United States General Accounting Office

GAO

Report to the Chairman, Subcommittee on Human Resources and Intergovernmental Relations, Committee on Government Operations, House of Representatives

November 1994

BREAST CONSERVATION VERSUS MASTECTOMY

Patient Survival in Day-to-Day Medical Practice and in Randomized Studies

GAO/PEMD-95-9

GAO

United States
General Accounting Office
Washington, D.C. 20548

Program Evaluation and
Methodology Division

B-257065

November 15, 1994

The Honorable Edolphus Towns
Chairman, Subcommittee on Human Resources
and Intergovernmental Relations
Committee on Government Operations
House of Representatives

Dear Mr. Chairman:

As you know, the effectiveness of breast-conservation therapy (that is, lumpectomy and related treatments) is a topic of concern to many breast cancer patients and physicians. Experts who considered the results of randomized clinical studies in 1990 concluded that patient survival rates following mastectomy and breast-conservation therapy were "equivalent."[1] (See National Institutes of Health (NIH), 1991.) But a key question is: Have results been similar in day-to-day medical practice—with its less certain quality of treatments? To address this question, we developed a three-step analysis, the results of which are reported here, at your request.

The first step of our analysis consisted of examining 5-year survival results separately for single-center and for multicenter randomized studies (since the latter more closely resemble day-to-day medical practice, as explained below). In step 2, we examined database records for breast cancer patients treated outside randomized studies. Specifically, we analyzed a set of medical practice cases that had been selected to be comparable to the kinds of patients covered in the randomized studies.[2] (The main characteristics of the patient population examined here are age 70 or younger, node-negative, with tumors 4 cm or smaller.[3]) Step 3 consisted of quantitative comparisons across study designs and a consideration of the strength of the evidence.

Results in Brief

Our three-step analysis indicated that—for the kinds of patients we examined—the effectiveness of breast-conservation therapy has, on average, been similar to that of mastectomy in community medical

[1]In a randomized study, patients do not choose their own treatments. Rather, each patient is randomly assigned to one of two treatments—in this case, breast conservation or mastectomy—in order to ensure unbiased comparison of outcomes.

[2]We do not address here issues of generalizability to broader patient populations.

[3]Node-negative patients are those whose breast cancer has not spread to the lymph nodes beneath the arm.

practice as well as in randomized studies. Specifically, for medical practice cases, the adjusted 5-year survival rates (averaged across all selected patients) were 86.3 percent for breast-conservation patients and 86.9 percent for mastectomy patients. These results clearly correspond to the results of multicenter randomized studies (88 percent 5-year survival for breast conservation and 88 percent for mastectomy). Single-center studies reported somewhat higher survival for both treatment groups. Thus, on average, for breast cancer patients of physicians in regular medical practice who are similar to patients in randomized studies, there appears to be no appreciable risk associated with selecting breast-conservation therapy rather than mastectomy.

Background

Breast-conservation therapy involves a number of physician decisions not required for mastectomy, including the selection of patients for breast conservation, the amount of tissue to be removed from the area surrounding the tumor, the details of administering radiation, and so forth. (See Sacks and Baum, 1993; Winchester and Cox, 1992; Harris et al., 1990; NIH, 1991.) And since breast-conservation therapy involves radiation, its implementation would logically vary depending upon the availability of appropriate radiation equipment and expertise in operating that equipment. Breast-conservation therapy also requires "careful long-term breast monitoring" in order to identify and treat local recurrences in the breast that was subjected to lumpectomy (NIH, 1991).

All these treatment-implementation factors can potentially affect breast-conservation patients' survival—and may not be the same in randomized studies and in medical practice. At least, the typical treatments given in day-to-day medical practice could fall short of the presumably consistent and high-quality treatments provided by a single prestigious research center, such as the National Cancer Institute (NCI). Some randomized studies are conducted at single centers, while others are conducted at diverse sites (that is, multiple centers). To more closely approximate day-to-day medical practice, multicenter studies have, in some instances, intentionally involved "community surgeons."[4] For this reason—and also because the treatments given in multicenter studies may vary from one center to another—multicenter studies' results may more closely approximate results in medical practice than the results of

