Validity of Eight Integrated Healthcare Delivery Organizations' Administrative Clinical Data to Capture Breast Cancer Chemotherapy Exposure

Thomas Delate[1], **Erin J. Aiello Bowles**[2], **Roy Pardee**[3], **Robert D. Wellman**[4], **Laurel A. Habel**[5], **Marianne Ulcickas Yood**[6], **Larissa Nekhlyudov**[7], **Katrina A. Goddard**[8], **Robert L. Davis**[9], **Catherine A. McCarty**[10], **Adedayo A. Onitilo**[11], **Heather Spencer Feigelson**[12], **Jared Freml**[13], and **Edward Wagner**[14]

[1]Kaiser Permanente Colorado, Pharmacy Department, 16601 E. Centertech Pkwy, Aurora, CO 80011, tel: 303-739-3538, fax: 303-739-3574, tom.delate@kp.org

[2]Group Health Cooperative, Group Health Research Institute, 1730 Minor Ave, Suite 1600, Seattle, WA 98101, tel: 206-287-2708, fax: 206-287-2871, aiello.e@ghc.org

[3]Group Health Cooperative, Group Health Research Institute, 1730 Minor Ave, Suite 1600, Seattle, WA 98101, tel: 206-287-2900, fax: 206-287-2871, pardee.r@ghc.org

[4]Group Health Cooperative, Group Health Research Institute, 1730 Minor Ave, Suite 1600, Seattle, WA 98101, tel: 206-287-2900, fax: 206-287-2871, wellman.r@ghc.org

[5]Kaiser Permanente Northern California, Division of Research, 2000 Broadway Oakland, CA 94612, tel: 510-891-3715, fax: 510-627-2520, laurel.habel@kp.org

[6]Henry Ford Health System, Josephine Ford Cancer Center, 19675 Allen Road, Brownstown Twp, MI 48183, tel: 734-479-3311, fax: 734-479-3311, muyood@muyood.com

[7]Harvard Pilgrim Health Care Institute/Harvard Medical School, Department of Population Medicine and Department of Medicine, Harvard Vanguard Medical Associates, 133 Brookline Avenue, 6th Floor, Boston, MA 02215, tel: 617-509-9983, fax: 617-859-8112, larissa_nekhlyudov@harvardpilgrim.org

[8]Kaiser Permanente Northwest, Center for Health Research, 3800 N. Interstate Avenue, Portland, OR 97227, tel: 503-335-2400, fax: 503-335-2402, katrina.AB.goddard@kpchr.org

[9]Kaiser Permanente Georgia, Center for Health Research Southeast, 3495 Piedmont Road, NE, Atlanta, GA 30305, tel: 404-365-4106, fax: 404-365-4105, robert.l.davis@kp.org

[10]Essentia Institute of Rural Health, 502 East Second St., Duluth, MN 55805, tel: 218-786-2128, fax: 218-727-8159, cmccarty@eirh.org

[11]Marshfield Clinic, Marshfield Clinic Research Foundation, 3501 Cranberry Blvd., Weston, WI 54476., tel: 715-393-1400, fax: 715-393-1469, onitilo.adedayo@mcrf.mfldclin.edu

[12]Kaiser Permanente Colorado, Institute for Health Research, P.O. Box 378066, Denver, CO 80237, tel: 303-614-1214, fax: 303-614-1305, heather.s.feigelson@kp.org

[13]Kaiser Permanente Colorado, Pharmacy Department, 16601 E. Centertech Pkwy, Aurora, CO, 80011, tel: 303-764-4538, fax: 303-861-3498, jared.m.freml@kp.org

Corresponding Author Information: Thomas Delate, PhD, MS, Clinical Pharmacy Research Scientist, Pharmacy Dept., Kaiser Permanente Colorado, 16601 E. Centretech Pkwy., Aurora, CO 80011, Ph: 303-739-3538; Fax: 303-739-3574, tom.delate@kp.org.

[14]Group Health Cooperative, Group Health Research Institute, 1730 Minor Ave, Suite 1600, Seattle, WA, 98101, tel: 206-287-2877, fax: 206-287-2871, wagner.e@ghc.org

Abstract

Background—Cancer Research Network (CRN) sites use administrative data to populate their Virtual Data Warehouse (VDW). However, information on VDW chemotherapy data validity is limited. The purpose of this study was to assess the validity of VDW chemotherapy data.

Methods—This was a retrospective, cohort study of women   18 years with incident, invasive breast cancer diagnosed between January 1999 and December 2007. Pharmacy and procedure chemotherapy data were extracted from each site's VDW. Random samples of 50 patients stratified on trastuzumab, anthracyclines, and no chemotherapy exposure were selected from each site for detailed chart abstraction. Weighted sensitivities and specificities of VDW compared to abstracted data were calculated. Cumulative doses calculated from VDW data were compared to doses obtained from the medical chart review.

