

Evaluation of Three Algorithms to Identify Incident Breast Cancer in Medicare Claims Data

Heather T. Gold and Huong T. Do

Objective. To test the validity of three published algorithms designed to identify incident breast cancer cases using recent inpatient, outpatient, and physician insurance claims data.

Data. The Surveillance, Epidemiology, and End Results (SEER) registry data linked with Medicare physician, hospital, and outpatient claims data for breast cancer cases diagnosed from 1995 to 1998 and a 5 percent control sample of Medicare beneficiaries in SEER areas.

Study Design. We evaluate the sensitivity and specificity of three algorithms applied to new data compared with original reported results. Algorithms use health insurance diagnosis and procedure claims codes to classify breast cancer cases, with SEER as the reference standard. We compare algorithms by age, stage, race, and SEER region, and explore via logistic regression whether adding demographic variables improves algorithm performance.

Principal Findings. The sensitivity of two of three algorithms is significantly lower when applied to newer data, compared with sensitivity calculated during algorithm development (59 and 77.4 percent versus 90 and 80.2 percent, $p < .00001$). Sensitivity decreases as age increases, and false negative rates are higher for cases with in situ, metastatic, and unknown stage disease compared with localized or regional breast cancer. Substantial variation also exists by SEER registry. There was potential for improvement in algorithm performance when adding age, region, and race to an indicator variable for whether the algorithm determined a subject to be a breast cancer case ($p < .00001$).

Conclusions. Differential sensitivity of the algorithms by SEER region and age likely reflects variation in practice patterns, because the algorithms rely on administrative procedure codes. Depending on the algorithm, 3–5 percent of subjects overall are misclassified in 1998. Misclassification disproportionately affects older women and those diagnosed with in situ, metastatic, or unknown-stage disease. Algorithms should be applied cautiously to insurance claims databases to assess health care utilization outside SEER-Medicare populations because of uneven misclassification of subgroups that may be understudied already.

Key Words. Breast neoplasm, incidence, algorithm validation, registries, Medicare

Researchers studying the quality of cancer care in the United States have noted disparities by geography, race/ethnicity, and socioeconomic status. Prior studies to examine these differences have relied on large secondary databases and chart abstraction (Wennberg et al. 1987; Nattinger and Goodwin 1994; Harlan et al. 1995; Ayanian and Guadagnoli 1996; Michalski and Nattinger 1997; Earle et al. 2002; Smedley, Stith, and Nelson 2003; Gilligan 2005; Neuss et al. 2005). The disadvantages of these data sources are that chart abstraction is costly and time-consuming, and large administrative databases often have limited generalizability, making it expensive or difficult to analyze national patterns of care. The National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) population-based cancer registry is considered the reference standard for cancer case ascertainment in the United States, and collects limited treatment information. Linking these data with Medicare claims data further restricts the available study subjects to those ages 65 and older (Potosky et al. 1993). Researchers attempting to obtain data on a broader age and geographic range of subjects are often limited to data sets covering a single state (Ayanian et al. 1993; McClish et al. 1997; Hodgson et al. 2003; McClish, Penberthy, and Pugh 2003; McClish and Penberthy 2004; Penberthy et al. 2005), smaller, localized populations (Elston et al. 2005), or areas covered by passive surveillance systems, which may have lower rates of case ascertainment or incomplete data (Brewster et al. 1997; Yoo et al. 2002; Greenberg et al. 2003; Wang et al. 2005). Owing to these limitations, several researchers (Warren et al. 1999; Freeman et al. 2000; Nattinger et al. 2004; Ramsey et al. 2004) have developed algorithms to identify cancer cases using Medicare claims data to determine if broad cancer incidence and patterns-of-care studies can be performed solely with administrative claims data sources. Reliable methods to identify incident breast cancer using administrative data would permit the study of patterns and quality of care without the time and cost requirements of chart abstraction or linkage of claims and cancer registry data sources, and allow researchers to study populations not covered by existing surveillance systems. Health insurance claims data are available across the United States wherever health insurance is used. Thus, an effective algorithm would allow study of patterns and costs of care in larger and more diverse insured populations, including subjects under age 65, members of

Address correspondence to Heather Taffet Gold, Ph.D., Department of Public Health, Weill Medical College of Cornell University, 411 E, 69th Street, New York, NY 10021. Huong T. Do, M.A., is with the Department of Public Health, Weill Medical College of Cornell University, New York, NY.

health maintenance organizations (HMOs), and those in previously understudied racial/ethnic populations or regions.

