Published in final edited form as:

Med Care. 2014 October ; 52(10): e65–e73. doi:10.1097/MLR.0b013e318277eb6f.

Validating Billing/Encounter Codes as Indicators of Lung, Colorectal, Breast, and Prostate Cancer Recurrence using Two Large Contemporary Cohorts

Michael J. Hassett, MD, MPH*, Debra P. Ritzwoller, PhD*, Nathan Taback, PhD, Nikki Carroll, MS, Angel M. Cronin, MS, Gladys V. Ting, MSc, Deb Schrag, MD, Joan L. Warren, PhD, Mark C. Hornbrook, PhD, and Jane C. Weeks, MD, MSc

Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA (MJH, DS, and JCW); the Center for Health Research, Kaiser Permanente Northwest, Portland, OR (DPR); the University of Toronto, Toronto, ON (NT); the Dana-Farber Cancer Institute, Boston, MA (AMC and GVT); the National Cancer Institute, Bethesda, MD (JLW); and the Institute for Health Research, Kaiser Permanente Colorado, Denver, CO (MCH).

Abstract

Background—A substantial proportion of cancer-related mortality is attributable to recurrent, not de novo metastatic disease, yet we know relatively little about these patients. To fill this gap, investigators often use administrative codes for secondary malignant neoplasm or chemotherapy to identify recurrent cases in population-based datasets. However, these algorithms have not been validated in large, contemporary, routine care cohorts.

Objective—To evaluate the validity of secondary malignant neoplasm and chemotherapy codes as indicators of recurrence after definitive local therapy for stage I-III lung, colorectal, breast, and prostate cancer.

Research Design, Subjects & Measures—We assessed the sensitivity, specificity, and positive predictive value (PPV) of these codes 14- and 60-months after diagnosis using two administrative datasets linked with gold-standard recurrence status information: CanCORS/Medicare (diagnoses 2003-2005) and HMO/Cancer Research Network (diagnoses 2000-2005).

Results—We identified 929 CanCORS/Medicare patients and 5298 HMO/CRN patients. Sensitivity, specificity, and PPV ranged widely depending on which codes were included and the type of cancer. For patients with lung, colorectal, and breast cancer, the combination of secondary malignant neoplasm and chemotherapy codes was the most sensitive (75%-85%); no code-set was

Correspondence & Reprint Requests Michael J. Hassett, M.D., M.P.H., Center for Outcomes & Policy Research, Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215-5450 Telephone: (617) 632-6631; Fax: (617) 394-2747; michael_hassett@dfci.harvard.edu.

debra.ritzwoller@kp.orgnathan.taback@utoronto.canikki.m.carroll@kp.organgelm_cronin@dfci.harvard.edugladys_ting@dfci.harvard.edudeb_schrag@dfci.harvard.eduwarrenj@mail.nih.govmark.c.hornbrook@kp.orgjane_weeks@dfci.harvard.edu.

*Drs. Hassett and Ritzwoller contributed equally to this work and are considered co-first authors)

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

highly sensitive and highly specific. For prostate cancer, no code-set offered even moderate sensitivity (19%).

Conclusions—Secondary malignant neoplasm and chemotherapy codes could not identify recurrent cancer without some risk of misclassification. Findings based on existing algorithms should be interpreted with caution. More work is needed to develop a valid algorithm that can be used to characterize outcomes and define patient cohorts for comparative effectiveness research studies.

Introduction

Cancer is the second most common cause of death in the United States, accounting for approximately 600,000 lives lost per year.¹ While some cancer deaths are attributable to progression of the primary disease, many, if not the majority, are due to recurrent metastatic cancer that develops among patients who previously completed definitive therapy and were declared disease free.² Patients with recurrent metastatic cancer may differ significantly from those with primary metastatic cancer, because they previously received cancer-directed therapy, have co-morbid medical conditions from their prior disease or treatment, have a different distribution or extent of disease, or have different goals and values. Reliably identifying patients with recurrent cancer would facilitate efforts to compare the effectiveness of treatments given at the time of recurrence and would permit the use of recurrence-free-survival as an outcome in comparative effectiveness research studies.

Tumor registries generate extensive information about patients with de novo metastatic disease, but they cannot do as much to illuminate the experiences of patients with recurrent metastatic disease because they do not have robust mechanisms for identifying recurrence. Some tumor registries, such as those available from SEER (Surveillance, Epidemiology and End Results) and the Cancer Research Network (CRN), have been linked with administrative data and used to assess cancer patients' patterns-of-care and outcomes.³⁻⁹ Believing that these large, longitudinal, population-based datasets could also provide valuable information about recurrent cancer, investigators have developed algorithms that use the administrative codes for secondary malignant neoplasm and chemotherapy to identify patients with recurrent disease. While a number of studies have used these codes to indicate which patients have recurrent cancer¹⁰⁻¹³, few have assessed their validity as potential indicators of recurrence.

