Pragmatic and Observational Research

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

A validation of clinical data captured from a novel Cancer Care Quality Program directly integrated with administrative claims data

This article was published in the following Dove Press journal:
Pragmatic and Observational Research
26 August 2017
Number of times this article has been viewed

David M Kern[1]
John J Barron[1]
Bingcao Wu[1]
Alex Ganetsky[2]
Vincent J Willey[1]
Ralph A Quimbo[1]
Michael J Fisch[3]
Joseph Singer[1]
Ann Nguyen[4]
Ronac Mamtani[5]

[1]Health Economics and Outcomes Research, HealthCore, Inc, Wilmington, DE, [2]Department of Pharmacy, Hospital of the University of Pennsylvania, Philadelphia, PA, [3]Medical Oncology, AIM Specialty Health, Chicago, IL, [4]Oncology Solutions, Anthem, Inc, Indianapolis, IN, [5]Division of Hematology/Oncology, Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA, USA

Background: Data from a Cancer Care Quality Program are directly integrated with administrative claims data to provide a level of clinical detail not available in claims-based studies, and referred to as the HealthCore Integrated Research Environment (HIRE)-Oncology data. This study evaluated the validity of the HIRE-Oncology data compared with medical records of breast, lung, and colorectal cancer patients.

Methods: Data elements included cancer type, stage, histology (lung only), and biomarkers. A sample of 300 breast, 200 lung, and 200 colorectal cancer patients within the HIRE-Oncology data were identified for medical record review. Statistical measures of validity (agreement, positive predictive value [PPV], negative predictive value [NPV], sensitivity, specificity) were used to compare clinical information between data sources, with medical record data considered the gold standard.

Results: All 300 breast cancer records reviewed were confirmed breast cancer, while 197 lung and 197 colorectal records were confirmed (PPV =0.99 for each). The agreement of disease stage was 85% for breast, 90% for lung, and 94% for colorectal cancer. The agreement of lung cancer histology (small cell vs non-small cell) was 97%. Agreement of progesterone receptor, estrogen receptor, and human epidermal growth factor receptor 2 status biomarkers in breast cancer was 92%, 97%, and 92%, respectively; epidermal growth factor receptor and anaplastic lymphoma kinase agreement in lung was 97% and 92%, respectively; and agreement of KRAS status in colorectal cancer was 95%. Measures of PPV, NPV, sensitivity, and specificity showed similarly strong evidence of validity.

Conclusion: Good agreement between the HIRE-Oncology data and medical records supports the validity of these data for research.

Keywords: validation, administrative claims, breast cancer, lung cancer, colorectal cancer, oncology

Introduction

The use of administrative claims data to perform observational health outcomes research has substantially increased over the past decade (Figure S1). Claims data offer researchers the ability to capture large amounts of data over geographically diverse populations for a fraction of the time and cost of a prospective study.[1] Because the primary use of administrative claims data is for billing and reimbursements, the validity of diagnostic codes within claims data for the use of research has been studied extensively.[2–4] Researchers now have access to a number of validated claims-based algorithms to identify a wide range of disease states.[5–7]

However, one of the largest remaining limitations of using administrative claims data for research is the inability to capture detailed clinical information, which is particularly

Correspondence: David M Kern
Health Economics and Outcomes Research, HealthCore, Inc, 123 Justison Street, 2nd floor, Wilmington, DE 19148, USA
Tel +1 302 230 2102
Email dkern@healthcore.com

submit your manuscript | www.dovepress.com
Dovepress
http://dx.doi.org/10.2147/POR.S140579

important in cancer outcomes research. For example, information on cancer stage (eg, local or metastatic), histology (eg, adenocarcinoma or squamous cell carcinoma), and biomarkers (eg, hormone receptor status in breast cancer) is not routinely available in claims-based datasets, yet is among the most important factors for influencing treatment decisions and patient prognosis. Given improvements in cancer survival over the past few decades,[8,9] there is increasing importance in the study of cancer treatment effectiveness and outcomes. For high-quality oncology outcomes research to be performed, there is need for additional data sources to supplement claims data in order to provide researchers with a complete clinical profile of oncology patients.[10]

The Cancer Care Quality Program (CCQP), a novel program by Anthem, Inc health plans, is designed to align reimbursement with evidence-based, cost-effective oncology treatment.[11] A major element of the CCQP are the cancer treatment pathways ("pathways"), which are developed using evidence-based medicine. The objective of a pathway for a specific tumor type is to identify a subset of regimens supported by clinical evidence and practice guidelines with the goal of creating more consistent care and reducing variation in cost. Pathways are selected according to clinical benefit, safety/side effects, strength of national guideline recommendations, and cost of regimens.

