

### Use of Medicare Data to Identify Incident Breast Cancer Cases

Joan L. Warren, Ph.D., Gerald F. Riley, M.S.P.H., A. Marshall McBean, M.D., M.Sc.,  
and Rosemarie Hakim, Ph.D.

*Surveillance, Epidemiology and End Results (SEER) data from the National Cancer Institute (NCI) provide reliable information about cancer incidence. However, because SEER data are geographically limited and have a 2-year time lag, we evaluated whether Medicare data could provide timely information on cancer incidence. Comparing Medicare women hospitalized for breast cancer with women reported to SEER, Medicare data had high specificity (96.6 percent), yet low sensitivity (59.4 percent). We conclude that Medicare hospitalization data can identify incident cases for cancers that usually require inpatient hospitalization. For cancers that often only receive outpatient treatment, such as breast cancer, additional Medicare data, such as physician bills, are needed to understand the entirety of treatment practices.*

#### INTRODUCTION

Data collected from the SEER program maintained by the NCI are usually considered the "gold standard" used to estimate the incidence and treatment of cancers throughout the United States. Prior to 1992, there were five States (Connecticut, Hawaii, Iowa, New Mexico, and Utah) and four metropolitan areas (Seattle, San Francisco-Oakland, Detroit, and Atlanta)

participating in the SEER program. The geographic areas represent about 10 percent of the Nation's population (Miller et al., 1993), are concentrated in the western United States, and do not include large numbers of some demographic groups, such as African Americans, raising concerns about the representativeness of the data. Moreover, because the SEER areas are geographically limited, they may not capture regional variation in treatment practices for specific cancers, costs of care, and medical outcomes following treatment. In addition, there is a 2-year lag needed for NCI to obtain case reports from the State registries.

Given concern about the representativeness and timeliness of the SEER data, other data sources may be able to provide accurate and more current information regarding cancer incidence and treatment. A potential alternative source of information about the incidence of cancer in the population is the administrative data collected for insurance billing purposes, such as Medicare data. These data offer the opportunity for timely studies that include the entirety of the United States. The cancer diagnoses from administrative data for inpatient stays have been found to have high levels of sensitivity and specificity when compared with the medical record for the hospitalization (Fisher et al., 1992; Romano and Luft, 1992). Previous studies have used Medicare hospitalization bills to analyze whether incidence rates from Medicare data were comparable to inci-

---

Joan L. Warren, Gerald F. Riley, and Rosemarie Hakim are with the Office of Research and Demonstrations, Health Care Financing Administration (HCFA). A. Marshall McBean is with the School of Public Health, University of Minnesota. The opinions expressed herein are those of the authors and do not necessarily reflect the views of the University of Minnesota or HCFA.

dence rates from SEER data for five cancers—breast, colon, esophagus, lung, prostate, and uterus (Whittle et al., 1991; McBean, Warren, and Babish, 1994; McBean, Babish, and Warren, 1993). The comparability of the rates varied by type of cancer. McBean, Warren, and Babish (1994) and McBean, Babish, and Warren (1993) found that for those cancers that are usually treated in the hospital setting, such as esophagus, lung, and uterine, Medicare rates were comparable to rates from the SEER data. For colon and prostate cancer, which are often treated in the outpatient setting only, the rates calculated from Medicare hospitalization and SEER data were significantly different. These studies only utilized aggregate data from SEER and Medicare for comparisons and did not attempt to link files at the individual level to determine if the same persons were being identified from the two independent sources of data.

The purpose of this study was to determine if persons identified as having incident breast cancer from the Medicare data were also found in the SEER data for that year. If a method could be developed to identify specific women with incident breast cancer from administrative data, it would help researchers to identify cohorts to examine treatment practices for the Nation or for subgroups. Given our goal, we opted to develop an algorithm with stringent inclusion criteria, resulting in a high level of specificity. A risk of using an algorithm with high specificity is the potential loss of true cases that might not be identified and the potential bias that the cases identified might be systematically different from cases not identified. Therefore, we also included in our analysis an examination of whether cases appearing in the SEER file were found in the Medicare hospitalization data. Cases found only in the SEER data were compared with cases

found in both the Medicare and SEER data to determine if the cases found only in SEER data had different characteristics. Breast cancer was chosen because it is the leading type of cancer occurring in elderly women (Miller et al., 1993) and is treated in both the inpatient and outpatient setting.