In a 2002 study of 89 acute myeloid leukemia patients treated at a large cancer center, Earle and colleagues¹⁴ assessed the ability of inpatient procedure codes for the administration of chemotherapy to identify relapse after a treatment gap of at least four months (N=22 patients). The algorithm yielded a sensitivity of 86%, a specificity of 99%, and a PPV of 95%. In a study of 45 patients with node-positive breast cancer who received adjuvant chemotherapy as part of a clinical trial, Lamont and colleagues¹⁵ tested the ability of secondary malignant neoplasm codes to identify disease-free survival (N=12 patients). At five years of follow up, the algorithm yielded a sensitivity of 100% and a specificity of 97%, with an area under the receiver operator characteristic (ROC) curve of 0.97.

Both studies relied on small, selected populations of patients with specific cancer types. The validity of secondary malignant neoplasm and chemotherapy codes in less selected cohorts and across multiple cancer types remains uncertain. Our goal was to test and validate these two previously published recurrence detection algorithms using two larger, more representative samples. Further evaluation of the methods currently being used to identify patients who have recurrent cancer is needed to ensure that our beliefs about this population are accurate. If administrative data fail to identify all patients who truly have recurrence or inappropriately classify patients as having recurrence the resulting misclassification could lead to erroneous conclusions.

Methods

Potential Indicators of Recurrence

The codes that we tested as potential indicators of cancer recurrence included the secondary malignant neoplasm codes described by Lamont and colleagues¹⁵ and the chemotherapy codes described by Earle and colleagues¹⁴. The secondary malignant neoplasm codes included the International Classification of Diseases-Version 9-Clinical Modification (ICD-9-CM) diagnosis codes for disease involving specific organ sites (i.e., 197-198.82).¹⁵ As described by Lamont and colleagues, the code for secondary malignant neoplasm at the site of the original cancer was excluded as an indicator of recurrence (e.g., when looking for recurrence after a lung cancer diagnosis, the code for secondary malignant neoplasm of the lung [197.0] was excluded), because we could not determine whether it referred to recurrent disease, the original cancer, or a new primary. Also similar to the previous study, we conducted a sensitivity analysis in which we evaluated the impact of adding the secondary malignant neoplasm codes for lymph nodes and other non-specific sites (i.e., 196 and 198.89) **to see if they improved sensitivity**. The chemotherapy codes included generic codes for the administration of anti-cancer agents and specific codes for cytotoxic and targeted medications commonly used to treat cancer. Similar to the algorithm described by Earle and colleagues¹⁴, we included ICD-9-CM, Current Procedure Terminology (CPT), Healthcare Common Procedure Coding System (HCPCS), Diagnosis-Related Groups (DRG), revenue center, and Berenson-Eggers Type of Service (BETOS) codes [see Table, Supplemental Digital Content 1, which provides a complete list of all codes used]. To improve sensitivity, we added National Drug Codes (NDC); and to improve specificity we excluded hormonal therapy medications (because they are frequently used as part of prolonged adjuvant therapy or for non-oncologic indications) and some chemotherapy medications that are rarely used to treat recurrent cancer (e.g., topical 5-fluorouracil). To avoid the chance that chemotherapy given for adjuvant purposes would be construed as indicative of recurrence, we excluded chemotherapy occurring within 6 months of the index diagnosis for lung cancer and within 12 months for breast and colorectal cancer. These “blackout” periods were based on the duration of standard adjuvant chemotherapy regimens and a review of contemporary patterns-of-care. Finally, we assessed the performance characteristics of the two code-sets combined (i.e., either secondary malignant neoplasm or chemotherapy codes).

Data Sources

We assessed the validity of these codes using two datasets that linked administrative codes, which provided potential indicators of recurrence, with information abstracted from the medical record, which provided gold standard recurrence status. One dataset contained patients cared for in a fee-for-service setting (CanCORS/Medicare) and the other contained patients cared for in a health maintenance organization-based Cancer Research Network (HMO/CRN). We hypothesized that a code-based algorithm would be a more sensitive indicator of recurrence when applied to the fee-for-service versus the capitation environment.

CanCORS/Medicare—The Cancer Care Outcomes Research and Surveillance (CanCORS) Consortium is a large, prospective, population and health-system based study of the care provided to and outcomes experienced by patients from across the United States with lung and colorectal cancer diagnosed 2003-2005. The CanCORS dataset includes information from surveys of patients or their surrogates, and from the abstraction of patients' medical records. Details of the study design, population, and data collection procedures have been described elsewhere.¹⁶ After obtaining institutional review board permission, and with the consent of the Centers for Medicare and Medicaid Services, CanCORS records were linked with Medicare claim files from 2002-2007. The Dana-Farber/Harvard Cancer Center institutional review board provided oversight for the CanCORS/Medicare portion of this project.