The clinical data obtained from the CCQP were integrated with administrative claims data to provide a level of clinical detail not typically available in claims-based studies. Prior to using this new data source for cancer outcomes research, it is important to examine the quality of the data. This study examines the validity of the CCQP data relative to information abstracted from the medical records of breast, lung, and colorectal cancer patients.

Materials and methods

HealthCore Integrated Research Environment (HIRE)-Oncology data

Data from the CCQP were integrated with HIRE, and are referred to as the HIRE-Oncology data. The CCQP offers evidence-based cancer treatment information enabling physicians to compare planned cancer treatment regimens against evidence-based clinical criteria.[11] The CCQP has identified certain cancer treatment pathways, based on current clinical evidence, published literature, and national guideline recommendations, which have been shown to be efficacious, less toxic, and cost effective. The physicians participating in the CCQP receive additional reimbursement per patient for treatment planning and care coordination when prescribed treatment regimens align with the identified pathway, encouraging evidence-based quality care for the patients and value-based benefits for the physicians. Data are obtained when physicians request approval for this pathway-based enhanced reimbursement as well as prior authorization for the various cancer treatments. The clinical information is typically collated by nonclinical staff at the oncologists' office and entered either directly into the electronic system via a web portal by office staff or indirectly via a telephone conversation with health plan personnel. As of September 2015, the program was implemented in all 14 states where Anthem has commercial health plans.

Patient identification and data elements

Patients included in this study had commercial health plan coverage from Anthem at the time of their HIRE-Oncology record of interest and could be linked to HealthCore's administrative claims database. HIRE-Oncology data from June 23, 2014 through June 1, 2015 for patients with breast, lung, or colorectal cancer were used for this study. The data elements obtained for this study included cancer type, cancer stage (0, I, II, III, IV, or limited), biomarkers unique to each cancer type (breast: estrogen receptor [ER], progesterone receptor [PR], and human epidermal growth factor receptor 2 [HER2]; lung: epidermal growth factor receptor [EGFR] mutation, anaplastic lymphoma kinase [ALK] mutation; colorectal: KRAS gene mutation), and the histology of lung cancer (small cell vs non-small cell). Additionally, age and gender were captured from the HIRE-Oncology data.

Medical record abstraction

Medical records for a sample of 300 breast, 200 lung, and 200 colorectal cancer patients identified from the HIRE-Oncology data were collected. Patients in the HIRE-Oncology data with available histology and biomarker information received a higher priority for sampling in order to maximize the sample size within each endpoint. For each record, the oncologist's office identified on the date of the request was targeted for medical record collection. Medical records were obtained from the physician offices by a third party vendor and then transferred to HealthCore on a weekly basis. A trained redactor then reviewed each page of the record and using industry standard redaction software blacked out any standard Health Insurance Portability and Accountability Act (HIPAA) protected health information (PHI) of the patient and facility. Redacted records were then scanned, saved as .pdf files and underwent a thorough quality check to ensure no PHI remained visible.

The redacted medical records were transferred via a secure file transfer protocol to a blinded board certified oncology pharmacist and a registered nurse for abstraction of relevant information. The abstractors received a medical record abstraction form which asked the abstractor to identify the gender of the patient, the cancer type (breast, lung, colorectal, other, or unknown), stage of disease (0, I, II, III, IV, limited, unknown), biomarker status for each biomarker of interest (positive or mutation, negative or wild-type, conflicting results, equivocal [for HER2 status only], or unknown/test was not performed), menopausal status for those identified as having breast cancer, and disease histology for those with lung cancer. Information from an individual medical record was abstracted by exactly one abstractor. Completed abstraction forms were then sent back to the HealthCore research team. A copy of the blank medical record abstraction form is included in the Supplementary materials.

