1,2} Tumor registries generate extensive information about the first course of systemic therapy in patients, but they do not capture the full course of treatment, including the number of courses, discontinuation of therapy, or the use of multiple courses of therapy. Of the studies that have evaluated the receipt of systemic therapy, most did not extend beyond the first course, and many used only SEER-Medicare data and/or did not include oral chemotherapy agents (ie, those covered by Medicare Part D).^{1,3-10} Other studies have looked at second- or third-course therapies, but the algorithms were cancer specific or had strict inclusion or exclusion criteria.^{7-9,11,12}

In 2009, Kaiser Permanente Colorado (KPCO) added a medical oncology module to its Epic-based ambulatory integrated electronic health record (EHR; Epic Systems, Verona, WI). Although the addition of the oncology module improved the ability to evaluate entire courses of systemic therapy, it still had limitations. It did not include data about patients who received systemic therapy or pharmacy dispenses outside of KPCO (eg, contract providers who submitted claims data), and it did not include all oral therapies that were dispensed in outpatient pharmacies. In addition, the systemic therapy data for patients who received treatment before 2009 existed in separate files that contained National Drug Codes (NDCs), procedure codes, and Healthcare Common Procedure Coding System (HCPCS) codes.

The objective of this study was to construct and validate an algorithm that combined all data sources

to characterize entire courses of systemic therapy treatment in patients diagnosed with breast, colorectal, or lung cancer. Specifically, we wanted an algorithm that (1) could characterize not only first-course therapy but all subsequent courses the patient received after diagnosis and (2) would be agnostic to any data source. The algorithm was validated against gold standards of the KPCO Virtual Tumor Registry (VTR) and manual chart review.

METHODS

Data Sources

The primary data source for this analysis was the KPCO Virtual Data Warehouse (VDW). The VDW is the local research-ready database of KPCO that contains data consistent with the standards, formats, and definitions used in the Health Care Systems Research Network and the Cancer Research Network (CRN).^{13,14} The VDW contains administrative, EHR, and other data that have been extracted and loaded into relational tables linked through a common, unique identifier.¹⁵⁻¹⁹ Within the VDW, the VTR contains data consistent with the North American Association of Central Cancer Registries standard.^{19a} VTR data are obtained from manual reviews of the medical charts of patients by certified tumor registrars and include coded clinical data associated with inpatient and outpatient events, date of diagnosis, first-course treatment (eg, surgery, radiotherapy, chemotherapy), tumor characteristics, and more. VDW diagnosis and procedure files include coded diagnoses and procedures associated with inpatient and outpatient events. These codes are based on the International Classification of Diseases (9th revision [ICD-9] and 10th revision, clinical modification [ICD-10-CM]), HCPCS, and the Common Procedure Terminology codes (4th edition [CPT]). Greater than 90% of the VDW diagnosis and procedure data used in this analysis were derived from EHRs. These data capture the diagnoses and treatment associated with the systemic therapy events that take place predominantly in health plan–owned ambulatory infusion centers. Claims data included in these analyses were associated with cancer treatments administered by contract providers, including hospital-based care. The VDW pharmacy files capture NDC-based oral and other prescription drugs dispensed from both outpatient pharmacies (which includes oral therapies such as capecitabine) and KPCO-owned infusion centers.

Exact systemic therapy treatment regimens from the EHR were manually extracted for each patient for chart review comparisons. Each abstractor was

given the date of cancer diagnosis and abstracted any systemic therapy treatment given within 1 year of diagnosis. All abstractors were blinded to any VTR or electronic results. This project was determined to be exempt from human subjects research and was classified as quality work by the KPCO institutional review board.

