

# Underascertainment of Radiotherapy Receipt in Surveillance, Epidemiology, and End Results Registry Data

Reshma Jagsi, MD, DPhil<sup>1</sup>; Paul Abrahamse, MS<sup>2</sup>; Sarah T. Hawley, PhD<sup>2</sup>; John J. Graff, PhD, MS<sup>3</sup>; Ann S. Hamilton, PhD<sup>4</sup>; and Steven J. Katz, MD, MPH<sup>2,5</sup>

**BACKGROUND:** Surveillance, Epidemiology, and End Results (SEER) registry data have been used to suggest underuse and disparities in receipt of radiotherapy. Prior studies have cautioned that SEER may underascertain radiotherapy but lacked adequate representation to assess whether underascertainment varies by geography or patient sociodemographic characteristics. The authors sought to determine rates and correlates of underascertainment of radiotherapy in recent SEER data. **METHODS:** The authors evaluated data from 2290 survey respondents with non-metastatic breast cancer, aged 20 to 79 years, diagnosed from June of 2005 to February 2007 in Detroit and Los Angeles and reported to SEER registries (73% response rate). Survey responses regarding treatment and sociodemographic factors were merged with SEER data. The authors compared radiotherapy receipt as reported by patients versus SEER records. The authors then assessed correlates of radiotherapy underascertainment in SEER. **RESULTS:** Of 1292 patients who reported receiving radiotherapy, 273 were coded as not receiving radiotherapy in SEER (underascertained). Underascertainment was more common in Los Angeles than in Detroit (32.0% vs 11.25%,  $P < .001$ ). On multivariate analysis, radiotherapy underascertainment was significantly associated in each registry (Los Angeles, Detroit) with stage ( $P = .008$ ,  $P = .026$ ), income ( $P < .001$ ,  $P = .050$ ), mastectomy receipt ( $P < .001$ ,  $P < .001$ ), chemotherapy receipt ( $P < .001$ ,  $P = .045$ ), and diagnosis at a hospital that was not accredited by the American College of Surgeons ( $P < .001$ ,  $P < .001$ ). In Los Angeles, additional significant variables included younger age ( $P < .001$ ), nonprivate insurance ( $P < .001$ ), and delayed receipt of radiotherapy ( $P < .001$ ). **CONCLUSIONS:** SEER registry data as currently collected may not be an appropriate source for documentation of rates of radiotherapy receipt or investigation of geographic variation in the radiation treatment of breast cancer. *Cancer* 2012;118:333-41. © 2011 American Cancer Society.

**KEYWORDS:** breast cancer, SEER, radiation therapy, data quality, ascertainment.

The National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program began collecting cancer registry data in 1973. Currently, SEER collects data from regional cancer registries that cover 26% of the US population.<sup>1</sup> These data include information on cancer incidence, patient demographics, clinical and treatment factors, and survival, information of considerable relevance to those pursuing the agenda for comparative effectiveness research in health care.

SEER data have been used to answer a variety of research questions.<sup>2</sup> Several influential studies have relied upon SEER data alone to determine the appropriateness of care delivered to breast cancer patients, including rates of receipt of radiotherapy (RT) after breast-conserving surgery (BCS).<sup>3,4</sup> These studies have suggested underuse of RT as well as disparities in use by race, age, and geography.

**Corresponding author:** Dr. Reshma Jagsi, MD, DPhil, Department of Radiation Oncology, University of Michigan, UHB2C490, SPC 5010, 1500 East Medical Center Drive, Ann Arbor, MI 48109-5010; Fax: (734) 763-7370; [rjagsi@med.umich.edu](mailto:rjagsi@med.umich.edu)

<sup>1</sup>Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan; <sup>2</sup>General Medicine Division, University of Michigan, Ann Arbor, Michigan; <sup>3</sup>Cancer Institute of New Jersey, Robert Wood Johnson Medical School, New Brunswick, New Jersey; <sup>4</sup>Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California; <sup>5</sup>Department of Health Management and Policy, University of Michigan School of Public Health, Ann Arbor, Michigan

Presented in preliminary form at the American Society of Clinical Oncology Annual Meeting, Health Services Research Oral Presentation Session, June 7, 2010; Chicago, IL.

We thank our project staff: Barbara Salem, MS, MSW and Ashley Gay, BA (University of Michigan); Ain Boone, BA, Cathey Boyer, MSA, and Deborah Wilson, BA (Wayne State University); and Alma Acosta, Mary Lo, MS, Norma Caldera, Marlene Caldera, Maria Isabel Gaeta, and Meryl Leventhal (University of Southern California). (All of these individuals received compensation for their assistance.) We also thank the breast cancer patients who responded to our survey.