All study materials were handled in compliance with the HIPAA, and a limited dataset was used for all analyses, as defined by the Privacy Rule. The New England Institutional Review Board approved the protocol as well as granted an HIPAA waiver of authorization to obtain the medical records.

Statistical analysis

Data obtained from the oncologists' medical records were considered the gold standard in this analysis. Appropriate measures of validity were calculated according to the outcome being measured. The positive predictive value (PPV) and agreement were calculated for every outcome. For all variables other than cancer type and stage of disease, negative predictive value (NPV), sensitivity, and specificity were also calculated.

Agreement was measured as the proportion of patients for whom the value of a given outcome according to the medical record was the same as (ie, in agreement with) the value according to HIRE-Oncology. Sensitivity was defined as the proportion of patients who were identified as having a positive result in the medical records and also had a positive indication in the HIRE-Oncology data. Specificity was calculated as the proportion of patients who had a negative result in the medical records and also had a negative result in HIRE-Oncology. PPV was calculated as the proportion of positive results identified from HIRE-Oncology that were confirmed as positives from medical records. And, lastly, NPV was defined as the proportion of individuals with negative results in HIRE-Oncology who had a confirmed negative result in the medical record.

Observations without available data for a given data point were excluded from the analysis for which data were missing. Point estimates and 95% Clopper–Pearson (exact) confidence intervals were reported for each measure. All analyses were performed using SAS Enterprise Guide 7.1 (SAS Institute Inc, Cary, NC, USA).

Results

Overall agreement of cancer type and disease stage

The mean ages of patients with breast, lung, and colorectal cancer were 53, 60, and 57 years, respectively (Table 1). Females represented more than half of those with lung (53%) and colorectal (56%) cancer, and nearly all of the breast cancer patients (99%). All 300 breast cancer records reviewed were confirmed as breast cancer by the medical records (PPV =1.00), while 197 of the 200 lung cancer records and 197 of the 200 colorectal cancer records were confirmed (PPV =0.99 for each; 95% CI =[0.97–1.00]). The agreement of disease stage (the proportion of records for which the stage of disease between medical records and the HIRE-Oncology data matched) was 0.85 (0.81–0.89)

Table 1 Characteristics and key validation statistics for breast, lung, and colorectal cancer patients

	Cancer type		
	Breast	**Lung**	**Colorectal**
Cases identified in HIRE-Oncology, n	300	200	200
Cases confirmed from medical records, n	300	197	197
Age, mean (SD), years	53.0 (8.9)	59.9 (8.2)	57.3 (9.4)
Female, %	99.3%	47.2%	44.2%
Stage (according to HIRE-Oncology)			
0	1	1	2
I	51	1	1
II	107	4	5
III	43	21	22
IV	98	167	167
Limited	0	3	0
Data available in both HIRE-Oncology and medical records, n			
Stage	284	194	195
Biomarker 1	ER: 278	ALK: 74	KRAS: 114
Biomarker 2	PR: 261	EGFR: 69	
Biomarker 3	HER2: 282		
Histology		174	
Key validation statistics			
PPV of cancer type (95% CI)	1.00 (1.00–1.00)	0.99 (0.97–1.00)	0.99 (0.97–1.00)
Agreement of disease stage (95% CI)	0.85 (0.81–0.89)	0.90 (0.86–0.94)	0.94 (0.90–0.97)

Abbreviations: HIRE, HealthCore Integrated Research Environment; PPV, positive predictive value; CI, confidence interval; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; ALK, anaplastic lymphoma kinase; EGFR, epidermal growth factor receptor; SD, standard deviation.

for breast, 0.90 (0.86–0.94) for lung, and 0.94 (0.90–0.97) for colorectal cancer. The PPV of stage III breast cancer (0.61) was lower than that of stage I (0.94), stage II (0.90), or stage IV (0.86). The PPV of stage IV lung cancer (0.92) and colorectal cancer (0.97) was higher than other stages, though all stages had a PPV ≥0.80.