Results—The cohort included 13497 patients with 6456 (48%) chart-review eligible. Patients in the sample (N=400) had a mean age of 65 years. Trastuzumab, anthracycline, and other chemotherapy weighted sensitivities were 95%, 97%, and 100%, respectively; specificities were 99%, 99%, and 93%, respectively; positive predictive values were 96%, 99%, and 55%, respectively; and negative predictive values were 99%, 96%, and 100%. Trastuzumab and anthracyclines VDW mean doses were 873 mgs and 386 mgs, respectively, while abstracted mean doses were 1734 mgs and 369 mgs, respectively (R^2=0.14, p<0.01 and R^2=0.05, p=0.03, respectively).

Conclusions—Sensitivities and specificities for CRN chemotherapy VDW data were high and dosages were correlated with chart information.

Impact—The findings support the use of CRN data in evaluating chemotherapy exposures and related outcomes.

Keywords

Background

Adjuvant chemotherapy is the standard of care in breast cancer when the patient has an increased risk for relapse or progression after initial therapy (1). Randomized clinical trials (RCTs) have provided the bulk of chemotherapy exposure efficacy and safety information. However, RCTs' strict inclusion/exclusion criteria and standardized treatment protocols may limit assessments of the consequences of chemotherapy exposure for real-world patients with diverse personal, social, and health conditions (2). Information on chemotherapy exposures and outcomes may be obtained more cost-effectively with observational studies of well-characterized, large, heterogeneous populations where a range of treatments and subgroups can be studied (3).

Linked Surveillance Epidemiology and End Results (SEER) - Medicare observational data (4) have been validated (5–8) and used in studies of chemotherapy exposures (9–11). While SEER data contain information on cancer type and stage, diagnosis date, some patient characteristics, and whether the patient received first-line therapy (12), special data-collection efforts have been undertaken by the National Cancer Institute to supplement these data with information from random samples of cancer patients on in-office administrations of specific chemotherapy agents (Patterns of Care studies) (6). Medicare data, which contain

claims for facility and provider billing purposes, are linked to patients' SEER data to provide information on health services utilization, including dates and types of chemotherapy administered (12). However, SEER-Medicare data are not without limitations. The linked data typically have limited information on: use of oral chemotherapy agents, non-disabled patients <65 years of age, specific chemotherapy dosages, cancer recurrences, and managed care patients (13, 14). In addition, there may be a lag period of four years between use of health services and reporting of data, thus, limiting their use for timely investigations (14).

The Cancer Research Network (CRN) is a consortium of 14 non-profit research centers, based in integrated healthcare delivery organizations, within the HMO Research Network (15). Unlike SEER-Medicare, the CRN's virtual data warehouse (VDW) includes administrative data (i.e., information on members' enrollment, healthcare delivery, and reimbursement for services), managed care patients, and, importantly, information about oral and parenteral chemotherapy including specific agents and dosages (16). Two prior CRN studies have provided some evidence of the validity of the capture of chemotherapy exposure information from the VDW compared to medical record (chart) abstracted data (17, 18). However because chemotherapy data are collected differently across the CRN sites, it is imperative to assess the validity of the CRN's standardized chemotherapy information. In this study, CRN data were used to conduct a larger and more robust assessment of breast cancer chemotherapy exposure capture. Specifically, the validity of VDW administrative data was assessed in their capture of the receipt of chemotherapy, use of specific chemotherapeutic agents, and dosage information among a cohort of breast cancer patients who received healthcare at eight CRN sites.

Methods

Study Design & Setting

This was a retrospective, multi-site, cohort study to evaluate the validity of administrative clinical data to capture chemotherapy exposure. Data were obtained for breast cancer patients enrolled at eight CRN integrated healthcare delivery sites: Group Health Cooperative (GHC), Harvard Pilgrim Health Care and Harvard Vanguard Medical Associates, Henry Ford Health System, Marshfield Clinic, and Kaiser Permanente regions in Colorado, Georgia, Northern California, and the Northwest. These healthcare delivery sites had a combined membership of >5 million members in 2008. This study was approved by the GHC Institutional Review Board for GHC and five other sites that ceded review to GHC and, separately, by the Institutional Review Boards at Marshfield Clinic and Henry Ford Health System.