## METHODS

### Data Sources

The two sources of data for this study were the SEER data from NCI and Medicare hospitalization data maintained by HCFA.<sup>1</sup> The SEER data have been collected since 1973 and, for each patient, contain: age, race, and sex; each occurrence of a primary incident cancer (recurrent and metastatic sites are not collected); year of diagnosis; site of report (hospital, physician office, etc.); stage of disease; type of surgery performed; and for this file only, the person's unique Medicare identification number.

The goal of this study was to determine what portion of cases defined as incident from 1989 Medicare data appeared as incident in 1989 SEER data. Therefore, we wanted to retain as many SEER cases as possible for a potential match with Medicare data. We included all cases of female breast cancer reported by the five State registries participating in the SEER program, regardless of the age of the woman or the year of diagnosis. After identification of all cases of breast cancer from all years of SEER data, we created a subset of 1989 SEER cases. To adjust for any lag between the time a case was first diagnosed and reported to SEER and when the surgical treatment resulting in hospitalization

<sup>1</sup>Both SEER and Medicare data contain personal identifiers and are subject to stringent data confidentiality requirements. These data are available to the research community, with restrictions, although identifiers are removed prior to release.

might occur, cases that appeared in SEER within the 12 months of 1989 or the last two months of 1988 were considered 1989 cases. This decision was made because in a preliminary analysis, the date of diagnosis in the SEER file was 2 months before the date of hospitalization for 43 percent of Medicare 1989 incident breast cancer cases.

The Medicare hospitalization data, contained in the Medicare provider analysis and review (MEDPAR) file, are available from 1984 to the present and contain summarized records of all claims submitted by hospitals for inpatient services provided to Medicare beneficiaries. Our analysis focused on women with breast cancer reported by a hospital to Medicare with a discharge date in 1989. For 1989 hospitalizations, the MEDPAR records contain information about the beneficiary's age, race, sex, and State of residence; up to five diagnosis codes and three procedure codes classified according to the *International Classification of Diseases, 9th Revision, Clinical Modification* (ICD-9-CM) (U.S. Department of Health and Human Services, 1980); dates of admission and discharge from the hospital; and State of the hospital's location. For the same period, there is also Medicare enrollment data which reports if a beneficiary is eligible for Part A (hospital) services or enrolled in a health maintenance organization (HMO).

### **Algorithm to Identify Incident Cases**

The algorithm to identify incident cases of female breast cancer from Medicare hospitalization data is as follows:

- We used 100 percent of Medicare hospitalizations for 1989 to identify all women 65 years of age or over with one or more hospitalizations with a diagnosis of breast cancer (ICD-9-CM codes

174-174.9 and 233.0) appearing as the principal diagnosis.

- Incident cases of breast cancer were defined from the Medicare data as women who had no prior Medicare hospitalization with breast cancer appearing as any of the five diagnoses from 1984-88 or history of breast cancer (ICD-9-CM V10.3) appearing as any of the five diagnoses from 1984-89.
- We excluded the estimated 5 to 6 percent of Medicare beneficiaries who were enrolled in an HMO during 1989, because Medicare utilization records for HMO enrollees are incomplete. To ensure that the beneficiaries identified from the Medicare data were indeed residents of the SEER State, we limited our analysis to women whose Medicare records indicated that they resided in one of the five SEER States in 1989 and who were also treated in a hospital located in one of the five SEER States.

However, we collected Medicare data on all cases of female breast cancer nationally so that we could determine if women reported to SEER, but not residing in and treated in one of the five SEER States per Medicare data, were found in the Medicare data for the remaining parts of the Nation.

### **Data Linkage**

We used each individual's Medicare identification number to link cases in SEER with Medicare hospitalization data. Patients with breast cancer could have been found to be incident in both the 1989 Medicare and 1989 SEER data; misclassified from the 1989 Medicare as incident when they were actually prevalent cases based on appearing in SEER data for earlier years; found only in 1989 Medicare data, but not in any year of SEER data; or appearing only in 1989 SEER data, but not in the

cases identified through the algorithm applied to 1989 Medicare data.

