# ARTICLE

# Enhancing Breast Cancer Recurrence Algorithms Through Selective Use of Medical Record Data

Candyce H. Kroenke, Jessica Chubak, Lisa Johnson, Adrienne Castillo, Erin Weltzien, Bette J. Caan

**Affiliations of authors:** Kaiser Permanente Northern California, Division of Research, Oakland, CA (CHK, AC, EW, BJC); Group Health Research Institute, Seattle, WA (JC); Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, WA (LJ).

**Correspondence to:** Candyce H. Kroenke, ScD, MPH, Kaiser Permanente, Division of Research, 2000 Broadway, 5th Floor, Oakland, CA 94612 (e-mail: candyce.h.kroenke@kp.org).

## Abstract

**Background:** The utility of data-based algorithms in research has been questioned because of errors in identification of cancer recurrences. We adapted previously published breast cancer recurrence algorithms, selectively using medical record (MR) data to improve classification.

**Methods:** We evaluated second breast cancer event (SBCE) and recurrence-specific algorithms previously published by Chubak and colleagues in 1535 women from the Life After Cancer Epidemiology (LACE) and 225 women from the Women's Health Initiative cohorts and compared classification statistics to published values. We also sought to improve classification with minimal MR examination. We selected pairs of algorithms—one with high sensitivity/high positive predictive value (PPV) and another with high specificity/high PPV—using MR information to resolve discrepancies between algorithms, properly classifying events based on review; we called this "triangulation." Finally, in LACE, we compared associations between breast cancer survival risk factors and recurrence using MR data, single Chubak algorithms, and triangulation.

**Results:** The SBCE algorithms performed well in identifying SBCE and recurrences. Recurrence-specific algorithms performed more poorly than published except for the high-specificity/high-PPV algorithm, which performed well. The triangulation method (sensitivity = 81.3%, specificity = 99.7%, PPV = 98.1%, NPV = 96.5%) improved recurrence classification over two single algorithms (sensitivity = 57.1%, specificity = 95.5%, PPV = 71.3%, NPV = 91.9%; and sensitivity = 74.6%, specificity = 97.3%, PPV = 84.7%, NPV = 95.1%), with 10.6% MR review. Triangulation performed well in survival risk factor analyses vs analyses using MR-identified recurrences.

**Conclusions:** Use of multiple recurrence algorithms in administrative data, in combination with selective examination of MR data, may improve recurrence data quality and reduce research costs.

Breast cancer is the most common cancer affecting women in the United States (1). Though survival has improved dramatically, research is needed to reduce recurrence. Studies thus require the identification of second breast cancer events (recurrences and second primary breast tumors). Chart abstraction is the gold standard for event ascertainment but is often prohibitively expensive and inefficient. Because of this, researchers have attempted to develop algorithms using service utilization documented within administrative health care data, available through Medicare or health maintenance organizations (HMOs) (2–5), to identify events. However, the utility of algorithms in research has been limited by errors in their identification (6). Improving upon the utility of algorithms will depend on the ability to evaluate them in varied datasets (7).

Recently, Chubak et al. (5) published work to determine second breast cancer events (SBCEs) and recurrences from indications of second events and health care utilization using data from a western Washington-based HMO (Group Health

Cooperative) and cancer registry records; the algorithms were designed to be readily adaptable for use in other HMO settings.

We evaluated the performance of these algorithms in Kaiser Permanente Northern California (KPNC) data, but we also considered how these algorithms might be adapted and improved because HMOs are increasingly developing databases that can be harmonized. We used the concept of triangulation, borrowed from navigational techniques that help determine a third point in space based on information about location of two other distinct points in space (8) and adapted by social science researchers to improve validity of study findings by using multiple methodologies to measure a phenomenon (9), to help improve algorithms. "High sensitivity" algorithms capture more recurrences than "high specificity" algorithms, which more accurately capture nonrecurrences. If the algorithms are of high quality, the algorithms will be mostly concordant, though there will be a fraction of case patients for whom algorithm results will disagree. It is this group for whom medical records (MRs) might be used to resolve patient status. If two high-quality algorithms are each correct most of the time in their identification of outcomes and positive predictive value (PPV) is high for both algorithms, the two algorithms will be concordant on most outcomes and the need for MR review will be reduced.

