Published in final edited form as:

Med Care. 2013 October ; 51(10): e67–e73. doi:10.1097/MLR.0b013e31824def85.

Validation of Electronic Data on Chemotherapy and Hormone Therapy Use in HMOs

Debra P. Ritzwoller, PhD¹, Nikki Carroll, MS¹, Thomas Delate, PhD¹, Maureen O’Keeffe-Rossetti, MA², Paul A. Fishman, PhD³, Elizabeth T. Loggers, MD PhD³, Erin J. Aiello Bowles, MPH³, Jennifer Elston Lafata, PhD⁴, and Mark C. Hornbrook, PhD²

¹The Institute for Health Research, Kaiser Permanente Colorado, Denver, CO

²The Center for Health Research, Kaiser Permanente Northwest, Portland, OR

³The Group Health Research Institute, Seattle, WA

⁴Virginia Commonwealth University, Richmond, VA and the Henry Ford Health System, Detroit, MI

Abstract

Background—Most data regarding medical care for cancer patients in the United States comes from SEER-linked Medicare analyses of individuals aged 65 years or older and typically excludes Medicare Advantage enrollees.

Objectives—To assess the accuracy of chemo- and hormone therapy treatment data available via the Cancer Research Network’s Virtual Data Warehouse (VDW).

Research Design—Retrospective, longitudinal cohort study. Medical record-abstracted, tumor registry-indicated treatments (“gold standard”) were compared to VDW-indicated treatments derived from HMO pharmacy, electronic medical record, and claim-based data systems.

Subjects—Enrollees aged 18 years and older diagnosed with incident breast, colorectal, lung or prostate cancer from 2000 through 2008.

Measures—Sensitivity, specificity, and positive predictive value (PPV) were computed at six- and 12-months post-cancer diagnosis.

Results—Approximately 45% of all cancer cases (total N=23,800) were aged 64 years or younger. Overall chemotherapy sensitivity/specificities across the three health plans for incident breast, colorectal, lung, and prostate cancer cases were 95%/90%, 95%/93%, 93%/93%, and 85%/77%, respectively. With the exception of prostate cancer cases, overall PPV ranged from 86% to 89%. Small variations in chemotherapy data accuracy existed by cancer site and data source, while greater variation existed in hormone therapy capture across sites.

Conclusions—Strong concordance exists between gold standard tumor registry measures of chemotherapy receipt and CRN VDW data. HMO VDW data can be used for a variety of studies addressing patterns of cancer care and comparative effectiveness research that previously could only be conducted among elderly SEER-Medicare populations.

CORRESPONDING AUTHOR Debra P. Ritzwoller PhD, The Institute for Health Research, Kaiser Permanente Colorado, PO Box 378066, Denver, CO 80237-8066.

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.

Keywords

cancer; chemotherapy; tilization

Most published research on the patterns of utilization and costs of cancer treatment have been limited to the linked Surveillance, Epidemiology and End Results (SEER)-Medicare data, National Comprehensive Cancer Network (NCCN) studies, or randomized clinical trials (RCTs) (1–3). While several studies have documented the reliability of SEER-Medicare data to capture chemotherapy administration (4, 5), SEER-Medicare data do not include details on treatment patterns and outcomes of cancer care for patients aged younger than 65 years or those receiving their care in health maintenance organizations (HMOs). Furthermore, detailed data on the use of antineoplastic agents are often captured only for a select group of patients that are not representative of the broader population of cancer patients (5, 6). Generalizable data are needed to conduct comparative effectiveness research on treatment options commonly used in community-based oncology practices.

Such data are potentially available from HMOs that participate in the National Cancer Institute (NCI)-funded Cancer Research Network (CRN) (7–9). CRN HMOs compile and maintain data from internal, automated information systems in a distributed data network entitled the “Virtual Data Warehouse (VDW)” (10, 11). VDW encounter, claims, and electronic medical record data have, in turn, been linked to validated tumor registry data derived from the abstraction of HMO enrollee charts using methods similar to (or in some cases equivalent to) SEER registries (<http://seer.cancer.gov/>, last accessed 01/03/2012). Prior CRN research using “Virtual Tumor Registry” (VTR) data linked to VDW data has focused primarily on surveillance/screening (12–15) epidemiologic (16, 17) and survivorship studies (18, 19). However, the VDW also captures HMO enrollees’ use of antineoplastic agents and could therefore serve as a resource for population-based research on treatment patterns, diffusion of new therapies, and costs associated with the cancer treatment services provided in community settings for patients not currently represented by data derived from SEER-Medicare or RCTs.