### **Analysis of Non-Matched Cases**

For cases not found in both data bases in 1989, we searched for a possible reason. Cases identified as incident from the 1989 Medicare data that were found to be prevalent cases based on SEER information may have been misidentified as incident because Medicare never received a hospital bill in earlier years for the cancer treatment of the patient. Reasons that Medicare might never have received a hospital bill in earlier years include: the woman developed breast cancer prior to 1984, the first year of Medicare data available for this study; she was diagnosed before becoming a Medicare beneficiary at 65 years of age; or the cancer treatment did not result in a hospitalization.

Cases found in the 1989 Medicare data, but not found in any year of SEER, were sent for evaluation to the two largest State registries, Connecticut and Iowa. We wanted to determine if these cases represented incident cases not captured by the registries or if the cases were known to the registries but were not matched because the women did not have Medicare numbers identified. Approximately 94 percent of SEER cases have a Medicare identification number (Potosky et al., 1993).

For cases that were found only in the 1989 SEER data, we created a hierarchy of mutually exclusive possible reasons that these cases were not found in the Medicare data. These reasons include: factors that might result in Medicare never receiving a hospital bill for the case; cases that appeared in the Medicare data but did not meet the criteria developed for our algorithm; or mismatch due to our use of 14 months of SEER data, compared with 12 months of Medicare data, to adjust for

any time lag between when a woman was diagnosed and underwent hospital treatment. We compared the cases appearing only in SEER with cases appearing in both files to determine if there were differences in age group, race, site of report, stage of breast cancer, or type of surgery performed. Chi-square tests with 95-percent confidence intervals were calculated to assess if there were significant differences between the cases found only in SEER and cases found in both files.

### **RESULTS**

In 1989, there were 2,247 elderly Medicare women hospitalized with incident breast cancer identified from the algorithm (Table 1). We were able to match 2,106 (93.7 percent) of the incident cases to any year of SEER data, with 2,051 (91.3 percent) of all incident cases found in the 1989 Medicare file reported to SEER in 1989. Of women identified as having incident breast cancer from the Medicare data, 55 (2.4 percent) were really prevalent cases and 141 (6.3 percent) could not be matched to any year of the SEER data. Comparing SEER data with Medicare data revealed that of the 3,454 women who were 65 years of age or over who were reported in the 14 months of SEER data, 2,051 (59.4 percent) were matched with the 1989 Medicare incident cases. A residual of 1,403 cases appeared in SEER only.

Nineteen of the 55 women (34.5 percent) identified as prevalent by SEER data were reported to SEER prior to 1984, the first year of Medicare hospitalization data available. Nine women (16.4 percent) were diagnosed with breast cancer before they turned 65 years of age and became eligible for Medicare. An additional nine cases may not have generated a hospital bill because SEER data showed that no in-hos-

**Table 1**

**Comparison of Cases of Female Breast Cancer Identified From 1989 Medicare Hospitalization Data With Cases of Female Breast Cancer Reported in SEER Data for Women Identified as Medicare Beneficiaries**

Cases	Total for 5 SEER States		Connecticut		Hawaii		Iowa		New Mexico		Utah	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Total Incident Cases From 1989 Medicare Hospitalization Data	2,247	100.0	797	100.0	164	100.0	806	100.0	234	100.0	246	100.0
1989 Medicare Incident Cases Matched to Any Year of SEER Cases	2,106	93.7	736	92.3	159	97.0	752	93.3	230	98.3	229	93.1
Medicare Incident Cases Matched to 1989 SEER Cases <sup>1</sup>	2,051	91.3	713	89.5	157	95.7	736	91.3	225	96.2	220	89.4
Medicare Incident Cases Determined to Be Prevalent Cases From SEER Data	55	2.4	23	2.9	2	1.2	16	2.0	5	2.1	9	3.7
1989 Medicare Incident Cases Not Found in Any Year of SEER Data	141	6.3	61	7.7	5	3.0	54	6.7	4	1.7	17	6.9
Total Incident Cases From 1989 SEER Data <sup>1</sup>	3,454	100.0	1,305	100.0	259	100.0	1,203	100.0	356	100.0	331	100.0
1989 SEER Cases Matched to 1989 Medicare Incident Hospitalization Data	2,051	59.4	713	54.6	157	60.6	736	61.2	225	63.2	220	66.5
1989 SEER Cases Not Matched to 1989 Medicare Incident Hospitalization Data	1,403	40.6	592	45.4	102	39.4	467	38.8	131	36.8	111	33.5

<sup>1</sup> Incident cases from 1989 SEER also included November and December of 1988 to allow for a lag between when a woman was diagnosed and hospitalized for treatment.