**DOI:** 10.1002/cncr.26295, **Received:** January 14, 2011; **Revised:** February 22, 2011; **Accepted:** February 23, 2011, **Published online** June 29, 2011 in Wiley Online Library ([wileyonlinelibrary.com](http://wileyonlinelibrary.com))

The SEER program issues a standardized coding manual that indicates that all treatments administered as part of the first course (which is defined in detail in the manual and is no longer limited to a 4-month period of time) are to be considered for the radiation treatment summary field.<sup>5</sup> However, some studies have suggested that registry data may be incomplete, particularly for treatments like RT that may be delivered in the outpatient setting.<sup>6,7</sup> These studies have led to some increased caution in the use of SEER data alone, but have not convinced researchers to abandon publishing studies of RT use based on analyses of SEER data alone,<sup>8-10</sup> nor even always to acknowledge potential limitations. Furthermore, existing studies assessing the adequacy of ascertainment of treatments in SEER registry data have lacked adequate sample diversity by race, age, and geography to assess whether ascertainment varies by these subgroups. In addition, rates of RT underascertainment may have risen in recent years, with increasing use of chemotherapy before RT leading to the delay of RT beyond a time period readily ascertained by registrars.

In light of these gaps in the literature, we conducted a study comparing SEER data on RT receipt from 2 large regional registries to self-report by patients recently treated for breast cancer to answer 3 questions. First, how well do these SEER registries ascertain RT receipt in the current era; second, do different SEER sites differ in rates of RT ascertainment; and third, if RT underascertainment exists, does it vary systematically by clinical or sociodemographic factors?

## MATERIALS AND METHODS

### *Sampling and Data Collection*

Details of the study design have been published elsewhere.<sup>11-13</sup> Women in the metropolitan areas of Los Angeles and Detroit aged 21 to 79 years, diagnosed with stage 0 to 3<sup>14</sup> primary ductal carcinoma in situ or invasive breast cancer from June 2005 through February 2007 were eligible for sample selection. Latina (in Los Angeles) and African American (in both Los Angeles and Detroit) patients were oversampled.

Eligible patients were selected using rapid case ascertainment as they were reported to the Los Angeles Cancer Surveillance Program and the Metropolitan Detroit Cancer Surveillance System-SEER program registries. This method was used to obtain a representative sample of cases sooner after diagnosis than can be provided by routine ascertainment. We selected all African Americans in both sites and all Hispanic patients in Los Angeles

followed by a random sample of non-African American/non-Hispanic patients in both regions to achieve the target sample size.<sup>12</sup> Asian women in Los Angeles were excluded because these women were being enrolled in other studies.

Physicians were notified of our intent to contact patients, followed by a patient mailing of survey materials and \$10 to eligible subjects. The questionnaire was translated into Spanish,<sup>15</sup> and the Dillman method was used to encourage response.<sup>16</sup> The protocol was approved by the institutional review boards of the University of Michigan, University of Southern California, and Wayne State University.

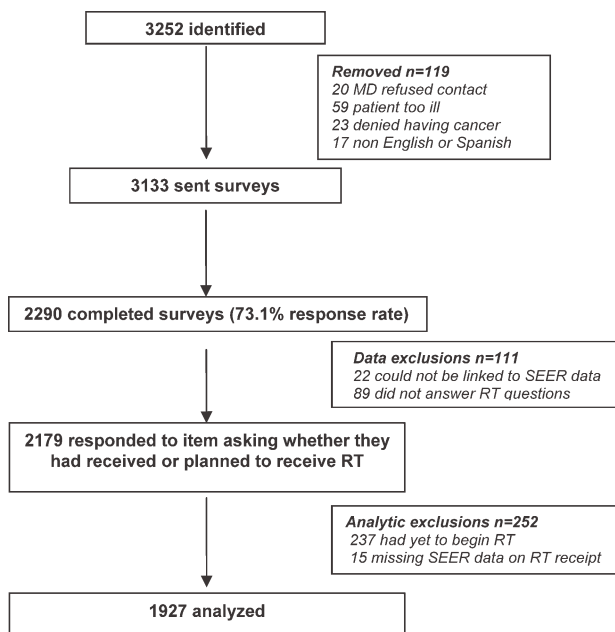
We accrued 3252 eligible patients, including approximately 70% of Latina and African American patients and 30% of non-Latina white patients diagnosed in Los Angeles and Detroit during the study period. After initial selection, another 119 were excluded because 1) physician refused permission to contact (n = 20); 2) patient did not speak English or Spanish (n = 17); 3) patient was too ill or incompetent to participate (n = 59); and 4) patient denied having cancer (n = 23). Of the 3133 eligible patients included in the final accrued sample, 2290 (73.1%) completed surveys, and 2268 (72.4%) were later able to be matched to quality-controlled incident cases ascertained by the SEER registries. On average, patients were surveyed 10 months after diagnosis.