We evaluated classification statistics for several Chubak algorithms (5) plus triangulation in 225 women with breast cancer from the Women's Health Initiative (WHI) Cancer Survivor cohort, as well as in 1535 women from the Life After Cancer Epidemiology (LACE) cohort, receiving health care in KPNC. We tested the triangulation method for recurrences (vs SBCEs) as the more stringent test and because we had recurrence data for both cohorts. We also examined associations between several well-known breast cancer recurrence risk factors and recurrence in the LACE cohort to compare these methods (single algorithms, triangulation) against the MR gold standard.

## Methods

### Study Populations

#### LACE Cohort

The LACE Study cohort consisted of 2264 women diagnosed with invasive breast cancer between 1997 and 2000 who were recruited primarily from the KPNC Cancer Registry (83%) and the Utah Cancer Registry (12%) between 2000 and 2002. Further details are provided elsewhere (10). Eligibility criteria included: 1) age 18 to 70 years at enrollment; 2) diagnosis of early-stage primary breast cancer (stage I ≥1 cm, II, or IIIA); 3) enrollment between 11 and 39 months postdiagnosis; 4) completion of breast cancer treatment (except adjuvant hormonal therapy); 5) freedom from recurrence at enrollment; and 6) no history of other cancers in the five years prior to enrollment. We included the 1535 from LACE who were part of the KPNC population. This study was approved by the institutional review boards of KPNC and the University of Utah.

#### WHI Cancer Survivor Cohort

The design of the WHI has been previously described (11–13). Briefly, the WHI observational study (OS) is a multiethnic cohort of 93 676 postmenopausal women enrolled between 1993 and 1998 at 40 geographically diverse clinical centers throughout the United States. Eligibility criteria included: 1) age 50 to 79 years, 2) postmenopausal status, 3) willingness to provide informed consent, and 4) at least a three-year life expectancy. The WHI clinical trials (CT) included 68 132 women with the same basic eligibility who agreed to participate in controlled clinical trials of diet or hormone therapy and then a calcium/vitamin D trial.

At study baseline, participants provided detailed information on various risk factors through self-administered questionnaires. Medical history has been updated annually in the OS and every six months in the CT, by mail, and/or telephone questionnaires. Consenting participants have continued to provide information annually. Within the WHI, cancer cases were either self-reported or identified as a primary or contributing cause on a death certificate and were adjudicated by centralized MR review. More recently, the WHI Cancer Survivor Cohort was developed to assemble data and specimens related to cancer survival and survivorship.

We specifically evaluated algorithms in the 225 women from the WHI with local or regional invasive breast cancer who were also members of the KPNC HMO population. Human Subjects Review Committees at each participating institution approved protocols and documents.

### Data Collection

#### Breast Cancer Characteristics

In LACE, information on clinical factors including American Joint Committee on Cancer tumor stage (4th edition), tumor size, number of positive lymph nodes, hormone receptor status, and treatment was obtained through KPNC electronic data sources. In the WHI, characteristics (histology, extent of disease, hormone receptor status, HER2 status) were coded using the Surveillance, Epidemiology, and End Results (SEER) coding system (14,15).

### Recurrences

#### LACE

Recurrences were ascertained by: 1) a mailed semi-annual or annual (after April 2005) health status questionnaire asking participants to report events occurring in the preceding six or 12 months, respectively; 2) examination of electronic databases (eg, chemotherapy/infusion) for records of utilization after the completion of initial treatment; 3) ICD-9 codes indicating a cancer diagnosis after the initial diagnosis; and 4) examination of mortality records. Recurrences included locoregional cancer recurrences and distant recurrences/metastases. Total second breast cancer events (SBCE) additionally included development of contralateral and ipsilateral breast primaries. Nonrespondents were called by telephone to complete questionnaires. Medical records were reviewed to verify all outcomes.

#### WHI

For the 225 WHI participants within KPNC, medical records were reviewed for all participants; women were defined as having a recurrence if there was documentation of metastases or a recurrence in pathology reports; if there was a computed tomography scan, a bone scan, or a positron emission tomography scan with documentation that metastases were because of breast primary; progress notes by a medical doctor indicating a recurrence; or an oncology consult and/or oncology progress notes with a recurrence noted. We did not evaluate second breast primaries in the WHI.

