NIH-PA Author Manuscript

Muddy Water? Variation in reporting receipt of breast cancer radiation therapy by population-based tumor registries

Gary V. Walker, M.D., M.P.H.[1], **Sharon H. Giordano, M.D., M.P.H.**[2], **Melanie Williams, Ph.D.**[3], **Jing Jiang, M.S.**[4], **Jiangong Niu, Ph.D.**[2], **Jill MacKinnon, PhD, CTR**[5], **Patricia Anderson, M.S.**[5], **Brad Wohler, M.S.**[5], **Amber H. Sinclair, Ph.D.**[6], **Francis P. Boscoe, Ph.D.**[6], **Maria J. Schymura, Ph.D.**[6], **Thomas A Buchholz, M.D.**[1], and **Benjamin D. Smith, M.D.**[1,✉]

[1]Department of Radiation Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas

[2]Department of Breast Medical Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas

[3]Texas Cancer Registry, Department of State Health Services, Austin, TX

[4]Division of Quantitative Sciences, The University of Texas M. D. Anderson Cancer Center, Houston, Texas

[5]Florida Cancer Data System, University of Miami School of Medicine, Miami, Florida

[6]New York State Cancer Registry, New York State Department of Health, Albany, NY

Abstract

Purpose—Receipt of radiation therapy (RT) is a key quality indicator in breast cancer treatment. Prior analyses using population-based tumor registry data have demonstrated substantial underuse of RT for breast cancer, but the validity of such findings remains debated. To address this controversy, we evaluated accuracy of registry RT coding compared to the gold standard of Medicare claims.

Methods and Materials—Using SEER-Medicare data, we identified 73,077 patients age 66 diagnosed with breast cancer from 2001-2007. Underascertainment (1-sensitivity), sensitivity, specificity, kappa, and chi-square were calculated for RT receipt determined by registry data vs. claims. Multivariate logistic regression characterized patient, treatment, and geographic factors associated with underascertainment of RT. Findings in the SEER-Medicare registries were compared to three non-SEER registries (Florida, New York, and Texas).

Results—In the SEER-Medicare registries, 41.6% (n=30,386) of patients received RT according to registry coding versus 49.3% (n=36,047) according to Medicare claims (P<0.001). Underascertainment of RT was more likely if patients resided in a newer SEER registry (OR 1.70, 95%CI 1.60-1.80;P<0.001), rural county (OR 1.34, 95%CI 1.21-1.48;P<0.001) or if RT was delayed (OR 1.006/day, 95%CI 1.006-1.007,P<0.001). Underascertainment of RT receipt in SEER registries was 18.7% (95% CI 18.6-18.8%), compared to 44.3% (95% CI 44.0-44.5%) in non-SEER registries.

✉Corresponding author: Benjamin D. Smith, M.D., Dept. of Radiation Oncology, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd., Unit 97, Houston, TX 77030, BSmith3@mdanderson.org.

Discussion—Population-based tumor registries are highly variable in ascertainment of RT receipt and should be augmented with other data sources when evaluating quality of breast cancer care. Future work should identify opportunities for the radiation oncology community to partner with registries to improve accuracy of treatment data.

Keywords

Breast Cancer; radiation therapy utilization; SEER

Introduction

The Institute of Medicine has long advocated the development of systems to measure and monitor the quality of cancer care received by the US population.(1) Such systems could identify populations who receive poor quality care, thereby enabling targeted interventions to improve care and outcomes. Population-based tumor registries represent one important avenue for measuring quality of cancer care. For example, in breast cancer, data from the Surveillance, Epidemiology, and End Results (SEER) population-based registry program have been used to determine whether radiation therapy (RT) is used appropriately after lumpectomy and mastectomy.(2-6) Such studies have largely concluded that RT is underutilized in breast cancer patients and, alarmingly, that underutilization has actually worsened in recent years.(3)

Inherent to the use of registry data is the assumption that receipt of RT is correctly ascertained by the reporting registry. Several prior studies have generally suggested good accuracy of RT ascertainment by SEER registries when compared against the gold standard of Medicare billing claims.(7, 8) However, such studies focused on the nine to thirteen SEER registries available at that time and did not evaluate the accuracy of RT coding in the expanded SEER program, which now includes 16 registries with linked Medicare billing claims, or in other population-based registries that do not participate in SEER. As registry data continues to be used by investigators to evaluate RT utilization and outcomes, a contemporary evaluation of the accuracy of RT coding is warranted.

Accordingly, we sought to evaluate the accuracy of registry RT ascertainment against the gold standard of Medicare billing claims in a contemporary cohort of breast cancer patients. To accomplish this, we used linked SEER-Medicare data, representing approximately 26% of the U.S. population, to (1) determine underascertainment, sensitivity, specificity, and kappa for registry RT receipt and (2) identify factors associated with underascertainment of RT. In addition, we partnered with the three largest non-SEER registries—Florida, New York and Texas, representing an additional 20% of the U.S.—to compare RT ascertainment among these registries to the SEER registries.