Patient Population

Female patients aged 18 years, diagnosed with incident invasive (local, regional, or distant summary stages) breast cancer between January 1, 1999 and December 31, 2007, and enrolled in one of the eight CRN healthcare delivery sites at time of diagnosis were included. Patients were enrolled continuously in their respective site during the 12 months prior to cancer diagnosis (membership gaps of 90 days were permissible) (N=13472). Due to its large patient population, a 10% random sample of eligible women diagnosed from 2001–2007 from Kaiser Permanente Northern California were included (chemotherapy data from 1999 and 2000 were incomplete and not included). Additionally, Harvard data only included women diagnosed from 2001–2006 due to delays in the linkage of their cancer registry data with administrative data.

In order to maximize the possibility that selected patients were eligible to receive chemotherapy treatment, only women with a tumor size greater than 2.0 cm and/or positive lymph nodes were eligible for chart review (n=6456). Stratified random samples of 50 patients from each of the eight sites (total n=400) were selected for detailed medical chart review. If a patient's record was unobtainable, a random substitute patient meeting the same abstraction criteria from the site was reviewed (see Table 1).

Study Outcomes

The primary outcomes were weighted sensitivities, specificities, positive predictive values (PPV), and negative predictive values (NPV) of the administrative data for identifying chemotherapy treatment compared to medical record data (the gold standard). Secondary outcomes were a calculation of the prevalences of chemotherapy treatments and an assessment of the sensitivities, specificities, PPVs, and NPVs for patients < and ≥ 65 years of age. In addition, an assessment of the variation in the PPVs and NPVs across sites was performed. Furthermore, an estimation of the cumulative doses of chemotherapy based on administrative data were compared to cumulative doses recorded in patient charts.

Administrative Data Collection

The CRN uses a federated database (the Virtual Data Warehouse [VDW]) where each site retains control of their administrative data stored in a common data structure (16). Thus, a programmer at one site can develop programming code that can be run at all sites to extract similar data. The VDW contains patients' health services procedure data (Healthcare Common Procedure Coding System (HCPCS) codes [including Current Procedural Terminology (CPT) codes] and International Classification of Diseases (ICD)-9 codes) and outpatient pharmacy data (National Drug Codes (NDC)). Each CRN site has a dataset with all of the NDCs that have ever been used at the site.

The VDW includes tumor registry data with information on each patient's cancer stage at diagnosis, date of cancer diagnosis, age at time of cancer diagnosis, laterality, lymph node involvement, and first-line of cancer treatment. Patient identifiers from the tumor registry were linked with administrative databases to obtain information on pharmacy dispensings, inpatient and outpatient diagnoses and procedures, along with patient characteristics at the time of cancer diagnosis. A comprehensive list of chemotherapy-related HCPCS, CPT, ICD-9, and NDC codes was developed from local, national, and CMS sources. (A list of all codes used in the analyses is available from the authors.) Trastuzumab and anthracycline-specific treatment data were identified with HCPCS and NDC codes. Non-specific chemotherapy treatments were identified using CPT, ICD-9, and HCPCS codes (e.g., 96410 - Chemotherapy administration, IV: infusion up to 1 hr). Common programming code was developed, run against each site's VDW (often with local modification because of site-specific differences, such as location and structure of infusion chemotherapy databases), and data were transferred to GHC for analysis.

Procedure and pharmacy administrative data independently were used to calculate cumulative dose for anthracyclines and trastuzumab. Only those patients with evidence of chemotherapy dosing in both their chart and administrative data source were included. Pharmacy dispensing and procedure data were extracted in the twelve months after cancer diagnosis. For each pharmacy dispensing, the days and amount of medication supplied were captured. In addition, information on the count per day of each individual procedure and pharmacy code populated in the administrative data was captured.

Three methods were employed to estimate cumulative dose from administrative data. (1) For trastuzumab, a standard dosing of 4mg/kg loading dose followed by 2mg/kg follow-up doses

was assumed (19). Patient's weight most proximal and prior to start of chemotherapy was captured and applied to this formula for each specific day (i.e., loading or follow-up) of treatment. These values were then summed across the follow-up period. (2) For anthracyclines, the concentration of drug in each NDC was multiplied by the amount of drug infused as recorded in the administrative data to obtain the treatment dose (e.g., [NDC 63323-0101-61 = Doxorubicin 2 mg/mL]×[Amount dispensed = 55 mL] = Dose of 110 mg). These values were then summed across the follow-up period. (3) For trastuzumab and anthracyclines, we summed the number of unique administrations in procedure and pharmacy data for each drug to be used as a proxy for cumulative dose.