Breast cancer biomarkers

The overall agreements of PR, ER, and HER2 statuses in breast cancer were 0.92, 0.97, and 0.92, respectively. The sensitivity, specificity, PPV, and NPV of ER status (n=278 with non-missing data) were all ≥0.94, indicating very good validity. For example, the PPV of ER data in the breast cancer cohort (PPV =0.98) indicates that 98% of cases identified in HIRE-Oncology as being ER+ were confirmed ER+ in the medical records. For PR status (n=261 with non-missing data), the sensitivity (0.96), specificity (0.88), PPV (0.91), and NPV (0.95) were all strong. Lastly for breast cancer, the HER2 status (n=282) showed validity measures similar to that of the other biomarkers, with a high sensitivity (0.93), specificity (0.94), PPV (0.94), and NPV (0.96).

Lung cancer biomarkers

Agreement across each of the specific lung cancer measures was high: 0.92 for ALK status, 0.97 for EGFR status, and 0.97 for disease histology (small cell vs non-small cell). The ALK status (n=74) showed very strong sensitivity (1.00), specificity (0.91), and NPV (1.00), but a much lower PPV (0.54). Validity measures of EGFR (n=69) were high according to the sensitivity (1.00), specificity (0.96), PPV (0.87), and NPV (1.00). Lastly, the histology data also showed strong sensitivity (0.85), specificity (0.99), PPV (0.92), and NPV (0.97).

Colorectal cancer biomarkers

The agreement of KRAS status (n=114) between the two data sources for colorectal cancer patients was 0.95, and had similar levels of sensitivity (0.93), specificity (0.96), PPV (0.93), and NPV (0.96). The sensitivity of the KRAS mutation data indicates that HIRE-Oncology identified 93% of all patients who had a KRAS mutation according to the medical records.

The complete set of results, including the confidence limits of each point estimate, can be found in Table 2. The raw data showing the cross tabulation of values obtained from the

Table 2 Complete validation results within confirmed breast (n=300), lung (n=197), and colorectal cancer (n=197) patients

Breast cancer (n=300)				Lung cancer (n=197)				Colorectal cancer (n=197)			
	Estimate	95% CI			Estimate	95% CI			Estimate	95% CI	
		Lower	Upper			Lower	Upper			Lower	Upper
Staging (n=284)				Staging (n=194)				Staging (n=195)			
Agreement	0.849	0.807	0.890	Agreement	0.902	0.860	0.944	Agreement	0.939	0.905	0.972
PPV – stage I	0.938	0.869	1.000	PPV – stage III	0.810	0.642	0.978	PPV – stage II	0.800	0.449	1.000
PPV – stage II	0.900	0.841	0.959	PPV – stage IV	0.921	0.880	0.962	PPV – stage III	0.864	0.720	1.000
PPV – stage III	0.610	0.460	0.759					PPV – stage IV	0.970	0.944	0.996
PPV – stage IV	0.862	0.792	0.932								
Estrogen receptor (n=278)				ALK mutation (n=74)				KRAS mutation (n=114)			
Agreement	0.971	0.952	0.991	Agreement	0.919	0.857	0.981	Agreement	0.947	0.906	0.988
Sensitivity	0.975	0.953	0.997	Sensitivity	1.000	1.000	1.000	Sensitivity	0.925	0.843	1.000
Specificity	0.963	0.921	1.000	Specificity	0.910	0.842	0.979	Specificity	0.960	0.915	1.000
PPV	0.985	0.968	1.000	PPV	0.539	0.268	0.810	PPV	0.925	0.843	1.000
NPV	0.939	0.887	0.991	NPV	1.000	1.000	1.000	NPV	0.960	0.915	1.000
Progesterone receptor (n=261)				EGFR mutation (n=69)							
Agreement	0.923	0.891	0.956	Agreement	0.971	0.931	1.000				
Sensitivity	0.958	0.925	0.991	Sensitivity	1.000	1.000	1.000				
Specificity	0.881	0.823	0.940	Specificity	0.964	0.916	1.000				
PPV	0.907	0.861	0.954	PPV	0.867	0.695	1.000				
NPV	0.946	0.903	0.988	NPV	1.000	1.000	1.000				
HER2 (n=282)				Histology, small cell or non-small cell (n=174)							
Agreement	0.922	0.891	0.953	Agreement	0.966	0.938	0.993				
Sensitivity	0.928	0.876	0.979	Sensitivity	0.846	0.708	0.985				
Specificity	0.939	0.904	0.974	Specificity	0.987	0.968	1.000				
PPV	0.938	0.889	0.986	PPV	0.917	0.806	1.000				
NPV	0.961	0.932	0.989	NPV	0.973	0.948	0.999				