Methods

Data Sources

The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute assembles information on cancer incidence and survival from 16 population-based tumor registries with a case ascertainment ratio of 97%.(9) Data collected include patient demographics, tumor characteristics, and treatment including RT receipt during the initial treatment course. The National Cancer Institute has linked SEER records to the Medical billing claims of Medicare beneficiaries. Medicare covers inpatient and outpatient medical care for approximately 95% of the US population aged 65 years and older.(10)

Distinct from the SEER-Medicare data, the Florida, New York, and Texas Cancer Registries have also linked their records to Medicare billing claims under the guidance of the National Cancer Institute, the Association of Schools of Public Health, and the National Program of Cancer Registries at the Centers for Disease Control and Prevention with support from their respective state departments of health. Data elements captured by these registries are similar in structure and format to the data collected by the SEER registries.

Description of Study Cohorts

From the SEER-Medicare data, we identified 127,308 women age 66 or older with a diagnosis of invasive breast cancer from 2001-2007. Patients were excluded if they had non-continuous Medicare Part A and B or health maintenance organization (HMO) coverage within 12 months before and after diagnosis, were diagnosed from autopsy or death certificate, presented with metastatic disease or unknown stage at diagnosis, or were diagnosed with a second cancer or died within one year of diagnosis, leaving 73,077 patients for the analytic cohort (Supplementary Table 1).(11)

A similar approach was used to create analytic cohorts for the Florida-Medicare (n=17,165), New York-Medicare (n=5,292), and Texas-Medicare cohorts (n=15,403), with the exception that the New York cohort only included patients diagnosed in 2004-2006 (Supplementary Table 1).

Receipt of Radiation

RT receipt was determined from both registry data and Medicare claims. Registry data were considered to indicate RT receipt if patients were coded as receiving "Beam radiation", "Combination of beam with implants or isotopes" or "Radiation, NOS -- method or source not specified" during the initial treatment course. Medicare claims were considered to indicate RT receipt if at least one claim for delivery of RT was present within one year of diagnosis (Supplementary Table 2). The period of one year was chosen as this was the timeframe established by the Commission on Cancer (CoC) of the American College of Surgeons (ACoS) as a quality measure for receipt of RT following breast conserving surgery (BCS). Receipt of brachytherapy, a newer form of breast cancer RT, was not considered in our definition of RT receipt as it was not a standard of care procedure from 2001-2007 and we were thus concerned that it might not be coded accurately by registries. The RT start interval was defined as time in days from the diagnosis date to the first Medicare claim for delivery of RT. As SEER reports only month and year of diagnosis, each patient was assigned a diagnosis date at the midpoint of the month of diagnosis. For patients whose RT started prior to the assigned diagnosis date (n=19), the RT start interval was reclassified as zero.

Key Covariates

Race, age and residence at the time of diagnosis were determined from registry data. "Urban" was defined as big metropolitan, metropolitan, or urban and "rural" was defined as less urban or rural using SEER definitions. Type of breast surgery was determined by selecting the most extensive surgery reported by either registry data or Medicare billing claims within 12 months of diagnosis (Supplementary Table 2). SEER registries were grouped as the original SEER-9 registries (Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah) versus the newer SEER registries (San Jose, Los Angeles, Rural Georgia, Greater California, Kentucky, Louisiana, and New Jersey).

Statistical Analysis

Data analysis was performed using Stata/SE 12.0 statistical software and SAS v9.2. For our first objective, we used the SEER-Medicare cohort to calculate sensitivity, specificity, positive predictive value, negative predictive value, and Cohen's kappa statistic for RT receipt coded by registries compared to the gold standard of Medicare billing claims. We defined underascertainment (1-sensitivity) as the number of cases where Medicare claims indicated that the patient received RT but registry data indicated that the patient did not receive RT, divided by the total number of cases where Medicare claims indicated that the patient received RT. Univariate predictors of RT receipt were tested using the Pearson X^2 test for categorical variables.

For our second objective, we used the SEER-Medicare cohort to evaluate factors associated with underascertainment of RT. To accomplish this, we first determined underascertainment for each ten day increment in RT start interval. Ordinary least squares regression estimated the association between underascertainment and RT start interval among the entire cohort and also among only those with stage I disease. We then used multivariate logistic regression to identify factors associated with RT underascertainment. Candidate covariates were included based on clinical significance or univariate $P < .25$. The model was iteratively refined to minimize colinearity. Goodness of fit was assessed using the method of Hosmer and Lemeshow.

For our third objective, underascertainment, sensitivity, specificity, positive predictive value, negative predictive value, and kappa were calculated for the Florida, New York, and Texas Cancer registries. Unadjusted logistic regression compared underascertainment for these three registries to the SEER registries. (Adjusted analyses could not be conducted as our existing data user's agreements prohibit data sharing across institutions.)

This study was granted exempt status by our Institutional Review Board. An alpha of 0.05 was used for all analyses. All statistical tests were two-sided.

Results

Descriptive characteristics of SEER-Medicare cohort

Of 73,077 patients identified in the SEER-Medicare cohort, median age at diagnosis was 75 years, 56.9% (n=41,581) underwent BCS, 37.8% (n=27,657) underwent mastectomy, and 18.0% (n=13,123) received chemotherapy (Table 1).