#### Mortality

In LACE, participant deaths were determined through KPNC electronic data sources, a family member responding to a mailed questionnaire, or a phone call to the family. Copies of death certificates were obtained to verify primary and underlying

causes of death (International Classification of Diseases, 9th revision).

In the WHI, attribution of cause of death was based on MR review by physician adjudicators at local clinical centers, with central final adjudication (14). The National Death Index was crosschecked at two- to three-year intervals for participants who were lost to follow-up or had ambiguous cause of death based on nonreceipt of medical records or death certificate.

## Statistical Analyses

### SBCE and Recurrence Algorithms

Chubak et al. developed algorithms within an HMO for SBCEs, including both recurrences (defined as recurrences and second ipsilateral primaries) and second contralateral primaries, as well as specifically to identify recurrences. We implemented several "high sensitivity" and "high specificity" algorithms (described in Table 1) and computed sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for each algorithm in each data set. We evaluated SBCE algorithms against SBCEs and recurrence-specific algorithms against recurrences. Though SBCE algorithms were not developed specifically to identify recurrences, we also used these algorithms to identify recurrences because recurrences comprise the great majority of SBCE. We developed algorithms assuming access to SEER data, evaluating them against recurrences separately in the LACE study and against recurrences in the WHI data. Chubak et al. included patients with stages I-II breast cancer; we included patients with stages I-IIIA breast cancer.

### Triangulation

To test the triangulation method, we selected pairs of algorithms—one with "high sensitivity" and another with "high specificity" (based on the Chubak report)—and used information from the MRs to settle discrepancies between algorithms. To reduce the need for MR examination in future research, we were most interested in the triangulation of high sensitivity/high PPV and high specificity/high PPV algorithms because algorithms

with high PPV more accurately identify recurrences. For didactic purposes, we also triangulated the "extremely high sensitivity" algorithms (eg, 3 and 10) against the high-specificity algorithms (eg, 2 and 9).

We computed cross-tabulations of recurrence outcomes (yes or no as identified by an algorithm) for each pair of algorithms and further cross-classified these data with recurrences correctly classified using MR data (the gold standard). For those recurrences identified by one algorithm and not the other, we used MR data to appropriately classify participants regarding their recurrence status. To clarify, all recurrences in LACE and WHI were confirmed by MR prior to this study, which provided the gold standard against which algorithms were evaluated. However, triangulation would ultimately result in the review of a fraction of all MR data. We defined MR% as the percent of cohort participants requiring MR abstraction, based on the percent of the population in which two algorithms provide discordant results. After appropriately classifying outcomes, we computed classification statistics.

False positives (FP) were those identified by both algorithms as recurrences but were identified by MRs as nonrecurrences. False negatives (FN) were identified as recurrences by the MR but not identified by either algorithm. True positives (TP) included those identified as recurrences by at least one algorithm and the MR. True negatives (TN) included those identified as nonrecurrences by at least one algorithm and by MR.

### Comparative Analyses

We used Cox proportional hazards models (SAS PROC PHREG; SAS Institute, Cary, NC) (16,17) to conduct analyses of several survival risk factors and recurrence, comparing results when recurrences were identified by high-quality single algorithms, triangulation of those algorithms, and MR data. Specifically, we compared analyses of variables well known to be related to breast cancer recurrence with large main effects, including stage and lymph node status, as well as variables that have been associated with recurrence but have effect sizes small enough that they might be missed with sufficient outcome misclassification