One of the challenges to using HMO automated data is lack of published validation studies assessing the quality of automated data against a consistent gold standard. Health information systems vary across HMOs (different hardware platforms and software vendors) and over time (upgrades to software and changes in vendors). Data derived from treatment, payment, and operations sources are not collected under research standards; data are entered by a multitude of clinicians and staff members with little redundant entry to assess reliability. Many uncontrolled factors underlying treatment, payment, and operations data create a need for formal evaluation of the validity and reliability of these data.

To validate the usefulness of CRN VDW data for the capture of chemo- and hormone therapy treatment among cancer patients aged 18 and older, we assessed the sensitivity, specificity, and positive predictive value (PPV) of indicators of first-course chemotherapy and hormone therapies found in the CRN VDW compared to a “gold standard” of chart-abstracted measures of first course treatment contained in the tumor registries of three CRN HMOs.

METHODS

Study Design and Setting

This was a retrospective, longitudinal cohort study conducted within three non-profit HMOs [the Colorado (Denver/Boulder) and Northwest (Portland, OR/Vancouver, WA) regions of

Kaiser Permanente, and Group Health Cooperative (Seattle, WA)]. Each is a member of the Cancer Research Network (NCI Cooperative Agreement No. U19 CA79689) and provides comprehensive health services to its members, primarily through closed panel delivery models by salaried physicians. In each of these HMOs, the majority of ambulatory cancer care is delivered in plan-owned facilities. This project was approved by the Institutional Review Boards of the three participating HMOs.

Data Sources

The primary data source for this analysis was the CRN's Virtual Data Warehouse (VDW). The VDW is a standardized data system in which CRN participating HMOs retain local control of their data, but a programmer at one site can write a program that can be run at all sites (10, 11). The Virtual Tumor Registry (VTR) contains standardized data derived from each CRN HMO's tumor registry. The VTR contains data consistent with the North American Association of Central Cancer Registries (NAACCR) standards (20) and includes information on each patient's cancer site, sequence, diagnosis date, American Joint Committee on Cancer (AJCC) stage, and receipt of primary chemotherapy, immunotherapy, hormone therapy, and surgical and radiotherapy, that are obtained from manual reviews of cancer patients' medical charts. Information from the VTR can be considered a gold standard for cancer treatment information because data contained within the VTR were collected by certified tumor registrars (CTRs) who review all elements of each patient's medical record, including pathology reports, imaging reports, infusion session summaries, surgical operation reports, physician orders (for chemotherapy and hormonal therapy), and clinician progress notes. For the three HMOs included in this study, two have internal registries staffed with fulltime CTRs; neither are SEER sites, but both report NAACCR standard cases and fields to their respective State registries. The third HMO receives tumor registry data feeds, in NAACCR format, from the SEER Cancer Surveillance System. All diagnosis and treatment data provided to health plan enrollees were available for medical record review by the CTRs. Some variation was observed across the CRN sites and over time with respect to the capture of specific tumor registry variables. For example, not all CRN VTRs captured variables including date of first chemotherapy or hormone therapy treatment for all years included in this study.

VDW diagnosis and procedure files include coded diagnoses and procedures associated with inpatient and outpatient encounters or events. These codes are based on International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) codes, Healthcare Common Procedure Coding System (HCPCS) codes, or the Fourth Edition of the Common Procedure Terminology codes (CPT-4). Greater than 90% of the VDW diagnosis and procedure data used in this analysis were derived from electronic medical records (EMRs). These data capture the diagnoses and treatment associated with the chemotherapy events that take place predominately 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. VDW pharmacy files capture national drug code (NDC) based prescription drugs dispensed from both outpatient pharmacies and HMO-owned infusion centers. VDW files (VTR, Procedure, Diagnosis, and Pharmacy) were linked by programmers at the individual CRN sites using individual medical record identifiers. These identifiers were then stripped from the analytic data sets prior to pooling data for final analyses. In this analysis, and consistent with Warren et al (4), we focused on the ability of the VDW to capture definitive first course chemotherapy and hormone therapy at 6 months following a patient's incident cancer diagnosis date from four VDW files: VTR, Pharmacy, Procedures, and Diagnoses during the period of 01/01/2000 through 12/31/2008. Given the potential for the delay in first course treatment, we repeated this analysis at 12 months post incident cancer diagnosis.