Abbreviations: PPV, positive predictive value; NPV, negative predictive value; CI, confidence interval; HER2, human epidermal growth factor receptor 2; ALK, anaplastic lymphoma kinase; EGFR, epidermal growth factor receptor.

medical records versus HIRE-Oncology for each endpoint of interest can be found in the Supplementary materials.

Discussion

This study examined the validity of cancer stage, histology, and biomarker data among patients with breast, lung, and colorectal cancers using a novel database – HIRE-Oncology – integrating provider reported clinical data with administrative claims. Results of this study found that relative to the gold standard medical record review, the data in the electronic record achieve a high measure of validity. We report agreement, sensitivity, specificity, NPV, and PPV measures generally >90% for cancer stage, histology, and biomarkers for three common cancers, suggesting that these data may be used for observational oncology research.

The use of administrative claims data has been a major advancement in cancer outcomes and health services research.[12] While claims data provide large cohorts of patients with diagnosed cancer, they lack crucial predictors of cancer outcomes such as cancer stage and biomarker status. For example, patients with late-stage (ie, metastatic) disease have worse outcome than patients with early-stage (ie, localized) disease. Additionally, patients who are biomarker positive (eg, hormone receptor in breast cancer or EGFR in lung cancer) have better prognosis and response to targeted therapy versus cytotoxic chemotherapy.[13] Failure to account for each of these variables may confound the relationship between cancer treatment and outcome, leading to biased results. Further, the uncertain validity of codes for metastatic cancer in claims data remains a major limitation,[14,15] and manual medical record review may not be feasible. The necessity of having reliable and timely data for cancer outcomes research is becoming more important as the oncology treatment space rapidly changes; there are as many as 836 new medications and vaccines for cancer in various stages of development, with 80% of them being potential first in class therapies.[16] Thus, the ability to link clinical with claims data to accurately classify cancer stage and biomarkers is a unique and important strength of these data and central to the conduct of high-quality oncology research. The capacity to integrate various data highlights the importance of fully identifiable research databases such as the HIRE, which can be linked not only with the data specific to this study, but can be also integrated with any other identifiable data sources which have been approved for research purposes and where appropriate permissions have been obtained.

Since the CCQP utilizes clinical data transmitted by oncology practices to the health plan for the purposes of pathway-based enhanced reimbursement or prior authorization activities, we felt it important to assess the validity of these data due to a number of factors. First, these data are typically transmitted to the health plan by nonclinical office staff. Although electronic health records (EHRs) may make it easier to collect all the necessary clinical information, there is the potential for misinterpretation or simple error with the transcription of the information. In addition, prior research has demonstrated that physicians may use deception to obtain approvals from health plan payers.[17,18] The high measure of validity we found between both data sources helps to reassure that these concerns did not adversely impact the HIRE-Oncology data.

These data may prove to be an improvement over the use of other data sources, which have previously been shown to be effective tools in supplementing claims-based research but have significant limitations, namely cancer registries[19] and EHRs.[20] Inherent limitations of registries are that not all researchers have access to the registry data, the data in the registry may not be complete, there may be a lag in the data due to annual updating, and/or the registry may be limited to a specific subset of patients being studied, thereby limiting the generalizability of the study.[21–23] The integration of EHR data with administrative claims has recently become popular in outcomes research, but there are limitations to the relatively new data source, such as a high variation in the validity of data across different clinical variables,[24] variation across different EHR systems,[25] and the current lack of a uniform data quality assessment.[26,27] Thus, the HIRE-Oncology data may be a valuable alternative to these data sources.