For this study, we use the linked SEER-Medicare data to evaluate three published algorithms designed to identify incident breast cancer cases using inpatient, outpatient, and physician insurance claims data. We assess algorithm validity based on more recent claims data by population subgroup (i.e., by age, race, stage, and region). We implement these algorithms and compare them based on standard diagnostic characteristics, including Receiver-Operating Characteristic (ROC) curve analysis, sensitivity, and specificity.

METHODS

Data

We obtained hospital inpatient, outpatient, and physician claims for all breast cancer cases identified in nine SEER registry regions and a 5 percent "control" sample of Medicare beneficiaries in the same SEER areas without breast cancer (but who may have other cancer types) from the linked SEER-Medicare database. The data also include demographic and Medicare entitlement information on all subjects and diagnosis and treatment information for all breast cancer cases. The reference standard for case identification is the SEER registry, which ascertains a very high rate of cases via hospital, physician, and laboratory reporting and death certificates. The current 17 SEER registries capture 98 percent of cases within the registry areas and maintain a 95 percent followup rate on reported cases (Surveillance Implementation Group 1999). These registries currently represent 26 percent of the U.S. population and tend to be more urban and higher socioeconomic areas (Nattinger, McAuliffe, and Schapira 1997).

Subjects included in this study are women residing in the first nine registry areas of the SEER program during any year from 1995 to 1998 who are 65 years or older as of January of the index year and alive for the entire index year. Depending on the algorithm, patients who were ever members of an HMO or were not continuously enrolled in Medicare Parts A and B for either (1) the entire calendar year (criterion A) or (2) the calendar year plus the first 3 months of the following year (criterion B) were excluded because their Medicare claims records likely would not capture all of their health care utilization. (The number of cases excluded ranged from 44 to 68 depending on the year.) Each year of data is analyzed independently, resulting in sample sizes of 66,183–73,995, depending on the year and the algorithm inclusion

criteria. Incident breast cancer cases account for approximately 12 percent of the sample subjects in each year.

Algorithms

Each of the three algorithms (Warren et al. 1999; Freeman et al. 2000; Nattinger et al. 2004) uses a different combination of diagnosis and procedure codes to identify incident cases and exclude prevalent cases. Freeman et al. (2000) used 1990–1992 data from the linked SEER-Medicare database to determine which breast cancer diagnosis and procedure codes are predictors of incident breast cancer in 1992. Their sample included inpatient, outpatient, and physician claims for breast cancer patients and a 5 percent sample of noncancer controls in the nine SEER registry areas who were ages 65–74 in 1992 and not excluded using criterion A. They used a logistic regression model with an outcome variable set to 1 if the subject was a SEER-identified incident case and 0 if the subject was a control, and independent indicator variables for the presence of 36 breast cancer diagnosis and procedure codes. They then used these predictor variables in four different combinations, and used the coefficients from the models to calculate the probability that a subject was a breast cancer case. They evaluated the sensitivity and specificity of their models at different probability cutpoints and estimated an ROC curve and the area under the ROC curve (AUC). We only evaluate their best model, Model 4, which includes the 19 significant predictor variables.

The second algorithm, developed by Nattinger et al. (2004), also used the linked SEER-Medicare data, although they used 1995–1996 data to identify 1995 incident breast cancer cases. The subjects in their study were women ages 65 or older who were not excluded under criterion B. Nattinger et al. applied a combination of clinical insight and statistical analysis to create a four-part algorithm. The first step requires a potential case to have a breast cancer diagnosis *and* procedure code (which do not have to be on the same claim) in the inpatient, outpatient, or physician claims. If this was met, the second step requires that the potential case have both (1) a mastectomy claim or a lumpectomy/partial mastectomy claim with a claim for radiotherapy with a breast cancer diagnosis, and (2) at least two outpatient or physician claims on different dates with a primary diagnosis of breast cancer. If step 2 is not passed, subjects are entered into a logistic regression derived criterion (step 3), which requires the patient to meet one of four combinations of breast cancer-related billing codes to be classified as a case. If the subject passes step 2 or 3, she goes to step 4 to rule out prevalent cancer cases using the 3 prior years of claims

data. Nattinger's algorithm separately applies two reference standards: SEER and SEER plus those cases passing step 2. In our analysis, we apply only the model that uses SEER alone as the reference standard.