**Table 1.** Algorithms to detect second breast cancer events and cancer recurrences using HMO medical record data, developed by Chubak et al. in the Group Health Cooperative HMO in western Washington state*

| Named in this paper | Location in Chubak et al. | Description | Uses SEER variables |
|---|---|---|---|
| Second breast cancer event | In main paper | | |
| Algorithm 1† | Figure 1 | High sensitivity | Yes |
| Algorithm 2† | Figure 2 | High specificity, PPV | Yes |
| Algorithm 3† | Figure 3 | Extremely high sensitivity | Yes |
| Algorithm 4 | Figure 4 | High sensitivity | No |
| Algorithm 5 | Figure 5 | High specificity, PPV | No |
| Algorithm 6 | Figure 6 | Extremely high sensitivity | No |
| Recurrence-specific | In supplementary analysis | | |
| Algorithm 7† | Supplementary Figure 2 | High sensitivity | Yes |
| Algorithm 8† | Supplementary Figure 3 | High specificity | Yes |
| Algorithm 9† | Supplementary Figure 4 | High specificity and PPV | Yes |
| Algorithm 10† | Supplementary Figure 5 | Extremely high sensitivity | Yes |
| Algorithm 11 | Supplementary Figure 6 | High specificity and PPV | No |
| Algorithm 12 | Supplementary Figure 7 | Extremely high sensitivity | No |

* Chubak et al. (5) outlined several approaches to defining second breast cancer events and recurrences using Surveillance, Epidemiology, and End Results (SEER) data and utilization codes. HMO = health maintenance organization; NPV = negative predictive value; PPV = positive predictive value; SEER = Surveillance, Epidemiology, and End Results Program.
† In this paper, we focused on algorithms making use of SEER data, ie, Algorithms 1, 2, 3 and 7, 8, 9, and 10.

(eg, alcohol intake and physical activity). Person-years of follow-up were counted from the date of diagnosis until the date of death, recurrence, or end of follow-up, whichever came first. We chose recurrence as the outcome of greatest research interest. We conducted tests of proportionality with variable by time interactions. All statistical tests were two-sided; the criterion for statistical significance was a P value of less than .05.

## Results

In the data reported in Chubak, 77.1% of SBCEs were recurrences and 22.9% were second contralateral primary breast cancers. In LACE data, 84.7% of SBCEs were recurrences and 15.3% were second primary breast cancers.

Classification statistics for SBCE algorithms (algorithms 1, 2, 3) in LACE were highly similar to published statistics. By contrast, with the exception of algorithm 9, which performed similarly to the published report, other recurrence-specific algorithms evaluated (algorithms 7, 8, 10) performed somewhat more poorly than reported (findings in LACE and WHI in Tables 2 and 3; published values in Table 3). This could be partially explained by the different approaches to defining recurrence. However, only 15% of second primaries (8 of the 1535 women) in LACE had an ipsilateral second primary, suggesting that statistics would not change markedly if we recategorized these as recurrences.

In further testing, we found the SBCE algorithms were more sensitive (range = 81.7%-88.5%) in identifying recurrences than were recurrence-specific algorithms (range = 57.1%-78.8%). Of the high-sensitivity algorithms, SBCE algorithm 1 performed best overall in LACE and WHI/KPNC in identifying recurrences, with reasonably high sensitivity (LACE: 88.5%, WHI: 86.7%), specificity (LACE: 90.7%, WHI: 89.7%), PPV (LACE: 65.2%, WHI: 56.5%), and NPV (LACE: 97.6%, WHI: 97.8%) (Tables 2 and 3). Of the high-specificity algorithms, algorithms 2 and 9 performed well; recurrence-specific algorithm 9 performed best overall in identification of recurrences (LACE: sensitivity = 74.6%, specificity = 97.3%, PPV = 84.7%, NPV = 95.1%; WHI: sensitivity = 63.3%, specificity = 95.9%, PPV = 70.4%, NPV = 94.4%) (Tables 2 and 3). In comparing algorithms, specificity and NPV were generally high, with the greatest tradeoffs occurring between sensitivity and PPV.

Classification improved when we triangulated any combination of algorithms; statistics matched or surpassed the best values of either algorithm. The percentage of cases requiring MR review varied from 4.4% to 25.7%; when we triangulated two algorithms with high PPV (eg, algorithms 1 and 2, 1 and 9, or 7 and 9), less MR review was required. When we triangulated recurrences using a high-sensitivity algorithm with lower PPV and a high-specificity/high-PPV algorithm, there was less overlap in the identification of recurrences and an opportunity to improve sensitivity beyond the higher sensitivity algorithm (eg, triangulation of algorithms 2 and 3 or 9 and 10). However, this improvement was modest and it resulted in a considerably higher need for MR examination to resolve discrepancies between algorithms. We found similar results when triangulating algorithms in the WHI/KPNC data (Table 4).