Strengths of this study included stratified random sampling from all subjects within HIRE-Oncology and rigorous validation of cancer data against standard medical record review; however, there are several potential limitations of this study. As with most validation studies, we cannot exclude the possibility of misclassification bias. For example, the PPV for recorded ALK mutation among lung cancer patients was only 0.54. This result must be interpreted with caution given the relatively wide confidence intervals and the relatively low prevalence of positive ALK mutations in the cohort. Importantly, all other validity measures (agreement, sensitivity, specificity, and NPV) of ALK status were >90% as were all validity measures for the majority of other variables from HIRE-Oncology; thus, the impact of any misclassification bias on the results of our study was likely low. It is also noteworthy that if ALK status (among other biomarkers) is not necessary for the treatment requested in the CCQP then the field is not required to be completed, hence the

relatively low numbers identified in this validation. More data are needed before a definitive conclusion can be made regarding the validity of a positive ALK status. We examined histology for lung cancer data only, as we needed to ensure we could differentiate small cell from non-small cell lung cancer, as future research will require this for evaluating various treatments. In colorectal cancer, the vast majority of cases (>95%) were adenocarcinomas,[28] so we did not feel that it was necessary to validate a measure with such low variation. Although breast cancer histology is more varied, we did not examine this as part of our validation as histology type is not typically a factor in the selection of systemic therapy. We did not compare the HIRE-Oncology recorded stage to tumor registry stage. However, tumor registry data are limited to the patient's stage at initial diagnosis and do not contain information for patients who develop metastatic disease over time. We were also unable to calculate overall sensitivity for any breast, colon, or lung cancer diagnosis, as the data sampled included only a subset of these cancer patients enumerated within HIRE-Oncology. Likewise, as this research was specific to common cancers in patients from a single health care system, these data may not be fully representative of the broader US population.

While medical records were considered the gold standard in this validation study, the information in medical records may be missing or incomplete. For example, HER2 status was present for all 300 breast cancer patients in HIRE-Oncology, but was missing in the medical records from 18 patients. Furthermore, targeting records with available biomarker results may have led to oversampling of stage IV metastatic disease (eg, presence of EGFR/ALK mutation in lung cancer); however, the majority of lung and colorectal cancer cases in the overall HIRE-Oncology data are stage IV (77% and 75%, respectively), and thus the results are likely representative of the data as a whole.

Conclusion

The study findings suggest that the clinical data entered by participating oncology practices as part of the CCQP are accurate relative to medical records. The good agreement between the HIRE-Oncology data and the gold standard of medical records supports the validity of these data, and suggests the potential to increase efficiency and reduce costs associated with future observational research. These data can enhance claims-based studies by providing real-world clinical data directly integrated with health care utilization data that may not otherwise be available to researchers on a national level.

Disclosure

Authors DMK, JJB, VJW, RAQ, and JS are employees of HealthCore, Inc, a subsidiary of Anthem, Inc. BW was an employee of HealthCore, Inc at the time of the study. AN is an employee of Anthem, Inc. MJF is an employee of AIM Specialty Health. RM was supported by the National Institutes of Health (NIH) / National Cancer Institute (NIC) grant K23CA187185. The authors report no other conflicts of interest in this work.

