Title

An Evaluation of the Utility of Big Data to Supplement Cancer Treatment Information: Linkage Between IQVIA Pharmacy Database and the Surveillance, Epidemiology, and End Results Program.

Journal

JNCI Monographs, 2020(55)

ISSN

1052-6773

Authors

Tran, Quyen
Warren, Joan L
Barrett, Michael J
et al.

Publication Date

2020-05-01

DOI

10.1093/jncimonographs/lgz036

Peer reviewed

J Natl Cancer Inst Monogr (2020) 2020(55): lgz036

doi: 10.1093/jncimonographs/lgz036
Article

ARTICLE

An Evaluation of the Utility of Big Data to Supplement Cancer Treatment Information: Linkage Between IQVIA Pharmacy Database and the Surveillance, Epidemiology, and End Results Program

Quyen Tran, Joan L. Warren, Michael J. Barrett, Dave Annett, Melissa Marth, Rosemary D. Cress, Dennis Deapen, Sally L. Glaser, Scarlett L. Gomez, Stephen M. Schwartz, Christopher I. Li, Xiao-Cheng Wu, Lindsey Enewold, Linda C. Harlan, Donna R. Rivera, Deborah M. Winn, Lynne Penberthy, Kathleen A. Cronin

See the Notes section for the full list of authors' affiliations.
Correspondence to: Quyen Tran, PharmD, 10903 New Hampshire Avenue, Silver Spring, MD 20993 (e-mail: Quyen.tran1@fda.hhs.gov).

Abstract

Abstract Oral anticancer medications (OAMs) are increasingly utilized. We evaluated the representativeness and completeness of IQVIA, a large aggregator of pharmacy data, for breast cancer, colon cancer, chronic myeloid leukemia, and myeloma cases diagnosed in six Surveillance, Epidemiology, and End Results Program (SEER) registries between 2007 and 2011. Patient's SEER and SEER-Medicare data were linked and compared with IQVIA pharmacy data from 2006 to 2012 for specific OAMs. Overall, 67.6% of SEER cases had a pharmacy claim in IQVIA during the treatment assessment window. This varied by location, race and ethnicity, and insurance status. IQVIA consistently identified fewer cases who received an OAM of interest than SEER-Medicare. The difference was least pronounced for breast cancer agents and most pronounced for myeloma agents. The IQVIA pharmacy database included a large portion of persons in the SEER areas. Future studies should assess receipt of OAMs for other cancer sites and in different SEER registries.

The role of oral anticancer medications (OAMs) in the treatment of cancer has been increasing in the past decade with currently over 50 United States Food and Drug Administration-approved OAMs on the market (1). This number is expected to continuously rise, with approximately 25.0–30.0% of oncology pipeline drugs being developed as oral agents (2). OAMs have changed the landscape of oncology treatment by providing patients with greater autonomy and convenience of how, when, and where they take their medication as compared with intravenous medication (3). In contrast to intravenous anticancer agents, OAMs are readily available and predominately dispensed at pharmacies instead of inpatient and outpatient facilities.

Cancer surveillance, the ongoing and systematic collection and analysis of information on new cancer cases, extent of disease, treatment, survival, and mortality, plays an important and crucial role in understanding the burden and effects of cancer in the United States. Traditionally, the main source of information for cancer surveillance has been hospitals. However, with the increase in OAM utilization, relying on hospital data to capture cancer treatment is no longer adequate because most

Received: 9 October 2019; **Revised:** 12 December 2019; **Accepted:** 19 December 2019

hospital data do not include information about OAM use. To obtain information about OAM use, other data sources need to be explored. One possible data source is large aggregators of electronic pharmacy data. IQVIA is one such source; they collect electronic pharmacy transactions from various pharmacy sources (eg, retail, mail order, and specialty pharmacies). However, there has been no assessment of the extent to which pharmacy data can enhance information about OAM use at the population level. The availability of such data could be used to understand the patterns of OAMs usage and adoption of new agents. In addition, the availability of pharmacy data regarding OAM utilization may provide needed information about medication adherence because reducing nonadherence is an ongoing public health priority that has health and economic impacts (4).

We performed a data linkage to the IQVIA pharmacy data to evaluate the completeness and representativeness of these data for the capture of OAMs for persons with cancer diagnosed in six population-based cancer registries that participate in the National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results Program (SEER). The aims of the study were to 1) describe the number of persons in SEER data who were also included in the IQVIA pharmacy data, 2) assess the differences in demographic and clinical characteristics between persons in the SEER data who were and were not included in the IQVIA data, and 3) evaluate the validity of selected OAMs reported in IQVIA data compared with SEER-Medicare and the Fred Hutchinson Cancer Research Center's Breast Cancer Risk and Various Outcomes study (BRAVO) data.

Methods

Data Sources

Four data sources were used in this study: IQVIA (formerly known as IMS Health and QuintilesIMS) longitudinal pharmacy data, patient-level information from the participating SEER registries, the SEER-Medicare linked dataset, and BRAVO data (5–8).

IQVIA

The analysis included a linkage of persons captured in the IQVIA pharmacy data with persons reported in the SEER data. IQVIA is an information, services, and health technology company drawing information from more than 100 000 suppliers. Specifically, IQVIA obtains pharmacy data from multiple sources, including retail, mail order, and specialty pharmacies. IQVIA currently covers approximately 90.0% of dispensed prescriptions in the United States although at the time of this analysis, coverage was reportedly 70.0% in 2015. IQVIA obtains information about prescription claims, including a subset of those paid out of pocket (cash) from collaborating commercial entities. This information includes personally identifiable information (PII) (eg, first and last name, sex, date of birth, and address), drug, dose, day supply, date of prescription, and date of prescription filled. IQVIA does not maintain any PII in their database. Instead, a trusted third party of IQVIA assigns an IQVIA ID to each patient in the IQVIA database using encryption software. Only the IQVIA IDs are shared with IQVIA, resulting in a deidentified database.

SEER Registries

The SEER registries collect demographic information, clinical and tumor characteristics, mortality outcomes, and cause of death for all incident cancers reported for individuals who reside in one of the registries' defined geographic areas. The SEER registries collect data primarily from hospitals, with limited information about treatment provided outside the hospital. Therefore, we used the linked SEER-Medicare data to obtain information about OAM utilization from Medicare claims.