In the best combination for recurrence, in triangulating algorithms 7 and 9, classification (sensitivity = 81.3%, specificity = 99.7%, PPV = 98.1%, and NPV = 96.5%) surpassed that achieved by either algorithm alone (algorithm 7, sensitivity = 57.1%, specificity = 95.5%, PPV = 71.3%, NPV = 91.9%; algorithm 9, sensitivity = 74.6%, specificity = 97.3%, PPV = 84.7%, NPV = 95.1%). Triangulating these algorithms would result in an MR review of 10.6% of cases (Table 4).

### Comparative Analyses

We compared analyses of risk factors and recurrence identified by high-sensitivity algorithm 7, high-specificity/high-PPV algorithm 9, triangulation of algorithms 7 and 9, or MR data. For analyses of lymph node status, results obtained using the single algorithms 7 (4+ positive nodes, odds ratio [OR] = 2.27, 95% confidence interval [CI] = 1.32 to 3.89), algorithm 9 (OR = 3.18, 95% CI = 1.90 to 5.34), and triangulation (OR = 3.08, 95% CI = 1.81 to 5.23) were reasonably close to results obtained using MR data (hazard ratio [HR] = 2.85, 95% CI = 1.81 to 4.51). Analytic results for stage, physical activity, and alcohol and recurrence based on the single algorithms and triangulation approximated results obtained in survival analyses using MR data (Table 5). Results obtained using triangulation better approximated results obtained in gold standard analyses, which often, though not always, fell between results obtained with single algorithms. For example, physical activity of more than 2.5 hours per week was associated with a nonsignificant decrease in risk of recurrence in gold standard analyses (HR = 0.87, 95% CI = 0.64 to

**Table 2.** Evaluation of Chubak breast cancer algorithms for recurrence in the LACE study (n = 1535)*

| Single Chubak algorithms | MR Yes Alg Yes | MR No Alg Yes | MR Yes Alg No | MR No Alg No | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|---|---|---|---|
| Single algorithms | | | | | | | | |
| SBCE | | | | | | | | |
| Algorithm 1 | 266 | 76 | 41 | 1152 | 86.6 | 93.8 | 77.8 | 96.6 |
| Algorithm 2 | 244 | 31 | 63 | 1197 | 79.5 | 97.5 | 88.7 | 95.0 |
| Algorithm 3 | 257 | 344 | 50 | 884 | 83.7 | 72.0 | 42.8 | 94.6 |
| Recurrences | | | | | | | | |
| Algorithm 1 | 223 | 119 | 29 | 1164 | 88.5 | 90.7 | 65.2 | 97.6 |
| Algorithm 2 | 206 | 69 | 46 | 1214 | 81.7 | 94.6 | 74.9 | 96.3 |
| Algorithm 3 | 206 | 395 | 46 | 888 | 81.7 | 69.2 | 34.3 | 95.1 |
| Algorithm 7 | 144 | 58 | 108 | 1225 | 57.1 | 95.5 | 71.3 | 91.9 |
| Algorithm 8 | 201 | 73 | 51 | 1210 | 79.8 | 94.3 | 73.4 | 96.0 |
| Algorithm 9 | 188 | 34 | 64 | 1249 | 74.6 | 97.3 | 84.7 | 95.1 |
| Algorithm 10 | 186 | 288 | 66 | 995 | 73.8 | 77.6 | 39.2 | 93.8 |

* Alg = algorithm; LACE = Life After Cancer Epidemiology study; MR = medical record; NPV = negative predictive value; PPV = positive predictive value; SBCE = second breast cancer event.