HMO/CRN—The CRN is a consortium of 14 health care organizations affiliated with the HMO Research Network and the National Cancer Institute (<http://crn.cancer.gov/>). The CRN's Virtual Data Warehouse (VDW) was the primary data source for this analysis. Described in detail elsewhere^{17,18}, the VDW is a distributed data system using standard data formats and definitions where each HMO retains local control of its data, but a central programmer can direct extraction and analysis of data across all participating sites. It includes data on patient demographics, health plan enrollment, SEER-compatible tumor characteristics, and cancer treatments. For this analysis, we used data from the two HMO/CRN sites that have “in-house” tumor registries staffed by certified tumor registrars who collect high quality data regarding recurrence status. Both sites, Kaiser Permanente Colorado (KPCO), Denver, CO and Kaiser Permanente Northwest (KPNW), Portland, OR, provide comprehensive health services through closed panel delivery models with salaried physicians. Institutional review boards from both participating CRN sites approved this project.

Study Populations & Data Extraction

For both datasets we identified patients with newly diagnosed stage I-III lung, colorectal, breast, and prostate cancer. Patients with locally advanced or primarily progressive disease and those who did not complete definitive local-regional therapy were excluded, because we wanted to focus on patients for whom recurrence detection would be relatively easy (i.e., we only wanted to test the validity of these codes in situations that maximized their potential sensitivity and specificity). So, patients with stage IIIb lung cancer, those who did not have definitive local-regional therapy within 12 months of their index cancer diagnosis, and those

who were never noted to be disease-free were excluded. Patients who were missing medical record abstraction data, those who had unknown recurrence status, and those who had documentation of recurrent disease in the medical record before they had definitive local therapy were also excluded. Patients with no documented recurrence who died from causes other than cancer were considered non-recurrent.

For the CanCORS/Medicare sample, patients had to be continuously enrolled in Medicare Parts A & B from the month of cancer diagnosis to 18 months later or death (whichever came first), where continuous enrollment was defined as no more than 2 consecutive months without enrollment. Patients enrolled in Medicare Advantage plans were excluded out of concern that their claims data were incomplete. Patients had to have at least one Medicare claim to ensure they were seeking care for their index cancer. Medical record abstraction data were used to characterize patients' gold-standard recurrence status. The suggestion of recurrence in a radiology report, without confirmation in a clinician's note, was not considered sufficient to indicate recurrence. Characteristics of the patients (e.g., age, race/ethnicity, education, income, co-morbid diagnoses^{19,20}, etc.) and their diagnoses (e.g., stage, grade, etc.) were derived from the medical record abstraction and patient/surrogate survey files. Administrative codes were extracted from the Medicare hospital (MedPAR), physician (Carrier), and Outpatient facility claims files (N.B.: NDC codes were not available from the Medicare files during the years encompassed by this analysis). To assign a date to each Medicare code the "claim through date" was used for all files except MedPAR where the discharge date was used.

The HMO/CRN database, unlike CanCORS/Medicare, continuously enrolls patients. For this analysis, the HMO/CRN cohort included incident cancer cases diagnosed between January 1, 2000 and December 31, 2005. Cases were linked to enrollment, census, pharmacy, procedure, and diagnosis data from 12 months prior to the incident cancer diagnosis until death, disenrollment, or the end of the study period (December 31, 2008). Gold **standard** recurrence status was ascertained by tumor registrars during routine chart surveillance efforts. Characteristics of the patients and their diagnoses were ascertained using the VDW tumor registry, census, pharmacy, procedure, and diagnosis databases. The majority of the HMO/CRN encounter codes were derived from events documented in KP HealthConnect (an EpicCare-based electronic medical record); a smaller proportion (<15%) were identified using claims data when services were delivered by contract providers. Health-care-event specific data were extracted from the VDW files. Like CanCORS/Medicare, this included ICD-9, CPT, HCPCS, and DRG codes; while revenue center and BETOS codes were not available for HMO/CRN patients, NDC codes were. Depending on the data source, the "procedure date," "encounter date," or "dispense date" was used to assign a time to each code.

Analysis

To evaluate the validity of these administrative codes as indicators of recurrence we calculated their sensitivity and specificity. For the CanCORS/Medicare cohort, we assessed recurrence 14 months after the index cancer diagnosis because the CanCORS chart abstraction process (which provided the gold-standard recurrence status) was designed to

capture medical records up to 15 months post-diagnosis. Longer follow up was available for the HMO/CRN cohort, so for these patients we assessed recurrence at both 14 and 60 months. For each time point, the cohort included patients who were followed for at least that number of months as well as any patients who developed recurrence or died prior to that time regardless the duration of follow-up. A patient was classified as having an indicator of recurrence if at least one relevant code was present on or before the final month of the assessment period.