[4]Of the three multicenter studies included in this report, one includes about 90 centers in the United States and Canada. Another, conducted in one European country, is also broad-based; although it began with only a few centers, eventually 20 hospitals were involved, and these 20 are responsible for about 50 percent of the breast surgeries conducted in that country. A third involves only eight hospitals, but these are located in different countries and different languages are involved.

single-center studies at prestigious institutions. But unlike medical practice, both single-center and multicenter studies stipulate that participating physicians follow a set of prespecified procedures. The question remains, then, as to whether or not breast conservation therapy has produced results similar to mastectomy in day-to-day medical practice.[5]

Randomized clinical studies are the "gold standard" of medical research. Random assignment essentially equates patients in the two treatment groups. Because the two groups should not differ on variables related to cancer survival, their outcomes can be directly compared, and any difference in survival can be attributed to the difference in treatment. In contrast, the statistical analysis of cases from a medical practice database represents a potential "window" on how well breast-conservation therapy has, in fact, worked in community medical practice. But the results of such analyses may be less conclusive because of their vulnerability to hidden selection bias.[6] (See Byar, 1980; Office of Technology Assessment, 1994.) Briefly, in day-to-day medical practice, patients and physicians freely choose between treatments; a database analyst must, therefore, attempt to control for the potentially differing characteristics of patients who received breast-conservation therapy and those who received mastectomy. In this report, we have made all possible efforts to minimize the impact of selection bias, as described below.

Scope and Methodology

The analyses presented here are based on a unique combination of meta-analysis (to summarize randomized studies' results),[7] statistical analysis of records from a medical practice database, and cross design comparison of results. To our knowledge, this is the first time such an approach has been used in the area of breast cancer treatment.

[5]One previous analysis of a medical practice database has been reported (Lee-Feldstein, 1994). That study found that survival rates following breast conservation were at least as good as those following mastectomy. Unlike the analyses reported here, that study covered just one county in California and did not include controlled comparisons to randomized studies' results.

[6]We use the term "selection bias" to indicate both (1) a tendency for patients with better prognoses to select (or be selected for) a particular treatment—or the process by which this occurs—and (2) the resultant distortion of an estimated treatment effect. "Hidden selection bias" refers to the continued distortion of an estimated treatment effect that may remain **after** the analyst has used statistical procedures to adjust for known, measured sources of bias.

[7]Meta-analysis refers to the quantitative summary of results across several individual studies that have addressed essentially the same research question. Often, the treatment effects observed in individual randomized studies are statistically combined. (See Dickersin and Berlin, 1992; Ellenberg, 1988; Louis, Fineberg, and Mosteller, 1985.)

In all analyses presented here, breast-conservation therapy is defined as including lumpectomy, nodal dissection, and radiation.[8] With respect to time frame, the randomized studies enrolled and treated patients from 1972 to 1989, and the medical practice cases selected for this analysis were diagnosed from 1983 to 1985.[9] Because of limitations in the medical practice database (discussed in appendix I), all our analyses use the outcome criterion of 5-year survival and examine node-negative patients only.[10]

The medical practice data were drawn from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) database. SEER archives records for almost all cancer patients residing in five states—Connecticut, Hawaii, Iowa, New Mexico, and Utah—and four metropolitan areas—Atlanta, Detroit, San Francisco-Oakland, and Seattle -Puget Sound. (See Hankey et al., 1992.)

Our analysis consisted of three major steps.

- In step 1, we performed a meta-analysis to summarize randomized studies' results and obtain summary figures that can be compared to medical practice results. We conducted meta-analyses separately for the single-center studies and for the more generalizable multicenter studies to determine if similarity of survival following breast-conservation therapy and mastectomy holds for both kinds of studies.
- In step 2, we obtained information on the survival of breast-conservation and mastectomy patients in day-to-day medical practice. Specifically, from the SEER database, we drew records for a relatively homogeneous set of patients who, on the basis of several characteristics, were comparable to those enrolled in randomized studies. For this group of SEER patients, we conducted an analysis of survival following breast-conservation therapy and mastectomy. SEER results were adjusted for tumor size and several other variables so that patients who had received breast-conservation therapy would be "matched" to those who had received mastectomy. The matching was intended to minimize the effects of differing characteristics of patients who received breast-conservation therapy and mastectomy. In

[8]Implementations of lumpectomy, radiation, and nodal dissection do vary. Notably, lumpectomy ranges from removal of the tumor itself to quadrantectomy—removal of one-quarter of the breast. Nodal dissection refers to the removal of the lymph nodes beneath the arm that is adjacent to the breast in which the tumor is located. Some or all of the nodes may be removed.