SEER-Medicare

The SEER-Medicare data result from a linkage of persons in the SEER data to Medicare enrollment information. For SEER patients aged 65 years or older, 95.0% have been linked to the Medicare enrollment file (9). Medicare Part B and D claims were used to obtain information about specific OAMs. Part B claims were available only for patients enrolled in fee-for-service (FFS). Part D claims were available for patients enrolled in FFS and health maintenance organizations (HMOs). Since the introduction of Medicare Part D in 2006, enrollment in a Part D plan has increased from 53.0% to 72.0% of all eligible Medicare beneficiaries in 2015 (10). All claims include dates of service and codes for specific drugs using either Health Care Procedure Codes or National Drug Codes (NDCs). We used the linked SEER-Medicare data as a comparator source for identifying OAMs.

Breast Cancer Risk and Various Outcomes Study

BRAVO includes data from approximately 3000 women aged 20–69 years old with breast cancer living in Washington and New Mexico. Data were abstracted from medical records and obtained from telephone interviews with the study participants about their medical history, breast cancer treatment, lifestyle, and family cancer history. We used the BRAVO data as another comparator source for identifying breast cancer OAMs. BRAVO includes breast cancer cases who are younger than 65 years, allowing for an assessment not available with the SEER-Medicare data.

Study Sample and Eligibility Criteria

Persons included in the analysis resided in one of the six SEER registries that participated in the study: Louisiana, Seattle-Puget Sound, Greater California, Los Angeles, San Francisco-Oakland, and San Jose-Monterey. These SEER registries were selected because IQVIA penetration was estimated to be high in these regions. Persons in the analysis were required to have been diagnosed with in situ and invasive female breast cancer, chronic myeloid leukemia (CML), colon cancer, or multiple myeloma between 2007 and 2011. These cancer sites were chosen because OAMs are commonly utilized in the first course of therapy. For breast and colon cancer, only patients with a first or only primary of the breast or colon, respectively, were included. Individuals with unknown or incomplete SEER data about first and last name, date of birth, sex, address, ZIP code, and month of diagnosis were excluded. Autopsy-only and death certificate–only cases were excluded. The final analytic dataset included 198 648 case patients.

For the comparison of OAMs identified from the IQVIA data with those identified in the SEER-Medicare data, persons were required to have matched and have at least one prescription claim in the IQVIA pharmacy database during the treatment assessment window, also known as matched and active in IQVIA, and have continuous Medicare enrollment during the treatment assessment window. The treatment assessment window was defined as the 3 months before to the 12 months following the cancer diagnosis date. The agents that were evaluated varied by cancer site. For breast cancer, we focused on tamoxifen (Nolvadex), anastrozole (Arimidex), exemestane (Aromasin),

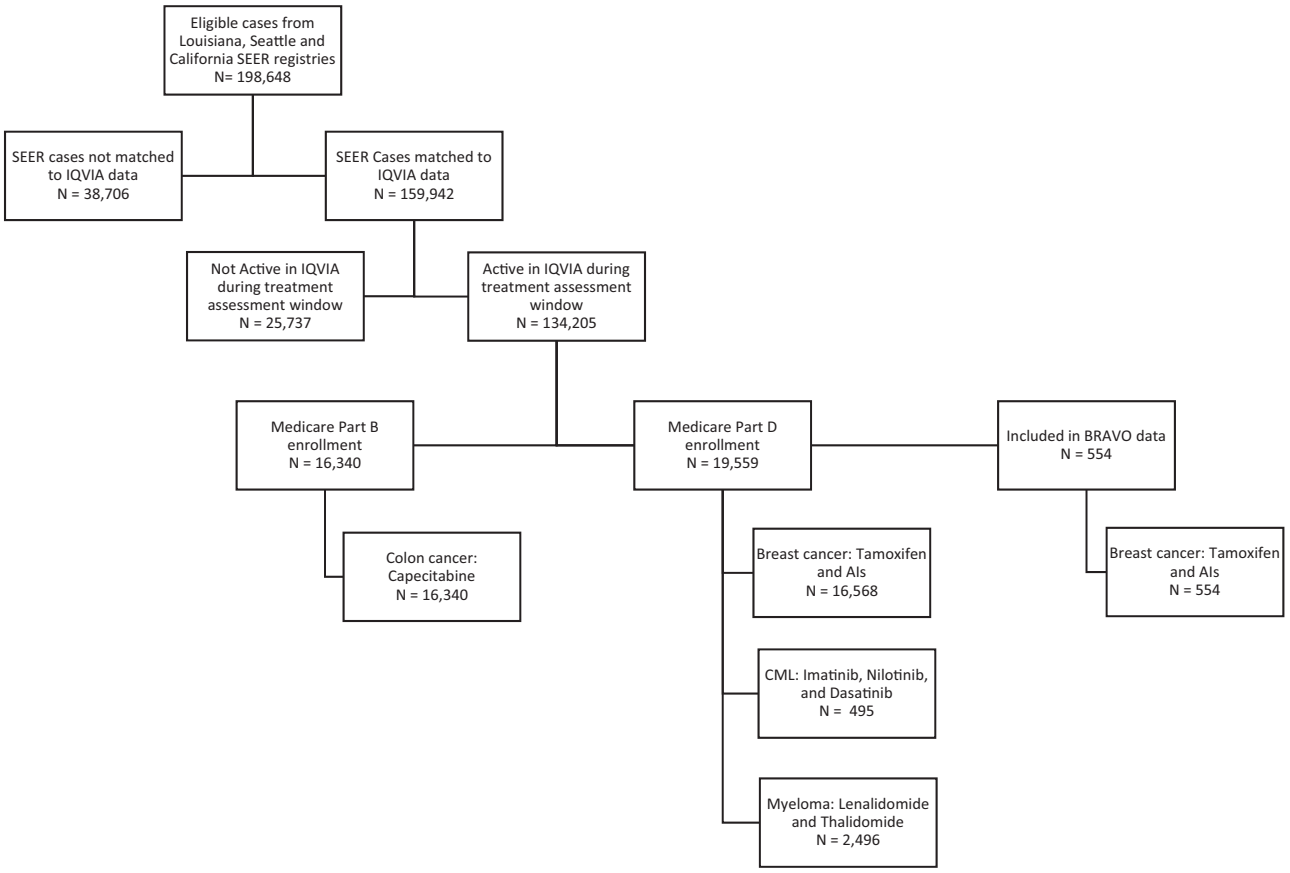

Figure 1. Creation of study cohort for the comparison of oral anticancer medications (OAMs) for the treatment of chronic myeloid leukemia (CML), colon cancer, and myeloma between IQVIA pharmacy data, SEER-Medicare data and the Fred Hutchinson Cancer Research Center's Breast Cancer Risk and Various Outcomes study (BRAVO) data. Active in IQVIA was defined as having a pharmacy transaction of any kind, not specifically for an oncology agent, captured in IQVIA. Treatment assessment window was defined as 3 months before to 12 months following cancer diagnosis date. For the Medicare component, cases were required to have continuous Medicare coverage during the treatment assessment window. AIs = aromatase inhibitors, including letrozole, anastrozole, and exemestane; SEER = Surveillance, Epidemiology, and End Results Program.