Study Sample

Cancer cases were identified in the VTR as individuals diagnosed with breast, colorectal, lung, or prostate cancer between 01/01/2000 and 12/31/2007, aged 18 or older at diagnosis, and eligible for care in one of the three HMOs for the year prior to and including the date of diagnosis.

Identification of Chemotherapy

Cancer therapy “look-up” tables were created by extracting codes from published research (4, 6, 21), NCI websites (22, 23), plan-specific formularies, and expert clinician opinion (from medical Oncologists and pharmacists with doctoral level training practice primarily or exclusively in oncology settings). These look-up tables contain over 2,000 NDC and 300 procedure and diagnosis treatment-related codes. Please see Tables 4, 5, and 6, Supplemental Digital Content 1, which contains the lists of codes included in this study. In addition, these tables are accessible via the CRN public website (<http://crn.cancer.gov/>). Using these chemotherapy and hormone specific NDC, procedure and diagnosis codes, variables were created and stratified by class of product (chemotherapy vs. hormone therapy). These lookup tables were then mapped to the VDW pharmacy, procedure, and diagnosis files using individual medical record identifiers for all incident breast, colorectal, lung or prostate cancer cases identified from the VTR (as noted above). If one or more chemotherapy-related or hormone therapy-related event was captured in the VDW files within the first six months following the incident diagnosis date then the patient was flagged as receiving therapy at six months. Similarly, if a chemotherapy-related event occurred within the first 12 months following the incident diagnosis date, then the patient was flagged as receiving therapy at 12 months. The first date of a VDW captured treatment event in one or more of the VDW files (pharmacy, procedure, or diagnosis) that occurred after the incident cancer diagnosis date noted in the VTR was considered the first chemotherapy or hormone therapy event. The chart-abstracted VTR data was extracted from each health plan from the “first course of chemotherapy (or hormone therapy),” variable. If the “first course chemotherapy” variable was coded as 01, 02, or 03 (chemotherapy, NOS; chemotherapy, single agent; and chemotherapy, multiple agents), the patient was flagged as having received chemotherapy. Similarly, if the “first course hormone therapy” was coded as 01 (“hormone therapy including NOS and antihormone”), then the patient was flagged as having received hormone therapy treatment.

Data Analysis

For each cancer case, we compared results from the VTR to results from the VDW for chemotherapy and hormone therapy. Due to the small number of cases across all cancers receiving immunotherapy (n=111), we excluded immunotherapy from this analysis. For six and 12 months post cancer diagnosis, we computed sensitivity (proportion of those who were flagged in the VTR as having received treatment and in the VDW as having received treatment), specificity (proportion of those who were flagged in the VTR as NOT receiving treatment and in the VDW as not receiving treatment), and PPV (proportion of those flagged as receiving treatment in the VDW who were correctly identified as receiving treatment by the VTR) by cancer site. We then stratified the data by stage, HMO and year. Finally, we examined variation in the sensitivity and specificity of the VDW data source (pharmacy versus procedure and diagnosis files). The pharmacy files capture dispensing of drugs to the infusion unit or outpatient pharmacy, while the procedure and diagnosis files capture administration of infusions.

RESULTS

During the eight-year study period, 8,543 breast cancer cases, 4,399 colorectal cases, 4,713 lung cases, and 6,145 prostate cancer cases were diagnosed in the participating CRN HMOs. The proportions of breast, colorectal, lung and prostate cancer cases diagnosed prior to age 65 were 59%, 37%, 32%, and 42%, respectively. The distributions of incident cancer cases by health plan, cancer site, stage, and year of diagnosis, as well as the proportions of cases indicated by the VTR as receiving chemotherapy or hormone therapy are shown in Table 1. The majority of breast and prostate cancer cases, in total and by health plan, were initially diagnosed at early stages, while colorectal cases were more evenly distributed across all stages. Lung cancers were predominantly diagnosed at later stages. While not shown here, the distribution of cancer cases by health plan and diagnosis year was relatively stable, with a slight linear increase (from 11.6% in 2000 to 14.2% in 2007) over the observation period. We were unable to link 0.3% (N=83) of the cancer cases identified in VTR to the VDW files.