Identification of Systemic Therapy Agents From the VDW

As described previously,^{1,15,20} the cancer systemic therapy look-up tables contain more than 8,000 NDCs and 300 procedure codes and diagnostic treatment-related codes. The systemic therapy tables that identify the NDCs, procedure codes, and diagnostic treatment-related codes used in this analysis are publicly available on the CRN website.^{20a} These look-up tables were linked to VDW pharmacy, infusion, procedure, and diagnosis tables to identify all systemic therapy agents received by each patient. All systemic therapy agents identified from the VDW tables were then compiled into one data set and sorted by a patient identifier, the date of administration of the systemic therapy agent, and the generic name of the agent administered. The resulting data set contained unique patient–day–systemic therapy agent observational data.

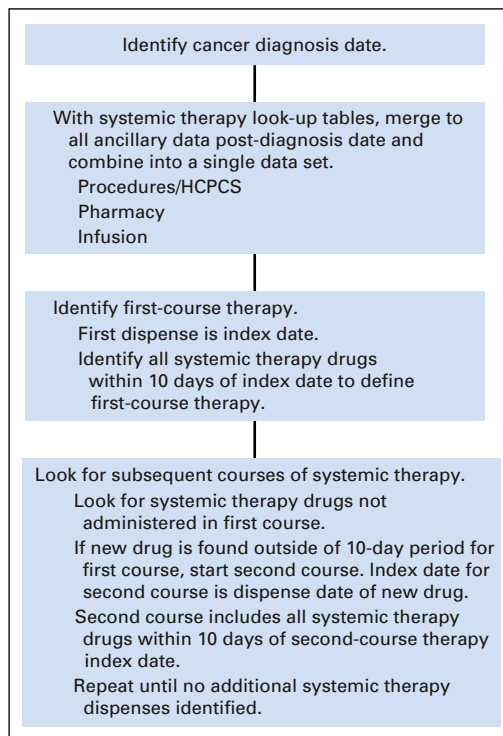
Definitions for Algorithm

The first date of a VDW-captured treatment event in one or more the VDW files (pharmacy, procedure, infusion, or diagnosis) that occurred within 180 days of the incident cancer diagnosis dated in the VTR was considered the first systemic therapy event, and the patient was flagged as a recipient of first-course systemic therapy. All systemic therapy agents received within 10 days of the first systemic therapy event were considered the first-course systemic therapy regimen.^{1,21}

Changes in systemic therapy agents were evaluated longitudinally up to 1 year after cancer diagnosis. Consistent with other published studies, a change in systemic therapy agents indicated a new course of therapy—that is, any addition of one or more systemic therapy agents was considered a new course of treatment.^{5,7,8,12} Discontinuation of a single agent from a regimen was not considered a change in a course of therapy.

For specific treatment plans that administer drugs outside of a 10-day window, the algorithm would look ahead in the data for receipt of the additional systemic therapy drugs (Fig 1). There were such adjustments in the algorithm for two specific

Fig 1. Algorithm processing flow diagram. HCPCS, Healthcare Common Procedure Coding System.



breast cancer and two specific colorectal cancer treatment plans.

Cancer Cohort

The cohort used to construct our algorithm included patients identified in the VTR as diagnosed with stages I to IV breast, colorectal, or lung cancer between 2005 and 2014 and observed through 2015. Patients with previous or subsequent cancer diagnoses were excluded.

Sampling Strategy

A randomized, iterative sampling strategy was implemented, and the performance of each iteration of the algorithm was assessed separately for each cancer. Patients within each cancer were assigned and sorted by a random number. The first 100 randomly ordered patients were pulled to compose the first group and determine baseline results. The algorithm was applied, and the performance was assessed; algorithm results were compared with the chart review and the VTR. Modifications were made to the algorithm on the basis of these comparisons. The next 50 randomly ordered patients were pulled to compose the second group. The refined algorithm was applied, and the performance was assessed. The next 50 randomly ordered patients composed the third group, and the refined algorithm was applied. Performance measures showed that no additional modifications

were needed, so the algorithm was applied to a fourth group of the next 50 randomly ordered patients to confirm results. Again, performance measures indicated that no modifications were needed to the algorithm. Because the quality and completeness of the data (eg, treatment protocols, plans) increased after implementation of the EHR oncology module, we pulled two additional groups of randomly ordered patients whose treatments were completed after 2009. Performance specifically related to entire courses of therapy was determined in these groups (groups 5 and 6). Figure 2 shows a diagram of this iterative sampling strategy and performance assessment.