The third and final algorithm tested here was developed by Warren et al. (1999) using 1992 hospital and physician claims data for all Medicare eligible women residing in one of five SEER state registries (Connecticut, Hawaii, Iowa, New Mexico, and Utah) who were age 65 or older as of January 1, 1992 and were not excluded under criterion A. The authors identified the women from this sample who were linked to the SEER registry with incident breast cancer in 1992, excluding as prevalent cases women who had a breast cancer diagnosis code or history of breast cancer in any claim from previous years. Two models were developed, the first using only breast cancer diagnosis codes to classify cases and the second using diagnosis and procedure codes. Although the authors show that procedure codes used in the second model are significant predictors of incident breast cancer, values for model sensitivity and specificity are only provided for Model 1, which is what we use for comparison.

Analytic Methods

Applying each of these three algorithms to the linked SEER-Medicare data for each year from 1995 to 1998, we calculate the sensitivity, specificity, and misclassification rates. We assess how well the algorithms predict breast cancer incidence in our data based on age, stage, race, and geography (i.e., SEER region) using a one-sample test of proportions. Misclassification rates are calculated for cases by adding false negatives and false positives and dividing the sum by the sample size. In addition, we evaluate the AUC for the Freeman model to identify if the model achieves >90 percent sensitivity and specificity at any probability cutpoint as stated in the original article (Freeman et al. 2000). Finally, we explore via logistic regression and using the likelihood ratio test whether adding demographic variables to each algorithm improves algorithm predictive value, because demographic variables may add to the ability of procedure and diagnosis codes to identify new cancer cases. All analyses are conducted using *Stata* (versions 8.2 and 9.1, College Station, TX) and algorithms are implemented in *SAS* (version 9.1, Cary, NC).

RESULTS

The data we use are for more recent years, 1995–1998, compared with the data used in the published algorithms (Table 1). Our total sample is smaller

Table 1: Sample Size by Algorithm

<i>Data Year</i>	<i>Total Sample Size</i>			<i>Number of Cases</i>		
	<i>Nattinger</i>	<i>Warren</i>	<i>Freeman</i>	<i>Nattinger</i>	<i>Warren</i>	<i>Freeman</i>
1995	71,839	73,995	73,995	8,391	8,746	8,746
1996	70,202	72,422	72,422	8,197	8,561	8,561
1997	67,854	70,346	70,346	8,397	8,785	8,785
1998	66,183	68,220	68,220	8,335	8,699	8,699
Reported sample size	132,584	659,260	47,560	7,700	3,230	3,339

Note: Sampling criteria are the same for the Warren and Freeman algorithms, thereby yielding the same sample size for analyses of both.

compared with two of the algorithms’ reported sample sizes, although our number and percentage of cases is substantially higher than all three algorithms’ data sets.

Sensitivity of two of three algorithms applied to our data is significantly lower at 59 and 77.4 percent, compared with the sensitivity obtained by the algorithm developers, 90 and 80.2 percent, respectively (Table 2). Substantial variation exists in sensitivity and specificity by age and SEER region. Sensitivity decreases as age increases (Table 2). False negative rates are higher for cases with in situ, metastatic, and unknown stage disease compared with localized or regional breast cancer (Table 3). Overall misclassification ranges from 2.5 to 5.2 percent. (Data not shown.) There also is substantial variation by SEER registry. For example, Warren’s algorithm applied to 1998 data yields a sensitivity of 70.4 percent (confidence interval [CI]: 68.0–72.7 percent) in the Detroit registry, 74.1 percent (CI: 71.6–76.5 percent) in the Connecticut registry, and 77.5 percent (CI: 75.1–79.7 percent) in the Iowa registry. The number of false positives is very small by year in smaller registries, making any inference difficult. Differences by race are insignificant, also possibly due to small sample size. The overall variation in specificity is statistically significant, but the impact is minimal in terms of misclassification bias.

Positive predictive value (PPV), the probability a subject is a true case given the algorithm is positive, was 82.6 percent (CI: 78.3–86.3 percent) for Nattinger’s algorithm, 47.2 percent (CI: 44.2–50.3 percent) for Warren’s, and 93.2 percent (CI: 88.8–95.9 percent) for Freeman’s when applied to 1995 data. Only the Warren’s algorithm had a significant change in PPV by 1998, when the algorithm improved to 56.5 percent (CI: 52.7–60.1 percent). PPV also varied by race, age, and region.