A total of 41.6% (n=30,386) of patients were coded as receiving RT according to SEER registries versus 49.3% (n=36,047) according to Medicare claims (P<0.001) (Table 1). For the entire cohort, underascertainment of RT was 18.7% (n=6,738/29,309, 95% CI 18.6-18.8%). The overall kappa was 78.6% (95% CI 78.1%-79.0%) with sensitivity of 81.3% (95% CI 81.2%-81.5%) and specificity of 97.1% (95% CI 96.9%-97.3%) (Table 2). SEER-9 registries had an underascertainment of 14.1% (95% CI 13.9%-14.2%) compared to 21.8% (95% CI 21.7%-22.0%) for the newer SEER registries (P<0.001). Neither sensitivity nor underascertainment (P=0.43) changed with year of diagnosis.

Predictors of Underascertainment in SEER-Medicare Cohort

Delay in the start of RT was associated with an increase in underascertainment, (0.12%/day, 95% CI 0.11-0.13%/day, $R^2 = 0.94$) (Figure 1). This increase persisted even when analyzing only patients with stage I disease (n=25,400, 0.12%/day, 95% CI 0.10-0.13%/day, $R^2 = 0.88$). In adjusted analysis, factors associated with underascertainment included residence in a newer SEER registry (OR 1.70, 95%CI 1.60-1.80;P<0.001) or rural county (OR 1.34,

95%CI 1.21-1.48;P<0.001) and delay in start of RT (OR 1.006/day, 95%CI 1.006-1.007,P<0.001) (Table 3). Advanced age and treatment with mastectomy were also associated with higher likelihood of underascertainment, but year of diagnosis and race were not.

Comparison of Florida, New York, and Texas Registries to the SEER Registries

Sensitivity of RT reporting was 48.4% for Florida, 56.1% for Texas, and 81.1% for New York, compared to a range for 72.6% to 94.4% for the SEER registries (Table 4). Underascertainment was 51.6% for Florida, 43.9% for Texas, and 18.9% for New York, compared to a range of 5.6% to 27.4% for the SEER registries. In comparison to the SEER registries, underascertainment was more likely for Florida (OR 4.63, 95% CI 4.41-4.86; P<0.001) and Texas (OR 3.41, 95% CI 3.23-3.60; P<0.001) but not for New York (OR 1.01, 95% CI 0.92-1.12; P=0.79).

Discussion

In this unique cohort of breast cancer patients representing nearly half the US population, we found significant variation in the ascertainment of RT by cancer registries. While some registries demonstrated extremely high sensitivity exceeding 90%, sensitivity for other registries was below 60%. Further, we found that the original SEER-9 registries had the most accurate data, while two non-SEER registries had the least accurate data. In addition, we found that rural residence and increased RT start interval negatively impacted ascertainment of RT receipt by tumor registries. These results illustrate that in certain settings RT can be coded with good accuracy, but that caution is generally needed when evaluating studies of RT utilization that rely on registry data alone.

For example, influential studies reported in *The Lancet* and *Journal of Clinical Oncology* used tumor registry data to argue that the likelihood of inappropriate local-regional management of breast cancer has increased over time, as BCS gained acceptance in breast cancer management.(2, 3) These studies defined appropriate management as total mastectomy with axillary lymph node dissection or BCS with axillary lymph node evaluation and RT. However, because these studies used registry data alone, it is likely that underascertainment of RT by registries resulted in inappropriately low rates of suitable local-regional management reported by these studies. Similarly, our findings call into question other studies of SEER registries alone which reported underuse (4, 12), rural disparities,(6) and geographic variation (3, 5, 13) in RT receipt for breast cancer.

Prior literature has evaluated ascertainment of RT by tumor registries in comparison to Medicare claims. In a cohort of women diagnosed in 1992-1993, Du et al evaluated accuracy of the SEER-9 registries and reported underascertainment of 18.7%.(8) In a similar study evaluating women diagnosed in 1991-1996, Virnig et al reported underascertainment of only 13.5%.(7) In our study underascertainment was 14.1% for the original SEER-9 registries compared to 21.8% for the newer SEER registries, suggesting that underascertainment remains a persistent issue, particularly with the newer registries.

Various studies have also evaluated ascertainment of RT by registries in comparison to medical record review or patient self-report. (14, 15) For example, a comparison of data from the California Cancer Registry to medical record review found that underascertainment was 25.6% from 1992-1996.(16) In addition, a survey of 2,290 breast cancer patients residing in the catchment areas for the Los Angeles or Detroit registries found that underascertainment was 32.0% in Los Angeles and 11.3% in Detroit.(17) In this study, underascertainment was significantly associated with registry, income, mastectomy receipt, chemotherapy receipt and diagnosis at a hospital not accredited by the ACoS. Our study

complements and expands upon this prior literature by revealing that efforts to improve RT ascertainment should focus on newer registries, delayed radiation treatment, and rural regions.

Several underlying factors likely account for the measured variation in ascertainment of RT by tumor registries. For example, registries have various policies in place regarding how they obtain information regarding receipt of RT. These policies are most often dependent on registry funding and staffing Some registries actively survey facilities that deliver RT to ascertain treatment, while other registries must passively wait for reporting from treating facilities.(17) As certified tumor registrars typically work within ACoS accredited hospitals, it is also conceivable that regional variation in the proportion of patients treated at non-ACoS accredited hospitals or freestanding centers may also impact RT ascertainment. As highlighted in our analysis, registries with a more rural population may have more variation in ascertainment of RT, which could explain some of the difference between the Florida/Texas and the New York registries.