Chemotherapy Concordance

Sensitivity across all three HMOs was highest for breast (95.5%), followed by colorectal (95.1%), lung (93%), and prostate cancer (85.5%) (Table 2). Specificity for breast, colorectal, lung, and prostate was 89.8%, 92.6%, 92.6%, and 76.6%, respectively. Positive predictive values were generally high (>86%), except for prostate cases (6.3%). Sensitivity and specificity varied slightly by health plan with overall measures of concordance the lowest for HMO A. Minimal variation in sensitivity and specificity was noted by stage or year of diagnosis.

When the observation period post-diagnosis was extended to 12 months, the sensitivity increased by 0.6 percentage points for colorectal and lung cancer cases and 0.4 percentage points for breast cancer cases (no change was observed for prostate cancer cases). Specificity decreased for breast, colorectal, lung, and prostate cancer cases by 0.4, 2.3, 1.5, and 2.5 percentage points, respectively (data not shown).

Table 3 describes the changes in the sensitivity and specificity when the analysis was limited to either only VDW pharmacy NDC data, or only procedures and diagnosis codes associated from outpatient or inpatient encounters derived from the EMRs or claims data. Using pharmacy data only, the largest percentage point decline in sensitivity was found for HMO A across all four cancers (34.4 – 54.6 percentage points), and for prostate cancers at HMO B (42.9 percentage point decline). However, specificity improved from 18.8 to 33.6 percentage points for the same cases. Using only diagnosis and procedure codes, minimal changes were found for HMO A, but large percentage point declines in sensitivity across all four cancer sites were found for HMO C, with no improvement in specificity. Moderate declines in the sensitivity (7.8 – 14 percentage points) were found for breast, colorectal and lung cases at HMO B.

Hormone Therapy Concordance

Hormone treatment was noted in the VTR data for 4,903, 107, 33, and 1,706 breast, colorectal, lung, and prostate cancer cases, respectively. While specificity was high for all cancer sites and HMOs, consistent with the small number of cases receiving hormone therapy, sensitivity was lowest for colorectal (49.5%) and lung cancer (48.5%) as compared to 72.3% for breast and 78% for prostate cancer (Table 4). In contrast to the chemotherapy analysis, hormone treatment sensitivity increased 19.9 and 8.2% percentage points for breast and prostate cancer, respectively, when the observation period was extended to 12 months post diagnosis (data not shown). Concordance across cancers fell slightly when the analysis

was limited to pharmacy data, with the exception of breast cancer sensitivity. When the analysis was limited to procedure and diagnosis data, the sensitivity fell 0.9 and 46.1 percentage points for breast and prostate cancer, respectively. Due to the small number of colorectal and lung cancer patients receiving hormone therapy, we did not estimate changes in sensitivity or specificity by source of data.

CONCLUSIONS

This is the first evaluation of the consistency between gold-standard tumor registry data and automated data derived from the CRN VDW with respect to chemotherapy and hormone therapy use for HMO enrollees aged 18 years or older diagnosed with one of the four most common cancers in the United States. We found that CRN VDW data associated with coded diagnoses, procedures and prescription drug use, when used in combination, have both high (>90%) sensitivity and specificity for identifying chemotherapy use among breast, colorectal, and lung cancer cases relative to chemotherapy treatment data derived from the chart-abstracted tumor registries. This analysis extends the work of Aiello and colleagues (24) conducted in these same HMOs, which compared chemotherapy for ovarian cancer patients, based on CPT-4, NDC, and ICD-9-CM codes, to tumor registries and found an overall sensitivity of 89%.

Differences found in sensitivity and specificity by HMO and by data source (pharmacy vs. other) could be explained by the differences in source data systems. Two of the three HMOs have “in-house” tumor registries employing trained full-time CTRs to conduct comprehensive chart audits, while the third populated their VTR from automated SEER data files. The SEER derived data may be less complete than the “in-house” registries because SEER abstractors do not have continuous access to medical records, which increases the barriers to additional updates beyond the initial chart review. Also consistent with the findings of Aiello and colleagues (24), we found that using only one source of data (e.g., pharmacy only) was insufficient for capturing pharmacologic treatment data. Two of the three HMOs had equal capture of pharmacologic events from both the VDW pharmacy and utilization files (capture of both NDC and HCPCS or CPT-4 codes) for over 90% of the cases receiving chemotherapy, while the third captured chemotherapy events predominately from encounter based diagnoses and procedure files—less than 50% of the cases had a chemotherapy-related NDC code. This likely reflects differences in the organization of pharmacy systems within the plan-owned infusion centers.