Table 2: Sensitivity and Specificity of Algorithms

Algorithm Source	Sensitivity (%)			Specificity (%)			
	Reported	1995	1998	Reported	1995	1998	
Nattinger	80.11–80.26	79.6*	77.4*	99.95	99.9*	99.9*	
Warren	62.0	76*	73.7*	99.9	99.6*	99.7*	
Freeman	90.0	58.7*	59.0*	99.86	100*	100*	
	<i>Ages</i> 65–69	<i>Ages</i> 70–74	<i>Ages</i> 75–79	<i>Ages</i> 80+	<i>Ages</i> 65–69	<i>Ages</i> 70–74	<i>Ages</i> 75–79
By age group, using 1998 data							
Nattinger	81.6*	77.7*	76.6*	74.4*	99.9 [†]	99.9 [†]	99.9 [†]
Warren	78.6*	74.5*	71.7*	70.8*	99.7*	99.6*	99.7*
Freeman	63.2*	62.6*	59.6*	51.4*	100 [‡]	100 [‡]	100 [‡]
	<i>White</i>	<i>Black</i>	<i>Asian</i>	<i>White</i>	<i>Black</i>	<i>Asian</i>	
By race, using 1998 data							
Nattinger	77*	80.5	85.7*	99.9*	100 [‡]	100 [‡]	
Warren	73.7*	72.2*	79*	99.7*	99.7*	99.7 [†]	
Freeman	58.7*	62.2*	64.4*	100 [‡]	100 [‡]	100 [‡]	

* $p < .00001$ for equality of proportions compared with originally published value (for all subjects).

[†] $p < .007$ for equality of proportions compared with originally published value (for all subjects).

[‡]Unable to test for equality of proportions due to estimated specificity of 100%.

Note: Readers may want to divide the p -values by the number of comparisons to address the multiplicity of outcomes in this study.

There was a significant improvement in identifying cases using a multivariate model with an indicator variable for whether the algorithm determined a subject to be a breast cancer case and variables for age, region, and race ($p < .00001$), and AUC improved, as well (Table 4). However, the indicator variable of whether a subject is a breast cancer case according to the

Table 3: Number and Percentage of Cases Incorrectly Classified by Algorithm as Not Being a Case, by Algorithm and Stage, 1998 Data

	<i>In Situ</i>	<i>Localized</i>	<i>Regional</i>	<i>Metastatic</i>	<i>Unknown</i>
Nattinger	300 (23.6%)	935 (19.0%)	297 (18.4%)	204 (59.6%)	144 (74.6%)
Warren	349 (27.1%)	1,228 (24.4%)	384 (22.7%)	185 (43.0%)	139 (57.0%)
Freeman	528 (40.9%)	1,903 (37.8%)	665 (39.3%)	274 (63.7%)	200 (82.0%)

*Note: It is impossible to calculate the percentage of noncases that were false negatives by stage, because there is no stage information for the true negatives.

Table 4: Area under the ROC Curve and Model Fit Using Logistic Regression, 1998 Data

<i>Variable</i>	<i>Odds Ratio</i>	<i>95% Confidence Interval</i>		<i>p-Value</i>
Nattinger model				
Simple model				
Indicator from algorithm	4,222.56	3,157.84	5,646.27	<i>p</i> < .0001
AUC	0.89			
Multivariate model*				
Indicator from algorithm	4,487.05	3,348.85	6,012.10	<i>p</i> < .0001
Age 70–74	1.32	1.15	1.52	<i>p</i> < .0001
Age 75–79	1.36	1.18	1.57	<i>p</i> < .0001
Age 80+	1.22	1.06	1.40	<i>p</i> < .0001
Black	0.71	0.58	0.88	<i>p</i> < .0001
Asian	0.33	0.23	0.48	<i>p</i> < .0001
Other race	0.54	0.41	0.73	<i>p</i> < .0001
AUC	0.92			
LR test	1,277.71			<i>p</i> < .00001
Warren model				
Simple model				
Indicator from algorithm	979.99	836.82	1,147.7	<i>p</i> < .0001
AUC	0.87			
Multivariate model†				
Indicator from algorithm	1,038.36	884.15	1,219.5	<i>p</i> < .0001
Age 70–74	1.30	1.15	1.47	<i>p</i> < .0001
Age 75–79	1.41	1.24	1.59	<i>p</i> < .0001
Age 80+	1.21	1.07	1.36	<i>p</i> < .0001
Black	0.87	0.73	1.04	<i>p</i> = .12
Asian	0.39	0.29	0.54	<i>p</i> < .0001
Other race	0.59	0.46	0.77	<i>p</i> < .0001
AUC	0.91			
LR test	1,533.9			<i>p</i> < .00001
Freeman model				
Simple model				
Indicator from algorithm	6,576.53	3,812.08	11,345.69	<i>p</i> < .0001
AUC	0.79			
Multivariate model‡				
Indicator from algorithm	7,524.63	4,358.27	12,991.41	<i>p</i> < .0001
Age 70–74	1.14	1.03	1.27	<i>p</i> = .012
Age 75–79	1.19	1.07	1.33	<i>p</i> < .0001
Age 80+	1.18	1.07	1.30	<i>p</i> < .0001
Black	0.79	0.67	0.92	<i>p</i> < .0001
Asian	0.41	0.32	0.54	<i>p</i> < .0001
Other race	0.56	0.45	0.70	<i>p</i> < .0001
AUC	0.86			
LR test	2,632.29			<i>p</i> < .00001