We also used the HMO/CRN cohort to calculate the positive predictive value (PPV), defined as the proportion of patients who had recurrence among those who had codes indicative of recurrence. The PPV analysis was done only for the HMO/CRN cohort, because the CanCORS/Medicare cohort was smaller and the selective censoring of non-recurrent cases suggested that the 14-month CanCORS/Medicare sample might not be representative of the broader population. We explored trends in sensitivity, specificity, and PPV across the different data sources, cancer types, stages, and time points.

Results

We identified 929 CanCORS/Medicare and 5298 HMO/CRN patients, all of whom received definitive local therapy for stage I-III lung, colon, breast, or prostate cancer and had administrative data linked to gold-standard recurrence status. Major reasons for exclusion included having stage IV disease, not having definitive local therapy within 12 months, and, among CanCORS/Medicare patients, being enrolled in a managed care organization (Table 1). There were 115 confirmed recurrences in the CanCORS/Medicare cohort and 621 confirmed recurrences in the HMO/CRN cohort. Compared to CanCORS/Medicare patients, HMO/CRN patients had greater cumulative recurrence rates consistent with their longer median follow-up (14 months for CanCORS/Medicare **versus** 60 months for HMO/CRN). Not surprisingly, the CanCORS/Medicare cohort contained a smaller proportion of <65 year-olds, and a larger proportion of patients who were non-Hispanic Black, had less than a high school education, had an estimated annual income of less than \$20,000, and had a higher co-morbidity index (Table 2).

Among patients with lung, colorectal and breast cancer, the combination of secondary malignancy and chemotherapy codes consistently offered the highest sensitivity; chemotherapy codes alone offered the lowest sensitivity and secondary malignancy codes alone were the most specific (Table 3). No one set of codes consistently demonstrated both the highest sensitivity and the highest specificity. For patients with prostate cancer, none of the code-sets offered a sensitivity >19% and all of the code-sets offered a specificity >83%. Comparing the pattern and magnitude of results for the 14- versus 60-month time points and for the CanCORS/Medicare **versus** HMO/CRN datasets, the findings did not appear to change substantially. For patients with lung, colorectal, and breast cancer, adding the ICD-9 codes for secondary malignant neoplasm involving lymph nodes and other non-specific sites (i.e., 196; 198.89) to the core set of secondary malignant neoplasm codes increased sensitivity, but decreased specificity (see Figure 1 – **left panel**). When the results for the combination of secondary malignancy and chemotherapy codes were stratified by stage at

diagnosis, we found that as stage increased sensitivity increased and specificity decreased (see Figure 1 – **right panel**).

Not surprisingly, among lung, colorectal, and breast cancer patients from the HMO/CRN cohort, the most specific code-set generated the highest PPV's (94%, 94%, and 87%, respectively for secondary malignancy codes at 60 months) (Figure 2). The most sensitive code-set, the combination of secondary malignancy and chemotherapy codes, yielded PPV's of 72% for lung cancer, 53% for colorectal cancer, and 30% for breast cancer. The PPV tended to be higher when recurrence was more prevalent (i.e., lung **versus** colorectal or breast cancer; 60-month **versus** 14-month analysis; stage III **versus** stage I at diagnosis [stage-specific PPV's not shown]). No code-set offered a PPV >31% for prostate cancer.

Conclusion

We performed a validation study of secondary malignant neoplasm and chemotherapy codes as potential indicators of recurrence for patients with lung, colorectal, breast, and prostate cancer. Excluding patients who did not complete definitive local therapy or who had primary progressive disease allowed us to focus on patients for whom we thought recurrence detection would be less challenging. Even so, for patients with lung, colorectal, and breast cancer secondary malignancy codes offered a sensitivity of only 56%-74% and chemotherapy codes offered a sensitivity of only 28%-62% across different time-points and datasets. The combination of secondary malignant neoplasm or chemotherapy codes yielded more desirable sensitivities (75%-85%) and specificities (70%-88%), but there was still room for improvement. No code-set detected recurrent prostate cancer reliably. To our surprise, there was no consistent difference in the results from the fee-for-service environment of Medicare compared with the capitation environment of the CRN.

There are several reasons why our findings may differ from those published previously. It is not surprising that chemotherapy codes were better at identifying patients who received re-induction chemotherapy for AML compared to patients with recurrent solid organ malignancies.¹⁴ Many patients with recurrent lung, colorectal, breast, and prostate cancer do not receive chemotherapy at the time of recurrence; some are treated with another modality of therapy (e.g., hormonal therapy, radiation therapy, surgery, etc.) and others receive no cancer-directed therapy at all. Our cohort differed substantially from that described by Lamont and colleagues¹⁵, which only included patients with node-positive breast cancer who were 65 years old and had participated in a clinical trial. Even when we restricted our analysis to patients with stage II-III breast cancer the secondary malignant neoplasm codes still appeared to be less sensitive (65%-78%) than previously reported. Documentation of the administrative codes used as potential indicators of recurrence may have been more likely among clinical trial participants, patients receiving care at institutions that offer clinical trials, those who were 65 years old, or those who were treated in the more distant past. Several studies have assessed the ability of administrative codes to identify patients who developed either recurrent or second primary cancer.^{21,22} While these studies reported sensitivities and specificities that were high, the use of a combined outcome (i.e., recurrence or second primary cancer) makes it hard to assess how well these algorithms would detect recurrence alone.

