Exploring the Surgeon Volume–Outcome Relationship Among Women With Breast Cancer

Ann Butler Nattinger, MD, MPH; Purushottam W. Laud, PhD; Rodney A. Sparapani, MS; Xu Zhang, PhD; Joan M. Neuner, MD, MPH; Mary Ann Gilligan, MD, MPH

Background: A relationship between higher surgeon volume and lower mortality has been described for breast cancer, but selection bias has not been rigorously evaluated. We studied potential bias in the surgeon volume–outcome relationship by comparing the relationship of surgeon volume to breast cancer mortality and to mortality from other causes of death.

Methods: We conducted an observational cohort study from tumor registry and Medicare claims data on 12,216 women, 66 years or older, with stage I or II breast cancer, who were operated on by 1,856 surgeons. Breast cancer mortality and other-cause mortality were determined from death certificate sources and surgeon volume from Medicare claims.

Results: Treatment by a high-volume surgeon was associated with younger patient age, white race, less comorbidity, and residence in a more affluent zip code. Patients treated by low-, medium-, and high-volume surgeons had small differences in breast cancer mortality (17.4, 15.7, and 13.0 deaths per 1,000 person-years, respectively; P = .03) but larger differences in non-breast cancer mortality (46.0, 36.8, and 31.7 deaths per 1,000 person-years, respectively; P < .001). After adjustment for multiple patient and disease factors, women treated by high-volume surgeons, compared with those treated by low-volume surgeons, were not less likely to die of breast cancer (relative risk, 0.94; 95% confidence interval, 0.76-1.16) but were significantly less likely to die of other causes (relative risk, 0.86; 95% confidence interval, 0.75-0.98).

Conclusions: The surgeon volume–outcome relationship for these patients with breast cancer was attributable not to mortality from breast cancer but to other causes of death. The lack of specificity of this relationship raises the possibility of selection bias as an explanatory factor.

Arch Intern Med. 2007;167(18):1958-1963

A positive relationship between higher volumes of cases and better health outcomes has been described for a number of different surgical procedures and has led to the implementation of policy strategies to regionalize certain procedures. Policies to promote regionalization are grounded in the belief that the explanation for positive volume-outcome relationships is greater proficiency of the high-volume physician or hospital, with respect to either physician technical skill or better organized care processes. This is an attractive explanation with intuitive appeal. However, alternative explanations are possible. One such alternative is selection bias (ie, the possibility that high-volume physicians or hospitals attract patients at lower risk of an adverse outcome). The studies of volume-outcome relationships have been observational, and selection bias is always a threat to observational studies.

There are several different types of selection bias that require consideration. These include selection based on disease severity, selection based on other medical conditions (comorbidity), and selection based on socioeconomic status, a factor that is known to influence mortality and other outcomes. With respect to cancer, the available studies have typically controlled for disease severity, and some have controlled for comorbidity or socioeconomic status using administrative data. The volume relationships are typically reported to persist after controlling for these factors. Unfortunately, it is not possible to control fully for comorbidity using administrative data. The ability to control for socioeconomic factors using administrative data is also limited, with such factors generally available at the community level, if at all.

In this study, we used breast cancer as an opportunity for studying potential bias in the surgeon volume–outcome relation-
ship. Breast cancer is a useful condition for study for several reasons. Population-based data sources are available, reliable information is available regarding extent of disease, and data are available permitting the analysis of overall and disease-specific mortality. Three previous studies of the surgeon volume–outcome relationship for breast cancer found that patients of high-volume (≥ 15-30 patients annually) surgeons had about a 15% reduction in overall mortality, but these studies did not examine disease-specific mortality. We believed that if the breast cancer surgeon volume–outcome relationship were mediated by better quality of cancer care, the relationship of volume to breast cancer–specific mortality should be stronger than the relationship of volume to overall mortality. Analogously, one would expect little or no relationship of surgeon volume to non–breast cancer mortality. If, however, selection bias accounted for some or all of the relationship of volume to overall mortality, then the relationship of volume to breast cancer mortality would be less specific.

### METHODS

#### SOURCES OF DATA

A database, provided by the National Cancer Institute, of linked information from the Surveillance, Epidemiology, and End Results (SEER) Tumor Registry and Medicare administrative data was used to identify patients. The population-based SEER registries collect information on demographic characteristics, extent of disease, and initial treatment for incident cancer patients. A linkage between SEER and Medicare data was successfully completed for 95% of patients with breast cancer 65 years or older in SEER records.