Notes: LR test refers to the likelihood ratio test, which uses a χ^2 test to identify if the Simple Model is statistically significantly different from the Multivariate Model in which it is nested. AUC is the area under the receiver operating characteristic (ROC) curve, as calculated from the logistic model.

*Indicator variables for 6/10 registries are significant at *p* < .04.

†Indicator variables for 7/10 registries are significant at *p* < .05.

‡Indicator variables for 7/10 registries are significant at *p* < .001.

algorithm has the most quantitative impact, as the odds ratio of 4,487.05 with the Nattinger's algorithm, for example, is three orders of magnitude larger than the odds ratios for any of the demographic variables. All covariates were significant in the model except some region effects and black race in the Warren-algorithm model.

Finally, in our assessment of the algorithm by Freeman et al., we identified whether there existed a probability cutpoint that would yield a sensitivity and specificity of >90 percent simultaneously, the criterion the authors used to determine their cutpoint, and we did not find such a probability. The point on the ROC curve that yields the highest simultaneous sensitivity and specificity is when the probability cutpoint is .00588, yielding a sensitivity of 87.71 percent and a specificity of 87.74 percent.

DISCUSSION

The purpose of this project was to see how well published algorithms identify breast cancer cases in more recent claims data overall and by population subgroup (i.e., by age, race, stage, and region). Algorithm sensitivity is lower for the 1998 data compared with the 1995 data, indicating that published algorithms may need to be updated due to changing patient characteristics or patterns of care. Differential sensitivity of the algorithms by SEER region likely reflects geographic variation in practice patterns, because two of the algorithms rely on administrative procedure codes. Rates of misclassification range from nearly 3 percent to just over 5 percent in 1998, with false negatives highest in Freeman's algorithm and lowest using Nattinger's method. Misclassification disproportionately affects older women and those diagnosed with in situ, metastatic, or unknown-stage disease. Subjects of older age are more likely to have comorbid conditions, and subjects with metastatic disease are more likely to be facing imminent death. These two categories and those with in situ (the least severe) breast cancer therefore do not receive as aggressive treatment (Ballard-Barbash et al. 1996; Yancik et al. 2001; Bouchardy et al. 2003; Gold and Dick 2004), leading to a smaller pool of breast cancer-related claims that the algorithms can use to identify cases.

Because the addition of age, race, and region variables to the algorithms' case indicator variable improves the probability of correctly identifying incident breast cancer cases, using demographic information may enhance case identification. As an example, when applying Nattinger's algorithm, age categories could be incorporated into step 3, with older women requiring fewer

procedure codes to pass this step, as they may be less likely to receive aggressive treatment. Thus including these variables in the models may account for differences in treatment patterns due to age, region, and race, even though the demographic variables themselves are not indicators of cancer. Region variables may only be meaningful for the SEER areas and not for other studies where distinct regions are not well defined, however. It is also possible that the improved results are due to overfitting the model. We do not have an additional validation data set to test our findings.

PPV varies widely across the algorithms but improves over time with Warren's algorithm, although PPV is still lowest for this algorithm. PPV figures must be considered cautiously because our sample includes all breast cancer cases but only a 5 percent random sample of Medicare beneficiaries without breast cancer, yet we know PPV depends on disease prevalence. We present PPV to identify trends over time, but the absolute values may not be as meaningful.

The strength of this work is that our analyses include later years of data to represent more recent patterns of care (i.e., a shift to outpatient care), and we provide a head-to-head comparison of three algorithms using the newer data. We use a 5 percent random sample of nonbreast-cancer controls provided to us and assume that this is representative of the population without breast cancer. Otherwise, our results may be misleading.