Statistical and Sensitivity Analysis

Patient demographics and characteristics were evaluated at the date of cancer diagnosis. Variables were reported as a percentage of the group. All analyses were descriptive and were conducted with SAS 9.4 (SAS Institute, Cary, NC). No tests of statistical significance were planned or performed because of the descriptive nature of the study.

The algorithm was evaluated for three different elements: (1) the ability to identify receipt of any systemic therapy (yes/no), (2) the ability to accurately identify the first-course therapy regimen, and (3) the ability to accurately identify the number and exact regimens of all courses of therapy up to 1 year after cancer diagnosis. For each cancer occurrence, we compared the results from the algorithm to both the VTR and the chart review; the VTR was the gold standard in the comparison with the VTR, and chart review results were the gold standard in the comparison with the chart review. Sensitivity, false positive rate, and accuracy by cancer site were computed.

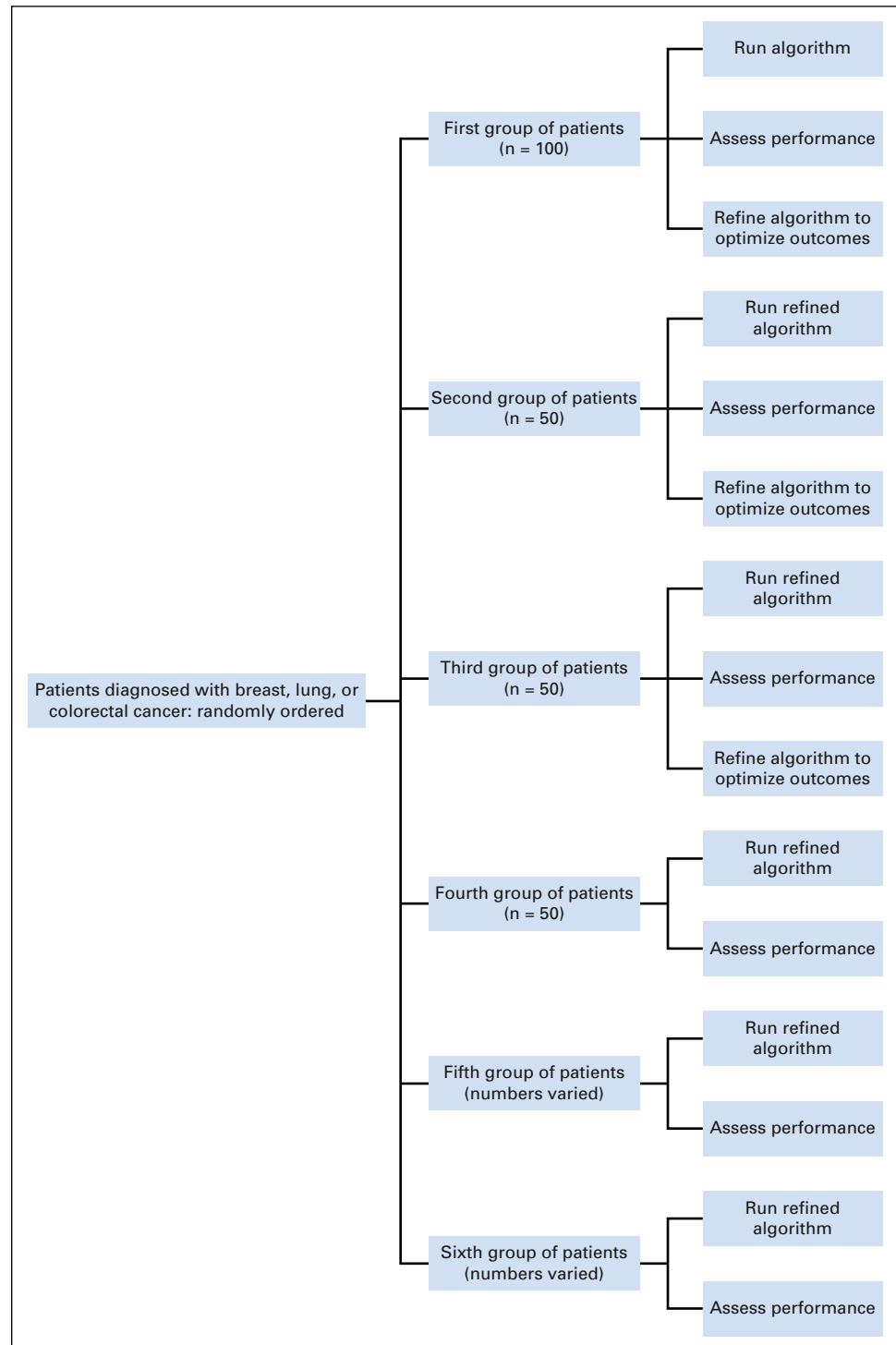
Algorithm Refinement

Specific components of the algorithm were examined. These included deletion of specific CPT or HCPCS codes that generated false-positive results, deletion of administrations of systemic therapy drugs that were prescribed for a noncancer indication, and adjustment of the algorithm for specific treatment plans that administer drugs outside of the 10-day window.

RESULTS

A total of 450 patients were randomly assigned into six groups for colorectal and lung cancer. A total of 449 patients were randomly assigned into six groups for breast cancer. One patient was excluded from the breast cancer cohort, because

Fig 2. Randomized sampling strategy for refinement and evaluation of the algorithm.