NOTE: SEER is Surveillance, Epidemiology and End Results program.

SOURCES: National Cancer Institute; Health Care Financing Administration, Office of Research and Demonstrations; Data from the Division of Health Information and Outcomes.

**Table 2**  
**Validation of the Status of Cases of Breast Cancer Identified From 1989 Medicare Hospitalization Data, but Not Appearing in the 1989 SEER Data for Connecticut and Iowa, by Type of Diagnosis or Procedure**

Cases	All Cases Sent to the Two Registries for Validation		Breast Cancer Diagnosis Only		Breast Cancer Diagnosis and Procedure		Breast Cancer Diagnosis and Mastectomy	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Total Cases in Category	115	100.0	13	100.0	11	100.0	91	100.0
Case Found in Registry Data With Breast Cancer in 1989 <sup>1</sup>	98	85.2	7	53.8	9	81.8	82	90.1
Case Found in Registry Data With Breast Cancer in Earlier Year	7	6.1	3	23.1	1	9.1	3	3.3
Case Not Found in Registry Data and the Hospital Record								
Showed Incident Case Not Captured by the Registry	1	0.9	0	0.0	0	0.0	1	1.1
Case Not Found in Registry Data for Other/Unknown Reason	9	7.8	3	23.1	1	9.1	5	5.5

<sup>1</sup> Incident cases from 1989 SEER also included November and December of 1988 to allow for a lag between when a woman was diagnosed and hospitalized for treatment.

NOTE: SEER is Surveillance, Epidemiology and End Results program.

SOURCES: National Cancer Institute; Health Care Financing Administration, Office of Research and Demonstrations; Data from the Division of Health Information and Outcomes.

pital surgery was performed. There were 18 prevalent cases (32.7 percent) for which we could identify no possible reason as to why the algorithm misidentified prevalent cases as incident cases.

There were a total of 115 Connecticut and Iowa women who appeared in the Medicare data but did not appear in any year of SEER data. After being sent for review, 105 cases (91.3 percent) had already been identified by the Connecticut and Iowa registries (Table 2). These cases could not be matched to Medicare hospitalization data because the SEER file did not have a Medicare identification number for these persons. Ninety-eight of the 115 cases (85.2 percent) were diagnosed in 1989. Of these cases, 7 had in the Medicare data a principal diagnosis of breast cancer without a procedure, 9 cases had a principal diagnosis of breast cancer and a breast procedure other than mastectomy, and 82 had a principal diagnosis of breast cancer and a mastectomy. Only 1 of the 115 cases was

found to be an incident case not previously captured by the registry.

Of the 1,403 cases found only in the SEER data, the most frequent explanation was related to a lag in time between diagnosis and treatment. Three hundred nine cases (22 percent) were found in November or December 1988, and 76 (5.4 percent) appeared in the Medicare hospitalization files in 1990 (Table 3). Another 299 cases (21.3 percent) had factors that might result in Medicare not receiving a hospital bill for the patient. For example, 118 cases (8.4 percent) were enrolled during 1989 in HMOs, which are not required to submit bills for hospitalization to Medicare. An additional 236 cases (16.7 percent) were reported in the 1989 Medicare hospitalization data, but were excluded from the incident cohort because of criteria for the algorithm. Ninety-three cases (6.6 percent) were omitted because breast cancer appeared a secondary diagnosis. We could not explain

Table 3

**Examination of Selected Reasons That Cases of Female Breast Cancer Reported in 1989 SEER Data<sup>1</sup> Were Not Identified by the Algorithm Used to Identify Incident Cases of Female Breast Cancer From 1989 Medicare Hospitalization Data**