Medical Record Data Collection

Each CRN site has access to their patients' electronic and/or paper medical charts. Fifty patients from each site's cohort were selected randomly within stratification groups for detailed chart abstraction (Table 1). Stratification was based on trastuzumab, anthracyclines, and no chemotherapy exposure to ensure representation of women who did and did not receive these drugs. As this study was one aim of a larger study to assess cardiotoxicity of breast cancer chemotherapy treatment, stratification on prevalent (occurring up to 12 months before breast cancer diagnosis), incident (occurring anytime after breast cancer diagnosis through study end), and no heart failure/cardiomyopathy outcomes (based on ICD-9 codes: 398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 422.90, 425.4, 425.9, and 428.xx) (20) was undertaken independently to ensure selecting patients who did and did not have one of these diagnoses. Following diagnosis, patient chart data were censored at the time of death, disenrollment from care delivery site, or one year after cancer diagnosis, whichever came first.

All abstractors were trained in use of the abstraction tool and blinded to patient sampling-stratum. Abstractors reviewed charts depending on availability. Information on patient diagnosis date, stage, characteristics (e.g., age at diagnosis, race/ethnicity), HER2 testing, and chemotherapy treatment (e.g., date of initiation, chemotherapy agent(s), and cumulative dose) were abstracted.

Data Analysis

The patient and tumor characteristics were compared among the chart reviewed patients (n=400) (*Chart Reviewed* patients), chart review eligible patients (n=6456) (*Chart Review Eligible* patients), and the entire cohort of patients (N=13472) (*Entire Cohort* patients). Since some sites captured chemotherapy in pharmacy databases, procedure data or both, both pharmacy and procedure administrative data were combined and counted as a single exposure if codes indicated that the same type of chemotherapy was administered on the same day. We categorized receipt of chemotherapy as ever/never for trastuzumab, any anthracycline (only epirubicin and doxorubicin were dispensed among cohort patients with the vast majority receiving doxorubicin), and other chemotherapy agents.

We calculated prevalences (with 95% confidence intervals (CI)) of ever receiving chemotherapy treatment. Prevalence was calculated as the percent of patients with chemotherapy treatment information in the data among patients in the respective patient groups. Prevalence was stratified on other chemotherapy, trastuzumab chemotherapy, and anthracycline chemotherapy.

In order to generalize the Chart Reviewed patients back to the Chart Review Eligible patients, the inverse probability of verification given the sampling stratification as indicated by the VDW was calculated (Table 1). Calculated weights were scaled to the random sample size (n=400) to provide standard errors relative to the size of the validation cohort.

Weighting was only applicable to women eligible for the chart review (n=6456). Sensitivity was calculated as the percent of cases with chemotherapy treatment information in both the administrative and medical record data among all patients with chemotherapy treatment information in the medical record data. Specificity was calculated as the percent of cases with no chemotherapy treatment information in both administrative and medical record data among all patients with no chemotherapy treatment information in the medical record data. Sensitivity, specificity, NPV, and PPV were calculated using weighted logistic regression where the population weights were standardized to reflect the size of the sample (n=400). Data were analyzed using SAS version 9.2 (SAS, Inc., Cary, NC). All statistical tests were two-sided and p-values <0.05 were considered statistically significant.

Sensitivity & specificity were stratified on age (<65 years vs. 65 years) to assess data validity of likely Medicare eligible and non-eligible age categories. The cumulative doses estimated from administrative data for trastuzumab and the anthracyclines were assessed for correlation with doses obtained from chart review using Spearman's correlation coefficient. The correlations were assessed for only those observations where both the chart dose and the administrative dose were not missing.

Results

Among the participating sites, a total of 13472 patients diagnosed with incident breast cancer during the study period were identified and 6456 (48%) were eligible for chart review. Of these, a total of 400 (6%) patients were chosen by stratified sampling for manual chart review. Approximately 50% of Chart Reviewed patients were <65 years of age even though the sample was enhanced for patients diagnosed with heart failure/cardiomyopathy. Most patients were white, recipients of cancer surgery, stage II or III, and had positive lymph node involvement (Table 2).

Trastuzumab, anthracycline, and other chemotherapy exposure was identified in administrative data for 20% (n=80), 38% (n=152), and 21% (n=85), respectively, of the Chart Reviewed patients. Correspondingly, trastuzumab, anthracycline, and other chemotherapy exposure was noted in the medical charts for 18% (n=72), 42% (n=158), 10% (n=38), respectively, of these patients. When the administrative data were weighted to the Chart Review Eligible patients (n=6456), Chart Reviewed and Chart Review Eligible patient samples had similar exposure prevalences (7% vs. 7%, 55% vs. 51%, and 14% vs. 19% for trastuzumab, anthracycline, and other chemotherapy exposure in the Chart Reviewed and Chart Review Eligible samples, respectively).