no systemic therapy information was available in the VTR. Demographic characteristics of patients for all groups are listed in [Table 1](#). Of note, almost half of the patients were 65 years of age or younger at the time of diagnosis, and cancer diagnoses were distributed evenly across American Joint Commission on Cancer stages I through IV for all three cancer sites.

Algorithm performance measures for the identification of first-course systemic therapy relative to the gold standard of the VTR and chart review are listed in [Table 2](#). Baseline results showed a high sensitivity for all three cancers; however, the false positive rate was also high and ranged from 4.3% for colorectal cancer to 13.0% for breast cancer compared with the VTR and from 0% for lung cancer to 9.1% for breast

Table 1. Demographic Characteristics of Cohorts

Characteristic by Cancer Population	Patients by Group					
	Group 1*	Group 2	Group 3	Group 4	Group 5	Group 6
Breast cancer						
Total No. of patients	99	50	50	50	100	100
Female sex, %	99	100	98	100	100	100
Ethnicity, %						
White	74	84	68	64	80	75
Nonwhite	26	16	32	36	20	25
Age, years, %						
< 55	47	42	40	32	31	25
55-64	27	26	34	30	31	24
65-74	18	22	14	24	26	34
≥ 75	7	10	12	14	12	17
AJCC stage, %						
I	46	38	44	42	62	57
II	30	34	34	34	29	24
III	11	18	18	20	6	34
IV	12	10	4	4	3	17
Colorectal cancer						
Total No. of patients	100	50	50	50	100	100
Female sex, %	45	48	38	40	45	40
Ethnicity, %						
White	71	66	72	80	71	74
Nonwhite	29	34	28	20	29	26
Age, years, %						
< 55	39	37	22	20	34	27
55-64	21	16	30	16	17	22
65-74	25	24	22	38	25	31
≥ 75	15	22	26	26	24	20
AJCC stage, %						
I	14	12	16	18	24	26
II	15	18	20	26	13	20
III	48	46	36	28	28	12
IV	23	24	28	28	19	22
Lung cancer						
Total No. of patients	100	50	50	50	100	100
Female sex, %	48	38	52	60	61	60
Ethnicity, %						
White	78	64	82	76	80	81
Nonwhite	22	36	18	24	20	19
Age, years, %						
< 55	15	10	6	10	5	6
55-64	23	38	20	18	19	23
65-74	28	32	44	42	34	35
≥ 75	34	20	30	30	42	36