[9]To the extent that patients or treatments have changed since this time frame, results may also differ.

[10]That is, data limitations meant that it was not possible to cover node-positive patients or to examine such outcomes as disease recurrence, quality of life, or longer term survival.

addition, a sensitivity analysis was performed to check for selection bias on life-threatening factors unrelated to cancer (such as heart-disease).

- In step 3, we compared (1) the summary results for the single-center and multicenter randomized studies to (2) the results of our analysis of cases selected from the SEER medical practice data. We also considered the logic of our analyses and, in particular, whether the resulting evidence was sufficient to conclude that—in day-to-day medical practice— breast-conservation therapy has been followed by survival similar to that observed for mastectomy. Throughout step 3, we drew on the principles of "cross design synthesis."[11]

In this report, we use the term "similar" when the observed difference between the survival rates (1) is not statistically significant **and** (2) has an absolute value of less than 1.5 percentage points.[12] Conversely, when a comparison of survival rates shows a difference of 1.5 percentage points or larger—**and** that difference is also statistically significant—we state that one rate is higher (or lower) than the other.[13]

Analysis of Single-Center and Multicenter Studies

Step 1 (the analysis of randomized studies) began with the identification of relevant single-center and multicenter studies through bibliographic searches and a survey of U.S. breast cancer researchers.[14] Our inclusion criteria were as follows:

- randomization of enrolled patients to alternative treatments—breast-conservation therapy or mastectomy;
- breast-conservation therapy that included lumpectomy, nodal dissection, and radiation;[15]

[11]See GAO, 1992; Droitcour, Silberman, and Chelimsky, 1993.

[12]We do not use the term "equivalent" because we do not believe it is possible to prove that survival rates following two treatments are absolutely identical based on probabilistic study results. An additional, more technical reason for avoiding the word "equivalent" in this context is discussed in appendix I.

[13]When a difference between survival rates is 1.5 percentage points or larger but not statistically significant, we term the result a nonsignificant pattern—that is, inconclusive owing to the lack of statistical significance. (See appendix I.)

[14]For the United States, our intention was to be comprehensive. For studies conducted outside the United States, we did not attempt to include those that were unpublished or that had not been published in English.

[15]Lumpectomy with nodal dissection plus radiation is the form of breast-conservation therapy recommended by the NIH Consensus Development Conference. (See NIH, 1991.) Two English studies, which did not include nodal dissection, were thereby excluded. These studies are notable in that their results indicated that breast-conservation therapy was less effective than mastectomy. (See appendix I.)

- no confounding treatments (such as the administration of an additional therapy to one treatment group);
- availability of 5-year survival rates by treatment group among node-negative patients (either previously published in a scholarly research journal or provided at our request); and
- published in English (if a non-U.S. study).

Six studies—three single-center and three multicenter studies—met these criteria. (See table 1.) Almost 2,500 node-negative breast cancer patients were enrolled and treated in these randomized studies.

Table 1: Six Randomized Studies Comparing Breast Conservation and Mastectomy

Short name of study	Years of patient enrollment	Number of node-negative patients[a]	Formal study name (published data source)	Other data source[b]
Single-center				
U.S.-NCI	1979 to 1987	141	U.S. National Cancer Institute (Lichter et al., 1992; Straus et al., 1992)	Seth Steinberg, NCI
Milan	1973 to 1980	520	National Cancer Institute in Milan[c] (Veronesi et al., 1986a; 1986b; 1981)	Umberto Veronesi, National Cancer Institute in Milan
French	1972 to 1980	121	Institut Gustave-Roussy (IGR) (Sarrazin et al., 1989; 1984; 1983)	Daniele Sarrazin and R. Arriagada, IGR
Multicenter				
Danish	1983 to 1989	577	Danish Breast Cancer Cooperative Group (DBCG) (Blichert-Toft et al., 1992; 1988)	Knud West Andersen, DBCG
EORTC	1980 to 1986	475	European Organization for Research and Treatment of Cancer (van Dongen et al., 1992a; 1992b)	J.A. van Dongen, Netherlands Cancer Institute; Francoise Mignolet, EORTC Data Center
U.S.-NSABP[d]	1976 to 1984	639	National Surgical Adjuvant Breast Project (Stablein, 1994a; 1994b)	Donald Stablein and Boris Freidlin, EMMES Corp.