Our analysis highlights a number of challenges with identifying patients who have recurrent cancer. First, secondary malignant neoplasm and chemotherapy codes alone offered disappointingly low sensitivities. To improve sensitivity, it may be necessary to add additional indicators, such as imaging, radiation therapy, hormonal therapy, or hospice codes. Some codes, such as those for imaging and hospice, could be particularly helpful at identifying patients who receive no cancer-directed therapy at the time of recurrence. Second, in light of the observation that cancer-type affected the performance characteristics of billing/encounter codes, developing one recurrence detection algorithm that works for all cancer types may be impractical. An alternate approach would be to start with an all-encompassing list of codes for common solid organ cancers, and selectively exclude those that reduce specificity and do not increase sensitivity on a cancer-by-cancer basis. Third, identifying patients who have recurrent disease is not enough; it is also important to characterize the time when recurrence occurs. Finally, it is hard to identify a specific threshold for sensitivity and specificity that should be considered good enough. The optimal threshold may depend on the goal of the analysis, including whether it is more important to optimize sensitivity or specificity. If used alone, an algorithm with a low PPV could result in misclassification such that efforts to make comparisons across groups would yield biased estimates of statistical parameters such as odds ratios.

Our results are limited by the potential for inaccuracies when classifying patients' gold-standard recurrence status. Given the stringent definitions used to identify gold-standard recurrence status, the risk of inappropriately classifying patients as non-recurrent according to the gold-standard is greater than the risk of inappropriately classifying patients as recurrent. Findings from the CanCORS/Medicare cohort are limited by relatively short follow-up, however it was reassuring that the 14- and 60-month findings from the HMO/CRN cohort were similar. Some of the codes used to indicate recurrence can also be used to document initial cancer care (e.g., adjuvant chemotherapy, regional lymph node involvement, etc.). We tried to address this issue by excluding chemotherapy codes that occurred within 6 months after a lung cancer diagnosis or within 12 months after a colorectal/breast cancer diagnosis; the optimal "black-out" window is unknown. Unlike one of the original studies¹⁵, we assessed the performance of administrative codes as indicators of recurrence alone, not recurrence or death. Given the small number of deaths in our cohort and the pattern of code documentation, changing the outcome to include death would likely have had either no impact or a negative effect on sensitivity and specificity. The large sample size, different cancer types, and various practice settings included in our analysis are important strengths.

Investigators who use secondary malignant neoplasm or chemotherapy codes to identify cancer recurrence should be aware of their potential for misclassification. The optimal thresholds for sensitivity and specificity can be determined only by investigators in consideration of their individual research goals, but there does appear to be room to improve the performance of currently available recurrence detection methods. While secondary malignant neoplasm and chemotherapy codes may be suboptimal for some investigators, they do offer potential and could serve as the foundation for a more sophisticated recurrence algorithm that incorporates additional indicators and accounts for the primary cancer type²³.

Rather than providing an absolute characterization of recurrence status (yes **versus** no), it could be more useful to develop a flexible tool that generates a probability of having recurrence in a defined time period. Continued efforts to develop an algorithm that identifies recurrent cancer using administrative data are worthwhile, because while success may not be assured the alternative, asking registries to track recurrence status, would likely be cost prohibitive. A well-validated, claims-based algorithm could provide valuable information about the experiences and outcomes of the many patients with recurrent cancer and help realize the full potential of administrative databases for comparative effectiveness research.

LIST OF SUPPLEMENTAL DIGITAL CONTENT

Supplemental Digital Content 1 – Table – Provides a complete list of all codes used.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

This work was supported by a grant from the National Cancer Institute (RC2 CA148185-01 to JCW and DPR) and an NCI cooperative agreement (U19 CA79689, Cancer Research Network). The American Society of Clinical Oncology (Career Development Award) and Susan G. Komen for the Cure (Career Catalyst Award) provided salary support to MJH. The work of the CanCORS consortium was supported by grants from the National Cancer Institute (NCI) to the Statistical Coordinating Center (U01 CA093344) and to the Primary Data Collection and Research Centers (Dana-Farber Cancer Institute/Cancer Research Network U01 CA093332; Harvard Medical School/Northern California Cancer Center U01 CA093324; University of Iowa U01 CA01013; University of North Carolina U01 CA093326); and by a Department of Veterans Affairs grant to the Durham VA Medical Center VA HSRD CRS-02-164. The funders did not have any involvement in the design of the study; the collection, analysis, or interpretation of the data; the writing of the article; or the decision to submit the article for publication.