**Table 3.** Evaluation of Chubak breast cancer algorithms for recurrence in WHI/KPNC data (n = 225) vs published values*

| | Published | | | | Kaiser (recurrence) | | | |
|---|---|---|---|---|---|---|---|---|
| Algorithm | Sensitivity | Specificity | PPV | NPV | Sensitivity | Specificity | PPV | NPV |
| Algorithm 1 | 96 | 95 | 74 | 99 | 86.7 | 89.7 | 56.5 | 97.8 |
| Algorithm 2 | 89 | 99 | 90 | 98 | 76.7 | 92.3 | 60.5 | 96.3 |
| Algorithm 3 | 99 | 81 | 43 | 100 | 83.3 | 81.5 | 41.0 | 97.0 |
| Algorithm 9 | 69 | 99 | 86 | 97 | 63.3 | 95.9 | 70.4 | 94.4 |
| Algorithm 10 | 99 | 81 | 37 | 100 | 76.7 | 85.6 | 45.1 | 96.0 |

* KPNC = Kaiser Permanente Northern California population; NPV = negative predictive value; PPV = positive predictive value; WHI = Women's Health Initiative Study.

**Table 4.** Classification statistics for recurrence with triangulation of algorithms, with modest medical record review

| | MR Yes Alg* Yes | MR No Alg Yes | MR Yes Alg No | MR No Alg No | Sensitivity | Specificity | PPV | NPV | MR% |
|---|---|---|---|---|---|---|---|---|---|
| LACE | | | | | | | | | |
| Alg 1,2,MR | 223 | 69 | 29 | 1214 | 88.5 | 94.6 | 76.4 | 97.7 | 4.4 |
| Alg 3,2,MR | 233 | 62 | 19 | 1221 | 92.5 | 95.2 | 79.0 | 98.5 | 25.7 |
| Alg 7,9,MR | 205 | 4 | 47 | 1279 | 81.3 | 99.7 | 98.1 | 96.5 | 10.6 |
| Alg 10,9,MR | 216 | 25 | 36 | 1258 | 85.7 | 98.1 | 89.6 | 97.2 | 21.5 |
| Alg 1,9,MR | 223 | 34 | 29 | 1249 | 88.5 | 97.3 | 86.8 | 97.7 | 7.8 |
| WHI/KPNC | | | | | | | | | |
| Alg 1,2,MR | 26 | 15 | 4 | 180 | 86.7 | 92.3 | 63.4 | 97.8 | 3.6 |
| Alg 3,2,MR | 27 | 12 | 3 | 183 | 90.0 | 93.8 | 69.2 | 98.4 | 14.7 |
| Alg 7,9,MR | 22 | 2 | 8 | 193 | 73.3 | 99.0 | 91.7 | 96.0 | 10.7 |
| Alg 10,9,MR | 25 | 4 | 5 | 191 | 83.3 | 97.9 | 86.2 | 97.4 | 16.0 |
| Alg 1,9,MR | 26 | 8 | 4 | 187 | 86.7 | 95.9 | 76.5 | 97.9 | 8.4 |

* MR% is defined as the percent of cohort participants requiring medical record abstraction, based on the percent of the population in which two algorithms provide discordant results. Alg = algorithm; KPNC = Kaiser Permanente Northern California population; LACE = Life After Cancer Epidemiology study; MR = medical record; NPV = negative predictive value; PPV = positive predictive value; SBCE = second breast cancer event; WHI = Women's Health Initiative study.

1.16) and with apparently decreased risks of recurrence defined by algorithm 7 (HR = 0.70, 95% CI = 0.50 to 0.97), algorithm 9 (HR = 0.91, 95% CI = 0.66 to 1.26), and by triangulation (HR = 0.83, 95% CI = 0.60 to 1.16).

## Discussion

The Chubak breast cancer algorithms performed well in KPNC data using SEER-based variables. SBCE algorithms were generally superior to the recurrence-specific algorithms for identifying both SBCEs and recurrences, though the high-specificity/high-PPV recurrence-specific algorithm 9 was the single best algorithm for identifying recurrences. Triangulating recurrences using a high-sensitivity/high-PPV and a high-specificity/high-PPV algorithm, with modest (≤10%) medical record examination, enabled us to match or surpass the highest sensitivity and specificity of each of two single algorithms, improve both PPV and NPV, and produce optimal analytical results. The development of recurrence algorithms and selective use of medical records may lead to high-quality data when full or substantial MR abstraction is infeasible.

