

Improving patient data quality by integrating oncology practice and state cancer registry tumor staging information: Feasibility and future value

Gregory Hess^{1,2,*} and Sally T. Haislip³

¹ University of Pennsylvania, USA

² Symphony Health Solutions (SHS) Horsham, PA, USA

³ Georgia Cancer Specialists, Affiliated with Northside Hospital, Atlanta, Georgia, USA

Abstract

Background: The transition in oncology to electronic charting offers the potential to improve the quality of patient care and value of observational research. Data fields that are more complete, have common standards, and are searchable are critical to help meet these goals. As a key data field, and proof-of-concept we studied the additional gain in recorded stage and agreement in cancer staging by adding 'missing' stage information into an oncology practice's electronic medical records (EMR) from a state cancer registry. **Methods:** In this observational study, patient records were matched and compared between a practice-based (EMR) database (Georgia Cancer Specialists [GCS]) and a state cancer registry (Georgia Comprehensive Cancer Registry [GCCR]). Impact on recorded cancer stage following a merge of the EMR and registry data was assessed. Eligible patients had ≥ 1 visit to any GCS practice site during the study period (1/1/2005-12/31/2008) and had a diagnosis of a primary, malignant solid neoplasm (except brain or spine). **Results:** The final sample included 38,248 patients from GCS files, with 13,486 matched to patients with a solid malignant tumor in the GCCR files. There were 3,424 (25%) patients without staging information prior to GCCR integration, which was reduced to 12% after GCCR integration - a relative gain of 52%. Differences between initial GCS stage and initial GCCR stage occurred in 45% of the sample, and varied by cancer type. **Conclusions:** Adding information from external data sources can help create more complete patient records. The concept is feasible and has the potential to improve data quality. Patient data collected in different systems for different reasons will often be discordant.

Keywords: integrating oncology practice; cancer registry; tumor staging information

Introduction

Cancer staging describes the extent or severity of an individual's cancer based on the extent of the primary tumor and the degree of spread of cancer to local or distant body sites [1]. Staging is important in planning treatment strategies, evaluating outcomes, estimating the patient's prognosis, and identifying clinical trials that might be appropriate for the patient [1-2]. Staging also provides a basis for exchanging information about patients and evaluating and comparing results of clinical trials. These benefits have motivated the development of international cancer staging standards such as the TNM (Tumor Nodes Metastases) standard of the American Joint Committee on Cancer (AJCC) and the International Union Against Cancer (UICC) [3]. Cancer staging is critical to optimizing outcomes in cancer treatment and research.

The transition in oncology to electronic charting with searchable electronic medical records (EMRs) that document cancer stage, offers the potential to improve the quality and value of cancer care data in research and patient care. By providing easy access to comprehensive patient records, the EMR can facilitate integration of clinical information and thereby improve healthcare decision-making as well as

evaluation of ongoing treatment response. Although EMR data have potential in this regard, they can be incomplete, unstructured, and inaccurate-characteristics that limit their usefulness in evaluating quality of care and conducting research [4]. Data are often poorly accessible because they are not entered into the appropriate structured EMR field and time-consuming, labor-intensive chart review of physician notes or pathology reports are necessary to find the data [5]. Although larger facilities may have designated tumor registrars who manage the data reporting process,

***Corresponding author:** Gregory Hess, MD, MBA, MSc., Sr. Fellow LDI, Health Economics & Policy, University of Pennsylvania, Chief Medical Officer & EVP, Symphony Health Solutions (SHS), Horsham, PA 19044, USA. Tel.: 610-574-7250; Fax: 215-444-8832; Email: Greg.Hess@Wharton.Upenn.edu

Received 20 August 2015 Revised 14 October 2015 Accepted 23 October 2015
Published 31 October 2015

Citation: Hess G, Haislip ST. Improving patient data quality by integrating oncology practice and state cancer registry tumor staging information: Feasibility and future value. J Cancer Res Ther. 2015; 3(10):118-123. DOI:10.14312/2052-4994.2015-17

Copyright: © 2015 Hess G, et al. Published by NobleResearch Publishers. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

many, especially smaller facilities or clinics, do not and this lack of a dedicated role likely slows data abstraction and reporting. EMRs often document stage at the moment of creating the record rather than stage at diagnosis; In oncology, staging is appropriately performed only once, at the time of diagnosis and prior to treatment.