We found the sensitivity of chemotherapy use capture could be improved only slightly (< 1 percentage point) by extending the observation period from 6 to 12 months, and the same change resulted in reduced specificity (0.4 to 2.5 percentage points). Differences in sensitivities could be due to delayed start of therapy for reasons related to other cancer care (e.g., surgery, radiation), while the drop in specificity may be associated with recurrence or progression.(25) Conversely, the sensitivity estimate for hormone therapy for breast cancer cases improved by almost 20 percentage points (72.3% to 92.2%) when the observation period was extended to 12 months. This is not surprising given that hormone therapy often is not started until after adjuvant chemotherapy treatment is complete, and suggests that future research related to hormone therapy that employs CRN VDW data should consider using a 12-month observation period.

In evaluating the PPV results, it is important to note that the VTR is limited to first treatment course, a restriction that cannot be applied to the coded diagnoses, procedures and prescription drug use VDW data. The chemotherapy PPV for prostate cancer cases was low, but the hormone therapy PPV was high (>90% overall). Chemotherapy is frequently only initiated in prostate cancer for advanced disease or with progression or recurrence (26, 27);

consistent with this, only 1.8% of cases (n= 111) in the VTR received primary chemotherapy treatment, while 27% of cases within the VDW pharmacy or procedure data received chemotherapy within 12 months of the diagnosis date. The hormone therapy PPV values are consistent with use of a variety of treatment approaches that can be initiated at various times post-diagnosis.(28)

There are several limitations associated with this study. We did not conduct chart reviews of false negative cases (indicated as having treatments from VTR data, but no VDW pharmacy, procedure, or diagnosis-related event). At least some of these cases likely reflect 3–5% of cancer patients (enrolled in these HMOs) who are participating in chemotherapy-related clinical trials. The health plan is not liable for the cost of the chemotherapy, nor is the trial drug necessarily part of the health plan's pharmacy formulary. Therefore, if the trial drug was administered at a cancer center or by another contract provider, it would be noted in the chart abstracted tumor registry, but no claim would be generated. If the trial drug was administered by the plan-owned infusion center, the trial drug may not be captured or included in the traditional pharmacy systems. In addition, the two health plans with internal tumor registries (non-SEER) generally perform chart abstraction on most cases within 6 to 8 months of the cancer diagnosis date. However, the CTRs at these plans have continuous access to the patient's EMR, allowing for abstraction updates in cases when first course treatment may be delayed by several months. This may result in a longer follow-up abstraction period relative to SEER registries. The potential for differences in SEER/non-SEER abstraction follow-up could also be the source of variation in specificities across health plans. Further, while we assumed that the first chemotherapy event following the diagnosis date of one of the four cancers of interest corresponded to that cancer, we did not limit cases to sequence 0 (first incident cancer). It is therefore possible that we are identifying cancer treatments associated with a previous cancer diagnoses. Future studies should also explore product, or regimen-level specificities, sensitivities, and PPVs, consistent with Warren and colleagues,(4) particularly as standardized electronic infusion therapy record systems are deployed across the majority of the CRN HMOs, further improving the quality and clarity of antineoplastic data.

Regardless of future data refinements, the results of the current multi-site analysis reveal that CRN VDW-based cancer treatment data are as sensitive and specific as Medicare fee-for-service claims data in identifying receipt of antineoplastic therapy—the dominant data source used for research related to cancer treatment and outcomes (4, 6, 29). Furthermore, this study highlights the ability of the CRN VDW to capture hormonal and oral chemotherapeutics at a time when Medicare Part D pharmacy data are not currently available via linked SEER-Medicare files. Finally, our study is one of a limited few that includes a representative population of cancer patients under the age of 65 years. This unique, comprehensive, and reliable data source sets the stage for patterns of care analyses and comparative effectiveness research in community-based oncology settings that will inform policy and medical practice by enabling better understanding of similarities and differences in cancer epidemiology, treatment patterns, and outcomes between aged and non-aged populations and between capitated and fee-for-service reimbursement systems.