As shown in Figure 1, 2179 patients responded to a question asking whether they had received or planned to receive RT. We excluded the 237 patients who indicated that they had yet to begin RT at the time of the survey; we also excluded the 15 patients missing SEER data on RT receipt, leaving 1927 patients for analysis.

### *Measures*

We measured RT receipt by asking: "Did you or are you going to have radiation therapy to treat your breast cancer?" We also asked about the timing of treatment (completed, started, or planned), as well as whether initiation of RT had been delayed for any reason. As noted above, those who reported that they planned to receive RT but had yet to start were excluded from analysis, so that the self-reported measure of RT receipt in this study was considered positive only for patients who reported already receiving radiation treatment.

We determined the final surgical procedure by asking about the initial surgical procedure after biopsy and whether subsequent procedures were performed. We also assessed age, race/ethnicity, comorbidities, insurance



**Figure 1.** Source of analyzed sample is shown. This flow diagram details the way in which the analytic sample was developed. MD, medical doctor; RT, radiotherapy; SEER, Surveillance, Epidemiology, and End Results.

status at time of diagnosis, total household yearly income at time of diagnosis, and educational attainment through separate survey questions. For age and race/ethnicity, we used SEER data for the few patients (<1%) missing data by self-report. We used SEER data for clinical information on tumor size and nodal status and to identify hospital of diagnosis, which we then categorized based upon whether that hospital was accredited by the American College of Surgeons.

To determine the RT receipt status in SEER registry data, we used the radsum variable, which is used to indicate any receipt of RT as part of initial therapy in the SEER database. Those whom SEER coded as 0 (none) or 7 (refused) were coded as not receiving RT; those whom SEER assigned codes 2 through 6 (codes for various modalities of RT) were coded as receiving RT; the few who were coded as 9 (unknown) and 8 (recommended; unknown if given) were excluded from analysis as noted above.

We defined underascertainment as patient report of RT receipt among patients coded in SEER as not having received RT.

### Analysis

We compared self-reported data on RT receipt to RT as reported in SEER registry records. We described the fre-

**Table 1.** SEER Registry Data Compared With Patient Self-Report of RT Receipt

SEER Status	Patient Reports RT Receipt	Patient Reports No RT Receipt
SEER Records RT Receipt	1019 (79%)	47 (7%)
SEER Records No RT Receipt	273 (21%)	588 (93%)
Total	1292 (100%)	635 (100%)

Abbreviations: RT, radiotherapy; SEER, Surveillance, Epidemiology, and End Results.

quency of RT underascertainment at each SEER site. We then calculated rates of underascertainment at each site after grouping patients by clinical and sociodemographic characteristics, as well as by treatment and hospital characteristics.

We performed univariate analyses using chi-square testing. We then regressed underascertainment within each SEER site on stage, age, surgery type, race/ethnicity, income, insurance, chemotherapy receipt, self-reported delay of RT initiation, and American College of Surgeons accreditation status of the diagnosing hospital as independent variables, adjusting for clustering by hospital. We evaluated all first-order interactions between significant variables, and none was significant except as reported. All results were weighted to account for the sampling design and differential nonresponse. Results are presented as unweighted values, with weighted percentages.

## RESULTS

Table 1 compares the RT receipt code in SEER to patient self-report. Of the 1292 patients who reported receiving radiation, 273 were coded as not receiving RT in SEER (underascertained). RT underascertainment was much more common in Los Angeles than in Detroit (32.0% vs 11.25%,  $P < .001$ ).

In Los Angeles, RT underascertainment was more frequent in patients with higher stage, multiple comorbidities, younger age, lower income, underinsurance, mastectomy receipt, chemotherapy receipt, delayed initiation of RT, and diagnosis at an American College of Surgeons unaccredited hospital (Table 2). In Detroit, patients with higher stage, mastectomy receipt, chemotherapy receipt, and diagnosis at an unaccredited hospital were associated with higher rates of RT underascertainment (Table 3).