References

1. Kochanek K, Xu J, Murphy SL, Minino AM, Kung HC. Deaths: Preliminary Data for 2009. National Vital Statistics Report. 2011; 59(4):1–51. http://www.cdc.gov/nchs/data/nvsr/nvsr59/nvsr59_04.pdf.
2. Howlader, N.; Noone, AM.; Krapcho, M., et al., editors. SEER Cancer Statistics Review, 1975-2009 (Vintage 2009 Populations). National Cancer Institute; Bethesda, MD: 2012. http://seer.cancer.gov/csr/1975_2009_pops09/, based on November 2011 SEER data submission, posted to the SEER web site, 2012
3. Jacobson M, Earle CC, Newhouse JP. Geographic variation in physicians' responses to a reimbursement change. N Engl J Med. Dec 1; 2011 365(22):2049–2052. [PubMed: 22047517]
4. Hu CY, Delclos GL, Chan W, Du XL. Assessing the initiation and completion of adjuvant chemotherapy in a large nationwide and population-based cohort of elderly patients with stage-III colon cancer. Med Oncol. Dec. 2011; 28(4):1062–1074.
5. Gray SW, Landrum MB, Lamont EB, McNeil BJ, Jaklitsch MT, Keating NL. Improved outcomes associated with higher surgery rates for older patients with early stage nonsmall cell lung cancer. Cancer. Jul 28, 2011
6. Chavez-MacGregor M, Zhao H, Fang S, Srokowski TP, Hortobagyi GN, Giordano SH. Complications associated with erythropoietin-stimulating agents in patients with metastatic breast cancer: a Surveillance, Epidemiology, and End Results-Medicare study. Cancer. Aug 15; 2011 117(16):3641–3649. [PubMed: 21656514]

7. Doubeni CA, Field TS, Buist DS, et al. Racial differences in tumor stage and survival for colorectal cancer in an insured population. *Cancer*. Feb 1; 2007 109(3):612–620. [PubMed: 17186529]
8. Field TS, Bosco JL, Prout MN, et al. Age, comorbidity, and breast cancer severity: impact on receipt of definitive local therapy and rate of recurrence among older women with early-stage breast cancer. *J Am Coll Surg*. Dec; 2011 213(6):757–765. [PubMed: 22014658]
9. McCahill LE, Single RM, Aiello Bowles EJ, et al. Variability in reexcision following breast conservation surgery. *JAMA*. Feb 1; 2012 307(5):467–475. [PubMed: 22298678]
10. Stokes ME, Thompson D, Montoya EL, Weinstein MC, Winer EP, Earle CC. Ten-year survival and cost following breast cancer recurrence: estimates from SEER-medicare data. *Value Health*. Mar-Apr; 2008 11(2):213–220. [PubMed: 18380633]
11. Schootman M, Jeffe DB, Gillanders WE, Aft R. Racial disparities in the development of breast cancer metastases among older women: a multilevel study. *Cancer*. Feb 15; 2009 115(4):731–740. [PubMed: 19130463]
12. Gooden KM, Howard DL, Carpenter WR, et al. The effect of hospital and surgeon volume on racial differences in recurrence-free survival after radical prostatectomy. *Med Care*. Nov; 2008 46(11):1170–1176. [PubMed: 18953228]
13. Cummings KC 3rd, Xu F, Cummings LC, Cooper GS. A comparison of epidural analgesia and traditional pain management effects on survival and cancer recurrence after colectomy: a population-based study. *Anesthesiology*. Apr; 2012 116(4):797–806. [PubMed: 22273991]
14. Earle CC, Nattinger AB, Potosky AL, et al. Identifying cancer relapse using SEER-Medicare data. *Medical Care*. 2002; 40(8 Suppl):IV–75–81.
15. Lamont EB, Herndon JE 2nd, Weeks JC, et al. Measuring disease-free survival and cancer relapse using Medicare claims from CALGB breast cancer trial participants (companion to 9344). *Journal of the National Cancer Institute*. 2006; 98(18):1335–1338. [PubMed: 16985253] [erratum appears in *J Natl Cancer Inst*. 2006 Nov 1;98(21):1584].
16. Ayanian JZ, Chrischilles EA, Fletcher RH, et al. Understanding cancer treatment and outcomes: the Cancer Care Outcomes Research and Surveillance Consortium. *Journal of Clinical Oncology*. 2004; 22(15):2992–2996. [PubMed: 15284250] [see comment][erratum appears in *J Clin Oncol*. 2004 Dec 15;22(24):5026].
17. Hornbrook MC, Hart G, Ellis JL, et al. Building a virtual cancer research organization. *J Natl Cancer Inst Monogr*. 2005; (35):12–25. [PubMed: 16287881]
18. Ritzwoller DP, Carroll N, Delate T, et al. Validation of Electronic Data on Chemotherapy and Hormone Therapy Use in HMOs. *Med Care*. Apr 23.2012
19. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases*. 1987; 40(5):373–383. [PubMed: 3558716]
20. Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW. Can comorbidity be measured by questionnaire rather than medical record review? *Medical Care*. 1996; 34(1):73–84. [PubMed: 8551813]
21. McClish D, Penberthy L, Pugh A. Using Medicare claims to identify second primary cancers and recurrences in order to supplement a cancer registry. *Journal of Clinical Epidemiology*. 2003; 56(8):760–767. [PubMed: 12954468]
22. Chuback 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
23. Kimmick GG, Camacho F, Hwang W, Mackley H, Stewart J, Anderson RT. Adjuvant Radiation and Outcomes After Breast Conserving Surgery in Publicly Insured Patients. *Journal of geriatric oncology*. Apr 1; 2012 3(2):138–146. [PubMed: 22712029]