and letrozole (Femara). For CML, the evaluation included imatinib (Gleevec), nilotinib (Tasigna), and dasatinib (Sprycel). For colon cancer, the assessment focused on capecitabine (Xeloda). For myeloma, the evaluation included thalidomide (Thalomid) and lenalidomide (Revlimid). These agents were specifically chosen because of their high utilization in the first course of therapy. For the evaluation of capecitabine use in colon cancer patients, persons were required to have continuous Medicare Part B because capecitabine is covered by Medicare Part B. For all other selected OAMs, persons were required to have continuous Medicare Part D enrollment during the treatment assessment window. For the evaluation of breast cancer OAMs, cases were also required to have hormone receptor positive tumors. There was a total of 16 340 colon cancer cases, 16 568 breast cancer cases, 495 CML cases, and 2496 myeloma cases (Figure 1).

For the comparison of breast cancer OAMs identified from the IQVIA data with those identified in the BRAVO data, cases were required to have matched and be active in IQVIA, as well as have hormone receptor positive tumors. There was a total of 554 cases.

Linkage Process Between SEER and IQVIA Data

Figure 2 depicts the data linkage process between the SEER registries and IQVIA data. The data linkage process consisted of five main steps. In step 1, a finder file containing PII for all eligible cases was transmitted from each participating SEER registry to an NCI contractor. The finder file included information on first and last name, date of birth, sex, address, and SEER case identification number. These files were matched to SEER, SEER-Medicare, and BRAVO data to obtain specific variables that were retained for the analysis. A full list of the data elements from SEER and SEER-Medicare data can be found in Supplementary Table 1 (available online). BRAVO data included the date of breast cancer hormonal therapy (HT) prescription filled. The NCI contractor then assigned a unique pilot study ID to each case. The NCI contractor used encryption software supplied by the trusted third party of IQVIA to create unique arbitrary values called hashed tokens. The hashed tokens were generated based on multiple combinations of each patient's PII. The output from the encryption software replaced PII with hashed tokens. The method was irreversible; PII could not be traced back from the hashed tokens.

In step 2, the NCI contractor sent a file that included only the hashed tokens and pilot study IDs to IQVIA's trusted third party. Using a similar process to that of the NCI contractor, IQVIA's trusted third party applied their encryption software to the PII for persons in the IQVIA data to produce hashed tokens. IQVIA's trusted third party used the hashed tokens to match persons in the SEER data to those in the IQVIA pharmacy database from 2006 to 2012 through a deterministic algorithm. The exact

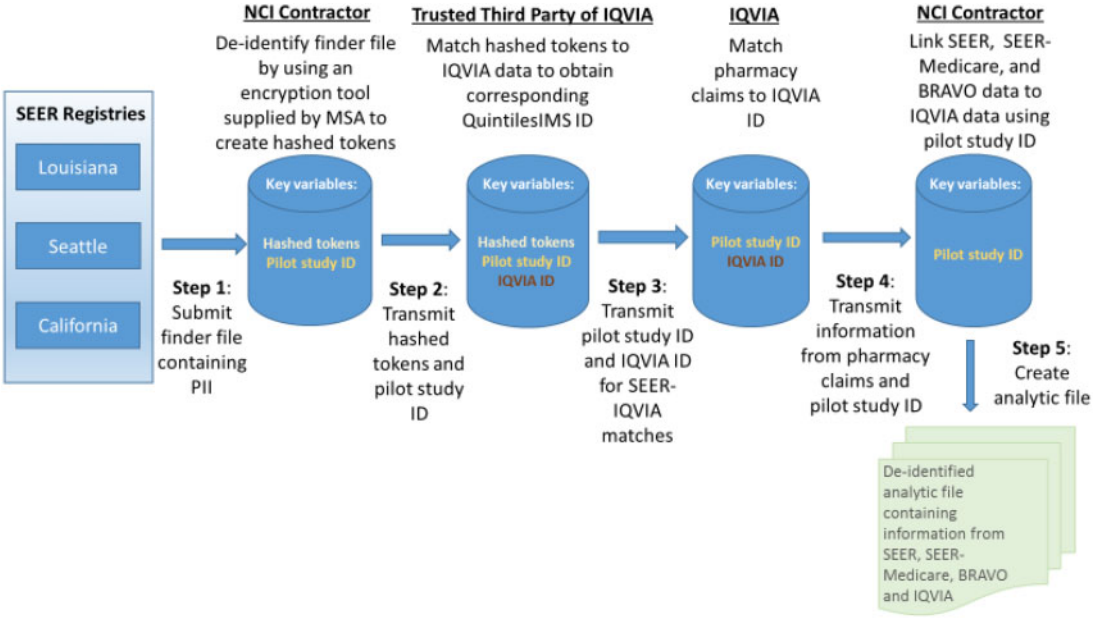

Figure 2. Data linkage process between Surveillance, Epidemiology, and End Results Program (SEER) and IQVIA. In step 1, each participating registry submits a finder file containing patient identifiable information (PII) for all eligible cases in their defined geographic location to the National Cancer Institute (NCI) contractor. In step 2, the NCI contractor retrieves and links desired data elements from the Surveillance, Epidemiology, and End Results program (SEER), SEER-Medicare, and the Fred Hutchinson Cancer Research Center's Breast Cancer Risk and Various Outcomes study (BRAVO) data for case patients in the finder files. Each case patient was assigned a unique pilot study ID. The NCI contractor then used an encryption tool supplied by IQVIA's trusted third party to generate arbitrary pseudo-random values called hashed tokens in place of PII. The file containing the hashed tokens and pilot study IDs was shared with IQVIA's trusted third party. In step 3, IQVIA's trusted third party matches SEER case patients captured in the pharmacy data of IQVIA using the hashed tokens. Every patient in the IQVIA database has a corresponding IQVIA ID. IQVIA's trusted third party shares the IQVOA IDs and pilot study IDs of each SEER-IQVIA match with IQVIA. In the fourth step, IQVIA links information about pharmacy claims that occurred from 2006 to 2012 for each IQVIA ID. IQVIA shares this information and the pilot study IDs with the NCI contractor. In the fifth step, the NCI contractor converges and links all the information from the four data sources (SEER, SEER-Medicare, BRAVO and IQVIA) using the pilot study IDs resulting in one de-identified analytic file.

matching algorithm is proprietary, but the data elements involved in the algorithm included name, date of birth, sex, address, and ZIP code.