Various strategies could be employed to improve RT coding by tumor registries. For example, passage of the Health Information Technology for Economic and Clinical Health (HITECH) Act establishes automated incident case reporting using electronic medical records as the new standard for cancer registries.(18) Compliance will be linked to the Medicare and Medicaid Electronic Health Record (HER) incentive program. Allowing registries to further access treatment information from electronic medical records or associated claims for cancer treatment could substantially improve RT ascertainment in a cost-effective manner.

Recently, a National Radiation Oncology Registry (NROR) has been created through collaboration between the Radiation Oncology Institute (ROI) and the American Society for Radiation Oncology (ASTRO). The purpose is, "To address the need for accurate, comprehensive, and clinically rich data, to determine national patterns of care, outcomes, and gaps in treatment quality, and to compare the effectiveness of different treatment modalities."(19) If the NROR could develop strategic data sharing agreements with local tumor registries, the information flowing into the NROR could then be used by registries to improve accuracy of RT ascertainment. Such arrangements would be in the best interest of patients and the field of radiation oncology, as the data provided would help to identify patient populations who remain at risk for inappropriate omission of RT and allow interventions to improve access to RT for such patients.

There are certain limitations of this analysis to consider. For example, we used receipt of RT based on Medicare claims data as the gold standard because our primary aim was to identify factors that contributed to underasceratainment of RT by tumor registries. Nevertheless, reliance on Medicare claims alone may also result in underascertainment of RT receipt, for example in patients who receive RT in military or Veterans Affairs Hospitals and thus do not generate Medicare billing claims. However, in our analysis, only 1.4% of patients were coded as receiving RT by SEER registries but not by Medicare claims, suggesting that this is a relatively uncommon event. In addition, it is possible that Medicare claims may inappropriately classify patients as receiving RT when such RT was not delivered as part of the initial treatment course. To mitigate this possibility, we excluded patients with distant disease who have the highest risk for early recurrence which would require RT and those with second cancers diagnosed within one year of the index diagnosis who may require RT for another indication. Nevertheless, it remains possible that Medicare claims could misclassify a small percentage of patients as having received RT when such treatment was not part of the initial treatment course. Such misclassification would bias results by decreasing the measured sensitivity of SEER registries. However, in our analysis, the

association between RT delay and underascertainment persisted even when analyzing only patients with stage I disease, who have the lowest risk of disease recurrence requiring RT outside the first treatment course.

In summary, population-based tumor registries, although critical to oncologic research, are highly variable in ascertainment of RT receipt and should generally be augmented with other data sources when evaluating quality of breast cancer care. Future work should identify opportunities to use electronic medical records and resources from the radiation oncology community to improve accuracy of registry treatment data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Hewitt, M.; Simone, J. National cancer policy board, institute of medicine: Ensuring quality cancer care. National Academy Press; Washington, DC: 1999.

2. Nattinger AB, Hoffmann RG, Kneusel RT, Schapira MM. Relation between appropriateness of primary therapy for early-stage breast carcinoma and increased use of breast-conserving surgery. Lancet. Sep 30; 2000 356(9236):1148–53. [PubMed: 11030294]

3. Freedman RA, He Y, Winer EP, Keating NL. Trends in racial and age disparities in definitive local therapy of early-stage breast cancer. J Clin Oncol. Feb 10; 2009 27(5):713–9. [PubMed: 19103731]

4. Baxter NN, Virnig BA, Durham SB, Tuttle TM. Trends in the treatment of ductal carcinoma in situ of the breast. J Natl Cancer Inst. Mar 17; 2004 96(6):443–8. [PubMed: 15026469]

5. Farrow DC, Hunt WC, Samet JM. Geographic variation in the treatment of localized breast cancer. N Engl J Med. Apr 23; 1992 326(17):1097–101. [PubMed: 1552910]

6. Dragun AE, Huang B, Tucker TC, Spanos WJ. Disparities in the application of adjuvant radiotherapy after breast-conserving surgery for early stage breast cancer: Impact on overall survival. Cancer. Jun 15; 2011 117(12):2590–8. [PubMed: 21656737]

7. Virnig BA, Warren JL, Cooper GS, Klabunde CN, Schussler N, Freeman J. Studying radiation therapy using SEER-medicare-linked data. Med Care. Aug; 2002 40(8 Suppl):49–54. IV.

8. Du X, Freeman JL, Goodwin JS. Information on radiation treatment in patients with breast cancer: The advantages of the linked medicare and SEER data. surveillance, epidemiology and end results. J Clin Epidemiol. May; 1999 52(5):463–70. [PubMed: 10360342]

9. Zippin C, Lum D, Hankey BF. Completeness of hospital cancer case reporting from the SEER program of the national cancer institute. Cancer. Dec 1; 1995 76(11):2343–50. [PubMed: 8635041]

10. Warren JL, Klabunde CN, Schrag D, Bach PB, Riley GF. Overview of the SEER-medicare data: Content, research applications, and generalizability to the united states elderly population. Med Care. Aug; 2002 40(8 Suppl):3–18. IV.