Triangulation performed well, though a small percent was still misclassified. In post hoc examination, we evaluated the distributions of survival risk factors (eg, age, race, disease severity, treatment, behavioral risk factors) for FNs, FPs, TPs, and TNs. It appeared that FNs were less likely to receive treatment than others with similar disease characteristics. Conversely, the FPs appeared equally or slightly more likely to receive treatment compared with those with similar disease characteristics (data not shown). This suggests that algorithms may be more likely to misclassify outcomes when patients receive more or less treatment relative to others with similar disease severity. Thus, variables leading to undertreatment may appear to be associated with a lower risk of outcomes than is true, whereas variables predicting more treatment than expected may appear to predict a higher risk of outcomes, leading to subsequent analytic biases proportional to the degree of misclassification.

Comparative analyses of single algorithms and triangulation produced highly similar results to those obtained using Cox hazards regression models with MR-identified recurrences. However, with the algorithms as currently developed, information on time to event would not be available. Further work will be needed to estimate time to event, or investigators may choose to conduct logistic regression analyses.

Evaluation of methods will depend on the goals of the investigator and availability of time and resources. Many investigators may choose the method that optimizes PPVs, given generally high values for NPVs and, specificity, to improve analytic accuracy. Others may choose options leading to the identification of the greatest number of possible events. When chart review is possible, single high-sensitivity algorithms can result in excellent classification (eg, algorithm 1 in combination with MR review, sensitivity = 88.5%, specificity = 100%, PPV = 100%, NPV = 97.8%). However, this approach would require a 22.3% MR chart review. Triangulation of algorithms 1 and 2, though classification is imperfect, would improve classification over either algorithm separately and require a smaller 4.4% MR review. One possible goal may be to use algorithms without chart review. Triangulation may facilitate sensitivity analyses as seen in Table 5. Investigators could evaluate results using two single algorithms (eg, high sensitivity/high PPV and high specificity/high PPV), obtaining a sense of the range of likely effect

ARTICLE

**Table 5.** Hazard ratios by prognostic factor using different methods of identifying recurrences*

| | Method of recurrence identification | | | |
|---|---|---|---|---|
| Prognostic factor | MR: LACE HR (95% CI) | Algorithm 7 (Chubak) HR (95% CI) | Algorithm 9 (Chubak) HR (95% CI) | Triangulation: Algorithms 7, 9 and 10% MR HR (95% CI) |
| Stage | | | | |
| I | Ref | Ref | Ref | Ref |
| II | 1.69 (1.11 to 2.59) | 1.32 (0.83 to 2.10) | 1.61 (0.99 to 2.62) | 1.56 (0.95 to 2.57) |
| IIIa | 4.01 (2.10 to 7.67) | 4.15 (2.01 to 8.57) | 3.77 (1.82 to 7.80) | 4.33 (2.09 to 8.96) |
| Positive nodes | | | | |
| 0 | Ref | Ref | Ref | Ref |
| 1–3 | 1.40 (0.93 to 2.12) | 1.19 (0.74 to 1.93) | 1.74 (1.10 to 2.77) | 1.60 (0.99 to 2.58) |
| 4+ | 2.85 (1.81 to 4.51) | 2.27 (1.32 to 3.89) | 3.18 (1.90 to 5.34) | 3.08 (1.81 to 5.23) |
| Alcohol intake | | | | |
| ≤0.5g/d | Ref | Ref | Ref | Ref |
| <6g/d | 1.03 (0.74 to 1.44) | 0.99 (0.65 to 1.37) | 1.08 (0.75 to 1.55) | 1.01 (0.69 to 1.47) |
| ≥6g/d | 1.17 (0.84 to 1.64) | 1.02 (0.69 to 1.50) | 1.12 (0.77 to 1.62) | 1.13 (0.77 to 1.64) |
| Physical activity | | | | |
| <2.5h/wk | Ref | Ref | Ref | Ref |
| >2.5h/wk | 0.87 (0.64 to 1.16) | 0.70 (0.50 to 0.97) | 0.91 (0.66 to 1.26) | 0.83 (0.60 to 1.16) |

* Models adjusted for age and simultaneously for risk factors. LACE = Life After Cancer Epidemiology cohort; MR = medical record; Ref = referent.

estimates or deriving a single estimate that falls between values obtained by single algorithms; this may be a reasonable assumption when triangulation results in very high PPVs, as it did with algorithms 7 and 9.