[a]The number of node-negative patients refers to those who **tested** node-negative.

[b]Additional information was provided to us through personal communications.

[c]Istituto Nazionale per lo Studio e la Cura dei Tumori, Milan, Italy.

[d]The number of node-negative patients listed for U.S.-NSABP is from recalculations, published in March 1994, which exclude data from a center at which fraud has been alleged. The specific number of node-negative patients was provided by Freidlin (1994). This number includes two treatment groups— (1) mastectomy and (2) lumpectomy with nodal dissection plus radiation. The lumpectomy group that did not receive radiation is excluded.

The **treatment effect**—that is, the effect of breast-conservation therapy relative to mastectomy—is represented by a comparison of survival following breast-conservation therapy to survival following mastectomy. (See table 2.)

Table 2: Treatment Effects Estimated in Six Randomized Studies[a]

Study[c]	5-year survival rates		Difference in rates	Odds ratio[b]	
	Breast conservation	Mastectomy	Breast conservation minus mastectomy	Estimate	Confidence interval
Single-center					
U.S.-NCI	93.9% (n = 74)	94.7% (n = 67)	–0.8%	0.85	.18 to 3.97
Milan	93.5% (n = 257)	93.0% (n = 263)	0.5%	1.04	.52 to 2.06
French	94.9% (n = 59)	95.2% (n = 62)	–0.3%	0.95	.18 to 4.90
Multicenter					
Danish	87.4% (n = 289)	85.9% (n = 288)	1.5%	1.13	.69 to 1.85
EORTC[d]	89% (n = 238)	90% (n = 237)	–1%	0.93	.51 to 1.69
U.S.-NSABP	89.0% (n = 330)	88.0% (n = 309)	1.0%	1.11	.68 to 1.81

[a]Node-negative patients only.

[b]We define the odds ratio as the odds of surviving (to not surviving) for breast-conservation therapy divided by the odds of surviving (to not surviving) for mastectomy. A ratio below 1 (such as 0.85) favors mastectomy; a ratio larger than 1 favors breast conservation. To calculate the odds ratios, the numbers of patients who died and survived in each study were estimated from percentages and rounded. To maintain consistency with the meta-analysis (shown in table 3), "effective n's" were used in calculations of the numbers who died and survived in the U.S.-NCI, Danish, and EORTC studies. (See appendix I.) Results may vary slightly because of rounding.

[c]The Milan and French studies did not have any patients who refused the assigned treatment because patients were randomized on the operating table (following tumor removal and determination of whether the tumor met size requirement for the study). The estimates from the Danish and the EORTC analyses are based on "intention-to-treat" analyses; that is, all patients were analyzed as having received the treatment to which they were assigned. (In the Danish study, 10 percent of randomized patients subsequently chose the opposite treatment; 3 percent of EORTC patients received the opposite treatment.) The estimates for the U.S. studies are based on those patients who accepted assigned treatments. (In the U.S.-NCI study, 6 percent withdrew following randomization; in the U.S.-NSABP analyses, 8 percent refused the assigned treatment.)

[d]The EORTC estimates are only available rounded to the nearest percentage point.

This comparison is made

- by subtracting the 5-year survival rate for mastectomy patients from the 5-year survival rate for breast-conservation patients to determine the difference between the rates; and
- by calculating the odds ratio (dividing the odds of surviving with breast-conservation therapy by the odds of surviving with mastectomy).[16]

As indicated in table 2, the breast-conservation and mastectomy treatment groups experienced similar survival rates in each of the studies, and the odds ratios are close to 1 (the point of equivalence).[17] The confidence intervals for the odds ratios all overlap 1, indicating no statistically significant difference in survival odds for the two treatments.[18] However, the confidence intervals surrounding these estimates are quite broad, indicating a lack of precision in the individual-study estimates. (The U.S.-NSABP figures in table 2 are taken from recalculations published by an NCI contractor in March 1994. The recalculations were published following charges of fraudulent data collection at one U.S.-NSABP center; they exclude the data from that center.)