11. Housman DM, Decker RH, Wilson LD. Regarding adjuvant radiation therapy in merkel cell carcinoma: Selection bias and its affect on overall survival. J Clin Oncol. Oct 1; 2007 25(28): 4503–4. author reply 4504-5. [PubMed: 17906216]

12. Smith BD, Smith GL, Haffty BG. Postmastectomy radiation and mortality in women with T1-2 node-positive breast cancer. J Clin Oncol. Mar 1; 2005 23(7):1409–19. [PubMed: 15735117]

13. Smith GL, Smith BD, Haffty BG. Rationalization and regionalization of treatment for ductal carcinoma in situ of the breast. Int J Radiat Oncol Biol Phys. Aug 1; 2006 65(5):1397–403. [PubMed: 16750316]

14. Bickell NA, Chassin MR. Determining the quality of breast cancer care: Do tumor registries measure up? Ann Intern Med. May 2; 2000 132(9):705–10. [PubMed: 10787363]

15. Du XL, Key CR, Dickie L, Darling R, Geraci JM, Zhang D. External validation of medicare claims for breast cancer chemotherapy compared with medical chart reviews. Med Care. Feb; 2006 44(2): 124–31. [PubMed: 16434911]

16. Malin JL, Kahn KL, Adams J, Kwan L, Laouri M, Ganz PA. Validity of cancer registry data for measuring the quality of breast cancer care. J Natl Cancer Inst. Jun 5; 2002 94(11):835–44. [PubMed: 12048271]

17. Jagsi R, Abrahamse P, Hawley ST, Graff JJ, Hamilton AS, Katz SJ. Underascertainment of radiotherapy receipt in surveillance, epidemiology, and end results registry data. Cancer. Jun 29.2011

18. CDC. cancer - NPCR - meaningful use of electronic health records [Internet]. cited 9/17/2012]. Available from: http://www.cdc.gov/cancer/npcr/meaningful_use.htm

19. Palta JR, Efstathiou JA, Bekelman JE, Mutic S, Bogardus CR, McNutt TR, et al. Developing a national radiation oncology registry: From acorns to oaks. Practical Radiation Oncology. 2012; 2(1):10–7. 0.

Summary

We found that registry data regarding receipt of radiation therapy (RT) for breast cancer is highly variable and depends on patient factors and the population-based registry charged with collecting this data. Studies relying on registry data alone should be cautious when reporting RT utilization.
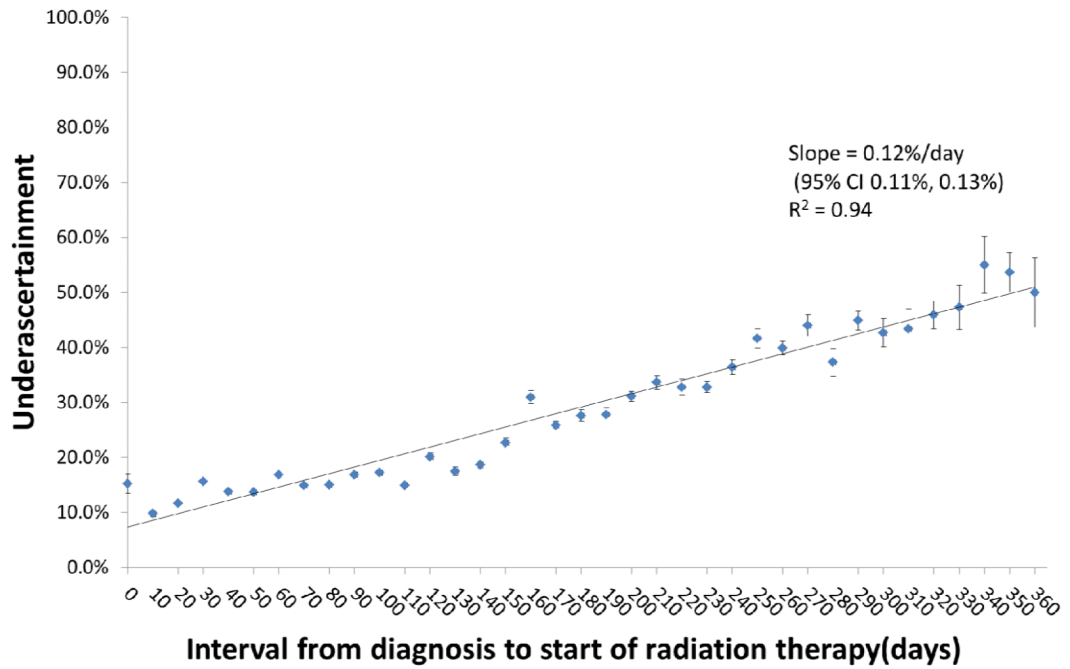
Figure 1.
Underascertainment of radiation therapy (RT) coding was determinted for each ten day increment in days from the diagnosis date to the first claim for delivery of RT. Ordinary least squares regression estimated the association between underascertainment and RT start interval. Vertical bars represent 95% confidence intervals.
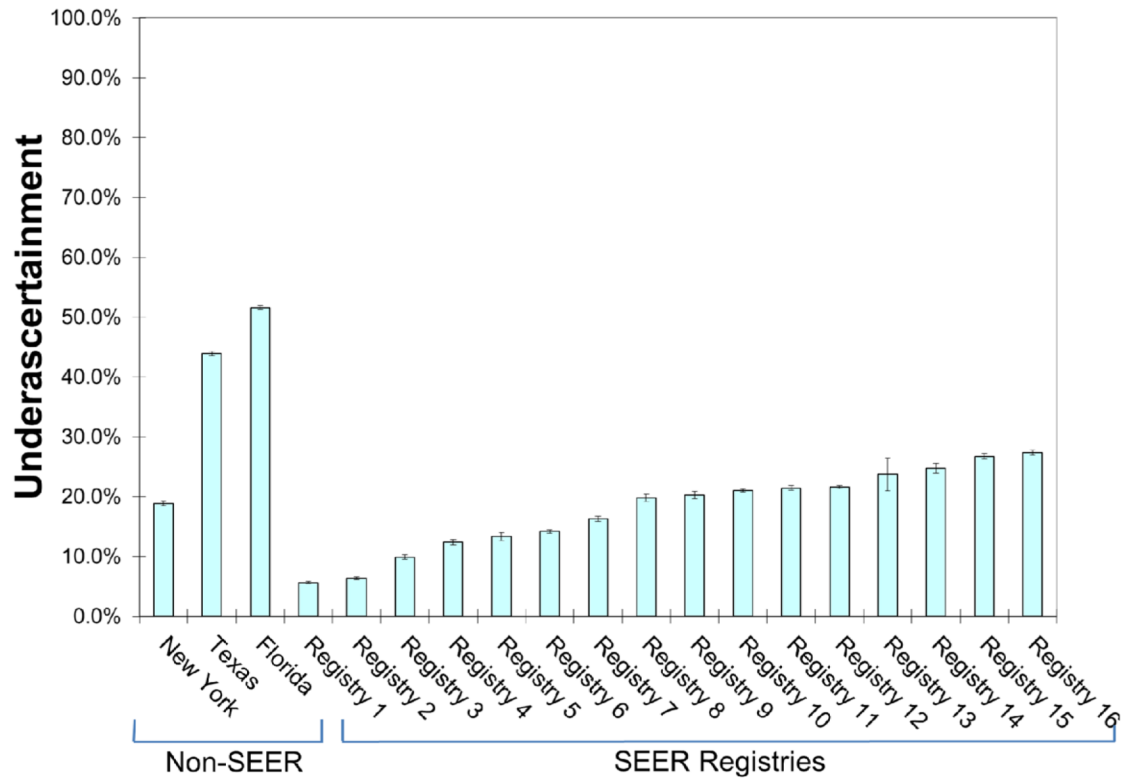
Figure 2.
Underascertainment of radiation therapy (RT) receipt, defined as as the number of cases where Medicare claims indicated that the patient received RT but registry data indicated that the patient did not receive RT, divided by the total number of cases where Medicare claims indicated that the patient received. Vertical bars represent 95% confidence intervals.