On multivariate analysis, as shown in Table 4, RT underascertainment was significantly associated in both registries ( $P$  values for Los Angeles, Detroit) with stage ( $P = .008$ ,  $P = .026$ ), income ( $P < .001$ ,  $P = .050$ ),

**Table 2.** Characteristics of Patients Who Reported RT Receipt and Rates of Underascertainment of RT by the Los Angeles Surveillance, Epidemiology, and End Results Registry

Characteristic	No.	Weighted %	% Underascertained <sup>a</sup>	P <sup>b</sup>
<b>Stage</b>				<.001
0 (DCIS)	135	22	26	
I	290	48	26	
II	139	21	42	
III	70	9	47	
<b>Comorbidity</b>				.001
None	257	41	31	
1	198	33	28	
2+	180	26	36	
<b>Age, years</b>				<.001
<50	165	22	43	
50-64	263	43	32	
65+	207	34	23	
<b>Race</b>				.18
Black	162	12	33	
White	184	65	29	
Hispanic	283	21	33	
<b>Income</b>				<.001
<\$20,000	130	13	34	
\$20,000-\$69,999	234	37	38	
\$70,000+	129	30	27	
Unknown	142	21	24	
<b>Insurance</b>				<.001
None	58	5	51	
Medicaid	64	7	35	
Medicare	149	26	27	
Other	321	62	33	
Unknown	58	5	51	
<b>Surgery type</b>				<.001
Breast conservation	548	87	28	
Mastectomy	87	13	51	
<b>Receipt of chemotherapy</b>				<.001
Yes	260	35	45	
No	369	65	24	
<b>Delay in initiating RT</b>				<.001
Yes	104	18	47	
No	486	82	28	
<b>ACoS accreditation of diagnosing hospital</b>				<.001
Yes	234	41	27	
No	387	59	33	

Abbreviations: ACoS, American College of Surgeons; DCIS, ductal carcinoma in situ; RT, radiotherapy.

<sup>a</sup>Percentage underascertained calculated within the weighted sample.

<sup>b</sup>P values for differences in the proportion of RT receipt by the categories shown; separate category included for "unknown" when unknown values exceeded 5% (income).

mastectomy receipt ( $P < .001$ ,  $P < .001$ ), chemotherapy receipt ( $P < .001$ ,  $P = .045$ ), and diagnosis at a hospital that was not American College of Surgeons accredited ( $P < .001$ ,  $P < .001$ ). In Los Angeles, additional significant variables included younger age ( $P < .001$ ), nonprivate insurance ( $P < .001$ ), and delayed receipt of RT ( $P < .001$ ).

## DISCUSSION

The 2 SEER registries included in our study differed substantially in both rates and correlates of RT underascertainment. RT underascertainment in Los Angeles was nearly 3× higher than in Detroit and was associated with age, insurance coverage, and delayed initiation of RT in

**Table 3.** Characteristics of Patients Who Reported RT Receipt and Rates of Underascertainment of RT by the Detroit SEER Registry

Characteristic	No.	Weighted %	% Underascertained <sup>a</sup>	P <sup>b</sup>
<b>Stage</b>				.02
0 (DCIS)	142	21	8	
I	251	40	8	
II	189	29	13	
III	68	11	14	
<b>Comorbidity</b>				.14
None	245	39	13	
1	202	31	8	
2+	210	31	11	
<b>Age, years</b>				.06
<50	162	25	14	
50-64	309	47	10	
65+	186	28	9	
<b>Race</b>				.23
Black	178	19	8	
White	446	76	12	
Other or unknown	33	6	6	
<b>Income</b>				.53
<\$20,000	87	12	11	
\$20,000-\$69,999	266	40	12	
\$70,000+	198	32	9	
Unknown	106	16	11	
<b>Insurance</b>				.50
None	10	2	0	
Medicaid	36	5	10	
Medicare	182	28	11	
Other	420	66	11	
Unknown	10	2		
<b>Surgery type</b>				<.001
Breast conservation	539	83	8	
Mastectomy	111	17	25	
<b>Receipt of chemotherapy</b>				<.001
Yes	310	50	15	
No	308	50	7	
<b>Delay in initiating RT</b>				.91
Yes	113	18	11	
No	507	82	11	
<b>ACoS accreditation of diagnosing hospital</b>				<.001
Yes	424	68	9	
No	201	32	16	

Abbreviations: ACoS, American College of Surgeons; DCIS, ductal carcinoma in situ; RT, radiotherapy.

<sup>a</sup>Percentage underascertained calculated within the weighted sample.

<sup>b</sup>P values for differences in the proportion of RT receipt by the categories shown; separate category included for "unknown" when unknown values exceeded 5% (income).

addition to variables that were significant at both locations (stage, income, mastectomy receipt, chemotherapy receipt, and hospital accreditation status). These results suggest that SEER registry data, collected by routine methods, may not be an appropriate source for docu-

menting rates of RT receipt by breast cancer patients or for investigating geographic variation in RT receipt.