A meta-analysis combining the results for node-negative patients across studies gives more precise estimates of the treatment effect. Table 3 shows meta-analysis results summarizing the treatment effect for single-center studies, multicenter studies, and both types of studies taken together. In addition, table 3 shows meta-analysis results calculated in two ways: (1) including the U.S.-NSABP recalculations published in March 1994 and (2) omitting U.S.-NSABP results entirely.[19]

[16]An odds ratio of 1 indicates the point of equivalence.

[17]Only one study (the Danish multicenter study) was characterized by a difference in survival rates (breast conservation versus mastectomy) as great as 1.5 percentage points, and in that instance, the difference favored breast conservation.

[18]Appendix I defines the confidence interval.

[19]We performed separate calculations with and without U.S.-NSABP because following the March 1994 publication of U.S.-NSABP recalculations (which omitted data from the center charged with fraud), NCI undertook a **multicenter audit** of that study—and the results of the multicenter audit had not been reported as of this writing.

Table 3: Meta-Analyses: Treatment Effects Estimated for Single-Center and Multicenter Studies[a]

Study category	5-year survival rate[b]		Difference in rates	Common odds ratio[c]	
	Breast conservation	Mastectomy	Breast conservation minus mastectomy	Estimate	Confidence interval
Single-center					
U.S.-NCI, Milan, French	93.7%	93.7%	0.0%	1.00[d]	.55 to 1.79
Multicenter					
Danish, EORTC, U.S.-NSABP	89%	88%	1%	1.07[e]	.79 to 1.44
Omitting U.S.-NSABP	88%	88%	0%	1.05[f]	.72 to 1.53
All six studies	90%	90%	0%	1.05[g]	.81 to 1.38
Five studies (omitting U.S.-NSABP)	91%	90%	0%[h]	1.03[i]	.75 to 1.42

[a]Node-negative patients only. In calculating the combined-studies survival rates and odds ratios, the number died and the number survived were estimated from percentages and rounded to the nearest whole number; results shown may vary slightly because of rounding. With respect to the presentation of combined-studies survival rates, the following rounding rule was applied: When results for all studies were available to the nearest tenth of a percentage point, results are reported to the nearest tenth of a percent. However, because one multicenter study's results were available only to the nearest whole percent, survival estimates involving this study were rounded to the nearest whole percent (to avoid implying a greater degree of precision than warranted).

[b]Weighted average survival rate. In calculating the weighted average survival rate for breast-conservation patients, the size of the total study (relative to the size of all relevant studies taken together) was used as the weight. The same is true for the calculation of the weighted average for patients who received mastectomy. Thus, a particular study's results had the same weight in calculations for survival following breast conservation and for survival following mastectomy. For the U.S.-NCI, Danish, and EORTC studies, "effective n's" were used. (See appendix I.) Results shown may vary slightly because of rounding in the use of these procedures.

[c]The odds ratio is defined here as the odds of surviving (to not surviving) for breast-conservation therapy divided by the odds of surviving (to not surviving) for mastectomy. A ratio below 1 favors mastectomy; a ratio larger than 1 favors breast-conservation therapy.

[d]Test for homogeneity of odds ratios: Breslow and Day (B-D) Statistic = 0.58; p = .97; no significant heterogeneity.

[e]B-D Statistic = 0.29; p = .87; no significant heterogeneity.

[f]B-D Statistic = 0.25; p = .62; no significant heterogeneity.

[g]B-D Statistic = 0.39; p = 1.00; no significant heterogeneity.

[h]Before rounding to the nearest whole percent, the 5-year survival estimates for the **five studies** combined were 90.5% (breast conservation) and 90.3% (mastectomy) with a difference between percentages of 0.2%. These figures rounded to 91%, 90%, and a difference of 0 percentage points. (The difference for all six studies—shown in the previous row of the table as 0%—was rounded from 0.4%)

[i]B-D Statistic = 0.33; p = .99; no significant heterogeneity.

Similar rates of 5-year survival characterized the breast-conservation therapy and mastectomy groups—not only in single-center studies (93.7 percent for breast-conservation patients and 93.7 percent for mastectomy patients), but also in multicenter studies, which may more closely approximate medical practice (88 percent for breast conservation and 88 percent for mastectomy, omitting U.S.-NSABP). Again, the odds ratios are close to 1, and the confidence intervals all overlap 1, indicating no statistically significant difference for any group of randomized studies.