Groups such as the National Cancer Board, the Institute of Medicine, and the American Society of Clinical Oncology have called for various initiatives to explore the feasibility and value of linking patient data from multiple sites and sources. A recent commentary [6] discussed the value in linking registry data with provider EMR data, but recognized the challenges associated with these efforts in the absence of standardized clinical narrative/text in EMRs, and lack of definitions for process and outcomes measures for integration of data. Although there are challenges, merging cancer staging information from an EMR database with information from other sources could enhance the usefulness of the data for clinical practice and clinical research by improving upon the comprehensiveness and accuracy of staging information relative to that in the EMR database alone [7]. For example, in the absence of a recorded stage in an EMR, other information sources could provide the best available information to guide a treatment plan, estimate prognosis, and/or inform public health research. Using Health Insurance Portability and Accountability Act and Health Information Technology for Economic and Clinical Health Act (HIPAA/HITECH)-compliant matching methods, combined with 'big data' from broad geographic areas, increased computing power, and advanced software, a number of initiatives have successfully demonstrated the ability to integrate disparate, patient-level data [8, 9]. The impact of merging information from multiple data sources, including EMRs, on the quantity and quality of cancer staging information has not been previously assessed. Likewise, while EMR data omissions vis-à-vis cancer staging are known to exist, their magnitude and nature have not been systematically studied. In the study described herein, records containing cancer stage data were matched and compared between a practice-based EMR database and a state cancer registry; In addition to demonstrating the feasibility of such linkage and serving as a further proof-of-concept to complement other studies and projects, the impact on data accuracy and comprehensiveness of merging the EMR data and the registry data was assessed.

Methods

This observational study was conducted to obtain descriptive data on the impact of merging the Georgia Cancer Specialists (GCS) EMR data and the Georgia Comprehensive Cancer Registry (GCCR) data on the overall completeness and accuracy of medical records. Specifically, GCS solid tumor neoplasms with missing stage from structured EMR template fields were targeted for improved staging. As a secondary objective, the degree of agreement between the GCS cancer stage data and the GCCR cancer stage data was assessed. The GCS and GCCR matched patient sample was assessed for possible discordance of stage values.

Data sources

GCS [10] was the source of the EMRs in this study. GCS is one of the largest oncology/hematology practices in the Southeastern United States and offers community-based medical oncology and hematology services as well as support services. GCS has approximately 30 offices in Georgia, 42 oncologists and >400 support staff providing care through approximately 190,000 annual patient visits (including 16,000 new patient visits per year). As with many practices, GCS stages cancer predominantly on a clinical basis utilizing the American Joint Committee on Cancer (AJCC) TNM system: (T) the extent of the tumor, (N) the extent of spread to regional lymph nodes, and (M) the presence of metastasis (M). Newly diagnosed cancer cases are reported to the GCCR State registry by area hospitals conducting the resections or biopsies for GCS patients.

GCCR [11] was the source of registry data for this study. GCCR is a statewide population-based cancer registry mandated to receive all cancer cases diagnosed among Georgia residents since January 1, 1995. Goals of the GCCR are to collect information on all newly diagnosed cancer cases; to calculate cancer incidence rates for Georgia; to make data available to the public and healthcare professionals; to identify and evaluate cancer morbidity and mortality trends and problems on an ongoing basis; to provide cancer incidence and mortality data to cancer control programs to assist them in developing strategies and evaluating their effectiveness; and to stimulate cancer control research. All healthcare providers in the state of Georgia are required to report specific information on newly diagnosed cancer in their patient population to the GCCR. This requirement applies to all facilities that provide diagnostic evaluations and/or treatment for cancer patients including but not limited to hospitals, outpatient surgical facilities, laboratories, radiation therapy facilities, medical oncology facilities, and physicians and physician's offices. In addition, the GCCR maintains reporting agreements with neighboring states so that Georgia residents who are diagnosed or treated in facilities out of state can be identified. The GCCR stages cancer by combining clinical and pathological stage for a "derived" stage.

Sample

Eligible patients had at least 1 visit to any of the GCS practice sites during the study period, which extended from January 1, 2005 to December 31, 2008 and had a diagnosis of a primary, malignant solid neoplasm with a respective International Classification of Diseases, 9th edition (ICD-9) code [12]. Patients with brain or spine as the primary neoplasm were excluded.

The GCS sample included 38,248 patients treated between January 1, 2005 and December 31, 2008. Patient's gender and the age reported at the time of their first visit during the study period were recorded for the study.