For step 3, IQVIA's trusted third party transmitted a file to IQVIA containing the pilot study ID and IQVIA ID for persons in the IQVIA data who matched to the SEER data. IQVIA retrieved and linked pharmacy claims data for each IQVIA ID. Specifically, IQVIA provided information on 1) pharmaceutical transactions for selected OAMs, including fill date, day supply, quantity, and type of payer and 2) whether any other pharmaceutical transactions were captured in their database for each month between 2006 and 2012.

In step 4, IQVIA's trusted third party sent a file with the pilot study IDs and corresponding pharmacy claims data for persons found in both the SEER and IQVIA data to the NCI contractor. Before receiving data from IQVIA, the NCI contractor deleted the crosswalk between the SEER case ID and pilot study ID.

For step 5, using the pilot study ID, the NCI contractor converged and created a deidentified analytic file that included data from SEER, SEER-Medicare, BRAVO, and IQVIA.

Assessment of Match Rates Between SEER and IQVIA

The linkage between persons in the SEER and IQVIA data was categorized into three mutually exclusive groups: nonmatched, matched but not active in IQVIA during the treatment assessment window, and matched and active in IQVIA during the treatment assessment window. Active in IQVIA was defined as a pharmacy claim of any kind, not specifically for oncology agents, captured in IQVIA during the treatment assessment window. The date of cancer diagnosis was ascertained from the SEER registries.

Overall counts and percentages of linkage match results by demographic and clinical characteristics were calculated. Characteristics examined included cancer site, registry (Louisiana, Seattle, and California), sex, age at diagnosis, race and ethnicity, insurance status, year of diagnosis, stage per the American Joint Committee on Cancer's Cancer Staging Manual (6th edition), and tumor grade (well differentiated, moderately differentiated, poorly differentiated, and anaplastic). Race and ethnicity were categorized into five distinct groups: non-Hispanic (NH) white, NH black, NH Asian or Pacific Islander, non-Hispanic American Indian or Alaska Native, and Hispanic. Insurance status was ascertained from SEER registries at the time of diagnosis and was classified into four categories: uninsured, Medicaid (including Indian or Public Health Service), insured (including private insurance and Medicare), and unknown insurance status. For SEER cases aged 65 years or older who had continuous Medicare Part B and/or D coverage during the treatment assessment window, distributions of linkage match results by type of Medicare coverage (Medicare Part B vs D) and payment model were calculated. SEER cases were grouped into three mutually exclusive payment model categories: FFS, HMO, and blend. FFS and HMO categories were defined as continuous FFS or HMO Medicare coverage during the treatment assessment window. Blend coverage reflected persons with continuous Medicare coverage during the treatment assessment window who had a mix of FFS and HMO coverage.

We used multivariable logistic regression to assess associations between case characteristics and the dependent variable of whether a person matched and was active in IQVIA during the treatment assessment window. Odds ratios (OR) and 95% confidence intervals (95% CI) were calculated (SAS, 9.3, Cary, NC). The SEER cases were grouped into those aged 64 years and younger and 65 years and older to assess three additional attributes only available for the 65 years and older population: Medicare Part B and/or D enrollment during the treatment assessment window, Medicare payment model (FFS, HMO, and blend), and Charlson comorbidity index.

Assessment of Capture Rate in SEER-Medicare, IQVIA, and BRAVO Data for Identification of Specific OAMs

Receipt of oral agents was identified in SEER-Medicare and IQVIA through NDCs. In addition to NDCs, we used Health Care Procedure Codes to identify receipt of capecitabine in Medicare Part B data. All other agents were examined in Medicare Part D data. Cases were required to have at least one qualifying transaction during the treatment assessment window to be identified as having received an agent. See Supplementary Tables 2 and 3 (available online) for a list of the drug codes used to identify claims for specific OAMs and a summary of the data sources used to evaluate OAM utilization among cases who matched and were active in IQVIA during the treatment assessment window, respectively.

Receipt of breast cancer oral agents was identified in BRAVO through self-report or data abstraction of medical records. Cases were defined in BRAVO as receiving an HT of interest if the case reported use of an HT of interest or medical records identified the case received the agent of interest within 1 year of cancer diagnosis date.

IQVIA capture rate for the comparison between IQVIA and Medicare was defined as the number of cases who had a pharmacy transaction in IQVIA for a specific agent divided by the number of cases who had a transaction for the specific agent recorded in either IQVIA or Medicare. IQVIA capture rate for the comparison between IQVIA and BRAVO was defined as the number of cases who had a pharmacy transaction in IQVIA for a specific agent divided by the number of cases who had a transaction for the agent recorded in IQVIA and cases identified to have received the agent in BRAVO. Similarly, the Medicare capture rate was defined as the number of cases who had a Medicare claim for a specific agent divided by the number of cases who had a transaction for a specific agent recorded in either IQVIA or Medicare. Although Medicare is generally considered complete for individuals eligible for coverage, Medicare was not considered as the gold standard because there were a few instances where transactions captured in IQVIA were not captured in Medicare. The BRAVO capture rate was defined as the number of cases identified to have received a specific agent in BRAVO divided by the number of cases who had a transaction for the agent recorded in IQVIA and cases identified to have received the agent in BRAVO.

Results

Table 1 shows the number and percent of cases that fell into the three categories by case demographics, tumor characteristics, and cancer registry. Overall, 67.6% of the cases were found to be in the IQVIA data and active during the treatment assessment window. Match rates were lower for colon cancer than for other cancer sites (colon: 61.5% vs breast: 69.6%; CML: 67.5%; and myeloma: 68.9%). Cancer site-specific distribution of linkage match results by demographic and clinical characteristics are presented in Supplementary Tables 4–7 (available online).

Large differences were seen by registry in terms of match rate and being active in the IQVIA data, with 65.2%, 71.6%, and 80.4% of cases active in IQVIA during the treatment assessment window for California, Louisiana, and Seattle, respectively. The California registries were presented together because no differences in match rate were observed between the three California registries. To assess the match rate difference between the three geographic locations, we examined the type of Medicare coverage a patient aged 65 years and older had at diagnosis: FFS, HMO, or a blend of the two. California and Seattle had lower match rates for cancer patients with HMO coverage, although the effect was more pronounced in California. The match rates in Louisiana did not greatly vary by type of insurance coverage (Supplementary Figure 1, available online).