Figure 1. Sensitivity and Specificity of Administrative Codes for Recurrent Lung, Colorectal, and Breast Cancer

Left Panel: Comparing the core set of secondary malignant neoplasm codes with the core set plus the addition of codes for lymph nodes and non-specific sites. Right Panel: Stage stratified results for the combination of either secondary malignant neoplasm or chemotherapy codes. The CanCORS dataset did not include breast cancer patients, so these columns are noted as not applicable (NA). Prostate cancer results were substantially lower and varied little by type of code or stage, so are not displayed.

Figure 2. Positive Predictive Values for the HMO/CRN Sample

Positive predictive values (PPV's) at 14- and 60-months using secondary malignancy codes, chemotherapy codes, or the combination (either secondary malignancy or chemotherapy).

Table 1**Cohort Assembly**

	Lung	Colorectal		
CanCORS participants with medical record abstraction data and linked Medicare claims data	2402	1917		
Stage I-III cancer (excluding IIIb lung cancer)	1054 (44%)	1538 (80%)		
Had definitive local therapy within 12 months of diagnosis	660 (63%)	1459 (95%)		
Recurrence status (yes or no) recorded in medical record following definitive local therapy [*]	563 (85%)	1182 (81%)		
No death from cancer diagnosis without documented recurrence	555 (99%)	1175 (99%)		
No HMO enrollment [†]	405 (73%)	801 (68%)		
No more than 2 consecutive months without both Part A and Part B Medicare enrollment in the time window [†]	312 (77%)	622 (78%)		
At least 1 Medicare claim from 30 days pre-diagnosis to 14 months post-diagnosis	309 (99%)	620 (99%)		
Any recurrence documented in CanCORS	59 (19%)	56 (9%)		

	Lung	Colorectal	Breast	Prostate
Age 21 years old diagnosed between 2000 and 2005 with a first non-metastatic cancer	1043	1525	4076	2197
Stage I-III (excluding IIIb lung cancer)	687 (66%)	1441 (94%)	3266 (80%)	2197 (100%)
Had definitive local therapy within 12 months of diagnosis	546 (79%)	1420 (99%)	3220 (99%)	1830 (83%)
Documented free of disease after completing initial therapy	375 (69%)	1166 (82%)	2782 (86%)	1159 (63%)
No death from cancer diagnosis without documented recurrence	333 (89%)	1088 (93%)	2726 (98%)	1151 (99%)
Any recurrence documented in tumor registry	129 (39%)	191 (18%)	212 (8%)	89 (8%)

^{*} Patients were excluded because they had unknown recurrence status after definitive local therapy, were noted to have had a recurrence before definitive local therapy, or were last noted to be recurrence-free before definitive local therapy.

[†] In the period from diagnosis to the earlier of death, 12/31/2007, or 14 months post-diagnosis.

Table 2

Patient Characteristics

	CanCORS/Medicare Sample			HMO/CRN Sample		
	Lung n=309 %	Colorectal n=620 %	Lung n=333 %	Colorectal n=1088 %	Breast n=2726 %	Prostate n=1151 %
Age at Diagnosis						
<50			8	9	21	4
50-54			7	8	14	12
55-59	13	7	11	11	15	19
60-64			15	12	14	25
65-69	24	22	21	15	12	22
70+	63	71	38	45	24	18
Sex[*]						
Male	52	50	48	49	NR	100
Female	48	50	52	51	NR	-
Race / Ethnicity[†]						
Non-Hispanic White	85	73	77	71	81	80
Non-Hispanic Black	6	18	4	2	2	4
Hispanic or Asian	NR	5	6	5	4	4
Other / Unknown	NR	4	13	22	13	12
Education[‡]						
<High school	21	27	(12)	(14)	(15)	(12)
High school, tech or some college	26	20	(50)	(51)	(53)	(50)
College or higher	17	23	(38)	(35)	(32)	(38)
Other / Unknown	36	30	-	-	-	-
Annual Income (US\$)[§]						
<20,000	27	26	(<1)	-	(<1)	(<1)
20,000 to <40,000	28	22	(17)	(14)	(12)	(10)
40,000 to <60,000	14	10	(45)	(47)	(41)	(40)
60,000	9	13	(34)	(37)	(44)	(47)