The Medicare program funds most inpatient and outpatient services for more than 97% of US residents 65 years or older. We used the Medicare Provider Analysis and Review File, which contains inpatient hospital claims; the Outpatient Standard Analytic File, which contains claims for outpatient facility services; the 100% Carrier Standard Analytic File, which contains claims for physician services; and the denominator file, which contains information on beneficiary entitlement and zip code of residence.

The US census data were linked with the SEER-Medicare database to determine the population density of the county in which breast cancer operations were performed and to determine socioeconomic characteristics of the zip codes in which patients resided.

The institutional review board of the Medical College of Wisconsin approved the study design.

#### STUDY POPULATION

Initial selection criteria included women 66 years or older, who were diagnosed as having invasive stage I or II breast cancer from January 1, 1994, through December 31, 1996, were eligible for Medicare parts A and B and not in a health maintenance organization for at least 1 year before diagnosis and after diagnosis for at least 4 months or until death, and had a Medicare surgeon claim within 4 months after diagnosis. These selection criteria yielded 13 030 women who were followed up through August 31, 2000. Women were excluded if their breast cancer operation was performed outside the SEER catchment areas (n=611) or if a valid surgeon unique provider identification number was not present on the Medicare claim (n=203). These criteria resulted in a final study cohort of 12 216 women.

#### PATIENT CHARACTERISTICS

The patient’s age at diagnosis was determined from the Medicare files, and race was determined from the SEER files. The patient’s zip code of residence was determined from the Medicare files and was linked to US census data to determine the per capita income of the zip code in which the patient resided, to estimate neighborhood socioeconomic status. Comorbidity was determined from the inpatient and ambulatory Medicare claims for the year before the breast cancer diagnosis using the adaptation of Klabunde et al of the Charlson comorbidity index.

Tumor characteristics were determined from the SEER registry data, including the size and grade of the primary tumor, the status of the axillary lymph nodes, and the hormone receptor status.

#### DETERMINATION OF VOLUME

For the determination of volume, a broader cohort of patients was used than for the study cohort described in the preceding paragraphs. For determining volume, women were required to be 65 years or older at diagnosis of a breast cancer any stage during the study years and were required to have an operation identified on a Medicare claim within 1 month before and 4 months after the SEER month of diagnosis. For each surgeon and hospital operating on a study cohort patient, the number of operations in this broader cohort was counted.

#### OUTCOMES

Date of death was determined from Medicare files. The cause of death was determined from SEER files, derived from a review of death certificates. The cause of death was unknown for 99 subjects, who were censored in the overall mortality analyses but censored in the cause-specific analyses.

#### STATISTICAL ANALYSIS

The mean annual Medicare volume was estimated for each surgeon. The surgeons were categorized into high-, medium-, and low-volume categories that had similar numbers of patients but still required that the high-volume surgeons have at least twice the volume of the low-volume surgeons. This resulted in categories of fewer than 5 annual Medicare cases, 5 to fewer than 10 annual cases, or 10 or more annual cases. To control for selection bias in the choice of high-, medium-, or low-volume surgeons, we performed a propensity score analysis. This is a method of dealing with possible bias in observational studies. We developed a model to predict the patient’s propensity for treatment by a low-, medium-, or high-volume surgeon. We used a trichotomous logistic regression for the propensity model and then formed the propensity groups based on planar tertiles of bivariate predicted probabilities. Candidate factors for the propensity model included available demographic and extent-of-disease factors. Factors that were significant predictors of propensity for undergoing treatment by a high-volume surgeon (all at the P ≤ .01 level) were white race, being married, residing in a rural area or a metropolitan statistical area of fewer than 250 000 persons, residing in a zip code with higher per capita income, having less comorbidity, having a known hormone receptor status, and having an unknown axillary lymph node status. The propensity groups markedly reduced differences between the surgeon volume groups...
with respect to all the factors in Table 1. In most but not all cases, the differences were rendered statistically insignificant within propensity groups.

To evaluate the relationship of surgeon volume and survival, Cox proportional hazards survival analysis was used. The patient was the unit of analysis. Clustering of outcomes within propensity groups. In most but not all cases, the differences were rendered statistically insignificant with respect to all the factors in Table 1. In most but not all cases, the differences were rendered statistically insignificant within propensity groups.