Acknowledgments

Funding for this research was provided by NCI Grant No. R01 CA114204, Medical Care Burden of Cancer: System and Data Issues, PI: Mark C. Hornbrook, PhD; NCI Cooperative Agreement No. U19 CA79689, Increasing Effectiveness of Cancer Control Interventions (Cancer Research Network), PI: Edward H. Wagner, MD; and NCI Grant No. RC2 CA148185, Building CER Capacity: Aligning CRN, CMS, and State Resources to Map Cancer Care, Co-PIs: Jane C. Weeks, MD and Debra P. Ritzwoller, PhD.

The following staff members provided data processing support for this study: KP Colorado: Addie L. Clark and Stephanie Latimer; KP Northwest: Donald J. Bachman, and Arthur Dixon; Group Health: Cristi Hanson.

REFERENCES

1. Warren JL, Klabunde CN, Schrag D, et al. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. *Med Care*. 2002; 40(8 Suppl):IV–18.
2. Schrag D, Rifas-Shiman S, Saltz L, et al. Adjuvant chemotherapy use for Medicare beneficiaries with stage II colon cancer. *J Clin Oncol*. 2002; 20(19):3999–4005. [PubMed: 12351597]
3. Schrag D, Cramer LD, Bach PB, et al. Age and adjuvant chemotherapy use after surgery for stage III colon cancer. *J Natl Cancer Inst*. 2001; 93(11):850–857. [PubMed: 11390534]
4. Warren JL, Harlan LC, Fahey A, et al. Utility of the SEER-Medicare data to identify chemotherapy use. *Med Care*. 2002; 40(8 Suppl):IV–61.
5. Lamont EB, Lauderdale DS, Schilsky RL, et al. Construct validity of medicare chemotherapy claims: the case of 5FU. *Med Care*. 2002; 40(3):201–211. [PubMed: 11880793]
6. Lamont EB, Herndon JE, 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*. 2005; 97(14):1080–1083. [PubMed: 16030306]
7. Wagner EH, Greene SM, Hart G, et al. Building a research consortium of large health systems: the Cancer Research Network. *J Natl Cancer Inst Monogr*. 2005; (35):3–11. [PubMed: 16287880]
8. Vogt, TM.; Wagner, EW. Health care systems as research platforms: The Cancer Research Network. Bethesda, MD: Oxford University Press; 2005.
9. National Cancer Institute. The HMO Cancer research Network: Capacity, collaboration, and investigation. Washington D.C.: U.S. Department of Health and Human Services, National Institutes of Health; 2010.
10. 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]
11. Go AS, Magid DJ, Wells B, et al. The Cardiovascular Research Network: a new paradigm for cardiovascular quality and outcomes research. *Circ Cardiovasc Qual Outcomes*. 2008; 1(2):138–147. [PubMed: 20031802]
12. Lash TL, Fox MP, Buist DS, et al. Mammography surveillance and mortality in older breast cancer survivors. *J Clin Oncol*. 2007; 25(21):3001–3006. [PubMed: 17548838]
13. Haque R, Quinn VP, Habel LA, et al. Correlates of screening sigmoidoscopy use among men in a large nonprofit health plan. *Cancer*. 2007; 110(2):275–281. [PubMed: 17559138]
14. Owusu C, Buist DS, Field TS, et al. Predictors of tamoxifen discontinuation among older women with estrogen receptor-positive breast cancer. *J Clin Oncol*. 2008; 26(4):549–555. [PubMed: 18071188]
15. Haque R, Achacoso NS, Fletcher SW, et al. Treatment of ductal carcinoma in situ among patients cared for in large integrated health plans. *Am J Manag Care*. 2010; 16(5):351–360. [PubMed: 20469955]
16. Aiello EJ, Buist DS, Wagner EH, et al. Diffusion of aromatase inhibitors for breast cancer therapy between 1996 and 2003 in the Cancer Research Network. *Breast Cancer Res Treat*. 2008; 107(3):397–403. [PubMed: 17393300]
17. Feigelson H, Rahm AK, Funk K, et al. OS2-11: Validation of KRAS mutation testing across five certified laboratories. *Clin Med Res*. 2011 Nov.9(3–4):141.
18. 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*. 2011; 213(6):757–765. [PubMed: 22014658]
19. Haque R, Yood MU, Geiger AM, et al. Long-term safety of radiotherapy and breast cancer laterality in older survivors. *Cancer Epidemiol Biomarkers Prev*. 2011 Oct; 20(10):2120–2126. [PubMed: 21878589]
20. North American Association of Central Cancer Registries. [Accessed 8-26-11] NAACCR Strategic Management Plan. <http://www.naacr.org/Home.aspx>