The GCCR sample comprised patients diagnosed with a primary, malignant solid neoplasm from January 1, 2004, through December 31, 2008, and matched by demographics (name, age, gender) and cancer cell organ type to patients in the GCS records. The index date for the

GCCR sample was one year earlier than that for GCS based on the observation that the majority of newly diagnosed patients would have a practitioner visit and/or treatment within one year of diagnosis.

Procedures

Patients in the GCCR file were compared to those in the GCS file for a match using probabilistic matching on a series of up to 7 fields: names, including last, first, middle initial, suffix, maiden name (if applicable), date of birth, and social security number.

Cancer registries record a patient's disease based on the International Classification of Diseases for Oncology (ICD-O) codes (topography), rather than on the basis of ICD-9 codes which are recorded in clinical practices such as GCS. Subsequently, ICD-9 codes from GCS were converted

to the corresponding Surveillance, Epidemiology, and End Results (SEER) Site Recode values. The site recode variables are based on the ICD-O-3 primary site and histology data fields that SEER makes available to facilitate research and registry operations. Once the 5-digit site recodes were obtained, sub-aggregate recode values (i.e., the first 4 digits of the recode value) were used to group tumors. Records were excluded if the GCCR primary tumor site did not match the GCS primary tumor site. For those patients and tumors that matched in each file, the first recorded stage (i.e., from the earliest reporting source) was identified in each file. The distribution of recorded stage (Table 1), including stage X, and those with unknown stage was analyzed in each file. For the patients with cancer data in both GCCR and GCS files, agreement and discordance of the first recorded stage was compared by date.

Table 1 Comparison matrix for differing GCS and GCCR initial TMN Stages reported for all primary tumors*.

Patient volume	GCCR initial stage								
	Total	Stage 0	Stage I	Stage II	Stage III	Stage IV	Stage X	Occult	No stage data
Total	6100	388	1225	1187	1126	1030	1021	71	85
Stage 0	24		16	6	-	-	2	-	-
Stage I	465	65		162	33	21	165	7	12
Stage II	732	31	255		192	47	187	15	11
Stage III	785	4	124	239		208	184	15	11
Stage IV	862	12	124	167	311		187	22	40
Stage X	466	80	105	129	104	40		1	7
Limited	64	-	8	2	38	9	6	1	-
Extensive	55	-	3	-	9	42	1	-	-
Occult	-	-	-	-	-	-	-		-
No stage data	2671	197	592	482	439	663	290	20	

Abbreviations: GCS = Georgia Cancer Specialists; GCCR = Georgia Comprehensive Cancer Registry; TNM = Tumor Nodes Metastases; * = Patients with multiple primary cancers contribute data to multiple cells in the table.

Results

Sample

The number of unique patients represented in the GCS files and treated between January 1, 2005, and December 31, 2008, was 38,248. Of those patients, 13,486 were matched to patients with a solid malignant tumor in the GCCR files and constituted the final sample. Reasons that no GCCR matches could be found for the remaining patients in the GCS files included 1) missed cases (i.e., facilities not compliant in reporting new cases to the registry), 2) late-stage or second-course therapy (i.e., original primary tumor was diagnosed elsewhere and/or patient was treated at GCS, as such cancers are not reportable per population-based registry rules), or 3) presence of cancers not typically diagnosed or treated in hospitals (e.g., hematological, dermatological, or urogenital cancers diagnosed, treated or passively followed in community oncology clinics).

Integration of missing stage data

Of the 13,486 patients in the sample, the number of patients with no staging information in the EMR data prior to GCCR integration was 3,424(25%). After GCCR integration,