Accurate identification of breast cancer cases has many implications for studying quality and costs of care. For true positive cases, we have all the information on subjects and would be able to study their treatment/surveillance patterns and costs of care. For false positive subjects, we would be evaluating care patterns of noncases to estimate health care utilization for breast cancer patients, thereby yielding underestimates of cancer costs and/or low compliance rates. For example, subjects without breast cancer should not be compliant with posttreatment mammography guidelines. We would therefore undercount the utilization of followup mammography in breast cancer patients. For true negative subjects, we would not anticipate any added error in our estimates. False negatives, however, would lead to a host of lost information, especially if they are differentially misclassified. We expect that the cases the algorithms miss would have fewer breast cancer-related claims due to less extensive or aggressive treatment, so they may more likely be early stage, older, facing imminent death, or with comorbid illness, and possibly of minority race. If one used the algorithms to identify cases for quality of care assessment, it could appear that there is less variation in care than actually exists, particularly for the vulnerable populations one might aim to study.

In assessing costs (i.e., reimbursed charges) of care using these algorithms, one would in effect overestimate average costs because the lower costs associated with less aggressive treatment would not balance out the high costs of advanced disease with its more involved treatment. Also, cancer-staging information is not available in claims data, so studies that are stage-treatment specific would be hard to conduct without linkage to tumor registry data. Previous research has shown cancer-stage identification to be difficult with claims data (Cooper et al. 1999). Important algorithm limitations to note are that Freeman's algorithm was developed for 65–74-year olds, Warren's was applied only to registries of entire states (not metropolitan areas), and none of the algorithms were designed to detect cases of in situ disease.

Because our study did not account for Medicaid claims data, there was concern that Medicare claims data for beneficiaries with state buy-in (SBI) coverage may be incomplete. Our findings did not bear this out, however (data not shown). A higher proportion of the older old in our sample does have SBI coverage (e.g., for 1998 data, almost 24 percent of those aged 80 and older have a full year of SBI coverage compared with 9.6 percent of those ages 65–69), but we could find no significant differences in rates of false negatives by SBI status within age groups for any of the algorithms for 1998 ($p > .12$ for all comparisons). We do note that 7 percent of white compared with 33 percent of black subjects had a full year of SBI coverage, but sample sizes are too small to draw meaningful conclusions about possible effects on algorithm performance. State buy-in coverage may act as a proxy of low-income status in our study sample, but likely does not directly affect the completeness of the utilization data, which challenges the notion that data for dual eligibles may be incomplete by using Medicare claims alone. In this study, the use of Medicare claims data appeared to be adequate to identify incident cases of breast cancer in SBI beneficiaries.

Some authors of the published algorithms recommended caution in using their algorithm to identify incident breast cancer cases, while others are more enthusiastic. We are not yet aware of any studies in which a researcher has used an algorithm alone to identify breast cancer cases. An important advancement in this field would be to refine an algorithm, which could be used to identify cases of recurrent cancer, information which most registries do not collect. Until the algorithms are refined, researchers probably should use the algorithms in isolation of cancer registry information only if they highlight the limitations of the method and there is no alternative. For other diseases, diagnosis and procedure codes may be more relevant to identify patient cohorts. In breast cancer, such codes often are used for patients undergoing diagnostic

testing to rule out disease or before a definitive cancer diagnosis (e.g., breast abnormality of some sort, rather than breast cancer). In addition, cancer stage, which can greatly affect treatment received, cannot be determined from diagnosis and procedure codes. The next question is: how good does an algorithm need to be in order to be confident in its application to new data? As with any diagnostic test, the algorithms yield trade-offs between sensitivity and specificity. Future work should explore the biases of algorithm misclassification in assessing use and costs of health care services. In the meantime, algorithms should be applied very cautiously to insurance claims databases to assess health care utilization and costs of breast cancer care outside SEER-Medicare populations.

ACKNOWLEDGMENTS

This work was funded by the American Cancer Society (Grant Number MRSGT-4-002-01-CPHPS) and was presented at the 27th Annual Meeting of the Society for Medical Decision Making in October 2005. The interpretation and reporting of the Linked SEER-Medicare Database are the sole responsibility of the authors. The authors acknowledge the efforts of the Applied Research Program, NCI; the Office of Information Services, and the Office of Strategic Planning, CMS; Information Management Services (IMS) Inc.; and the SEER Program tumor registries in the creation of the SEER-Medicare database. We appreciate the comments of two anonymous reviewers.