To evaluate the relationship of surgeon volume and survival, Cox proportional hazards survival analysis was used. The patient was the unit of analysis. Clustering of outcomes within propensity groups. In most but not all cases, the differences were rendered statistically insignificant within propensity groups.

The results reported reflect surgeon volume for longer than 10 months of follow-up; for earlier follow-up times, the volume-outcome relationship was in the same direction but larger in magnitude.

The 12,216 women who had undergone surgical treatment of stage I or II breast cancer were operated on by 1,856 different surgeons, of whom 1,325 were low-volume, 384 were medium-volume, and 147 were high-volume surgeons (Table 1). Patients treated by high-volume surgeons were slightly younger, were more likely to be white, and resided in more affluent areas than those treated by low-volume surgeons. The patients of high-volume surgeons had lower levels of comorbid illness, had tumors that were on average 1 to 2 mm smaller, and were more likely to be treated in high-volume hospitals. The relationship of surgeon and hospital volume was moderate, with a Spearman rank correlation coefficient of 0.37. The point was raised previously that if the extent-of-disease information is missing differentially across volume groups, then inclusion of extent-of-disease factors in the model could bias the results. However, differences by volume groups in

### Table 1. Characteristics of Women Operated on for Early-Stage Breast Cancer, by Surgeon Volume Category

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Overall</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of patients</td>
<td>12,216</td>
<td>4,524</td>
<td>4,456</td>
<td>3,236</td>
</tr>
<tr>
<td>No. of surgeons</td>
<td>1,856</td>
<td>1,325</td>
<td>384</td>
<td>147</td>
</tr>
<tr>
<td>Age, mean, y</td>
<td></td>
<td>75.9</td>
<td>75.6</td>
<td>75.3</td>
</tr>
<tr>
<td>White race</td>
<td></td>
<td>84.0</td>
<td>89.7</td>
<td>93.3</td>
</tr>
<tr>
<td>Per capita income of zip code (mean, × $1000)</td>
<td>17.2</td>
<td>17.8</td>
<td>19.9</td>
<td>18.1</td>
</tr>
<tr>
<td>Comorbidity index</td>
<td></td>
<td>0</td>
<td>1</td>
<td>&gt; 2</td>
</tr>
<tr>
<td>Lymph node status</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ER and PR status</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hospital volume, cases</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Abbreviations: ER, estrogen receptor; NA, data not applicable; PR, progesterone receptor.

### RESULTS

The 12,216 women who had undergone surgical treatment of stage I or II breast cancer were operated on by 1,856 different surgeons, of whom 1,325 were low-volume, 384 were medium-volume, and 147 were high-volume surgeons (Table 1). Patients treated by high-volume surgeons were slightly younger, were more likely to be white, and resided in more affluent areas than those treated by low-volume surgeons. The patients of high-volume surgeons had lower levels of comorbid illness, had tumors that were on average 1 to 2 mm smaller, and were more likely to be treated in high-volume hospitals. The relationship of surgeon and hospital volume was moderate, with a Spearman rank correlation coefficient of 0.37. The point was raised previously that if the extent-of-disease information is missing differentially across volume groups, then inclusion of extent-of-disease factors in the model could bias the results. However, differences by volume groups in

©2007 American Medical Association. All rights reserved.
missing extent-of-disease factors were small and inconsistent in direction (Table 1).

**RELATIONSHIP OF SURGEON VOLUME AND MORTALITY**

During the 50-month median follow-up, 2753 women (22.5%, or 54.4 per 1000 person-years) died; of these women, 760 (6.2%, or 15.6 per 1000 person-years) died of breast cancer and 1894 (15.3%, or 38.8 per 1000 person-years) died of other causes. For 99 women, no cause of death was available. In unadjusted analyses (Table 2), the difference in mortality rates between patients of high- and low-volume surgeons was 18.7 deaths per 1000 patient-years overall, 4.4 deaths per 1000 patient-years for breast cancer mortality, and 14.3 deaths per 1000 patient-years for other-cause mortality. The survival curve for the patients of medium-volume surgeons was within the boundaries of the high- and low-volume curves throughout the period of observation (Figure).