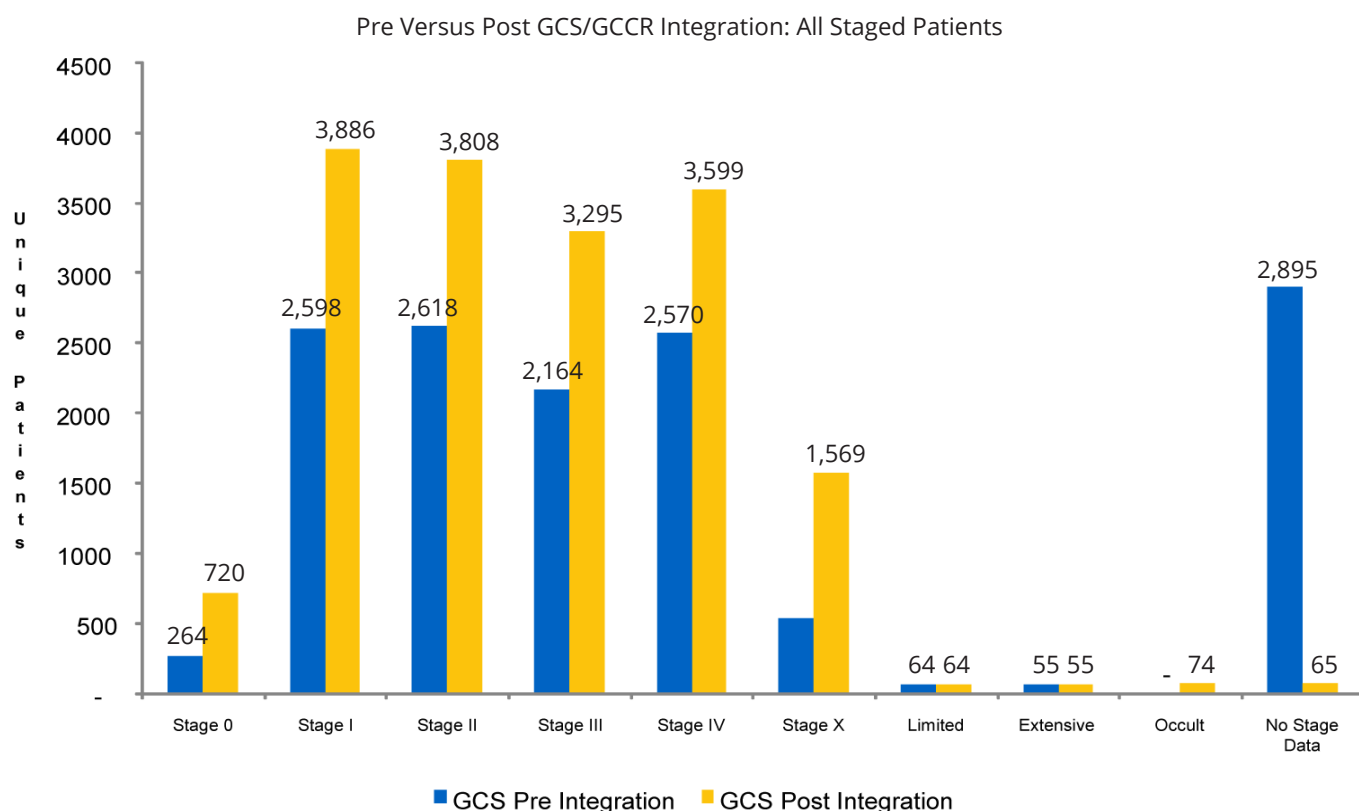
the number of patients with no staging information was reduced to 12%— a relative gain in staged patients of 52% (Figure 1). Gains in the numbers of patients with staging information were observed across all cancer stages (Figure 1).

Discordance of GCS stage Versus GCCR stage

Of the 13,486 patients in the sample, discordance between initial GCS stage and initial GCCR stage was observed in 6100 (45%) (Table 1). Conflicting values appear to originate from variations in coding schemas, differences in dates for first recoded stage, changes in methodology over time, and multiple sites of care reporting, including interpretative differences in findings. Discordance rates for initial cancer stage varied by cancer primary site such that discordance was highest for bone and connective tissue cancers and liver cancer, and lowest for breast cancer and colorectal cancer (Table 2).

Discussion

Integration of registry data can reduce the number of patients in a clinical practice with unknown stage. In this



Unknown or Stage X Pre vs. Post GCS/GCCR Integration	Unique Patients	Total Unstaged	Relative Gain in Staged Patients
GCS Stage Prior to GCCR Integration with Unknown or Stage X. N = 13,486	3,424	25%	
GCS Stage Post GCCR Integration with Unknown or Stage X. N = 13,486	1,634	12%	52%

Figure 1 Results of Georgia Cancer Specialists (GCS) and the Georgia Comprehensive Cancer Registry (GCCR) integration.

Table 2 Discordance by tumor type.

Tumor type	GCS/ GCCR matches	Same initial stage in GCS and GCCR	% with same initial stage in both sources
Total patients	13,486	7,444	55%
Anal cancer	99	32	32%
Bladder cancer	230	49	21%
Bone & connective tissue cancer	85	3	4%
Breast cancer	5,709	3,775	66%
Colorectal cancer	1,867	1,141	61%
Gastrointestinal cancers; Other	532	211	40%
Genitourinary cancer	109	49	45%
Gynecological cancer	170	81	48%
Head & neck cancer	301	88	29%
Liver cancer	145	24	17%
Lung cancer	3,042	1,510	50%
Ovarian cancer	2,012	91	43%
Pancreatic cancer	444	188	42%
Prostate cancer	453	148	33%
Renal cancer	221	89	40%

Abbreviations: GCS = Georgia Cancer Specialists; GCCR = Georgia Comprehensive Cancer Registry.