SEER data alone have long been used to evaluate the appropriateness of breast cancer treatment, including RT receipt. Nearly 2 decades ago, a seminal analysis of SEER

**Table 4.** Logistic Regression Models of RT Underascertainment

Characteristic	Los Angeles			<i>P</i> <sup>a</sup>	Detroit			<i>P</i> <sup>a</sup>
	Odds Ratio	95% Confidence Interval			Odds Ratio	95% Confidence Interval		
<b>Stage</b>				.008				.026
0	1.00				1.00			
I	0.82	0.64	1.03		0.67	0.35	1.28	
II	1.32	0.94	1.83		0.69	0.30	1.56	
III	0.76	0.49	1.21		0.26	0.09	0.70	
<b>Age, years</b>				<.001				.36
<50	3.01	2.09	4.31		1.34	0.54	3.30	
50-64	1.91	1.38	2.66		0.95	0.42	2.16	
65+	1.00				1.00			
<b>Race</b>				.14				.13
White	1.00				1.00			
Black	0.84	0.63	1.12		0.57	0.32	1.04	
Hispanic <sup>b</sup>	0.78	0.61	0.99					
Other					0.59	0.2	1.71	
<b>Income</b>				<.001				.050
<\$20,000	1.12	0.79	1.60		1.15	0.48	2.71	
\$20,000-\$69,999	1.66	1.31	2.11		1.96	1.19	3.24	
\$70,000+	1.00				1.00			
Unknown	1.02	0.76	1.37		1.42	0.72	2.78	
<b>Insurance</b>				<.001				.87
None	1.94	1.23	3.05		<0.001	<0.001	>999	
Medicaid	1.26	0.86	1.87		0.82	0.31	2.17	
Medicare	2.19	1.60	3.00		0.70	0.30	1.67	
Other	1.00				1.00			
<b>Surgery type</b>				<.001				<.001
Breast conservation	1.00				1.00			
Mastectomy	2.07	1.50	2.85		4.93	2.84	8.57	
<b>Chemotherapy receipt</b>				<.001				.045
Yes	1.84	1.43	2.37		1.86	1.02	3.41	
No	1.00				1.00			
<b>Delay initiating RT</b>				<.001				.15
Yes	1.93	1.52	2.44		0.66	0.38	1.16	
No	1.00				1.00			
<b>Diagnosed at ACoS accredited hospital</b>				<.001				<.001
Yes	1.00				1.00			
No	1.30	1.10	1.55		2.05	1.40	3.00	

Abbreviations: ACoS, American College of Surgeons; RT, radiotherapy.

<sup>a</sup>*P* values for group variables are reported from Wald tests; standard errors were adjusted for hospital clustering.

<sup>b</sup>There were too few Hispanic patients in Detroit to support a separate category. Thus, Hispanics in Detroit were included in the "other" race category.

data from 1983 to 1986 documented rates of RT receipt after BCS that varied significantly by geography, race, and age.<sup>3</sup> Another landmark study of SEER data from 1983 to 1995<sup>4</sup> found a decrease in the use of appropriate primary therapy for early stage breast cancer over time (with only 78% of women receiving appropriate primary therapy in 1995), driven by an apparent increase in use of BCS that was not followed by RT or axillary surgery.

However, several studies have since raised questions about the completeness of registry data, especially for treatments such as RT, which are often delivered in the outpatient setting. Bickell and Chassin took data collected as part of a quality improvement project on 365 cases of stage I-II breast cancer diagnosed between 1994 and 1996 at 3 New York hospitals and compared them with data in the hospitals' tumor registries, finding that only 58% of

RT was captured by the registries.<sup>6</sup> Malin et al compared California Cancer Registry data with data abstracted from medical records of 304 patients in the PacificCare of California health plan who were diagnosed with breast cancer from 1993 to 1995 in Los Angeles; they found that only 72.2% of RT was captured by the registry.<sup>7</sup> Given the sample size and the finding that studied patients were older and less diverse than the Los Angeles population, the study's ability to detect sociodemographic differences was limited. Systematic differences in ascertainment by disease stage were observed, however, and the authors noted that patients with more advanced disease more often received treatment in the ambulatory setting that was less likely to be reported to the registry.