REFERENCES

- Ayanian, J. Z., and E. Guadagnoli. 1996. "Variations in Breast Cancer Treatment by Patient and Provider Characteristics." *Breast Cancer Research and Treatment* 40 (1): 65–74.
- Ayanian, J. Z., B. A. Kohler, T. Abe, and A. M. Epstein. 1993. "The Relation between Health Insurance Coverage and Clinical Outcomes among Women with Breast Cancer." *New England Journal of Medicine* 329 (5): 326–31.
- Ballard-Barbash, R., A. L. Potosky, L. C. Harlan, S. G. Nayfield, and L. G. Kessler. 1996. "Factors Associated with Surgical and Radiation Therapy for Early Stage Breast Cancer in Older Women." *Journal of the National Cancer Institute* 88 (11): 716–26.
- Bouchardy, C., E. Rapiti, G. Fioretta, P. Laissue, I. Neyroud-Casper, P. Schafer, J. Kurtz, A. Pascal Sappino, and G. Vlasos. 2003. "Undertreatment Strongly

- Decreases Prognosis of Breast Cancer in Elderly Women." *Journal of Clinical Oncology* 21 (19): 3580–7.
- Brewster, D. H., J. Crichton, J. C. Harvey, and G. Dawson. 1997. "Completeness of Case Ascertainment in a Scottish Regional Cancer Registry for the Year 1992." *Public Health* 111 (5): 339–43.
- Cooper, G. S., Z. Yuan, K. C. Stange, S. B. Amini, L. K. Dennis, and A. A. Rimm. 1999. "The Utility of Medicare Claims Data for Measuring Cancer Stage." *Medical Care* 37 (7): 706–11.
- Earle, C. C., P. J. Neumann, R. D. Gelber, M. C. Weinstein, and J. C. Weeks. 2002. "Impact of Referral Patterns on the Use of Chemotherapy for Lung Cancer." *Journal of Clinical Oncology* 20 (7): 1786–92.
- Elston, L. J., J. Simpkins, L. Schultz, G. A. Chase, C. C. Johnson, M. U. Yood, L. Lamerato, D. Nathanson, and G. Cooper. 2005. "Routine Surveillance Care after Cancer Treatment with Curative Intent." *Medical Care* 43 (6): 592–9.
- Freeman, J. L., D. Zhang, D. H. Freeman, and J. S. Goodwin. 2000. "An Approach to Identifying Incident Breast Cancer Cases Using Medicare Claims Data." *Journal of Clinical Epidemiology* 53 (6): 605–14.
- Gilligan, T. 2005. "Social Disparities and Prostate Cancer: Mapping the Gaps in Our Knowledge." *Cancer Causes and Control* 16 (1): 45–53.
- Gold, H. T., and A. W. Dick. 2004. "Variations in Treatment for Ductal Carcinoma In Situ in Elderly Women." *Medical Care* 42 (3): 267–75.
- Greenberg, M. L., R. D. Barr, B. DiMonte, E. McLaughlin, and C. Greenberg. 2003. "Childhood Cancer Registries in Ontario, Canada: Lessons Learned from a Comparison of Two Registries." *International Journal of Cancer* 105 (1): 88–91.
- Harlan, L., O. Brawley, F. Pommerenke, P. Wali, and B. Kramer. 1995. "Geographic, Age, and Racial Variation in the Treatment of Local/Regional Carcinoma of the Prostate." *Journal of Clinical Oncology* 13 (1): 93–100.
- Hodgson, D. C., W. Zhang, A. M. Zaslavsky, C. S. Fuchs, W. E. Wright, and J. Z. Ayanian. 2003. "Relation of Hospital Volume to Colostomy Rates and Survival for Patients with Rectal Cancer." *Journal of the National Cancer Institute* 95 (10): 708–16.
- McClish, D., and L. Penberthy. 2004. "Using Medicare Data to Estimate the Number of Cases Missed by a Cancer Registry: A 3-Source Capture–Recapture Model." *Medical Care* 42 (11): 1111–6.
- McClish, D., L. Penberthy, and A. Pugh. 2003. "Using Medicare Claims to Identify Second Primary Cancers and Recurrences in Order to Supplement a Cancer Registry." *Journal of Clinical Epidemiology* 56 (8): 760–7.
- McClish, D. K., L. Penberthy, M. Whittemore, C. Newschaffer, D. Woolard, C. E. Desch, and S. Retchin. 1997. "Ability of Medicare Claims Data and Cancer Registries to Identify Cancer Cases and Treatment." *American Journal of Epidemiology* 145 (3): 227–33.
- Michalski, T. A., and A. B. Nattinger. 1997. "The Influence of Black Race and Socioeconomic Status on the Use of Breast-Conserving Surgery for Medicare Beneficiaries." *Cancer* 79 (2): 314–9.
- Nattinger, A. B., and J. S. Goodwin. 1994. "Geographic and Hospital Variation in the Management of Older Women with Breast Cancer." *Cancer Control* 1 (4): 334–8.