Overall, the weighted sensitivities of administrative data to capture chemotherapy exposure were high (>92%) (Table 3). Specificities, PPVs, and NPVs were consistently high (>93%) across data chemotherapy types except for the PPV of other chemotherapy exposure (55%). When examining the sensitivities and specificities of administrative data to capture chemotherapy exposure by likely Medicare eligible and non-eligible age category groups, sensitivities and specificities similarly were high (>91%) (Table 4). Across sites, PPVs ranged from 82% to 100%, 98% to 100%, 82% to 100%, and 38% to 94% for trastuzumab only, anthracyclines only, both anthracyclines and trastuzumab, and other chemotherapies, respectively. Across sites, NPVs ranged from 98% to 100%, 92% to 100%, 98% to 100%, and 100% to 100% for trastuzumab only, anthracyclines only, both anthracyclines and trastuzumab, and other chemotherapies, respectively.

Weighted estimates of the cumulative doses varied by chemotherapy type (Table 5). There were a substantial number of patients with limited trastuzumab exposure noted in the chart and, thus, the dose distribution was skewed. Cumulative dose estimates obtained from

pharmacy administrative data were modestly correlated and significantly associated with those obtained from medical chart review. Estimates obtained from procedural count administrative data exhibited statistically significant correlations with chart obtained data that appear to capture more accurately cumulative doses.

Discussion

We examined the validity of chemotherapy-related administrative data in capturing information on chemotherapy exposure in a random sample of 400 breast cancer patients who had received care at one of eight CRN sites across the US. We found that CRN administrative data were able to identify chemotherapy exposure with a high degree of certainty, including in patients who are not Medicare eligible. Our findings suggest that the data from the CRN are sufficiently accurate to undertake observational studies of chemotherapy effects in cancer patients.

Our findings support previous work by Aiello Bowles and colleagues who reported that the sensitivity of CRN sites' administrative data to identify chemotherapy in ovarian cancer patients was approximately 90% (17). In addition, our findings are comparable to those reported by Du and colleagues and Warren and colleagues who reported sensitivities of 91% and 88%, respectively, when comparing SEER-Medicare administrative claims data to medical chart abstracted data (21, 6). Our findings are important as they demonstrate the validity of CRN administrative data to capture chemotherapy use and provide an alternative/ supplement to such data available from SEER-Medicare. We found also a high sensitivity for individual drugs, potentially providing more detailed treatment data than SEER-Medicare since Du and colleagues, using SEER-Medicare data, confirmed receipt of a specific chemotherapy agent in only 22% of cohort who had received the agent according to chart-abstracted data (21).

We report that among those patients whose chart-abstracted data provided no evidence of receipt of chemotherapy, the specificity of administrative data was >99% for trastuzumab and anthracycline exposures. While the specificity for other chemotherapy agents was slightly lower, it still was very high. Our findings are similar to Du and colleagues who reported that SEER-Medicare data correctly identified 99% of patients where chart abstraction data found no evidence of chemotherapy receipt (21) and indicate that CRN administrative data are able to identify with a high degree of certainty patients who did not receive chemotherapy.

With sub-analysis by likely Medicare-eligible or not categories (i.e., age <65 years vs. 65 years), we found very high sensitivities (>93%) and specificities (>91%) in both age groups. The lowest specificity and PPV that we identified were for "other" chemotherapy use among patients aged 65 years and among all age groups. Examination of variability across sites revealed small variation in PPVs and NPVs except for other chemotherapies where we found wider variation. Since the goal of our study was to identify the use of anthracyclines and trastuzumab, we did not specify a complete set of possible codes to identify the "other" agents that may be used in clinical practice. Further enhancement of codes used for the assessment of other chemotherapies is being undertaken by the CRN to provide the most thorough assessment of chemotherapy use.

As future comparative effectiveness, costing, and other retrospective research will need to use cumulative chemotherapy dosing information in their analyses, we attempted to calculate cumulative chemotherapy doses for our study patients from administrative data using methodologies specific to trastuzumab and the anthracyclines. While we found minor correlations between pharmacy administrative data estimates and chart-tabulated cumulative

doses, scatter plot analyses (data not shown) indicated a wide distribution of the cumulative doses. However when using procedural count data to estimate cumulative dose, we found more reasonable correlations between administrative estimated and chart-tabulated data.

The mean doses we do report are within reason for what a typical invasive breast cancer patient (at 1.8 m^2) might receive during a course of treatment for the anthracyclines (e.g., approximately a 432 mg cumulative dose over five administrations). However, the mean doses we report for trastuzumab appear to be lower than anticipated for a cumulative dose of trastuzumab (approximately a 6000 mg cumulative dose over 50 administrations) (19). The lower cumulative doses we report may be related to the limited follow-up time allowed for by the study design and/or limited patient tolerability compared to that seen in RCTs.