Despite these studies, researchers have continued to use SEER data for evaluation of RT receipt. For example, a study published this year used SEER data to study rates of RT receipt among patients with locally advanced breast cancer, by race and surgery type.<sup>8</sup> The authors concluded that "rates of RT were low for all populations"; although they considered several possible explanations for this finding, they did not mention the possibility of RT underascertainment. They did state, "We considered RT use as a single surrogate marker of quality cancer care, but there are certainly others. Rates of breast reconstruction and adherence to hormonal or systemic therapy guidelines are all potential surrogate markers of quality cancer care, but these data fields are either limited or unavailable in the SEER database." However, they did not consider the possibility that the RT field in the SEER database might also have limitations. Ironically, in their introduction, the authors cited Malin's study, but only in support of the statement, "Adherence to RT guidelines improves overall and disease-specific survival and has been used as a surrogate marker of quality BCa care."

Other recent studies have acknowledged concerns about limitations of SEER registry data but have dismissed these largely based upon comparisons to merged SEER-Medicare data. For example, when Du and Gor published an analysis of RT receipt based on SEER data from 1992 to 2002,<sup>10</sup> they referenced a study using merged SEER-Medicare data on women aged 65 to 74 years diagnosed with breast cancer in 1992, finding that among 2784 women whom SEER recorded as not receiving RT, Medicare identified only 377 (13.5%) as receiving RT.<sup>17</sup> They also referenced a study<sup>18</sup> of SEER-Medicare data from 1991 to 1996 in patients 65 and older that found 94% agreement between SEER and SEER-Medicare for RT receipt in breast cancer patients, with

only minimal variation between individual SEER registries. Similarly, when Freedman and colleagues conducted a study using SEER data from 1988 to 2004, they included an appendix assessing RT underascertainment using the SEER-Medicare dataset from 1992 to 2002.<sup>9</sup> They found 91% agreement between the 2 data sources and concluded that "it is unlikely that our findings would be explained by problems ascertaining radiation therapy by the SEER registries." Sufficiently reassured to consider SEER data alone, they found a decrease in RT after BCS from 79% in 1988 to 66% in 2004, with differences by race, SEER site, and age. These rates are markedly lower than those reported by our patients in Los Angeles and Detroit in 2006.<sup>19</sup>

More recently, Dragun and colleagues published an interesting analysis that documented a 66% overall rate of radiation receipt, and significantly lower receipt in rural/Appalachian populations, in the Kentucky Cancer Registry (KCR).<sup>20</sup> These researchers discussed potential limitations in registry data but noted that the KCR "has been awarded the highest level of certification by the North American Association of Central Cancer Registries for an objective evaluation of completeness, accuracy, and timeliness every year since 1997. The KCR is also part of the . . . SEER program, which has the most accurate and complete population-based cancer registry in the world." Unfortunately, North American Association of Central Cancer Registries accreditation does not consider accuracy of coding of treatment receipt, including radiation receipt; it only confirms that there is accuracy, completeness, and timeliness with respect to identification of incident cases of cancer and demographic characteristics.<sup>21</sup> Thus, although the study findings may well reflect a true problem with undertreatment in settings where health care facilities are more limited, one must exercise caution in drawing firm conclusions unless treatment information in the Kentucky registry has been validated in ways not discussed in that article, as RT underascertainment may also be more likely in such settings.

Of note, the current study shows that RT underascertainment appears to be more frequent among younger patients not represented in SEER-Medicare comparisons and can occur even in SEER registries holding North American Association of Central Cancer Registries accreditation for high-quality incident case ascertainment. Moreover, more breast cancer patients receive chemotherapy before RT today than in the time periods of studies comparing SEER to SEER-Medicare data. Although the first course of treatment is no longer defined by SEER as a

4-month period from diagnosis, increased time between diagnosis and RT because of the administration of chemotherapy increases the difficulty for registrars to ascertain radiation receipt. In light of the findings of our current study, showing substantial RT underascertainment in 1 of the largest SEER registries, we believe that future studies should not use the SEER dataset alone to determine rates or correlates of RT receipt until the quality of the data in the other SEER registries are investigated more closely.

Of note, SEER itself recognizes the limitations in registry data collected by routine methods and so regularly also conducts Patterns of Care studies focused on different cancer sites.<sup>22,23</sup> These studies involve reabstraction of treatment information from hospital records and requests to physician offices to capture therapy administered in the outpatient setting. Additional analysis using these methods would be valuable to assess rates of RT underascertainment in other registry sites before further research relies upon SEER data alone to assess RT receipt. Certainly, the findings of the current study are sobering.