Table 1

Baseline demographics and radiation therapy coding in the SEER-Medicare registries

| | | All patients N (%) | Radiation Receipt | | | |
| | | | Medicare claims no | | Medicare claims yes | |
			SEER No N (%)	SEER Yes N (%)	SEER No N (%)	SEER Yes N (%)
All patients		73077 (100.0%)	35953 (49.2%)	1077 (1.5%)	6738 (9.2%)	29309 (40.1%)
Age	Median	75 Years	77 Years	72 Years	73 Years	74 Years
	66-69	15593 (21.3%)	6033 (38.7%)	389 (2.5%)	1778 (11.4%)	7393 (47.4%)
	70-74	18728 (25.6%)	7810 (41.7%)	280 (1.5%)	2015 (10.8%)	8623 (46.0%)
	75+	38756 (53.0%)	22110 (57.0%)	408 (1.1%)	2945 (7.6%)	13293 (34.3%)
Race	White	64315 (88.0%)	31353 (48.7%)	888 (1.4%)	5984 (9.3%)	26090 (40.6%)
	Black	4868 (6.7%)	2591 (53.2%)	113 (2.3%)	441 (9.1%)	1723 (35.4%)
	Other/Unknown	3894 (5.3%)	1967 (50.5%)	76 (2.0%)	306 (7.9%)	1481 (38.0%)
Year of Diagnosis	2001	10957 (15.0%)	5560 (50.7%)	112 (1.0%)	956 (8.7%)	4329 (39.5%)
	2002	10866 (14.9%)	5383 (49.5%)	129 (1.2%)	1079 (9.9%)	4275 (39.3%)
	2003	10618 (14.5%)	5042 (47.5%)	169 (1.6%)	925 (8.7%)	4482 (42.2%)
	2004	10558 (14.4%)	5090 (48.2%)	165 (1.6%)	969 (9.2%)	4334 (41.0%)
	2005	10286 (14.1%)	5089 (49.5%)	179 (1.7%)	886 (8.6%)	4132 (40.2%)
	2006	9861 (13.5%)	4846 (49.1%)	182 (1.8%)	899 (9.1%)	3934 (39.9%)
	2007	9931 (13.6%)	4943 (49.8%)	141 (1.4%)	1024 (10.3%)	3823 (38.5%)
Registry*	SEER-9	28919 (39.6%)	13885 (48.0%)	437 (1.5%)	2053 (7.1%)	12544 (43.4%)
	SEER-Other	44158 (60.4%)	22068 (50.0%)	640 (1.4%)	4685 (10.6%)	16765 (38.0%)
County of residence	Urban	66701 (91.3%)	32079 (48.1%)	1040 (1.6%)	6184 (9.3%)	27398 (41.1%)
	Rural	>6361† (>8.7%)	3871 (60.9%)	>46† (>0.7%)	554 (8.7%)	1910 (30.0%)
	Unknown	<11† (<0.02%)	<11†	<11†	<11†	<11†