Note: Includes patients with Stage X, no stage data and unknown stage; Patients were assigned to the first, initial recorded stage in the instance that the patient had multiple stages in either GCS or GCCR; Patients may have had more than 1 primary tumor and were counted in all multiple primary tumor types but only once in the total; Laterality was not used to determine stage values.

study, the integration of registry data with EMR data reduced the proportion of patients with unknown stage in the GCS EMR data by approximately 52%. However, a significant proportion of patients with 'unknown stage' were identified as having tumors with 'Stage X,' which from a research perspective presents significant challenges for analyses by stage. Stage X resulted from missing values for T, N or M in the EMR staging template. The T and N values are readily accessible from pathology reports earlier in the staging process while M values may be delayed due to the necessity of waiting for imaging results. Stage X commonly resulted from failure to update the staging template with the M value once it was known. For treatment purposes, stage was often observed within the non-searchable EMR text notes or scanned reports. The results should be interpreted with the knowledge that the data may not be generalizable to other state registries and other clinical outpatient practices. Further study of integration of staging information from multiple sources is warranted in other geographic and clinical settings.

The value of integrating multiple types of data to increase the completeness of information and provide cross-validation has been increasingly recognized as evidenced by a number of public and private initiatives, and demonstrated in several studies [4, 13-17]. Recently, the American Society of Clinical Oncology (ASCO) initiated CancerLinQ to "aggregate and analyze a massive web of real-world cancer care data..." [18]. As an illustration of the value of more complete information in the area of oncology, a study by Polsky et al. 2009 assessed the likely outcome in a claims-based only analyses of erythropoietin stimulating agents (ESA) costs by conducting propensity-score matching with and without baseline hemoglobin (Hb), which are recorded in electronic medical records (EMRs) [13]. The study found that claims-only studies could produce biased cost estimates and that adding data from an EMR, (e.g., Hb), materially altered the results. From a clinical and statistical perspective, the prima facie logic and value of having more data points and a more complete picture of a patient's tests results and treatments are self-evident if outcomes are to be improved.

The GCCR and the GCS differ in the staging information recorded for the same patients. The variations reflect several fundamental differences. Multiple stages, as well as histology, pathology, and clinical and derived stage based on ICD-0 codes at the initial date of diagnosis were recorded in GCCR. In contrast, one stage was generally recorded in GCS using clinical staging based on ICD-9 codes proximal to the initial visit. Staging was most often recorded for treated patients. Not all of the GCS patients with stage information were confirmed as treated although most probably did receive therapy.

Conflicting stage values were observed between the GCCR and the GCS in approximately 45% of patients' records. Conflicting values appear to originate from variations in coding schemas, differences in first date of staging, changes in methodology over time, and multiple sites of care reporting, including interpretative differences in findings. The sources of conflicting stage values point to specific areas that can be targeted for improvement in

cancer staging and care. Greater harmonization of staging methods and timing will reduce variability in recorded stage between clinical practices and registries in the future. Groups such as the North American Association of Central Cancer Registries for example have "ongoing efforts to coordinate and effectively transition from the Collaborative Staging system to use of the AJCC TNM staging standard with related biomarkers and prognostic factors..." [19].

Conclusion

This study demonstrates the feasibility and utility, from a research perspective, of merging information from multiple data sources in order to improve the quantity and quality of cancer staging information. Optimally, if such data integration is to be useful, cancer registries and clinical practices would exchange information on a routine basis, harmonize standards and coding schemas, automate labor-intensive manual processes, and ultimately provide more complete information at the patient's point of care. The opportunities for, and barriers to, linking cancer registry and EMR data have recently been summarized, and this information may provide a framework for future data integration [6]. Although there are still significant lags before registry data become available for research, and optimally for patient care in the future, technological advances and increased "meaningful" adoption of EMRs for an estimated 90% of US oncology clinics by 2014, are shortening the time to data availability. The network of information exchange would ideally include hospitals, community practices, pharmacies, and radiation centers to create the most complete patient-level, longitudinal records to advance each institution's mission and collectively improve cancer care in the United States. In the future, rapid access to more complete, integrated, timely data has the potential to not only improve research, but most importantly to improve patient care.

Conflict of interest

The authors declare no conflict of interest.