Cases Explained By Reason	All Cases Found Only in SEER Data for the 5 SEER States	
	Number	Percent
Total Number of Cases Appearing in SEER Only	1,403	100.0
Total Cases Potentially Explained as to Why Not Matched	920	65.6
Medicare Might Have No Hospital Bill		
Beneficiary Enrolled in an HMO at Any Time in 1989	118	8.4
Beneficiary Not Covered by Part A for 12 Months in 1989 (for Persons 66 Years of Age or Over)	112	8.0
Medicare Was the Secondary Payer, After Other Insurance	62	4.4
Reporting Source Was Other Than the Hospital	7	0.5
Algorithm Excluded Cases Appearing in Medicare Data		
Beneficiary Resided in Non-SEER State per Medicare Data	28	2.0
Beneficiary Treated Outside SEER State per Medicare Data	89	4.9
Beneficiary Appeared in 1989 Medicare Data With Breast Cancer as a Secondary Diagnosis	93	6.6
Beneficiary Appeared in 1989 Medicare Data With Breast Procedure, But No Breast Cancer Diagnosis	9	0.6
Beneficiary Had a Medicare Hospitalization With a Diagnosis of Breast Cancer Prior to 1989 and Also Had a Primary Breast Cancer Reported to SEER Prior to 1989	37	2.6
Reasons Related to Lag in Time of Diagnosis and Treatment		
Case Appeared in November or December 1988 SEER File Only <sup>1</sup>	309	22.0
Beneficiary Appeared in the 1990 Medicare Hospital Data With a Diagnosis of Breast Cancer	76	5.4

<sup>1</sup> The 1989 SEER also included the last 2 months of 1988 to allow for a lag between when a woman was diagnosed and hospitalized for treatment.

NOTES: SEER is Surveillance, Epidemiology and End Results program. HMO is health maintenance organization.

SOURCES: National Cancer Institute; Health Care Financing Administration, Office of Research and Demonstrations; Data from the Division of Health Information and Outcomes.

why 483 cases (34.7 percent) appeared in SEER only.

We compared characteristics between cases found in the SEER data only and the cases found in both SEER and Medicare data to determine if there were significant differences between the two groups. As shown in Table 4, cases found only in the SEER data were more likely to be 85 years of age or over, have in situ or unstaged cancer, and to be reported by a source other than the hospital. In addition, cases appearing only in SEER had a higher proportion of women having no surgery, biopsy only, or breast conserving surgery, although almost 50 percent underwent simple or modified radical mastectomy.

## DISCUSSION

We are encouraged that the algorithm used to identify incident cases from the Medicare data has a high level of specificity. At first, we found that 91.3 percent of cases defined as incident cases from the Medicare data were matched to the SEER data of the same year. With additional information from two of the five State registries, we were able to classify as incident or prevalent the 141 cases not found in SEER data. The addition of cases identified as incident by the registry increased the portion of true incident cases found in 1989 Medicare data to 96.6 percent.

The number of prevalent cases misidentified as incident from the Medicare data was low. Including the portion of cases identified

**Table 4**  
**Comparison of Selected Characteristics Between Elderly Women**  
**With Breast Cancer Identified From Both the 1989 Medicare and 1989 SEER Data**  
**and Women Identified Only in the 1989 SEER Data**

Measure	Cases Appearing In Both the Medicare and SEER Data (n=2,106)	Cases Appearing Only in the SEER Data (n=1,403)	Chi-Square p Value
	Percent	Percent	
<b>Age Group</b>			
65-74	55.0	51.9	0.00
75-84	36.5	32.8	
85 or Over	8.5	15.4	
<b>Race</b>			
White	93.1	93.4	0.14
Black	1.1	1.7	
Other	5.8	4.8	
<b>Case-Reporting Source</b>			
Hospital	99.9	97.5	0.00
Other	0.1	2.5	
<b>Stage of Breast Cancer</b>			
In Situ	5.5	10.8	0.00
Localized	60.2	53.5	
Regional	28.6	21.3	
Distant	4.6	6.1	
Unstaged	1.2	8.3	
<b>Type of Surgery</b>			
None or Biopsy Only	2.1	15.3	0.00
Breast Conserving Surgery	15.3	34.9	
Simple Mastectomy	7.4	3.5	
Modified Radical Mastectomy	75.1	46.1	
Radical Mastectomy	0.2	0.2	

NOTE: SEER is Surveillance, Epidemiology and End Results program.