(continued on following page)

Table 1. Demographic Characteristics of Cohorts (continued)

Characteristic by Cancer Population	Patients by Group					
	Group 1*	Group 2	Group 3	Group 4	Group 5	Group 6
AJCC stage, %						
I	8	14	12	24	23	21
II	2	4	8	6	7	7
III	32	24	38	28	15	16
IV	58	58	42	42	55	56

Abbreviation: AJCC, American Joint Commission on Cancer.

*Data from one patient in the breast cancer group were deleted because no information was entered into the Virtual Tumor Registry.

cancer compared with the chart review. The percentage matches for exact agents in first-course therapy were 73.7% for breast cancer, 79.0% for colorectal cancer, and 96.0% for lung cancer.

Results of our evaluation of the algorithm to identify the entire course of systemic therapy in the subset of those patients who received treatment after implementation of the EHR oncology module are listed in Table 3. The percentage matches of agents in first-course therapy increased to 87.2% for breast cancer, 86.8% for colorectal cancer, and 100% for lung cancer. Matches of all regimens and all courses of therapy were 87.2% for breast cancer, 86.8% for colorectal cancer, and 95.7% for lung cancer.

We determined from the baseline assessments that two modifications needed to be made to the algorithm: (1) exclusion of any bevacizumab that was prescribed for ophthalmic indications and (2) exclusion of evaluation and management CPT codes that were coded during ambulatory visit consultations to determine treatment after a cancer diagnosis. Reassessment of the algorithm performance after refinement of the algorithm showed that the sensitivity slightly decreased for breast cancer and lung cancer but increased for colorectal cancer compared with the VTR and chart review. Improvements to the algorithm came in the form of a decreased false-positive rate for all cancers (decreased to 0%)

Table 2. Algorithm Performance Measures for First-Course Therapy Compared With Tumor Registry and Chart Review

Measure by Cancer Population	Algorithm Versus Tumor Registry				Algorithm Versus Chart Review			
	Group 1 (baseline)	Group 2 (modification 1)	Group 3 (modification 2)	Group 4 (confirmatory)	Group 1 (baseline)	Group 2 (modification 1)	Group 3 (modification 2)	Group 4 (confirmatory)
Breast cancer								
No. of patients	99	50	50	50	99	50	50	50
Sensitivity	100	97.6	100	100	100	97.1	100	100
False-positive rate	13.0	0	0	0	9.1	0	0	0
Accuracy	97.0	98.0	100	100	98.0	98.0	100	100
Match to first course	73.7	92.0	86.0	90.0	82.8	84.0	94.0	90.0
Colorectal cancer								
No. of patients	100	50	50	50	100	50	50	50
Sensitivity	98.7	100	97.2	100	97.4	100	97.2	100
False-positive rate	4.3	0	0	0	8.7	0	0	0
Accuracy	98.0	100	98.0	100	96.0	100	98.0	100
Match to first course	79.0	90.0	90.0	90.0	87.0	94.0	84.0	88.0
Lung cancer								
No. of patients	100	50	50	50	100	50	50	50
Sensitivity	98.8	97.3	100	97.4	100	100	100	100
False-positive rate	5.0	0	0	0	0	0	0	0
Accuracy	98.1	98.0	100	98.1	100	100	100	100
Match to first course	96.0	98.0	98.0	93.6	95.0	98.0	98.1	94.0

Table 3. Algorithm Performance Measures for All Courses of Therapy After EHR Oncology Module Implementation