21. Li EC, DeMartino J. Preliminary report: the development of the NCCN Comparative Therapeutic Index as a clinical evaluative process for existing data in oncology. *J Natl Compr Canc Netw*. 2010; 8(Suppl 5):S1–S9.
22. National Cancer Institute, Cancer Control and Population Sciences. [Accessed 8-26-11] Health Services and Economics: Procedure Codes for SEER-Medicare Analyses. http://healthservices.cancer.gov/seermedicare/considerations/procedure_codes.html
23. National Cancer Institute. [Accessed 8-26-11] CI Drug Dictionary. <http://www.cancer.gov/drugdictionary>
24. Aiello Bowles EJ, Tuzzio L, Ritzwoller D, et al. Accuracy and complexities of using automated clinical data for capturing chemotherapy administrations: implications for future research. *Med Care*. 2009; 47(10):1091–1097. [PubMed: 19648826]
25. Earle CC, Nattinger AB, Potosky AL, et al. Identifying cancer relapse using SEER-Medicare data. *Med Care*. 2002; 40(8 Suppl):IV–81.
26. Tannock IF, de WR, Berry WR, et al. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. *N Engl J Med*. 2004; 351(15):1502–1512. [PubMed: 15470213]
27. Petrylak DP, Tangen CM, Hussain MH, et al. Docetaxel and estramustine compared with mitoxantrone and prednisone for advanced refractory prostate cancer. *N Engl J Med*. 2004; 351(15):1513–1520. [PubMed: 15470214]
28. National Cancer Institute. [Accessed 8-26-11] Prostate Cancer Treatment (PDQ®). <http://www.cancer.gov/cancertopics/pdq/treatment/prostate/HealthProfessional/page4>
29. Du XL, Key CR, Dickie L, et al. External validation of medicare claims for breast cancer chemotherapy compared with medical chart reviews. *Med Care*. 2006; 44(2):124–131. [PubMed: 16434911]

Table 1

Sample Characteristics by Cancer Site

	Breast n = 8,543	Colorectal n = 4,399	Lung n = 4,713	Prostate n = 6,145
Mean Age at Diagnosis (SD)	61.8 (13.1)	68.3 (13.0)	69.0 (10.4)	66.9 (9.5)
Age at Diagnosis (%)				
18–55 years	30.9	16.8	9.8	9.4
56–64 years	28.1	20.3	22.2	32.4
65+ years	41.0	62.9	68.1	58.2
Race (%)				
Black	2.5	2.8	3.1	4.3
White	85.1	80.5	85.5	83.9
Other/unknown	12.5	16.8	11.5	11.8
Stage at diagnosis (%)				
I	53.0	23.9	19.7	1.9
II	33.8	30.3	5.6	79.5
III	9.1	26.0	28.4	8.7
IV	4.1	19.8	46.3	9.8
HMO (%)				
A	38.3	38.1	35.4	40.5
B	29.8	30.0	31.5	26.9
C	31.9	31.9	33.1	32.6
Chemotherapy	40.5	40.6	41.5	1.8
First course Treatment				
Hormone therapy	57.4	2.4	0.7	27.8
Indicated in VTR (%)				
Distribution				
2000	11.6	11.8	11.3	11.6
2001	12.3	12.4	11.4	11.8
2002	12.4	11.9	13.3	12.3
2003	11.6	12.4	12.2	12.8
2004	11.2	12.6	12.6	12.5
2005	12.8	12.3	12.9	11.2
2006	13.4	13.4	13.3	12.5
2007	14.6	13.2	13.0	15.4
2008	11.6	11.8	11.3	11.6