SOURCES: National Cancer Institute; Health Care Financing Administration, Office of Research and Demonstrations; Data from the Division of Health Information and Outcomes.

as prevalent by the registries only increased the number of prevalent cases from 2.4 to 2.8 percent. Using this algorithm with additional years of Medicare data would reduce further the number of prevalent cases. For example, 12 of the 55 prevalent cases occurred from 1980-84 and could have been removed if 4 additional years of Medicare hospitalization data were available.

Of the cases found only in the 1989 Medicare file, it is noteworthy that over 90 percent were found in the existing registry files for Connecticut and Iowa. The apparent cause of this is that women appearing only in the Medicare data were not being identified as Medicare beneficiaries when the SEER file with Medicare identifiers

was created in 1993. There does not appear to be a problem with incident cases being missed by the registries.

A major concern of our analysis is that a large number of cases reported to SEER did not appear in the Medicare hospitalization data. We initially theorized that much of the low sensitivity of the algorithm reflected our empirically driven decision to use 14 months of SEER data to match against 12 months of Medicare data. We took the cases appearing in only the SEER data and subset them to include only those reported to the registries in the last 2 months of 1988 (309 of 1,403). These cases were then compared with the 1988 Medicare hospitalization data to determine

how many have had a Medicare hospital bill during November and December 1988. Only 12 of the 309 cases appearing only in the SEER data had a Medicare hospitalization for breast cancer during the last 2 months of 1988.

Other reasons that cases appeared only in the SEER data reflect the limitations of working with administrative data. Relying on billing data to identify cases can only succeed if bills are submitted. We could identify for over 20 percent of the women reported only by SEER an administrative reason why Medicare might not receive a hospital bill (Table 3). Unfortunately, there is little that could be done to ascertain cases that were missed as a result of administrative factors.

On the other hand, some of the cases found only in the SEER data could have been found in the Medicare data if the algorithm were changed. Almost 17 percent of cases reported as being found only in the SEER data actually appeared in the Medicare data, but were excluded from the incident cohort as a result of criteria included in the algorithm. Before adopting the algorithm presented in this article, we performed a sensitivity analysis to assess a variety of algorithms to determine which would result in the best sensitivity and specificity using Medicare data. With the other varied algorithms, the specificity to identify true incident cases decreased to 80-85 percent, while the sensitivity of the Medicare data remained essentially unchanged (data not shown). Our inability to improve the sensitivity even with the use of more inclusive algorithms demonstrates that there may be cases reported to SEER that will never be found by looking only at the hospital data.

The lack of inpatient treatment may also explain the differences we noted between women found in both the Medicare and SEER data and women found only in SEER

data. Women with breast cancer who are not hospitalized will not have a hospital bill. Cases appearing in the SEER data only had a higher proportion of women 85 years of age or over, a group less likely to have surgery than younger elderly women (Bergman et al., 1991; Goodwin, Hunt, and Samet, 1993; Mann et al., 1988). Another reason that cases might appear only in the SEER file is that surgery was performed in an outpatient setting. This theory is supported by the higher portion in the SEER data of in situ cases and more women undergoing breast-conserving surgery, although the 50 percent of cases in the SEER-only file undergoing mastectomy cannot be overlooked.

## CONCLUSION

From our analysis, it is clear that hospital data can be very useful to identify incident cases of cancer for those cancers that are treated predominantly with an inpatient hospitalization. However, for cancers such as breast cancer that are often treated only in outpatient settings, additional data from the physician office and outpatient sites are needed to understand the entirety of treatment practices. Previous studies that have only utilized Medicare hospitalization data to describe geographic variation in the treatment of breast cancer did not capture services in the outpatient setting (Nattinger et al., 1996; Nattinger and Goodwin, 1994; Nattinger et al., 1992). Therefore, these studies may not reflect the entirety of the current treatment practices among elderly women in the United States.