More work is needed to determine if these algorithms and this method of triangulation optimally identify recurrences in different health care settings and in other types of administrative datasets, such as Medicare data. The code developed by Chubak and colleagues was readily applied within KPNC data. Given increasingly standardized data available across HMOs, our results suggest the use of algorithms may produce reasonably high-quality data; these methods could be readily reproduced in other HMO settings.

A study strength was the ability to compare results for algorithms against those obtained using gold standard methods in a well-established cohort and the ability to test in two datasets. A limitation to our study was a lack of readily usable diagnostics in the algorithms that would facilitate further testing and adaptation of algorithms. Further work should evaluate algorithms that don't rely on SEER data items. Finally, though work has been done to develop algorithms to detect breast cancer recurrence, the successful application of the triangulation method will depend on the development of algorithms in other cancers. Reducing the cost of cancer survivorship research using administrative data will, however, facilitate translational science.

## Funding

## Notes

## References

1. U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2011 Incidence and Mortality Web-based Report. Atlanta, GA: Department of Health and Human Services, Centers for Disease Control and Prevention, and National Cancer Institute; 2014.
2. Lash TL, Riis AH, Ostenfeld EB, Erichsen R, Vyberg M, Thorlacius-Ussing O. A validated algorithm to ascertain colorectal cancer recurrence using registry resources in Denmark. *Int J Cancer*. 2015;136(9):2210–2215.
3. Deshpande AD, Schootman M, Mayer A. Development of a claims-based algorithm to identify colorectal cancer recurrence. *Ann Epidemiol*. 2015;25(4):297–300.
4. Hassett MJ, Ritzwoller DP, Taback N, et al. Validating billing/encounter codes as indicators of lung, colorectal, breast, and prostate cancer recurrence using 2 large contemporary cohorts. *Med Care*. 2014;52(10):e65-e73.
5. Chubak J, Yu O, Pocobelli G, et al. Administrative data algorithms to identify second breast cancer events following early-stage invasive breast cancer. *J Natl Cancer Inst*. 2012;104(12):931–940.
6. Hassett MJ, Ritzwoller DP, Taback N, et al. Validating Billing/Encounter Codes as Indicators of Lung, Colorectal, Breast, and Prostate Cancer Recurrence Using 2 Large Contemporary Cohorts. *Med Care*. 2012;52(10):e65-e73.
7. Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care data algorithms. *J Clin Epidemiol*. 2012;65(3):343–349 e2.
8. Rothbauer P. Triangulation. In: Given L, editor. *The SAGE Encyclopedia of Qualitative Research Methods*. Thousand Oaks, CA: Sage Publications; 2008:892–894.
9. Bogdan RC, Biklen SK, editors. *Qualitative research in education: An introduction to theory and methods*. Boston, MA: Allyn & Bacon; 2006.
10. Caan B, Sternfeld B, Gunderson E, Coates A, Quesenberry C, Slattery ML. Life After Cancer Epidemiology (LACE) Study: a cohort of early stage breast cancer survivors (United States). *Cancer Causes Control*. 2005;16(5):545–556.
11. Matthews KA, Shumaker SA, Bowen DJ, et al. Women's health initiative. Why now? What is it? What's new? *Am Psychol*. 1997;52(2):101–116.
12. Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin Trials*. 1998;19(1):61–109.
13. Hays J, Hunt JR, Hubbell FA, et al. The Women's Health Initiative recruitment methods and results. *Ann Epidemiol*. 2003;13(9 Suppl):S18-S77.

14. Curb JD, McTiernan A, Heckbert SR, et al. Outcomes ascertainment and adjudication methods in the Women's Health Initiative. *Ann Epidemiol.* 2003;13(9 Suppl):S122-S128.

15. National Cancer Institute. Surveillance, Epidemiology, and End Results program. In; 2009.

16. Cox D. Regression models and life-tables. *J Royal Stat Soc (B).* 1972;34:187–220.

17. Cupples LA, D'Agostino RB, Anderson K, Kannel WB. Comparison of baseline and repeated measure covariate techniques in the Framingham Heart Study. *Stat Med.* 1988;7(1–2):205–222.