The match rate also varied by sex, race and ethnicity, age, and insurance status. The match rate was lower for males and nonwhite patients. Match rates decreased with age, especially for patients aged 75 years and older (Table 1). The match rate was highest for patients enrolled in Medicaid and lowest for those uninsured and with unknown insurance status (Medicaid: 74.7% vs uninsured: 59.4%; and unknown: 59.2%). The match rate did not vary by tumor stage or grade (data not shown).

Results from the multivariable regression in Table 2 were similar to the descriptive analysis in Table 1. Table 2 presents the odds ratios of being matched and active in IQVIA during the treatment assessment window among SEER patients aged younger than 65 years and 65 years and older. Both groups younger than 65 years and 65 years and older had higher odds of being matched and active in IQVIA during the treatment assessment window for NH whites, younger ages, and later diagnosis years. For both age groups, odds were lower for colon cancer and for cancer diagnosed in California. Patients younger than 65 years with Medicaid coverage at diagnosis were more likely to be matched and active compared with other insured patients (OR = 1.66, 95% CI = 1.59 to 1.72). For patients aged 65 years and older, those not enrolled in Medicare Part B or D during the treatment assessment window had a 32.0% lower likelihood of matching and being active in IQVIA (OR = 0.68, 95% CI = 0.64 to 0.72). Of the patients enrolled in Medicare during the treatment assessment window, those enrolled through an HMO were less likely to be matched and active in IQVIA than those enrolled in FFS (OR = 0.39, 95% CI = 0.37 to 0.42).

We compared the reporting of specific OAMs in the IQVIA data with the SEER-Medicare data among the cohort of cases aged 65 years and older who 1) matched and were active in IQVIA during the treatment assessment window and 2) had continuous enrollment in Medicare during the treatment assessment window (Table 3). Except for capecitabine, the Medicare capture rate was higher than the IQVIA capture rate for all selected OAMs. For Louisiana, the IQVIA capture rate ranged from 77.7% to 85.8%, 0.0% to 71.1%, and 25.0% to 56.5% for breast, CML, and myeloma agents, respectively. In Seattle, the IQVIA capture rate ranged from 75.8% to 85.7%, 33.3% to 100.0%, and 50.6% to 65.8% for breast, CML, and myeloma agents, respectively. In California, the IQVIA capture rate ranged from 79.8% to 83.9%, 75.0% to 82.6%, and 51.7% to 70.5% for breast, CML, and myeloma agents, respectively.

A similar trend was identified among patients younger than 65 years old. Supplementary Table 8 (available online) presents the comparison between IQVIA and BRAVO. Of the eligible women, 91.9% of the women who matched and were active in

Table 1. Demographic and tumor characteristics of SEER breast cancer, CML, colon cancer, and myeloma cases diagnosed between 2007 and 2011 identified as matched and nonmatched in the IQVIA pharmacy database (N = 198 648)

	Nonmatched	Matched		Total
		Not active in IQVIA during treatment assessment window	Active in IQVIA during treatment assessment window*	
	Count no. (%)	Count no. (%)	Count no. (%)	
Overall	38 706 (19.5)	25 737 (13.0)	134 205 (67.6)	198 648
Cancer site				
Breast	24 668 (17.9)	17 340 (12.6)	95 945 (69.6)	137 953
CML	564 (19.7)	365 (12.8)	1932 (67.5)	2861
Colon	11 412 (24.1)	6862 (14.5)	29 180 (61.5)	47 454
Myeloma	2062 (19.9)	1170 (11.3)	7148 (68.9)	10 380
Registry				
Louisiana	2769 (12.7)	3427 (15.7)	15 634 (71.6)	21 830
Seattle	1881 (8.7)	2360 (10.9)	17 400 (80.4)	21 641
California	34 056 (22.0)	19 950 (12.9)	101 171 (65.2)	155 177
Sex				
Female	31 244 (18.6)	21 519 (12.8)	115 307 (68.6)	168 070
Male	7462 (24.4)	4218 (13.8)	18 898 (61.8)	30 578
Age at diagnosis, y				
<45	3637 (16.9)	2619 (12.2)	15 240 (70.9)	21 496
45–54	7918 (18.3)	5416 (12.5)	29 954 (69.2)	43 288
55–64	9931 (19.8)	6 577 (13.1)	33 544 (67.0)	50 052
65–74	8392 (20.1)	5287 (12.7)	28 130 (67.3)	41 809
75+	8828 (21.0)	5838 (13.9)	27 337 (65.1)	42 003
Race and ethnicity				
NH white	21 194 (16.9)	15 651 (12.5)	88 587 (70.6)	125 432
NH black	4802 (25.8)	2772 (14.9)	11 010 (59.2)	18 584
NH API	5000 (22.8)	2770 (12.6)	14 159 (64.6)	21 929
NH AI/AN	161 (17.1)	127 (13.5)	652 (69.4)	940
Hispanic	6907 (23.4)	4136 (14.0)	18 521 (62.7)	29 564
Unknown	642 (29.2)	281 (12.8)	1276 (58.0)	2199
Insurance status				
Uninsured	817 (23.9)	571 (16.7)	2028 (59.4)	3416
Medicaid	3364 (13.1)	3151 (12.3)	19 211 (74.7)	25 726
Insured	33 081 (20.2)	21 168 (12.9)	109 637 (66.9)	163 886
Unknown	1444 (25.7)	847 (15.1)	3329 (59.2)	5620
Year of diagnosis				
2007	8057 (21.2)	5632 (14.9)	24 249 (63.9)	37 938
2008	8063 (20.3)	5013 (12.6)	26 609 (67.1)	39 685
2009	7575 (18.8)	4791 (11.9)	27 939 (69.3)	40 305
2010	7338 (18.6)	5132 (13.0)	27 029 (68.4)	39 499
2011	7673 (18.6)	5169 (12.5)	28 379 (68.9)	41 221

*To be considered active in IQVIA data, an individual had to have at least one prescription claim between the 3 months before and the 12 months following cancer diagnosis captured in the IQVIA pharmacy data. AI/AN = American Indian and Alaskan Natives; API = Asian Pacific Islander; CML = chronic myeloid leukemia; FFS = fee-for-service; HMO = health maintenance organization; NH = non-Hispanic; SEER = Surveillance, Epidemiology, and End Results Program.