- Nattinger, A. B., P. W. Laud, R. Bajorunaite, R. A. Sparapani, and J. L. Freeman. 2004. "An Algorithm for the Use of Medicare Claims Data to Identify Women with Incident Breast Cancer." *Health Services Research* 39 (6, part 1): 1733–49.
- Nattinger, A. B., T. L. McAuliffe, and M. M. Schapira. 1997. "Generalizability of the Surveillance, Epidemiology, and End Results Registry Population: Factors Relevant to Epidemiologic and Health Care Research." *Journal of Clinical Epidemiology* 50 (8): 939–45.
- Neuss, M. N., C. E. Desch, K. K. McNiff, P. D. Eisenberg, D. H. Gesme, J. O. Jacobson, M. Jahanzeb, J. J. Padberg, J. M. Rainey, J. J. Guo, and J. V. Simone. 2005. "A Process for Measuring the Quality of Cancer Care: The Quality Oncology Practice Initiative." *Journal of Clinical Oncology* 23 (25): 6233–9.
- Penberthy, L., D. McClish, C. Manning, S. Retchin, and T. Smith. 2005. "The Added Value of Claims for Cancer Surveillance: Results of Varying Case Definitions." *Medical Care* 43 (7): 705–12.
- Potosky, A. L., G. F. Riley, J. D. Lubitz, R. M. Mentnech, and L. G. Kessler. 1993. "Potential for Cancer Related Health Services Research Using a Linked Medicare—Tumor Registry Database." *Medical Care* 31 (8): 732–48.
- Ramsey, S. D., M. T. Mandelson, R. Etzioni, R. Harrison, R. Smith, and S. Taplin. 2004. "Can Administrative Data Identify Incident Cases of Colorectal Cancer? A Comparison of Two Health Plans." *Health Services and Outcomes Research Methodology* 5 (1): 27–37.
- Smedley, B. D., A. Y. Stith, and A. R. Nelson. 2003. "Introduction and Literature Review." In *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*, edited by B. D. Smedley, A. Y. Stith, A. R. Nelson, and the Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care, pp. 29–79. Washington, DC: National Academies Press.
- Surveillance Implementation Group. *Cancer Surveillance Research Implementation Plan*. 1999. Bethesda, MD: National Cancer Institute, National Institutes of Health.
- Wang, Y., M. Sharpe-Stimac, P. K. Cross, C. M. Druschel, and S. A. Hwang. 2005. "Improving Case Ascertainment of a Population-Based Birth Defects Registry in New York State Using Hospital Discharge Data." *Birth Defects Research Part A: Clinical and Molecular Teratology* 73 (10): 663–8.
- Warren, J. L., E. Feuer, A. L. Potosky, G. F. Riley, and C. F. Lynch. 1999. "Use of Medicare Hospital and Physician Data to Assess Breast Cancer Incidence." *Medical Care* 37 (5): 445–56.
- Wennberg, J. E., N. Roos, L. Sola, A. Schori, and R. Jaffe. 1987. "Use of Claims Data Systems to Evaluate Health Care Outcomes. Mortality and Reoperation following Prostatectomy." *Journal of American Medical Association* 57 (7): 933–6.
- Yancik, R., M. N. Wesley, L. A. Ries, R. J. Havlik, B. K. Edwards, and J. W. Yates. 2001. "Effect of Age and Comorbidity in Postmenopausal Breast Cancer Patients Aged 55 Years and Older." *Journal of American Medical Association* 285 (7): 885–92.
- Yoo, K. Y., H. R. Shin, S. H. Chang, K. S. Lee, S. K. Park, D. Kang, and D. H. Lee. 2002. "Korean Multi-center Cancer Cohort Study including a Biological Materials Bank (KMCC-I)." *Asian Pacific Journal of Cancer Prevention* 3 (1): 385–92.