	All patients N (%)		Radiation Receipt			
			Medicare claims no		Medicare claims yes	
			SEER No N (%)	SEER Yes N (%)	SEER No N (%)	SEER Yes N (%)
Stage						
I	46067	(63.0%)	21993 (47.7%)	678 (1.5%)	3706 (8.0%)	19690 (42.7%)
II	20408	(27.9%)	10781 (52.8%)	274 (1.3%)	2073 (10.2%)	7280 (35.7%)
III	6602	(9.0%)	3179 (48.2%)	125 (1.9%)	959 (14.5%)	2339 (35.4%)
Surgery						
Biopsy only	3839	(5.3%)	2179 (56.8%)	401 (10.4%)	229 (6.0%)	1030 (26.8%)
BCS	41581	(56.9%)	10928 (26.3%)	521 (1.3%)	5213 (12.5%)	24919 (59.9%)
Mastectomy	27657	(37.8%)	22846 (82.6%)	155 (0.6%)	1296 (4.7%)	3360 (12.1%)
Chemotherapy receipt						
Yes	13123	(18.0%)	4899 (37.3%)	207 (1.6%)	2326 (17.7%)	5691 (43.4%)
No	59954	(82.0%)	31054 (51.8%)	870 (1.5%)	4412 (7.4%)	23618 (39.4%)
Interval from diagnosis to start of radiation therapy (days)						
0-120	26840	(36.7%)	NA	NA	3972 (14.8%)	22868 (85.2%)
121-365	9207	(12.6%)	NA	NA	2766 (30.0%)	6441 (70.0%)
NA	37031	(50.7%)	35955 (97.1%)	1076 (2.9%)	NA	NA

Abbreviations: SEER, Surveillance, Epidemiology and End Results; BCS, breast-conserving surgery; NA, not available

*
The original SEER 9 registries include Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah. The SEER-other registries include those added in 1992 and later (San Jose, Los Angeles, Greater California, Rural Georgia, Kentucky, Louisiana, and New Jersey).

†
Cell sizes have been rounded to protect the confidentiality of the unknown group in accordance with our data user's agreement.

Table 2

Accuracy of registry coding compared to Medicare claims in the SEER-Medicare cohort[*]

		Under-ascertainment	Kappa	Specificity	NPV	PPV
All patients		18.7%	96.5%	97.1%	97.1%	84.2%
Age						
	66-69	19.4%	72.2%	93.9%	77.2%	95.0%
	70-74	18.9%	75.7%	96.5%	79.5%	96.9%
	75+	18.1%	81.8%−	98.2%	88.2%	97.0%
Race						
	White	18.7%	78.6%	97.2%	84.0%	96.7%
	Black	20.4%	76.6%	95.8%	85.5%	93.8%
	Other/Unknown	17.1%	79.8%	96.3%	86.5%	95.1%
Year of Diagnosis						
	2001	18.1%	80.4%	98.0%	85.3%	97.5%
	2002	20.2%	77.7%	97.7%	83.3%	97.1%
	2003	17.1%	79.4%	96.8%	84.5%	96.4%
	2004	18.3%	78.5%	96.9%	84.0%	96.3%
	2005	17.7%	79.2%	96.6%	85.2%	95.8%
	2006	18.6%	78.0%	96.4%	84.4%	95.6%
	2007	21.1%	76.4%	97.2%	82.8%	96.4%
Registry[†]						
	SEER-9	14.1%	91.4%	97.0%	87.1%	96.6%
	SEER-Other	21.8%	87.9%	97.2%	82.4%	96.3%
County of residence						
	Urban	18.4%	78.4%	96.9%	83.8%	96.3%
	Rural	22.5%	79.7%	99.1%	87.5%	98.1%
Stage						
	I	15.8%	81.0%	97.0%	85.6%	96.7%
	II	22.2%	76.5%	97.5%	83.9%	96.4%
	III	29.1%	67.2%	96.2%	76.8%	94.9%
Surgery						
	Biopsy only	18.2%	64.0%	84.5%	90.5%	72.0%
	BCS	17.3%	69.3%	95.4%	67.7%	98.0%
	Mastectomy	27.8%	79.2%	99.3%	94.6%	95.6%
Chemotherapy receipt						
	Yes	29.0%	62.2%	95.9%	67.8%	96.5%
	No	15.7%	82.2%	97.3%	87.6%	96.4%
Interval from diagnosis to start of radiation therapy (days)						
	0-120	14.8%	NA	NA	NA	NA
	121-365	30.0%	NA	NA	NA	NA

Abbreviations: SEER, Surveillance, Epidemiology and End Results; BCS, breast-conserving surgery; NPV, negative predictive value; PPV, positive predictive value; NA, not available

[*]Specificity, positive predictive value, negative predictive value, and cohen's kappa statistic were calculated for radiation therapy (RT) receipt coded by registries compared to the gold standard of Medicare billing claims. Underascertainment of RT receipt was defined as the number of cases where Medicare claims indicated that the patient received RT but registry data indicated that the patient did not receive RT, divided by the total number of cases where Medicare claims indicated that the patient received RT (1-sensitivity).