IQVIA were younger than 65 years old (data not shown). The IQVIA capture rate was consistently lower than the BRAVO capture rate for breast cancer OAMs; the capture rate ranged from 66.7% to 84.1% and from 93.3% to 96.2% for IQVIA and BRAVO, respectively.

Discussion

This study analyzed the linkage between breast cancer, CML, colon cancer, and myeloma cases diagnosed in three SEER geographic locations (Louisiana, Seattle, and California, which is composed of four registries) between 2007 and 2011 with available information from the IQVIA pharmacy database between 2006 and 2012. The match rate of SEER cases to IQVIA data varied by registry. California had the lowest match rate, compared with Seattle and Louisiana, for persons of all ages and for the Medicare population. In Seattle and California, the match rate was lower for persons in Medicare HMOs. This finding suggests matching to IQVIA is affected by location and HMO penetration. In 2016, the Medicare HMO penetration in California was 39.0% compared with 31.0% and 30.0% in Louisiana and Washington, respectively (11). Additionally, match rates in California were lower than in Seattle for persons in HMOs. This suggests that HMO-affiliated pharmacies in California may not be captured in the IQVIA pharmacy database. Many HMOs operate their own pharmacies, thus, making prescription distribution and reimbursement streamlined and closed from external entities that are commonly involved in the prescription claims processing.

Table 2. Odds ratio of being matched and active in IQVIA data during the treatment assessment window by demographic and patient characteristics among SEER breast cancer, CML, colon cancer, and myeloma cases diagnosed between 2007 and 2011

	<65 YO at diagnosis		≥65 YO at diagnosis	
	OR	95% CI	OR	95% CI
Age at diagnosis, y				
<45	1.00	Referent	—	—
45–54	0.91	0.88,0.95	—	—
55–64	0.81	0.78,0.84	—	—
65–74	—	—	1.00	Referent
75+	—	—	0.88	0.85,0.90
Race and ethnicity				
NH white	1.00	Referent	1.00	Referent
NH black	0.53	0.51,0.56	0.61	0.57,0.64
NH API	0.77	0.64,0.92	0.74	0.58,0.94
NH AI/AN	0.69	0.66,0.71	0.97	0.92,1.03
Hispanic	0.62	0.60,0.65	0.95	0.90,1.00
Unknown	0.64	0.56,0.71	0.91	0.79,1.05
Insurance status				
Insured	1.00	Referent	1.00	Referent
Medicaid	1.66	1.59,1.72	1.18	1.11,1.24
Uninsured	0.80	0.74,0.86	0.98	0.78,1.23
Unknown	0.81	0.74,0.88	0.80	0.73,0.87
Year of diagnosis				
2007	1.00	Referent	1.00	Referent
2008	1.15	1.11,1.20	1.21	1.15,1.27
2009	1.28	1.23,1.33	1.37	1.30,1.43
2010	1.22	1.17,1.27	1.32	1.26,1.39
2011	1.28	1.23,1.33	1.30	1.24,1.36
Primary site				
Breast	1.00	Referent	1.00	Referent
CML	1.12	0.98,1.27	1.31	1.14,1.52
Colon	0.68	0.65,0.70	0.73	0.71,0.76
Myeloma	1.31	1.18,1.44	1.55	1.42,1.70
Stage (AJCC 6th)				
0, I, and II	1.00	Referent	1.00	Referent
III and IV	1.22	1.17,1.26	1.12	1.08,1.17
Unknown or not applicable	0.97	0.90,1.04	0.82	0.77,0.88
Grade				
I and II	1.00	Referent	1.00	Referent
III and IV	1.06	1.03,1.09	1.03	0.99,1.07
Unknown or other	0.87	0.83,0.91	0.79	0.75,0.84
Registry				
Louisiana	1.64	1.56,1.72	1.01	0.96,1.06
Seattle	2.02	1.92,2.12	1.80	1.70,1.90
California	1.00	Referent	1.00	Referent
Medicare Part B and/or D during treatment assessment window				
Yes	—	—	1.00	Referent
No	—	—	0.68	0.64,0.72
Payment model of Medicare during treatment assessment window*				
FFS†	—	—	1.00	Referent
HMO‡	—	—	0.39	0.37,0.42
Blend§	—	—	0.92	0.84,1.01
Charlson comorbidity index				
0	—	—	1.00	Referent
1	—	—	0.95	0.89,1.01
2+	—	—	0.77	0.73,0.82
Unknown or not applicable	—	—	0.61	0.57,0.65

*Limited to SEER cases aged 65 years and older who had continuous Medicare Part B and/or Part D enrollment 3 months before to 12 months after cancer diagnosis. AI/AN = American Indian/Alaska Native; API = Asian Pacific Islander; CI = confidence interval; CML = chronic myeloid leukemia; FFS = fee-for-service; HMO = health maintenance organization; NH = non-Hispanic; SEER = Surveillance; Epidemiology and End Results Program; OR = odds ratio; YO = years old.

†FFS was defined as continuous Medicare Part D, FFS, enrollment 3 months before to 12 months postdiagnosis.

‡HMO was defined as continuous Medicare Part B and/or D enrollment through HMO for 3 months to before 12 months postdiagnosis.

§Blend was defined as continuous Medicare Part B and/or D enrollment through FFS or HMO for 3 months before to 12 months postdiagnosis (ie, patient has Medicare Part B and/or D enrollment through FFS payment model for 3 months before cancer diagnosis and Medicare Part B and/or D enrollment through an HMO payment model for 12 months post cancer diagnosis).