Finally, referring again to tables 2 and 3, the 5-year survival rates appear to be higher in single-center studies than in multicenter studies. This could be because of more effective treatments in single-center studies, varying tumor-size limits across the studies, or hidden cross-study differences in patient prognoses prior to treatment.[20] (Step 3 presents more precise comparisons of combined-treatments survival rates.)

Analysis of SEER Medical Practice Data

Because the purpose of this report is to determine whether the treatment effect in day-to-day medical practice corresponds to the treatment effects observed in the single-center and multicenter studies, we would ideally "compare like with like." Therefore, step 2 (analysis of the medical practice data) began with the selection of SEER patients who, on the basis of their characteristics, would have been covered by randomized studies.[21] Table 4 shows the specific criteria we used in selecting SEER cases; the resulting SEER dataset included 5,326 cases that we believe are at least roughly comparable to the participants in randomized studies.[22] (Appendix I assesses the kinds of patients who participated in randomized studies and discusses SEER cases lost to follow-up.)

[20]Two single-center studies (the Milan and French studies) had a 2-cm tumor-size limit, whereas the U.S.-NCI study and all three multicenter studies had a limit—either stated or in effect—of 4 cm. See appendix I for data on patient characteristics in the six randomized studies.

[21]That is, the SEER patients we selected would have met major formal—and informal—eligibility criteria of some or all of the randomized studies. (See table I.3 in appendix I.) As mentioned in the previous footnote, the majority of the randomized studies had a tumor-size limit of 4 cm; we therefore used the 4-cm limit in selecting SEER cases. One example of a potential difference between the selected SEER patients and those in randomized studies is that some of the former probably would not have accepted random assignment of treatment.

[22]All patients included in our analyses tested node-negative. That is, when the requirement that SEER patients be coded node-negative is **combined with the requirement for nodal dissection** (for both breast-conservation and mastectomy patients), the result is, in effect, the elimination of any patients who were coded node-negative on the basis of a clinical examination alone. The same is true for the node-negative patients from the randomized studies. (This is important because some patients who appear to be node-negative on the basis of a clinical examination later test node-positive following nodal dissection and laboratory tests.)

Table 4: Criteria Used to Select SEER Cases

Type of criterion	Specific criterion
Patient characteristics	Infiltrating or invasive early-stage cancer; no in situ cases[a]
	Node-negative
	Tumor neither invading skin nor attached to pectoral muscle
	Type of cancer: infiltrating duct carcinoma or adenocarcinoma (NOS)[b]
	Tumor 4 cm or smaller
	No previous cancer
	Age 70 or younger
Treatment the patient received	If breast conservation: lumpectomy, nodal dissection, and radiation
	If mastectomy: removal of breast (but not the pectoral muscle) plus nodal dissection but no radiation
Data completeness	Complete data on all relevant treatment, control, and outcome variables (See appendix I.)

[a]Early-stage cancer means that there are no known distant or regional metastases and no local spread beyond the breast, breast skin, and pectoral muscle. The term in situ (noninvasive) refers to: "cancer in its earliest stage, that is, confined to the place or site where it started . . . . Some in situ cancers are considered precancerous." (Altman and Sarg, 1992, p. 136).

[b]NOS, not otherwise specified. These two very similar types of cancer are denoted by codes 8500 and 8140 in the International Classification of Diseases for Oncology (Percy, van Holten, and Muir, 1990).

As described below, our statistical analysis of the selected SEER cases used "propensity-score" adjustments (Rosenbaum and Rubin, 1984) that essentially "matched" the kinds of patients who received breast-conservation therapy and mastectomy on demographic characteristics and tumor size. Using these adjustments, we found that, on average, similar patient survival followed the two treatments.

Treatment Effect for SEER Cases

To achieve matched groups of patients for the two treatments, the 5,326 SEER cases were first divided into five quintiles, as shown in table 5. Patients were assigned to these quintiles based on their propensity scores, which were calculated to indicate each patient's likelihood of receiving breast-conservation therapy.[23] Patients in the first quintile shown in table 5 have very low propensity scores; that is, they are the kinds of patients who were quite **unlikely** to receive breast-conservation therapy. (An example of a patient with an extremely low propensity score would be a woman in her sixties, living in Iowa, diagnosed in 1983—the earliest year examined

[23]Appendix I describes the propensity-score calculations.

here—with a tumor sized 3 to 4 cm.) By contrast, patients assigned to each successive quintile were more likely to receive breast-conservation therapy. (An example of a patient with a relatively high propensity score would be under 40 years old, non-Asian, living in the San Francisco-Oakland or the Seattle-Puget Sound area and diagnosed in 1985—the most recent year examined—with a very small tumor.)