Measure by Cancer Population	Group 1 (baseline)	Group 2 (modification 1)	Group 3 (modification 2)	Group 4 (confirmatory)	Group 5 (confirmatory)	Group 6 (confirmatory)
Breast cancer						
No. of patients	47	33	24	28	91	96
Match to first-course therapy	87.2	93.9	95.8	93.0	92.3	100
Match to No. of courses	87.2	93.9	91.7	96.0	92.3	100
All-regimen match	87.2	87.9	91.7	93.0	92.3	100
Colorectal cancer						
No. of patients	38	23	17	17	98	94
Match to first-course therapy	86.8	87.0	94.1	94.1	98.0	98.9
Match to No. of courses	89.5	82.6	94.1	88.2	96.9	95.7
All-regimen match	86.8	82.6	94.1	88.2	96.9	95.7
Lung cancer						
No. of patients	46	22	23	29	99	94
Match to first-course therapy	100	95.5	100	100	100	98.9
Match to No. of courses	97.8	95.5	100	96.6	97.9	95.7
All-regimen match	95.7	95.5	100	96.6	97.9	95.7

NOTE. Data shown as percent, except rows where No. of patients is noted.
Abbreviation: EHR, electronic health record.

compared with both the VTR and the chart review (Table 2). Specifically, for data after EHR oncology module implementation, the percentage matches in first-course therapy increased to 93.9% in breast cancer, increased to 87.0% in colorectal cancer, and slightly decreased to 95.5% in lung cancer. Matches on all regimens were 87.9% for breast cancer, 82.6% for colorectal cancer, and 95.5% for lung cancer (Table 3).

Evaluation of the refined algorithm on the second group of patients illuminated that the algorithm was not capturing several treatment plans that administer drugs outside of a 10-day window correctly. For example, patients with colorectal cancer are administered FOLFOX (fluorouracil, leucovorin, and oxaliplatin) or FOLFIRI (fluorouracil, irinotecan, leucovorin) and then, 4 weeks later, are administered bevacizumab. Breast cancer treatment plans include a regimen of cyclophosphamide, doxorubicin, and a taxane, in which the patient is administered cyclophosphamide plus doxorubicin and then, 4 to 6 weeks later, is administered paclitaxel and/or a maintenance treatment of trastuzumab. The algorithm was altered to look ahead for these additional agents after the initial agents were administered. Sensitivity remained extremely high and ranged from 97.2% to 100% across all cancers, whereas the false-positive rate remained at 0%. The percentage matches of first-

course therapy decreased slightly for breast cancer but stayed the same for colorectal and lung cancers (Table 2). Data compiled for treatment received after EHR oncology module implementation showed an increase in both the first-course therapy matches and the matches to all courses (Table 3).

We assessed performance in group 4 with no additional modifications to the algorithm. The percentage match of first-course therapy for breast cancer increased, it stayed the same for colorectal cancer, and it slightly decreased for lung cancer. Sensitivity and false-negative rates remained the same (Table 2). The subset of patients in group 4 who received treatment after EHR oncology module implementation showed an increase in the percentage match for all regimens in breast cancer and a slight decrease for both colorectal and lung cancers (Table 3).

Additional evaluations of the algorithm compared with data compiled after EHR oncology module implementation were assessed in groups 5 and 6 and are listed in Table 3. Across groups 4 through 6 (all confirmatory assessments of the algorithm), the algorithm matched exact first-course therapy 92.3% to 100% of the time and matched all courses of therapy in groups 5 and 6 92.3% to 100% of the time for breast cancer. The algorithm matched exact first-course therapy 94.1% to

98.9% of the time for colorectal cancer and matched all courses of therapy 88.2% to 96.9% of the time. The algorithm matched exact first-course therapy 98.9% to 100% of the time and matched all courses of therapy 95.7% to 97.9% of the time. Across all courses of systemic therapy, 52 distinct regimens for breast cancer, 66 distinct regimens for colorectal cancer, and 43 distinct regimens for lung cancer were detected (Data Supplement).