Table 3. OAM agreement and capture rate using the IQVIA and Medicare data by registry and cancer type, 2007–2011

SEER registry or type of OAM	IQVIA no Medicare no	IQVIA no Medicare yes	IQVIA yes Medicare no	IQVIA yes Medicare yes	IQVIA capture rate* %	Medicare capture rate† %
Louisiana	Count no. (%)	Count no. (%)	Count no. (%)	Count no. (%)		
Breast (N = 1898)						
Tamoxifen	1571 (82.8)	73 (3.8)	— (0.4)	247 (13.0)	77.7%	97.9%
Anastrozole	1072 (56.5)	117 (6.2)	17 (0.9)	692 (36.5)	85.8%	97.9%
Letrozole	1461 (77.0)	82 (4.3)	— (0.8)	340 (17.9)	81.2%	96.6%
Exemestane	1836 (96.7)	— (0.6)	— (0.1)	49 (2.6)	80.6%	98.4%
CML (N = 71)						
Imatinib	33 (46.5)	— (15.5)	— (0.0)	27 (38.0)	71.1%	100.0%
Dasatinib	68 (95.8)	— (4.2)	— (0.00	— (0.0)	0.0%	100.0%
Nilotinib	66 (93.0)	— (4.2)	— (0.0)	— (2.8)	40.0%	100.0%
Colon (N = 2095)						
Capecitabine	1917 (91.5)	62 (3.0)	45 (2.1)	71 (3.4)	65.2%	74.7%
Myeloma (N = 283)						
Lenalidomide	183 (64.7)	75 (26.5)	— (0.0)	25 (8.8)	25.0%	100.0%
Thalidomide	214 (75.6)	30 (10.6)	— (0.7)	37 (13.1)	56.5%	97.1%
Seattle						
Breast (N = 1933)						
Tamoxifen	1548 (80.1)	93 (4.8)	— (0.4)	285 (14.7)	75.8%	98.2%
Anastrozole	1305 (67.5)	118 (6.1)	— (0.7)	496 (25.7)	81.2%	97.8%
Letrozole	1551 (80.2)	71 (3.7)	— (0.7)	297 (15.4)	81.4%	96.3%
Exemestane	1884 (97.5)	— (0.4)	— (0.1)	41 (2.1)	85.7%	98.0%
CML (N = 48)						
Imatinib	30 (62.5)	— (2.1)	— (2.1)	16 (33.3)	94.4%	94.4%
Dasatinib	45 (93.8)	— (4.2)	— (0.0)	— (2.1)	33.3%	100.0%
Nilotinib	46 (95.8)	— (0.0)	— (0.0)	— (4.2)	100.0%	100.0%
Colon (N = 1860)						
Capecitabine	1672 (89.9)	47 (2.5)	57 (3.1)	84 (4.5)	75.0%	69.7%
Myeloma (N = 266)						
Lenalidomide	185 (69.5)	40 (15.0)	— (0.8)	39 (14.7)	50.6%	97.5%
Thalidomide	228 (85.7)	— (4.9)	— (0.8)	23 (8.6)	65.8%	94.7%
California						
Breast (N = 12 742)						
Tamoxifen	10 679 (83.8)	416 (3.3)	39 (0.3)	1608 (12.6)	79.8%	98.1%
Anastrozole	7008 (55.0)	1147 (9.0)	107 (0.8)	4480 (35.2)	80.0%	98.1%
Letrozole	10 434 (81.9)	447 (3.5)	66 (0.5)	1795 (14.1)	80.6%	97.1%
Exemestane	12 034 (94.4)	114 (0.9)	27 (0.2)	567 (4.4)	83.9%	96.2%
CML (N = 376)						
Imatinib	238 (63.3)	24 (6.4)	— (1.3)	109 (29.0)	82.6%	96.4%
Dasatinib	352 (93.6)	— (1.6)	— (0.0)	18 (4.8)	75.0%	100.0%
Nilotinib	354 (94.1)	— (1.3)	— (0.3)	16 (4.3)	77.3%	95.5%
Colon (N = 12 385)						
Capecitabine	10 729 (86.6)	243 (2.0)	667 (5.4)	746 (6.0)	85.3%	59.7%
Myeloma (N = 1947)						
Lenalidomide	1312 (67.4)	307 (15.8)	— (0.6)	316 (16.2)	51.7%	98.1%
Thalidomide	1584 (81.4)	107 (5.5)	18 (0.9)	238 (12.2)	70.5%	95.0%

*IQVIA capture rate was defined as the number of cases who had a transaction for a specific agent in IQVIA divided by the number of cases who had a transaction for a specific agent recorded in either IQVIA or Medicare. CML = chronic myeloid leukemia; OAM = oral anticancer medications; SEER = Surveillance, Epidemiology, and End Results Program; — = counts were less than 16.

†Medicare capture rate was defined as the number of cases who had a transaction for a specific agent in Medicare divided by the number of cases who had a transaction for a specific agent recorded in either IQVIA or Medicare.

Differences between the capture rate of Medicare and IQVIA were observed by registry and specific agent. The difference was most pronounced for the identification of CML and myeloma agents. OAMs for the treatment of breast cancer, CML, colon cancer, and myeloma have unique distribution systems. The breast cancer agents are in an open distribution system and are easily accessible at retail pharmacies. CML and colon cancer agents are also in an open distribution system, but their accessibility is restricted to specialty pharmacies. This is partially due to the restrictions placed on the coverage of these agents by insurance companies such as the requirement of prior authorization. Myeloma agents are provided through a closed distribution system and thus have more restrictive access. Interestingly, the cost of OAMs is higher as the distribution of the agent is restricted. For example, the cost of a 1-month supply of tamoxifen is $56.83 compared with $10 612.20 for

thalidomide (12). OAMs that are provided through specialty pharmacies or closed distribution systems may not be processed by IQVIA and thus not included in their data. This could explain why there were more patients with CML and myeloma who had claims for the specific drug in Medicare than in the IQVIA data. The IQVIA capture rate for CML and myeloma agents was considerably less in Louisiana than in Seattle and California. Contrastingly, for breast cancer agents, the IQVIA capture rate was consistent, ranging from 75.8% to 85.8%, for all registries.

Capecitabine was the only agent where the IQVIA capture rate was higher than the Medicare capture rate. Although the Medicare capture rate for capecitabine was 9.5% higher than the IQVIA capture rate in Louisiana, the Medicare capture rate for capecitabine was 5.3% and 25.6% lower than the IQVIA capture rate in Seattle and California, respectively. The reason for lower reporting of capacitance in the Medicare data needs further investigation.

Although the Medicare data had better capture rates for most OAMs, a limitation of the Medicare data is that it excludes most patients younger than the age of 65 years. The IQVIA data cover cases aged younger than 65 years and older than 65 years who do not have Medicare Part D coverage. Our findings show that IQVIA data include about 70.0% of the SEER population aged younger than 65 years. The large portion of the population included in the IQVIA data suggests that these data may be a potential resource to assess OAM use in the cancer population. Additionally, the comparison between IQVIA and BRAVO showed that IQVIA can considerably capture the receipt of HT among women aged 64 years or younger diagnosed with breast cancer. For the population younger than 65 years, the linkage between IQVIA and SEER identifies a potential source to enhance the capture of treatment information in SEER. However, it is important to consider that the 32.5% of patients not included in IQVIA data during the treatment assessment window may differ from those patients included in the IQVIA data. We found that that NH black, Hispanics, and uninsured patients were less likely to match to the IQVIA data. Additional research is needed to investigate if the use of OAMs in the 67.6% of the population that matched and were active in IQVIA during the treatment assessment window is similar to those not captured in the IQVIA data. Of note, reported prescription coverage of the IQVIA pharmacy data has increased since the time of the study analysis. Thus, the representativeness of IQVIA pharmacy data and IQVIA capture rates for OAMs may be greater than reported in this study.