Table 2

Chemotherapy Concordance at Six Months Post Cancer Diagnosis

Cancer Site	HMO			Total
	A	B	C	
Breast	N=3,274	N=2,548	N=2,721	N=8,543
Number with at least 1 chemotherapy related pharmacy, procedure, or diagnosis code	1,580	1,088	1,154	3,822
Sensitivity (%)	90.4	98.6	98.6	95.5
Specificity (%)	79.8	96.2	96.2	89.8
Positive Predictive value (%)	74.8	94.8	94.6	86.5
Colorectal	N=1,676	N=1,320	N=1,403	N=4,399
Number with at least 1 chemotherapy related pharmacy, procedure, or diagnosis code	736	568	589	1,893
Sensitivity (%)	90.4	97.3	98.0	95.1
Specificity (%)	83.0	100	98.4	92.6
Positive Predictive value (%)	75.4	100	97.8	89.8
Lung	N=1,667	N=1,485	N=1,561	N=4,713
Number with at least 1 chemotherapy related pharmacy, procedure, or diagnosis code	822	557	643	2,022
Sensitivity (%)	86.0	97.9	96.5	93.0
Specificity (%)	78.8	100	99.6	92.6
Positive Predictive value (%)	75.7	100	99.4	89.9
Prostate	N=2,487	N=1,653	N=2,005	N=6,145
Number with at least 1 chemotherapy related pharmacy, procedure, or diagnosis code	799	299	407	1,505
Sensitivity (%)	72.7	85.7	94.9	85.5
Specificity (%)	68.6	82.2	82.0	76.6
Positive Predictive value (%)	4.0	2.0	13.8	6.3

Table 3

Percentage Point Change in Sensitivity and Specificity When Chemotherapy Concordance Analyses are Limited to One VDW Data Source

VDW Data Source	Pharmacy—based on National Drug Codes	Encounter – based on Diagnosis and Procedure Codes		
		HMO		HMO
Cancer Site	A	B	C	A B C
Breast				
Sensitivity	-47.1	-4.6	-10.7	-3.4 -7.8 -72.6
Specificity	21.4	5.5	4.7	0.8 0.5 0.7
Colorectal				
Sensitivity	-44.6	-3.9	-9.5	-2.1 -14.0 -64.1
Specificity	19.8	0.0	1.6	0.2 0.0 0.1
Lung				
Sensitivity	-34.4	-7.4	-13.9	-0.7 -10.4 -57.9
Specificity	21.5	0.0	0.5	0.6 0.0 0.1
Prostate				
Sensitivity	-54.6	-42.9	-11.9	0.0 0.0 -49.2
Specificity	33.6	18.8	18.3	0.8 0.2 0.1

Table 4**Hormone Therapy Concordance at Six Months Post Cancer Diagnosis Date**

Cancer Site	HMO			
	A	B	C	Total
Breast	N=3,274	N=2,548	N=2,721	N=8,543
Number with at least 1 hormone therapy related pharmacy, procedure, or diagnosis code	1,633	1,616	1,654	4,903
Sensitivity (%)	57.6	85.4	73.9	72.3
Specificity (%)	86.8	98.1	97.0	92.7
Positive Predictive value (%)	81.3	98.7	97.5	93.0
Colorectal	N=1,676	N=1,320	N=1,403	N=4,399
Number with at least 1 hormone therapy related pharmacy, procedure, or diagnosis code	44	16	47	107
Sensitivity (%)	72.7	12.5	40.4	49.5
Specificity (%)	94.8	97.5	97.6	96.5
Positive Predictive value (%)	27.4	5.7	36.5	26.0
Lung	N=1,667	N=1,485	N=1,561	N=4,713
Number with at least 1 hormone therapy related pharmacy, procedure, or diagnosis code	6	5	22	33
Sensitivity (%)	100	20.0	40.9	48.5
Specificity (%)	89.3	95.4	93.8	92.7
Positive Predictive value (%)	3.3	1.5	8.6	4.5
Prostate	N=2,487	N=1,653	N=2,005	N=6,145
Number with at least 1 hormone therapy related pharmacy, procedure, or diagnosis code	682	566	458	1,706
Sensitivity (%)	90.2	87.3	48.3	78.0
Specificity (%)	93.7	99.9	99.5	97.2
Positive Predictive value (%)	84.4	99.8	96.5	91.5