The substantial differences in rates of ascertainment by the 2 registries likely reflect differences in the methods of surveillance. In the Detroit registry, surveillance is active, and radiation facilities are surveyed as part of the process of incident case identification, which also allows for updating of the radiation receipt variable. In the Los Angeles registry, where surveillance depends upon reporting by registrars, it is not surprising that rates of RT underascertainment are higher. Reporting of treatment received in the outpatient setting is particularly difficult for registrars to capture. Moreover, in California, the state law that established the registry system does not require capture of treatment given outside the reporting facility. Thus, it is not surprising that RT underascertainment is strongly associated with hospital accreditation status, as American College of Surgeons-accredited cancer programs are required to capture all first-course treatments regardless of location, in contrast to the more basic requirements of state law that govern other institutions. The independent association of numerous clinical and sociodemographic variables with ascertainment likely reflects the way in which differences in these factors affect the timing of care and the type and/or number of facilities within which these patients receive medical care.

This study has several strengths, including its large sample size and diverse population. It also has several potential limitations. First, it relied upon self-report as the

gold standard to which SEER data were compared. Although we recognize that there is no true gold standard, previous studies have supported the validity of self-report regarding RT and have documented very high correlations between self-report and medical record review.<sup>24-26</sup> Criterion validity is supported by the finding that the overwhelming majority of patients who reported receiving radiation went on to respond that they had received information regarding management of RT side effects and reported receiving  $\geq 5$  weeks of treatment, as well as the finding that self-reported receipt of radiation was highly correlated with clinical factors that direct treatment recommendations.

Second, although this is the first study to our knowledge of RT underascertainment, other than the SEER-Medicare studies, to include  $>1$  geographic location, the study was limited to 2 metropolitan SEER registries. It is therefore not possible to comment definitively upon the rates of RT underascertainment in other registries, particularly more rural registries. Nevertheless, these data are sufficient to conclude that RT underascertainment is not uniform across SEER sites and can be quite substantial.

The call for comparative effectiveness research has led to heightened interest in the analysis of population-based registry data. As increasing numbers of researchers begin to use registry data, it is critical to evaluate the quality of those data. The SEER regional registry network represents a golden opportunity to continue to build population-based translational cancer research with real value to patients and their clinicians. Indeed, recent changes and additions to the content of SEER data reflect the increasing complexity of clinical information and treatment modalities for cancer and the interest of stakeholders in enhancing the use of these data for the purpose of assessing quality of care and outcomes.<sup>27</sup> However, increased demand for breadth and depth of data against decreasing budgets for its collection may be counterproductive. This study provides only 1 example of the ways in which misuse of poor quality data may lead to spurious policy conclusions. One increasingly common strategy to improve data quality is for registries to partner with investigators to use supplemental research funding for special studies in cancer outcomes and effectiveness research. Another strategy might be to allow and encourage regional SEER registries to subspecialize in the collection of the more challenging and resource-intensive data elements related to the first course of therapy. This would create regional registries of excellence with particular strengths in certain cancers or treatment modalities. The increasing



complexity of cancer care and increasing demand to evaluate it motivate researchers to find creative solutions to ensure the highest validity and quality of data collected by regional cancer registries.

## FUNDING SOURCES

This work was funded by grants R01 CA109696 and R01 CA088370 from the National Cancer Institute (NCI) to the University of Michigan. R.J. was supported by a Mentored Research Scholar Grant from the American Cancer Society (MRSO-09-145-01). S.J.K. was supported by an Established Investigator Award in Cancer Prevention, Control, Behavioral, and Population Sciences Research from the NCI (K05CA111340).

The collection of Los Angeles County cancer incidence data used in this study was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the NCI's Surveillance, Epidemiology, and End Results (SEER) program under contract N01-PC-35139 awarded to the University of Southern California; contract N01-PC-54404 awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement 1U58DP00807-01 awarded to the Public Health Institute. The collection of metropolitan Detroit cancer incidence data was supported by the NCI SEER Program contract N01-PC-35145.

The ideas and opinions expressed herein are those of the author(s), and endorsement by the State of California, Department of Public Health, National Cancer Institute, and Centers for Disease Control and Prevention or their contractors and subcontractors is not intended nor should be inferred.

## CONFLICT OF INTEREST DISCLOSURES

The authors made no disclosures.