Relatively little, if any findings have been reported on cumulative dose calculated from SEER-Medicare data. Lamont and colleagues reported the percent of patients with pre-specified billing doses of 5-flurouracil chemotherapy and the total count of claims for 5-flurouracil but did not calculate a cumulative doses nor assess the validity of such doses (5). Similarly, Du and colleagues reported the number of chemotherapy-related claims for breast cancer patients but not the cumulative chemotherapy dose (21).

Claims data alone might not allow for the calculation of cumulative dose since HCPCS codes for specific chemotherapy agents provide a set milligram amount (e.g., HCPCS J9355: Trastuzumab, 10 mg) and, thus, would require multiple claims be present in the data (e.g., for a weekly dose of 110 mgs, 11 claims with code J9355 per week would be required to accurately calculate the dose). In addition, HCPCS codes may not provide the fine detail to allow for graduated doses (e.g., 5 mgs of trastuzumab). Pharmacy data may provide more detailed cumulative dose estimations with the use of quantity dispensed information. However, this information was typically missing, incomplete, or inadequate in our pooled data; thus, we relied on algorithms to determine dose. A number of CRN sites are now using the oncology-specific Beacon® software component within the HealthConnect® (Epic Systems Corp., Madison, WI) electronic medical record. Extracted data from this software provides detailed information on the actual per-patient dose of chemotherapy infused/ingested during an inpatient/in-office cancer care visit. Future evaluations will need to be undertaken to ensure the validity of these data.

Our study had several limitations. Due to resource constraints, we could only abstract a 12-month follow-up period for the chart-review data while we had a two-year follow-up for administrative data. Thus, the chart review follow-up may not have been adequate to capture the entire course of chemotherapy for all patients. In addition, some of our patients may have received oncology care outside the care delivery site or the site was a secondary insurance provider; although, this is unlikely as patients in our integrated health care delivery systems typically obtain their care at the respective sites. These and the lack of a definitive data on administered dose may have restricted our ability to calculate cumulative dose. However, we used a thorough approach to chart reviewing, the 400 patients were weighted to represent over 6000 eligible breast cancer patients, and we had a mix of patients who did and did not receive chemotherapy allowing us to calculate sensitivity, specificity, NPV, and PPV.

In conclusion, we report high-quality sensitivity, specificity, PPV, and NPV of CRN administrative data to identify breast cancer patients with 2+ cm tumors and/or positive lymph nodes who had and had not received chemotherapy, respectively. The sensitivity transcended likely Medicare-eligible age groups. We had mixed results calculating cumulative chemotherapy doses and identifying patients who received other chemotherapies. However, the information obtained provides a context for additional studies

to further refine dosing algorithms and identify additional dose and chemotherapy data sources. In total, these findings support the use of CRN administrative data to conduct large-scale population-based studies of both Medicare eligible and non-eligible patients to examine comparative effectiveness of chemotherapy treatment.

Acknowledgments

References

1. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet. 2005; 365:1687–1717. [PubMed: 15894097]

2. Avorn J. In defense of pharmacoepidemiology--embracing the yin and yang of drug research. N Engl J Med. 2007; 357:2219–2221. [PubMed: 18046025]

3. Spigel DR. The value of observational cohort studies for cancer drugs. Biotechnol Healthc. 2010 Summer;7:18–24. [PubMed: 22478817]

4. National Cancer Institute. [Accessed January 20, 2011] SEER-Medicare Linked Database. Available at: http://healthservices.cancer.gov/seermedicare/.

5. Lamont EB, Lauderdale DS, Schilsky RL, Christakis NA. Construct validity of Medicare chemotherapy claims: the case of 5FU. Med Care. 2002; 40:201–211. [PubMed: 11880793]

6. Warren JL, Harlan LC, Fahey A, Virnig BA, Freeman JL, Klabunde CN, et al. Utility of the SEER-Medicare data to identify chemotherapy use. Med Care. 2002; 40(8 Suppl):IV-55–IV-61.

7. Potosky AL, Riley GF, Lubitz JD, Mentnech RM, Kessler LG. Potential for cancer related health services research using a linked Medicare-tumor registry database. Med Care. 1993; 31:732–748. [PubMed: 8336512]

8. McClish DK, Penberthy L, Whittemore M, Newschaffer C, Woolard D, Desch CE, et al. Ability of Medicare claims data and cancer registries to identify cancer cases and treatment. Am J Epidemiol. 1997; 145:227–233. [PubMed: 9012595]

9. Dobie SA, Baldwin LM, Dominitz JA, Matthews B, Billingsley K, Barlow W. Completion of therapy by Medicare patients with stage III colon cancer. J Natl Cancer Inst. 2006; 98:610–619. [PubMed: 16670386]

10. Du XL, Chan W, Giordano S, Geraci JM, Delclos GL, Burau K, et al. Variation in modes of chemotherapy administration for breast carcinoma and association with hospitalization for chemotherapy-related toxicity. Cancer. 2005; 104:913–924. [PubMed: 15991239]

11. Elkin EB, Hurria A, Mitra N, Schrag D, Panageas KS. Adjuvant chemotherapy and survival in older women with hormone receptor-negative breast cancer: assessing outcome in a population-based, observational cohort. J Clin Oncol. 2006; 24:2757–2764. [PubMed: 16782916]

12. Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the SEER-Medicare data: content, research applications, and generalizability to the United States elderly population. Med Care. 2002; 40(8 Suppl):IV-3–IV-18.