**ADJUSTED SURVIVAL MODELS OF SURGEON VOLUME AND MORTALITY**

In the Cox proportional hazards models, adjusting for patient demographic factors, extent of disease, comorbidity, and hospital volume, the relationship of surgeon volume to overall mortality was significant, with a hazard ratio of 0.86 (95% confidence interval [CI], 0.77-0.97) for patients of high-volume surgeons and a hazard ratio of 0.94 (95% CI, 0.85-1.03) for patients of medium-volume surgeons. However, surgeon volume was not significantly related to breast cancer mortality (Table 3). In contrast, surgeon volume was significantly related to non-breast cancer mortality, with a 14% lower hazard of death among patients of high-volume surgeons (Table 3).

The extent-of-disease factors were much more strongly related to breast cancer-specific mortality than to non-breast cancer mortality (Table 3). The level of comorbid illness had a stronger association with non-breast cancer mortality than with breast cancer mortality. The most common cause of non-breast cancer mortality was cardiovascular diseases, which accounted for the deaths of 961 women (7.9%). The relationship of surgeon volume of breast cancer cases to death from cardiovascular disease was quite significant, with a hazard ratio of 0.79 (95% CI, 0.65-0.96) for patients of high-volume surgeons and 0.89 (95% CI, 0.76-1.05) for patients of medium-volume surgeons.

In this study of 12,216 women with early-stage breast cancer, those operated on by high-volume surgeons were substantially less likely to die than those treated by low-volume surgeons. This finding is similar to that reported by other investigators. The usual interpretation of such a finding is that volume is a marker for quality of care. However, our disease-specific mortality results call that interpretation into question. The patients of high-volume surgeons were not less likely to die of breast cancer. Rather, they were significantly less likely to die of other causes.

This finding is consistent with the possibility of a non-specific effect, perhaps selection bias, in the types of patients cared for by high-volume surgeons as an explanation for at least some of the observed surgeon volume-mortality relationship. We observed clear differences among the patients cared for by high-, medium-, and low-volume surgeons, particularly in terms of sociodemographic characteristics and level of comorbidity. These factors, and other unmeasured factors, could be confounding the surgeon volume-outcome relationship.

With respect to possible selection by socioeconomic status, the socioeconomic factors were available for this study only at the zip code level. Even with individual-level information, it is difficult, perhaps impossible, to...
control fully for socioeconomic status. While we used current statistical techniques to control for selection bias, it is likely that these methods could not completely control for such bias. In addition, 1 recent study found low socioeconomic status to be associated with impaired functional capacity and worse all-cause mortality, which could explain the relationship we observed between breast cancer surgery volume and cardiovascular mortality.

Other explanations might also exist for these findings. High-volume high-quality surgeons may refer to other high-quality physicians, accounting for the reduction in non–breast cancer mortality. Even if so, however, one would expect the surgeon volume to breast cancer mortality relationship to be at least as strong as the relationship of surgeon volume to other mortality. Another possible explanation is that the breast cancer mortality model has several specific extent-of-disease predictors, while the non–breast cancer mortality model has only a general comorbidity measure. However, although better specification of overall mortality predictors (ie, general comorbidity) might reduce or eliminate the relationship of surgeon volume to other mortality, such specification would be unlikely to strengthen the surgeon volume–breast cancer mortality relationship given that the available comorbidity information shows that the low-volume surgeons care for sicker patients (Table 1). Therefore, if selection bias based on comorbid illness accounts for our findings, better controlling for the factor would likely reduce the association of surgeon volume to overall mortality.

Another limitation is that the study included only patients 65 years or older, and the surgeon volume was determined by measuring the number of operations performed on women in this age group. However, about half of all incident breast cancer cases occur in women 65 years or older, and there is no reason to believe case volume would be systematically biased by measuring it in this age group. A number of other investigators have also used this approach to determining volume. The study of Medicare pa-