References

- [1] Greene FL, Sobin LH. The staging of cancer: a retrospective and prospective appraisal. *CA Cancer J Clin.* 2008; 58(3):180-190.
- [2] National Cancer Institute, US National Institutes of Health. Cancer Staging: questions and answers. Accessed March 1, 2015 from: <http://www.cancer.gov/about-cancer/diagnosis-staging/staging/staging-fact-sheet>
- [3] Greene F, Page D, Fleming I, Fritz A, Balch C, et al. *AJCC Cancer Staging Manual* (edition 6). New York, NY, Springer; 2002.
- [4] UnitedHealth Group: UnitedHealthcare creates adult national cancer care registry with data and analysis to support oncologists in the fight against cancer. Accessed March 1, 2015 from: <http://www.unitedhealthgroup.com/newsroom/articles/news/unitedhealthcare/2010/0209cancerregistry.aspx>
- [5] Kristianson KJ, Ljunggren H, Gustafsson LL. Data extraction from a semi-structured electronic medical record system for outpatients: a model to facilitate the access and use of data for quality control and research. *Health Informatics J.* 2009; 15(4):305-319.
- [6] Hiatt RA, Tai CG, Blayney DW, Deapen D, Hogarth M, et al. Leveraging state cancer registries to measure and improve the quality of cancer care: a potential strategy for California and beyond. *J Natl Cancer Inst.* 2015; 107(5):pii:djv047.
- [7] Liu WL, Kasl S, Flannery JT, Lindo A, Dubrow R. The accuracy of prostate cancer staging in a population-based tumor registry and its impact on the black-white stage difference (Connecticut, United States). *Cancer Causes Control.* 1995; 6(5):425-430.

-
- [8] Lau EC, Mowat FS, Kelsh MA, Legg JC, Engel-Nitz NM, et al. Use of electronic medical records (EMR) for oncology outcomes research: assessing the comparability of EMR information to patient registry and health claims data. *Clin Epidemiol*. 2011; 3:259-272.
 - [9] Hess G, Fonseca E, Lee A, Wang H, Dhawan R, et al. Characteristics of patients included in the Myelofibrosis Real-World Practice-Based Network Research (ENRICH) Data Platform. *Blood*. 2013; 122(21):1580.
 - [10] Georgia Cancer Specialists. Accessed March 1, 2015 from: <http://www.gacancer.com/about-gcs>
 - [11] Georgia Department of Community Health: Division of Public Health. Georgia Comprehensive Cancer Registry (GCCR). Accessed March 1, 2015 from: <http://dph.georgia.gov/georgia-comprehensive-cancer-registry>
 - [12] Centers for Disease Control. International Classification of Diseases, Ninth Revision (ICD-9): ICD-9 Diagnosis Codes for malignant neoplasms stated or presumed to be primary, of specified sites, except of lymphatic and hematopoietic tissue. Accessed March 1, 2015 from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Publications/ICD-9/ucod.txt
 - [13] Polsky D, Eremina D, Hess G, Hill J, Hulnick S, et al. The importance of clinical variables in comparative analyses using propensity-score matching: the case of ESA costs for the treatment for chemotherapy-induced anaemia. *Pharmacoeconomics*. 2009; 27(9):755-65.
 - [14] Pawlson LG, Scholle SH, Powers A. Comparison of administrative-only versus administrative plus chart review data for reporting HEDIS hybrid measures. *Am J Manag Care*. 2007; 13:533-538.
 - [15] Dresser MV, Feingold L, Rosenkranz SL, Coltin KL. Clinical quality measurement. Comparing chart review and automated methodologies. *Med Care*. 1997; 35(6):539-552.
 - [16] Ashley L, Jones H, Thomas J, Forman D, Newsham A, et al. Integrating cancer survivors' experiences into UK cancer registries: design and development of the ePOCS system (electronic Patient-reported Outcomes from Cancer Survivors). *Br J Cancer*. 2011; 105 Suppl 1:S74-81.
 - [17] Keyser DJ, Dembosky JW, Kmetik K, Antman MS, Sirio C, et al. Using health information technology-related performance measures and tools to improve chronic care. *Jt Comm J Qual Patient Saf*. 2009; 35(5):248-55.
 - [18] American Society of Clinical Oncology, Institute for Quality. Accessed March 1, 2015.
 - [19] Collaborative Stage Transition Newsletter Issue 4 – August 18, 2014. Accessed March 1, 2015 from: <http://www.instituteforquality.org/cancerlinq>