DISCUSSION

To our knowledge, this is the first algorithm that combines widely available structured EHR, claims, and administrative data to determine all courses of treatment across any cancer diagnosis. By using an iterative sampling strategy, we constructed an algorithm with a 0% false-positive rate, high sensitivity, and high accuracy that captured the entire course of treatment in patients after a cancer diagnosis. Although not a main outcome for this project, higher sensitivity and specificity were observed when chart review was the gold standard versus the VTR. Because of the window of time postdiagnosis (eg, 6 to 9 months) that tumor registrars manually review medical records, tumor registries often have incomplete ascertainment of systemic therapy data for patients whose treatment starts more than 6 months after an initial diagnosis.²

There were several strengths to this study. KPCO data have the ability to capture oral systemic therapies at a time when Medicare Part D pharmacy data are not currently available via linked SEER-Medicare files. Second, this study includes a representative population of patients with cancer who are younger than 65 years of age. Third, the algorithm was validated with data from patients diagnosed with three different cancer sites and at multiple stages of diagnosis. Fourth, the EHR data show the actual administration of systemic therapy agents, some of which are prescribed outside of standard protocols, that provides insights to

real-world use of systemic therapy for treatment of cancers. Last, the algorithm code is flexible enough that it may be used on any structured data, including SEER-Medicare data, the national Patient-Centered Clinical Research Network common data model, or the Food and Drug Administration–funded Sentinel distributed database that captures a date of administration and an NDC, HCPCS code, or ICD-9/ICD-10 identifier for each agent of interest.

We also note a few limitations to the algorithm development and validation. First, some claims from contract providers include only nonspecific systemic therapy codes, so the exact agent administered was not identified in all cases. Second, this study was limited by the number of chart reviews used for comparisons that could be completed within the scope of this project. Third, changes in courses of therapy were defined that may not be applicable to all community settings. However, we believe the algorithm is flexible enough to adjust for different definitions of a switch or discontinuation in systemic therapy. Fourth, if a patient was enrolled in a clinical trial, the specific agents administered to the patient may not have been captured. If the trial drug was administered by a KPCO infusion center, the trial drug may not have been captured or included in the traditional pharmacy system. Last, the scope of this project was limited to expansion and improvement of the initial algorithm and did not allow ascertainment of the number of cycles administered in each course of therapy.

In conclusion, we constructed a highly sensitive and accurate algorithm that was able to match entire courses of treatment in patients diagnosed with breast, colorectal, or lung cancers. Future work includes the identification of the number of cycles within each course of therapy and/or the dose of systemic agents received.

DOI: <https://doi.org/10.1200/CCI.17.00002>

Published online on ascopubs.org/journal/cci on June 9, 2017.

AUTHOR CONTRIBUTIONS

Conception and design: Nikki M. Carroll

Collection and assembly of data: Nikki M. Carroll, Kate M. Burniece, Jeff Holzman, Deanna B. McQuillan, Angela Plata

Data analysis and interpretation: Nikki M. Carroll, Debra P. Ritzwoller

Manuscript writing: All authors

Final approval of manuscript: All authors

Agree to be accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/site/ifc.

Nikki M. Carroll

No relationship to disclose

Kate M. Burniece

No relationship to disclose

Jeff Holzman

No relationship to disclose

Deanna B. McQuillan

No relationship to disclose

Angela Plata

No relationship to disclose

Debra P. Ritzwoller

No relationship to disclose

ACKNOWLEDGMENT

We thank William Harding, data specialist/SAS programmer.