	CanCORS/Medicare Sample		HMO/CRN Sample			
	Lung n=309 %	Colorectal n=620 %	Lung n=333 %	Colorectal n=1088 %	Breast n=2726 %	Prostate n=1151 %
Missing	22	29	(4)	(2)	(3)	(3)
Health Insurance						
Medicare	98	96	61	60	37	40
Medicaid/low income	16	15	2	2	1	<1
Private/HMO	66	57	62	65	79	80
Other supplemental	10	8	-	1	<1	-
Co-morbidity Index[§]						
0	22	43	48	63	75	77
1	38	31	35	22	19	17
2	28	15	10	9	4	4
3	12	10	7	6	2	2
Year of Diagnosis						
2000	-	-	15	16	16	16
2001	-	-	15	16	17	18
2002	-	-	18	16	18	20
2003	26	29	16	16	17	18
2004	67	68	19	18	16	17
2005	7	3	17	17	17	12
Stage						
I	72	31	59	30	58	1
II	17	36	17	38	34	89
III	11	33	24	32	8	11

* The vast majority of patients with breast cancer (>99%) were female, but a small number were male. The proportions of men and women with breast cancer were not reported to comply with the terms of the data use agreement stating that cells with 10 patients not be reported. The proportions were noted as NR, indicating that the values were not reported.

[†] The proportion of CanCORS/Medicare lung cancer patients of Hispanic or Asian and other/unknown race/ethnicity were not reported (NR) due to the aforementioned data use agreement.

[‡] Education data for the HMO/CRN cohort were obtained from block-level Census data; the results represent the mean proportion of people who attained each level of education across the census tracts in which the patients actually resided. The results are in parentheses to highlight this difference.

[§] Income data for the HMO/CRN cohort were obtained from block-level Census data. The results are in parentheses to highlight this difference.

// Co-morbidity Index derived using methods developed by Charlson, Katz and colleagues^{19,20}, excluding cancer diagnoses.

Table 3

Sensitivity and specificity of claims for cancer recurrence *

	CanCORS/Medicare 14 Months		HMO/CRN 14 Months		HMO/CRN 60 months	
	Sensitivity (95%CI)	Specificity (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)
Lung	N=48	N=140	N=65	N=260	N=123	N=146
Secondary Malignancy [‡]	63 [47-76]	93 [87-97]	57 [45-68]	98 [96-99]	66 [57-74]	97 [92-99]
Chemotherapy [‡]	48 [31-61]	76 [68-83]	51 [43-68]	90 [84-91]	28 [21-36]	91 [85-95]
Either [§]	77 [63-88]	70 [62-77]	85 [74-91]	88 [84-91]	85 [78-91]	72 [64-79]
Colorectal	N=42	N=307	N=73	N=968	N=184	N=640
Secondary Malignancy	71 [55-84]	91 [87-94]	56 [45-67]	98 [97-99]	74 [68-80]	99 [98-99]
Chemotherapy	38 [24-54]	89 [85-92]	48 [37-59]	76 [73-79]	59 [52-66]	80 [77-83]
Either	81 [66-91]	83 [78-87]	75 [64-84]	75 [72-78]	83 [77-88]	79 [76-82]
Breast	N=43	N=2547	N=195	N=1667		
Secondary Malignancy			58 [43-72]	100 [99-100]	60 [53-67]	99 [98-99]
Chemotherapy			53 [39-67]	78 [76-80]	62 [55-69]	79 [77-81]
Either			81 [67-90]	78 [76-80]	78 [72-83]	79 [76-80]
Prostate	N=17	N=1079	N=77	N=729		
Secondary Malignancy			0 [0-18]	99 [98-100]	10 [5-19]	98 [96-98]
Chemotherapy			6 [1-27]	98 [97-98]	3 [1-9]	98 [97-99]
Either			6 [1-27]	97 [96-99]	19 [12-30]	83 [80-86]

* The CanCORS/Medicare sample included patients with lung or colorectal cancer. The HMO/CRN sample included patients with lung, colorectal, breast or prostate cancer. The N for sensitivity represents the number of patients who recurred, whereas the N for specificity represents the number of patients who did not recur (based on data from the gold standard). Abbreviations: N=number of patients; HMO=health maintenance organization; CRN=Cancer Research Network.

[‡]The secondary malignancy codes include ICD-9 codes for secondary malignant neoplasm associated with a solid organ.

[‡]The chemotherapy codes include procedure codes for the administration of chemotherapy agents and medication codes for specific chemotherapy drugs. ICD-9, DRG, CPT, and HCPC codes were used for both datasets; revenue center and BETOS were also used for CanCORS/Medicare patients; NDC codes were also used for HMO/CRN patients.

[§]Either = secondary malignancy or chemotherapy codes.