For the time period of this study, 1989, the Medicare data available for outpatient and physician services were limited to 5 percent of beneficiaries and the data did not contain diagnoses. In 1991, HCFA altered data collection and the physician

bill format. Since that time, 100 percent of all physician bills are available and these bills now include an ICD-9-CM diagnosis. HCFA and NCI are presently updating the project to identify Medicare numbers for all persons reported to SEER. When the update is completed, there will be SEER and Medicare identification numbers available through 1993. These data will present the opportunity to develop algorithms for identifying incident cases of breast cancer from physician bills as well as hospital cases. In the interim, researchers who plan to use administrative data to evaluate treatment of breast cancer should be aware of the strengths and the limitations.

## ACKNOWLEDGMENTS

The authors wish to thank Drs. Arnold Potosky and Benjamin Hankey at the NCI for their assistance in contacting the State cancer registries, and Jack Flannery of the Connecticut Tumor Registry and Dr. Charles Lynch of the Iowa SEER Registry for reviewing specific cases identified by HCFA.

## REFERENCES

- Bergman, L., Dekker, G., van Leeuwen, F.E., et al.: The Effect of Age on Treatment Choice and Survival in Elderly Breast Cancer Patients. *Cancer* 67:2227-2234, May 1, 1991.
- Fisher, E.S., Whaley, F.S., Krushat, W.M., et al.: The Accuracy of Medicare's Hospital Claims Data: Progress Has Been Made, But Problems Remain. *American Journal of Public Health* 82(2):243-248, 1992.
- Goodwin, J.S., Hunt, W.C., and Samet, J.M.: Determinants of Cancer Therapy in Elderly Patients. *Cancer* 72(1):594-601, 1993.
- Mann, B.A., Samet, J.M., Hunt, W.C., et al.: Changing Treatment of Breast Cancer in New Mexico From 1969 Through 1985. *Journal of the American Medical Association* 259(23):413-417, 1988.
- McBean, A.M., Warren, J.L., and Babish, J.D.: Measuring the Incidence of Cancer in Elderly Americans Using Medicare Claims Data. *Cancer* 73(9):2417-2425, 1994.
- McBean, A.M., Babish, J.D., and Warren, J.L.: The Determination of Lung Cancer Incidence in the Elderly Using Medicare Claims Data. *American Journal of Epidemiology* 137(4):226-234, 1993.
- Miller, B.A., Ries, L.A.G., Hankey, B.F., et al. (eds): *SEER Cancer Statistics Review: 1973-1990*. NIH Pub. No. 93-2789. National Cancer Institute, National Institutes of Health. Washington. U.S. Government Printing Office, 1993.
- Nattinger, A.B., Gottlieb, M.S., Hoffman, R.G., et al.: Minimal Increase in the Use of Breast-Conserving Surgery From 1986 to 1990. *Medical Care* 34(5):479-489, 1996.
- Nattinger, A.B., and Goodwin, J.S.: Geographic and Hospital Variation in the Management of Older Women With Breast Cancer. *Cancer Control* : 334-338, July/August 1994.
- Nattinger, A.B., Gottlieb, M.S., Veum, J., et al.: Geographic Variation of Breast-Conserving Treatment for Breast Cancer. *New England Journal of Medicine* 326(17):1102-1107, 1992.
- Potosky, A.L., Riley, G.F., Lubitz, J.D., et al.: Potential for Cancer Related Health Services Research Using a Linked Medicare-Tumor Registry Database. *Medical Care* 31(8):732-747, 1993.
- Romano, P.S., and Luft, H.S.: Getting the Most Out of Messy Data: Problems and Approaches in Dealing With Large Administrative Data Sets. In Grady, M.L., Schwartz, H.A., eds.: *Medical Effectiveness Research Data Methods*. AHCPR Pub. No. 92-0056. Agency for Health Care Policy and Research. Rockville, MD. U.S. Government Printing Office, 1992.
- U.S. Department of Health and Human Services: *International Classification of Diseases, 9th Revision, Clinical Modification. Volumes 1-3*. DHHS Pub. No. (PHS) 80-1260. Washington. U.S. Government Printing Office, 1980.
- Whittle, J., Steinberg, E.P., Anderson, G.F., et al.: Accuracy of Medicare Claims Data for Estimation of Cancer Incidence and Resection Rates Among Elderly Americans. *Medical Care* 29(12):1226-1236, 1991.

---

Reprint Requests: Joan L. Warren, Ph.D., Office of Research and Demonstrations, Health Care Financing Administration, C3-24-07, Baltimore, Maryland 21244-1850.