13. Bach PB, Guadagnoli E, Schrag D, Schussler N, Warren JL. Patient demographic and socioeconomic characteristics in the SEER-Medicare database: applications and limitations. Med Care. 2002; 40(8 Suppl):IV-19–IV-25.

14. Potosky, AL. [Accessed January 20, 2011] The Linked SEER-Medicare Data and Cancer Effectiveness Research. Available at: http://www.iom.edu/~/media/Files/Activity%20Files/Disease/NCPF/2009-OCT-5/Potosky-TheLinkedSEER-MedicareDataandCancerEffectivenessResearch.pdf

15. Wagner EH, Greene SM, Hart G, Field TS, Fletcher S, Geiger AM, et al. Building a research consortium of large health systems: the Cancer Research Network. J Nat Cancer Inst Monogr. 2005; 35:3–11.

16. Hornbrook MC, Hart G, Ellis JL, Bachman DJ, Ansell G, Greene SM, et al. Building a virtual cancer research organization. J Natl Cancer Inst Monograph. 2005; 35:12–25. (see also http:// crn.cancer.gov/).

17. Aiello Bowles EJ, Tuzzio L, Ritzwoller DP, Williams AE, Ross T, Wagner EH, et al. Accuracy and complexities of using automated clinical data for capturing chemotherapy administrations: implications for future research. Med Care. 2009; 47:1091–1097. [PubMed: 19648826]

18. Buist DSM, Chubak J, Prout M, Yood MU, Bosco JLF, Thwin SS, et al. Referral, receipt and completion of chemotherapy in early stage breast cancer patients aged 65 and older at high-risk of breast cancer recurrence. J Clin Oncol. 2009; 27:4508–4514. [PubMed: 19687341]

19. [Accessed January 21, 2011] NCCN Guidelines™ Version 2.2011 Invasive Breast Cancer: National Comprehensive Cancer Network. Available at: http://www.nccn.org/professionals/ physician_gls/pdf/breast.pdf

20. Go AS, Lee WY, Yang J, Lo JC, Gurwitz JH. Statin therapy and risks for death and hospitalization in chronic heart failure. JAMA. 2006; 296:2105–2111. [PubMed: 17077375]

21. Du XL, Key CR, Dickie L, Darling R, Geraci JM, Zhang D. External validation of Medicare claims for breast cancer chemotherapy compared with medical chart reviews. Med Care. 2006; 44:124–131. [PubMed: 16434911]

Table 1

Stratification Schema for Random Sampling of Patients for Chart Reviewing at Each of Eight Sites.

		Outcomes					
	Prevalent CM/HF		Incident CM/HF		No CM/HF		Patient Count
Exposures	LVEF	No LVEF	LVEF	No LVEF	LVEF	No LVEF	
Both Anthracycline and Trastuzumab	1	1	2	2	3	1	10
Either Anthracycline or Trastuzumab	1	1	2	2	3	1	10
Neither Anthracycline nor Trastuzumab	3	3	6	6	9	3	30
Patient Count	5	5	10	10	15	5	50

CM – Cardiomyopathy, HF – Heart failure, LVEF – Left ventricular ejection fraction measured

Table 2

Characteristics by Chart Reviewed, Chart Review Eligible, and the Entire Cohort of Breast Cancer Patients