[†]The original SEER 9 registries include Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah. The SEER-other registries include those added in 1992 and later (San Jose, Los Angeles, Rural Georgia, Greater California, Kentucky, Louisiana, and New Jersey).

Table 3

Predictors of underascertainment of radiation therapy by registries in the SEER-Medicare Cohort[*]

		OR	95% CI	*P*
Age				
	66-69	1		
	70-74	1.06	0.99-1.14	0.11
	75-79	1.14	1.05-1.23	**0.001**
	80+	1.10	1.02-1.20	**0.02**
Race				
	White	1		
	Black	0.95	0.85-1.07	0.40
	Other	0.89	0.78-1.01	0.06
County of residence				
	Urban	1		
	Rural	1.34	1.21-1.48	**<0.001**
Year of Diagnosis				
	2001	1		
	2002	1.13	1.02-1.25	**0.02**
	2003	0.93	0.84-1.03	0.14
	2004	0.99	0.89-1.09	0.77
	2005	0.94	0.85-1.04	0.24
	2006	0.98	0.88-1.08	0.65
	2007	1.13	1.02-1.26	**0.02**
Registry[*]				
	SEER-9	1		
	SEER-Other	1.70	1.60-1.80	**<0.001**
Surgery				
	Biopsy only	1		
	BCS	1.11	0.95-1.30	0.19
	Mastectomy	1.23	1.04-1.45	**0.02**
Interval from diagnosis to start of radiation therapy (days)				
	Continuous	1.006	1.006-1.007	**<0.001**

Abbreviations: SEER, Surveillance, Epidemiology and End Results; OR, odds ratio; CI, confidence interval; BCS, breast-conserving surgery

[*] This table presents a multivariate logistic regression model conducted in patients who received radiation therapy (RT) according to their Medicare claims (n=36,047). The modeled outcome is underascertainment of RT, defined as the number of cases where Medicare claims indicated that the patient received RT but registry data indicated that the patient did not receive RT, divided by the total number of cases where Medicare claims indicated that the patient received RT. In this model, an OR > 1 indicates a higher of odds of underascertainment.

[†] The original SEER 9 registries include Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah. The SEER-other registries include those added in 1992 and later (San Jose, Los Angeles, Rural Georgia, Greater California, Kentucky, Louisiana, and New Jersey).

Table 4

Accuracy of radiation coding in SEER-Medicare registries compared to the Florida, New York, and Texas Registries[*]

Registry	Under-ascertainment	Kappa	Specificity	NPV	PPV
SEER Registries	18.7%	78.6%	97.1%	84.2%	96.5%
Registry 1	5.6%	88.6%	94.3%	93.8%	94.8%
Registry 2	6.3%	93.5%	99.2%	95.5%	98.9%
Registry 3	9.9%	85.7%	95.9%	89.7%	96.0%
Registry 4	12.4%	83.8%	96.9%	87.1%	97.0%
Registry 5	13.3%	78.3%	92.8%	83.7%	94.2%
Registry 6	14.2%	81.7%	96.2%	86.3%	96.1%
Registry 7	16.3%	83.2%	97.7%	90.1%	96.0%
Registry 8	19.8%	78.8%	97.8%	84.9%	97.0%
Registry 9	20.3%	78.6%	98.5%	83.7%	98.0%
Registry 10	21.0%	74.8%	97.1%	79.6%	97.0%
Registry 11	21.4%	73.2%	97.2%	76.9%	97.5%
Registry 12	21.6%	75.4%	96.5%	82.8%	95.4%
Registry 13	23.7%	80.6%	100.0%	88.4%	100.0%
Registry 14	24.7%	75.4%	97.9%	84.3%	96.4%
Registry 15	26.8%	68.3%	96.0%	76.3%	95.4%
Registry 16	27.4%	74.7%	99.1%	83.7%	98.4%
Non-SEER Registries					
New York	18.9%	76.6%	95.9%	82.5%	95.5%
Texas	43.9%	55.5%	97.7%	72.3%	95.5%
Florida	51.6%	43.0%	98.2%	58.8%	97.3%

Abbreviations: SEER, Surveillance, Epidemiology and End Results; BCS, breast-conserving surgery; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value Note: Names of SEER registries are suppressed in according with our Data User's Agreement to protect confidentiality of individual SEER registries.

[*] Specificity, positive predictive value, negative predictive value, and cohen's kappa statistic were calculated for radiation therapy (RT) receipt coded by registries compared to the gold standard of Medicare billing claims. Underascertainment of RT receipt was defined as the number of cases where Medicare claims indicated that the patient received RT but registry data indicated that the patient did not receive RT, divided by the total number of cases where Medicare claims indicated that the patient received RT (1-sensitivity).