### Table 3. Relationship of Surgeon Volume to Disease-Specific Mortality

<table>
<thead>
<tr>
<th>Patient Characteristic</th>
<th>Breast Cancer Mortality</th>
<th>Non–Breast Cancer Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgeon volume</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>1 [Reference]</td>
<td>1 [Reference]</td>
</tr>
<tr>
<td>Medium</td>
<td>1.00 (0.84-1.20)</td>
<td>0.91 (0.81-1.02)</td>
</tr>
<tr>
<td>High</td>
<td>0.94 (0.76-1.16)</td>
<td>0.86 (0.75-0.98)</td>
</tr>
<tr>
<td>Tumor size, cm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 2</td>
<td>1 [Reference]</td>
<td>1 [Reference]</td>
</tr>
<tr>
<td>2-5</td>
<td>2.38 (2.03-2.79)</td>
<td>1.39 (1.26-1.54)</td>
</tr>
<tr>
<td>&gt; 5</td>
<td>4.38 (2.90-6.63)</td>
<td>1.22 (0.86-1.74)</td>
</tr>
<tr>
<td>Lymph node status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>1 [Reference]</td>
<td>1 [Reference]</td>
</tr>
<tr>
<td>Positive</td>
<td>2.61 (2.39-3.31)</td>
<td>1.08 (0.95-1.22)</td>
</tr>
<tr>
<td>Hormone receptor status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>1 [Reference]</td>
<td>1 [Reference]</td>
</tr>
<tr>
<td>Positive for ER and/or PR</td>
<td>0.43 (0.36-0.51)</td>
<td>0.88 (0.76-1.03)</td>
</tr>
<tr>
<td>Tumor grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well or moderately differentiated</td>
<td>1 [Reference]</td>
<td>1 [Reference]</td>
</tr>
<tr>
<td>Poorly differentiated or undifferentiated</td>
<td>2.31 (1.94-2.74)</td>
<td>1.12 (1.01-1.26)</td>
</tr>
<tr>
<td>Comorbidity index</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1 [Reference]</td>
<td>1 [Reference]</td>
</tr>
<tr>
<td>1</td>
<td>1.23 (1.01-1.48)</td>
<td>1.73 (1.54-1.95)</td>
</tr>
<tr>
<td>≥ 2</td>
<td>1.44 (1.11-1.86)</td>
<td>3.58 (3.11-4.11)</td>
</tr>
<tr>
<td>Hospital volume, cases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 20</td>
<td>1 [Reference]</td>
<td>1 [Reference]</td>
</tr>
<tr>
<td>20 to &lt; 40</td>
<td>0.83 (0.68-1.01)</td>
<td>0.99 (0.88-1.12)</td>
</tr>
<tr>
<td>≥ 40</td>
<td>0.84 (0.68-1.03)</td>
<td>0.89 (0.78-1.02)</td>
</tr>
</tbody>
</table>

Abbreviations: See Table 1.

a Data are given as relative risk (95% confidence interval). Results from Cox proportional hazards models including 12 216 Medicare patients undergoing surgery for stage I or II breast cancer. Each model also adjusts for patient age, race; Surveillance, Epidemiology, and End Results site; zip code per capita income; propensity for care by a medium- or high-volume surgeon; and surgeon clustering.
tients permitted analysis of a large population-based cohort of patients and surgeons, an important strength.

In summary, this study found a significant relationship between surgeon volume and non–breast cancer mortality, but not between surgeon volume and breast cancer mortality. This finding is evidence of possible selection bias, which could be based on socioeconomic status or general comorbidity. It is possible that the beneficial effect of high surgeon volume on breast cancer mortality has been overstated. Future studies should further evaluate this possibility and rigorously consider whether selection bias could account for volume-outcome relationships for other health conditions.

Accepted for Publication: April 24, 2007.

Correspondence: Ann Butler Nattinger, MD, MPH, Department of Medicine, Medical College of Wisconsin, 9200 W Wisconsin Ave, FEOB Ste 4200, Milwaukee, WI 53226 (anattinger@mcw.edu).

Author Contributions: Study concept and design: Nattinger and Laud. Acquisition of data: Nattinger. Analysis and interpretation of data: Nattinger, Laud, Sparapani, Zhang, Neuner, and Gilligan. Drafting of the manuscript: Nattinger and Laud. Critical revision of the manuscript for important intellectual content: Nattinger, Laud, Sparapani, Zhang, Neuner, and Gilligan. Statistical analysis: Laud, Sparapani, and Zhang. Obtained funding: Nattinger. Study supervision: Nattinger and Laud.

Financial Disclosure: None reported.

Funding/Support: This study was supported by grant R01-CA081379 from the National Cancer Institute, Public Health Services (Dr Nattinger), and grant K08-AG021631 from the National Institute on Aging, Public Health Services (Dr Neuner).

Role of the Sponsor: The funding bodies had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript.

Additional Contributions: James S. Goodwin, MD, provided helpful comments on an earlier draft of the manuscript; the Applied Research Program, National Cancer Institute, the Office of Research, Development, and Information, Centers for Medicare & Medicaid Services, Information Management Services, Inc, and the SEER Program Tumor registries assisted in the creation of the SEER-Medicare database.

REFERENCES


©2007 American Medical Association. All rights reserved.