Characteristic	Chart Reviewed (n=400)	Chart Review Eligible (n=6456)	Entire Cohort (N=13472)
Mean Age at Diagnosis (SD, median)	65 (14, 65)	60 (14, 58)	61 (13, 61)
Age 65 Years (n, %)	207, 52%	2326, 36%	5521, 41%
Year of Cancer Diagnosis (n, %)			
1999 – 2001	119 (30%)	1903 (30%)	4144 (31%)
2002 – 2004	124 (31%)	2143 (33%)	4522 (34%)
2005 – 2006	157 (39%)	2360 (37%)	4806 (36%)
Race (n, %)			
Asian	13, 3%	228, 4%	489, 4%
American Indian/Alaskan Native	1, <1%	20, <1%	37, <1%
Black/African American	67, 17%	735, 11%	1336, 10%
Hawaiian/Pacific Islander	1, <1%	11, <1%	24, <1%
White	299, 75%	5328, 83%	11312, 84%
Unknown	19, 5%	134, 2%	274, 2%
Worst Stage at Cancer Diagnosis (n, %)			
1	48, 12%	520, 8%	6354, 47%
2	182, 46%	3683, 57%	3711, 28%
3	149, 38%	1037, 16%	1098, 8%
4	21, 5%	244, 4%	369, 3%
Unknown	0, 0%	972, 15%	1940, 14%
Lymph Nodes (n, %)			
Negative	165, 41%	2622, 41%	9663, 72%
Positive	235, 59%	3834, 59%	3834, 28%
Surgical Treatment (n, %)			
None	74, 19%	333, 5%	616, 5%
Breast-Conservation	147, 37%	2972, 46%	7872, 58%
Total Mastectomy	67, 17%	970, 15%	1871, 14%
Modified Radical Mastectomy	109, 27%	2126, 33%	3037, 23%
Other	1, <1%	17, <1%	32, <1%
Missing	2, <1%	38, <1%	44, <1%

SD - Standard deviation

Table 3

Weighted Estimates of the Accuracy of Administrative Chemotherapy Exposure Data Compared to Medical Chart Abstracted Chemotherapy Exposure Data (n=400)

Chemotherapy Type	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value (95% CI)	Negative Predictive Value (95% CI)
Trastuzumab Only	94.6% (76.9–98.9)	99.7% (98.1–99.9)	96.2% (77.8–99.4)	99.6% (98.0–99.9)
Anthracyclines Only	96.7% (93.3–98.4)	99.5% (96.0–99.9)	99.6% (96.8–99.9)	95.8% (91.7–98.0)
Anthracyclines and Trastuzumab	92.4% (69.4–98.5)	99.7% (98.1–99.9)	94.2% (70.3–99.1)	99.6% (98.1–99.9)
Other Chemotherapy	100% (88.1–100)[1]	93.4% (90.4–95.6)	55.0% (41.6–67.7)	100% (98.9–99.4)[1]

[1] Exact binomial confidence intervals were computed in cases where accuracy measures were 0 or 100%.

Table 4

Weighted Estimates of the Sensitivity and Specificity of Administrative Chemotherapy Exposure Data Compared to Medical Chart Abstracted Chemotherapy Exposure Data by Age Categories (n=400)

Chemotherapy Type	Sensitivity (95% CI)		Specificity (95% CI)	
	Age <65 Years	Age 65 Years	Age <65 Years	Age 65 Years
Trastuzumab Only	93.6% (73.5–98.7)	100% (39.8–100)[1]	99.6% (97.0–99.9)	99.9% (63.7–100)
Anthracyclines Only	97.0% (93.3–98.6)	94.7% (78.1–98.9)	99.2% (87.8–99.9)	99.6% (92.7–99.9)
Anthracyclines and Trastuzumab	91.6% (66.8–98.3)	100% (15.8–100)[1]	99.6% (97.1–99.9)	99.9% (64.1–100)
Other Chemotherapy	100% (85.2–100)[1]	100% (54.1–100)[1]	94.3% (90.5–96.7)	91.8% (85.9–95.4)

[1] Exact binomial confidence intervals were computed in cases where accuracy measures were 0 or 100%.

Table 5

Estimates of the Cumulative Mean Chemotherapy Dose Comparing Medical Chart Abstracted Exposure Data to Administrative Chemotherapy Exposure Data

Chemotherapy Type	Medical Chart Obtained	Administrative Data Obtained	Spearman Correlation (R^2)	P-Value
Mean Milligrams (± SD, Median) Dispensed Dose[1]				
Trastuzumab (n=74)	1734 mgs (± 2252 mgs, 0 mgs)	873 mgs (± 1223 mgs, 196 mgs)	0.38 (0.14)	<0.001
Anthracyclines (n=86)	369 mgs (± 178 mgs, 416 mgs)	386 mgs (± 277 mgs, 420 mgs)	0.23 (0.05)	0.03
Mean (± SD, Median) Count of Administered Cycles[2]				
Trastuzumab (n=59)	15.4 (± 10.5, 13)	20.2 (± 14.5, 18)	0.85 (0.77)	<0.001
Anthracyclines (n=123)	4.2 (± 1.4, 4)	5.8 (± 3.6, 4)	0.45 (0.20)	<0.001

[1]Only patients with chemotherapy dosing information available in both administrative pharmacy data and chart review data are included

[2]Only patients with chemotherapy dosing information available in both administrative procedure data and chart review data are included