References

1. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol.* 2005;58(4):323–337.
2. van Walraven C, Bennett C, Forster AJ. Administrative database research infrequently used validated diagnostic or procedural codes. *J Clin Epidemiol.* 2011;64(10):1054–1059.
3. Johnson EK, Nelson CP. Utility and pitfalls in the use of administrative databases for outcomes assessment. *J Urol.* 2013;190(1):17–18.
4. Quan H, Li B, Saunders LD, et al. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res.* 2008;43(4):1424–1441.
5. Khokhar B, Jette N, Metcalfe A, et al. Systematic review of validated case definitions for diabetes in ICD-9-coded and ICD-10-coded data in adult populations. *BMJ Open.* 2016;6(8):e009952.
6. Youngson E, Welsh RC, Kaul P, McAlister F, Quan H, Bakal J. Defining and validating comorbidities and procedures in ICD-10 health data in ST-elevation myocardial infarction patients. *Medicine (Baltimore).* 2016;95(32):e4554.
7. Whyte JL, Engel-Nitz NM, Teitelbaum A, Gomez Rey G, Kallich JD. An evaluation of algorithms for identifying metastatic breast, lung, or colorectal cancer in administrative claims data. *Med Care.* 2015;53(7):e49–e57.
8. Howlader N, Noone A, Krapcho M, et al. SEER cancer statistics review, 1975–2013. Based on November 2015 SEER data submission. Bethesda, MD: National Cancer Institute; 2016. Available from: http://seer.cancer.gov/csr/1975_2013/. Accessed April 26, 2017.
9. American Cancer Society. Managing cancer as a chronic illness. 2016. Available from: http://www.cancer.org/treatment/survivorshipduringandaftertreatment/when-cancer-doesnt-go-away. Accessed October 20, 2016.
10. Meyer A-M, Carpenter WR, Abernethy AP, Stürmer T, Kosorok MR. Data for cancer comparative effectiveness research. *Cancer.* 2012;118(21):5186–5197.
11. Anthem. Cancer care quality program treatment pathways. 2017. Available from: https://anthem.aimoncology.com/pdf/pathways/Cancer_Pathways_Clinical_Detail.pdf. Accessed June 16, 2017.
12. Meyer A-M, Basch E. Big data infrastructure for cancer outcomes research: implications for the practicing oncologist. *J Oncol Pract.* 2015;11(3):207–208.
13. Hsueh C-T, Liu D, Wang H. Novel biomarkers for diagnosis, prognosis, targeted therapy and clinical trials. *Biomark Res.* 2013;1(1):1.
14. Chawla N, Yabroff KR, Mariotto A, McNeel TS, Schrag D, Warren JL. Limited validity of diagnosis codes in Medicare claims for identifying cancer metastases and inferring stage. *Ann Epidemiol.* 2014;24(9):666–672.
15. Nordstrom BL, Simeone JC, Malley KG, et al. Validation of claims algorithms for progression to metastatic cancer in patients with breast, non-small cell lung, and colorectal cancer. *Front Oncol.* 2016;6:18.
16. Buffery D. Innovation tops current trends in the 2016 oncology drug pipeline. *Am Health Drug Benefits.* 2016;9(4):233–238.
17. Freeman VG, Rathore SS, Weinfurt KP, Schulman KA, Sulmasy DP. Lying for patients: physician deception of third-party payers. *Arch Intern Med.* 1999;159(19):2263–2270.

18. Novack DH, Detering BJ, Arnold R, Forrow L, Ladinsky M, Pezzullo JC. Physicians' attitudes toward using deception to resolve difficult ethical problems. *JAMA.* 1989;261(20):2980–2985.
19. Kurian AW, Lichtensztajn DY, Keegan THM, et al. Patterns and predictors of breast cancer chemotherapy use in Kaiser Permanente Northern California, 2004–2007. *Breast Cancer Res Treat.* 2013;137(1):247–260.
20. Bayley KB, Belnap T, Savitz L, Masica AL, Shah N, Fleming NS. Challenges in using electronic health record data for CER: experience of 4 learning organizations and solutions applied. *Med Care.* 2013;51:S80–S86.
21. Mallin K, Palis BE, Watroba N, et al. Completeness of American Cancer Registry treatment data: implications for quality of care research. *J Am Coll Surg.* 2013;216(3):428–437.
22. Cress RD, Zaslavsky AM, West DW, Wolf RE, Felter MC, Ayanian JZ. Completeness of information on adjuvant therapies for colorectal cancer in population-based cancer registries. *Med Care.* 2003;41(9):1006–1012.
23. Du XL, Key CR, Dickie L, et al. Information on chemotherapy and hormone therapy from tumor registry had moderate agreement with chart reviews. *J Clin Epidemiol.* 2006;59(1):53–60.
24. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. *BMJ.* 2003;326(7398):1070.
25. Chan KS, Fowles JB, Weiner JP. Review: electronic health records and the reliability and validity of quality measures: a review of the literature. *Med Care Res Rev.* 2010;67(5):503–527.
26. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc.* 2013;20(1):144–151.
27. Kahn MG, Callahan TJ, Barnard J, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *EGEMS (Wash DC).* 2016;4(1):1244.
28. Stewart SL, Wike JM, Kato I, Lewis DR, Michaud F. A population-based study of colorectal cancer histology in the United States, 1998–2001. *Cancer.* 2006;107(5 Suppl):1128–1141.