REFERENCES

1. Ritzwoller DP, Carroll NM, Delate T, et al: Patterns and predictors of first-line chemotherapy use among adults with advanced non-small cell lung cancer in the cancer research network. *Lung Cancer* 78:245-252, 2012
2. Aiello Bowles EJ, Tuzzio L, Ritzwoller DP, et al: Accuracy and complexities of using automated clinical data for capturing chemotherapy administrations: implications for future research. *Med Care* 47:1091-1097, 2009
3. Warren JL, Harlan LC, Fahey A, et al: Utility of the SEER-Medicare data to identify chemotherapy use. *Med Care* 40:IV-55-IV-61, 2002 (suppl)
4. Lamont EB, Lan L: Sensitivity of Medicare claims data for measuring use of standard multiagent chemotherapy regimens. *Med Care* 52:e15-e20, 2014
5. Bikov KA, Mullins CD, Seal B, et al: Algorithm for identifying chemotherapy/biological regimens for metastatic colon cancer in SEER-Medicare. *Med Care* 53:e58-e64, 2015
6. Du XL, Key CR, Dickie L, et al: External validation of Medicare claims for breast cancer chemotherapy compared with medical chart reviews. *Med Care* 44:124-131, 2006
7. Hess GP, Wang PF, Quach D, et al: Systemic therapy for metastatic colorectal cancer: Patterns of chemotherapy and biologic therapy use in US medical oncology practice. *J Oncol Pract* 6:301-307, 2010
8. Karve S, Lorenzo M, Liepa AM, et al: Treatment patterns, costs, and survival among Medicare-Enrolled elderly patients diagnosed with advanced stage gastric cancer: Analysis of a linked population-based cancer registry and administrative claims database. *J Gastric Cancer* 15:87-104, 2015
9. Ramsey SD, McCune JS, Blough DK, et al: Colony-stimulating factor prescribing patterns in patients receiving chemotherapy for cancer. *Am J Manag Care* 16:678-686, 2010
10. Lamont EB, Herndon JE II, Weeks JC, et al: Criterion validity of Medicare chemotherapy claims in Cancer and Leukemia Group B breast and lung cancer trial participants. *J Natl Cancer Inst* 97:1080-1083, 2005
11. Parikh RC, Du XL, Morgan RO, et al: Patterns of treatment sequences in chemotherapy and targeted biologics for metastatic colorectal cancer: Findings from a large community-based cohort of elderly patients. *Drugs Real World Outcomes* 3:69-82, 2016
12. Ramsey SD, Martins RG, Blough DK, et al: Second-line and third-line chemotherapy for lung cancer: Use and cost. *Am J Manag Care* 14:297-306, 2008
13. Health Care Systems Research Network (HCSRN): VDW Data Model 2017. <http://www.hcsr.org/en/Tools%20&%20Materials/VDW/>
14. National Cancer Institute: Cancer Research Network 2017. <https://www.crn.cancer.gov/>
15. Ritzwoller DP, Carroll N, Delate T, et al: Validation of electronic data on chemotherapy and hormone therapy use in HMOs. *Med Care* 51:e67-e73, 2013
16. 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* (35):3-11, 2005
17. National Cancer Institute: The HMO Cancer Research Network: Capacity, collaboration, and investigation. Washington, DC, US Department of Health and Human Services, National Institutes of Health, 2010
18. Hornbrook MC, Hart G, Ellis JL, et al: Building a virtual cancer research organization. *J Natl Cancer Inst Monogr* (35):12-25, 2005
19. 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)* 2:1049, 2014
- 19a. North American Association of Central Cancer Registries: <http://www.naaccr.org>
20. Delate T, Bowles EJ, Pardee R, et al: Validity of eight integrated healthcare delivery organizations' administrative clinical data to capture breast cancer chemotherapy exposure. *Cancer Epidemiol Biomarkers Prev* 21:673-680, 2012
- 20a. National Cancer Institute, Cancer Research Network: Cancer Therapy Look-Up Tables. <http://crn.cancer.gov/resources/codes.html>
21. Zhu J, Sharma DB, Gray SW, et al: Carboplatin and paclitaxel with vs without bevacizumab in older patients with advanced non-small-cell lung cancer. *JAMA* 307:1593-1601, 2012