## REFERENCES

- National Cancer Institute. Surveillance Epidemiology and End Results. About SEER. Available at: <http://seer.cancer.gov/about/>. Accessed on July 30, 2010.
- Yu JB, Gross CP, Wilson LD, Smith BD. NCI SEER public-use data: applications and limitations in oncology research. *Oncology*. 2009;23:288-295.
- Farrow DC, Hunt WC, Samet JM. Geographic variation in the treatment of localized breast cancer. *N Engl J Med*. 1992;326:1097-1101.
- Nattinger AB, Hoffman RG, Kneusel RT, Schapira MM. Relation between appropriateness of primary therapy for early-stage breast carcinoma and increased use of breast-conserving surgery. *Lancet*. 2000;356:1148-1153.
- SEER Program Coding and Staging Manual 2010. Available at: [http://seer.cancer.gov/manuals/2010/SPCSM\\_2010\\_maindoc.pdf](http://seer.cancer.gov/manuals/2010/SPCSM_2010_maindoc.pdf). Accessed February 11, 2011.
- Bickell NA, Chassin MR. Determining the quality of breast cancer care: do tumor registries measure up? *Ann Intern Med*. 2000;132:705-710.
- Malin JL, Kahn KL, Adams J, et al. Validity of cancer registry data for measuring the quality of breast cancer care. *J Nat Cancer Inst*. 2002;94:835-844.
- Martinez SR, Beal SH, Chen SL, et al. Disparities in the use of radiation therapy in patients with local-regionally advanced breast cancer. *Int J Radiat Oncol Biol Phys*. 2010;78:787-792.
- 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*. 2009;27:713-719.
- Du XL, Gor BJ. Racial disparities and trends in radiation therapy after breast-conserving surgery for early-stage breast cancer in women, 1992 to 2002. *Ethn Dis*. 2007;17:122-128.
- Morrow M, Jagsi R, Alderman A, et al. Surgeon recommendations and receipt of mastectomy for treatment of breast cancer. *JAMA*. 2009;302:1551-1556.
- Hawley ST, Griggs JJ, Hamilton AS, et al. Decision involvement and receipt of mastectomy among racially and ethnically diverse breast cancer patients. *J Natl Cancer Inst*. 2009;101:1-11.
- Hamilton AS, Hofer TP, Hawley ST, et al. Latinas and breast cancer outcomes: population-based sampling, ethnic identity, and acculturation assessment. *Cancer Epidemiol Biomarkers Prev*. 2009;18:2022-2029.
- Greene FL, Page DL, Fleming ID, et al. AJCC Cancer Staging Manual (ed 6). Philadelphia, PA: Lippincott Raven Publishers; 2002.
- Marin G, Van Oss Marin B. Research With Hispanic Populations. Newbury Park, CA: Sage Publications; 1991.
- Dillman DA: Mail and Telephone Surveys: The Total Design Method. New York, NY: John Wiley and Sons; 1997.
- Du X, Freeman JL, Goodwin JS. Information on radiation treatment in patients with breast cancer: the advantages of the linked Medicare and SEER data. *J Clin Epidemiol*. 1999;52:463-470.
- Virnig B, Warren JL, Cooper GS, et al. Studying radiation therapy using SEER-Medicare linked data. *Med Care*. 2002;40(8 suppl):IV-49-IV-54.
- Jagsi R, Abrahamse P, Morrow M, et al. Patterns and correlates of adjuvant radiotherapy receipt after lumpectomy and after mastectomy for breast cancer. *J Clin Oncol*. 2010;28:2396-2403.
- Dragun AE, Huang B, Tucker TC, Spanos WJ. Disparities in the application of adjuvant radiotherapy after breast-conserving surgery for early stage breast cancer. *Cancer*. 2011;117:2590-2598.
- North American Association of Central Cancer Registries. Certification Criteria. Available at: <http://www.naacr.org/Certification/Criteria.aspx>. Accessed on February 10, 2011.
- Hamilton AS, Albertson PC, Johnson, TK et al. Trends in treatment of localized prostate cancer using supplemented cancer registry data. *BJU Int*. 2011;107:576-584.
- Harlan LC, Abrams J, Warren JL, et al. Adjuvant therapy for breast cancer: practice patterns of community physicians. *J Clin Oncol*. 2002;20:1809-1817.
- Liu Y, Diamant A, Thind A, Maly RC. Validity of self-reports of breast cancer treatment in low-income, medically underserved women with breast cancer. *Breast Cancer Res Treat*. 2010;119:745-751.
- Maunsell E, Drolet M, Ouhoumane N, Robert J. Breast cancer survivors accurately reported key treatment and prognostic characteristics. *J Clin Epidemiol*. 2005;58:364-369.
- Phillips KA, Milne RL, Buys S, et al. Agreement between self-reported breast cancer treatment and medical records in a population-based breast cancer family registry. *J Clin Oncol*. 2005;23:4679-4686.
- Thornton M, ed. Standards for Cancer Registries, Volume II: Data Standards and Data Dictionary, Record Layout Version 12.1. 15th ed. Springfield, IL: North American Association of Central Cancer Registries; 2010.