This is the first study to our knowledge to assess 1) the completeness and representativeness of IQVIA pharmacy data compared with a population-based dataset and 2) the validity of IQVIA and Medicare Part D to identify OAMs. Previous studies evaluating the validity of Medicare claims to identify chemotherapy agents were limited to Medicare Part B reimbursable agents consisting predominately of intravenous agents ([13–17]). Additionally, this is the first study to our knowledge to develop a method to link SEER cases to IQVIA pharmacy data. Although this study is a one-time linkage, our analysis provides details about the process needed for future efforts to complete such a linkage.

There are several limitations of this study. The algorithm used to match SEER cases to IQVIA pharmacy data is proprietary. Due to the proprietary nature of the data linkage, there was no opportunity to adjudicate possible matches. To address this concern, a file of dummy cases was developed for a test linkage. We found that sex, last name, and last two digits of ZIP code were necessary in determining a match. However, the data linkage could not be assessed for linkage errors, such as false and missed matches. Assessment of the data linkage is important because linkage errors due to imperfect identifiers, such as use of nonunique identifiers, identifiers that are prone to errors or missing values, or are dynamic, may be a potential source of bias in results (18). Research to develop and identify appropriate approaches for data linkages should be conducted to enhance transparency and uniformity. A second limitation is that this study was conducted in only three geographic locations. The locations evaluated in the study were chosen because of the estimated high penetration of IQVIA. Therefore, the generalizability of the findings to other locations and population-based cancer registries may be limited, particularly for nonurban areas, because this study consisted of predominately metropolitan locations.

In conclusion, we evaluated the completeness and representativeness of IQVIA pharmaceutical data for cancer cases diagnosed in six SEER registries between 2007 and 2011 and for the receipt of specific OAMs as compared with available information from the SEER-Medicare and BRAVO data. Overall, the IQVIA pharmacy database included a large portion of persons in the SEER areas and a sizeable number of oral oncology agents. However, variation existed in the completeness of IQVIA data by patient characteristics, geographic location, and specific agents. The findings indicate specific subpopulations of patients and agents that were adequately captured in the IQVIA data and thus may serve as a resource for researchers wishing to study OAM utilization.

Notes

Affiliations of authors: Food and Drug Administration, Silver Spring, MD (QT); National Cancer Institute, Bethesda, MD (JLW, LE, LCH, DRR, DMW, LP, KAC); Information Mangement Services, Inc., Calverton, MD (MJB, DA); IQVIA, Rockledge, PA (MM); Cancer Registry of Greater California, Sacramento, CA (RDC); University of Southern California, Los Angeles, CA (DD); University of California San Francisco, Fremont, CA (SLG); Fred Hutchinson Cancer Research Center, Seattle, WA (SMS, CIL); Louisiana State University Health Sciences Center, New Orleans, LA (X-CW).

All authors declare no conflict of interest.

References

1. Food and Drug Administration. Hematology/oncology (cancer) approvals and safety notifications. http://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm279174.htm. Accessed May 20, 2016.
2. Avalere. New study finds that 10% of cancer patients abandoned oral anticancer drugs. http://www.avalerehealth.net/wm/show.php? c=&id=881. Published 2011. Accessed May 19, 2016.
3. Liu G, Franssen E, Fitch MI, Warner E. Patient preferences for oral versus intravenous palliative chemotherapy. *J Clin Oncol*. 1997;15(1):110–115.
4. Osterberg L, Blaschke T. Adherence to medication. *N Engl J Med*. 2005;353(5):487–497.
5. IQVIA. IQVIA – about us. https://www.iqvia.com/about-us. Accessed October 31, 2019.
6. National Cancer Institute. About the SEER Registries. https://seer.cancer.gov/registries/. Accessed October 31, 2019.
7. National Cancer Institute. SEER-Medicare: brief description of the SEER-Medicare Database. https://healthcaredelivery.cancer.gov/seermedicare/overview/. Accessed October 31, 2019.
8. Fred Hutch. Cancer Epidemiology Research Cooperative. https://www.fredhutch.org/en/research/divisions/public-health-sciences-division/research/epidemiology/cancer-epidemiology-research-cooperative.html. Accessed October 31, 2019.

9. Centers of Medicare and Medicaid Services. Office of Information Services: data from the 100 percent Denominator File. https://www.cms.gov/MedicareMedicaidStatSupp/06_2015.asp. Accessed June 5, 2016.
10. The Henry J. Kaiser Family Foundation. Medicare Part D at ten years: the 2015 marketplace and key trends. http://kff.org/medicare/report/medicare-part-d-at-ten-years-the-2015-marketplace-and-key-trends-2006-2015/. Published 2019. Accessed October 18, 2016.
11. The Henry J. Kaiser Family Foundation. Medicare advantage. http://kff.org/medicare/fact-sheet/medicare-advantage/. Published 2016. Accessed October 18, 2016.
12. Lexicomp Online®. Hudson, Ohio: Lexi-Comp, Inc. Updated December 3, 2017. http://online.lexi.com. Accessed December 12, 2017.
13. Lund JL, Stürmer T, Harlan LC, et al. Identifying specific chemotherapeutic agents in Medicare data: a validation study. *Med Care*. 2013;51:27–34.
14. Lamont EB, Lauderdale DS, Schilsky RL, Christakis NA. Construct validity of Medicare chemotherapy claims: the case of 5FU. *Med Care*. 2002;40(3):201–211.
15. Lamont EB, Herndon JE 2nd, 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.
16. Lamont EB, Lan L. Sensitivity of Medicare claims data for measuring use of standard multiagent chemotherapy regimens. *Med Care*. 2014;52:15–20.
17. Warren JL, Harlan LC, Fahey A, et al. Utility of the SEER-Medicare data to identify chemotherapy use. *Med Care*. 2002;40:55–61.
18. Harron KL, Doidge JC, Knight HE, et al. A guide to evaluating linkage quality for the analysis of linked data. *Int J Epidemiol*. 2017;46(5):1699–1710.