In table 5, 5-year survival estimates are shown separately for breast-conservation patients and for mastectomy patients in each quintile. **Within each quintile**, patients are homogeneous, and the survival rates for the two treatments represent an estimate of the treatment effect for that quintile. The bottom rows of table 5 show the overall survival rates used to calculate the treatment effect for all selected SEER cases taken together. These summary rates, which are termed "adjusted across quintiles," are clearly similar to each other: 86.3 percent for breast-conservation therapy and 86.9 percent for mastectomy.

Table 5: Treatment Effect Estimated for SEER Cases, by Quintile

Quintile	Treatment	Number of node-negative patients	5-year survival[a] Estimate	5-year survival[a] Standard error
1	Breast conservation	56	85.6%	4.7%
	Mastectomy	1,008	86.7%	1.1%
2	Breast conservation	106	82.8%	3.7%
	Mastectomy	964	83.4%	1.2%
3	Breast conservation	193	85.2%	2.6%
	Mastectomy	866	88.8%	1.1%
4	Breast conservation	289	88.7%	1.9%
	Mastectomy	778	87.3%	1.2%
5	Breast conservation	462	89.0%	1.4%
	Mastectomy	604	88.5%	1.3%
Adjusted across quintiles[b]	Breast conservation	1,106	86.3%	1.4%
	Mastectomy	4,220	86.9%	0.5%

[a]As described in appendix I, the estimates for each quintile are weighted averages, which were calculated to adjust for minor differences between breast-conservation and mastectomy patients within each quintile. Standard errors were calculated as specified in Mosteller and Tukey (1977).

[b]The survival percentages shown above for patients receiving breast-conservation therapy in each of the five quintiles were averaged, with each percentage receiving an equal (1/5) weight; survival percentages for patients who received mastectomy were combined in the same way. Again, standard errors were calculated as specified in Mosteller and Tukey (1977).

The adjusted breast-conservation rate (86.3 percent) was calculated by combining the five separate quintile survival rates for breast-conservation patients—**giving each of the five rates an equal weight of one-fifth**. The adjusted mastectomy rate (86.9 percent) was calculated using analogous procedures. Thus, the adjusted survival rates are based on "matched" treatment groups; that is, the kinds of patients who were unlikely to receive breast-conservation therapy contribute equally to the breast- conservation and the mastectomy survival estimates—as do the kinds of patients who were much more likely to receive breast-conservation therapy. In this way, selection bias on measured variables was minimized.

As shown in table 6, the difference between the adjusted 5-year survival estimates for breast-conservation and mastectomy patients is just six-tenths of a percentage point, the odds ratio is relatively close to 1, and the confidence interval overlaps 1, indicating no statistically significant difference.[24] Thus, on average, the two treatments appear to produce similar results in day-to-day medical practice.

Table 6: Treatment Effect Estimated for SEER Cases: Difference in Survival Rates and Odds Ratio[a]

5-year survival rate[b]		Difference in rates	Odds ratio	
Breast conservation	Mastectomy	Breast conservation minus mastectomy	Estimate	Confidence interval[c]
86.3%	86.9%	–0.6%	.94	.75 to 1.14

[a]The odds ratio is defined as the odds of surviving (to not surviving) for patients who received breast-conservation therapy divided by the odds of surviving (to not surviving) for mastectomy patients.

[b]Survival rates for breast-conservation therapy and for mastectomy are the adjusted rates from table 5.

[c]This confidence interval was constructed using an estimate of the standard error of the odds ratio that was calculated using Woolf's method (Kahn and Sempos, 1989, citing Woolf, 1955). This calculation provides a conservative estimate of the error relative to the more complex approach presented by Rubin and Thomas (1992).

[24]The size of the odds ratio depends, in part, on the general level of the survival percentages. For example, if survival were close to the 50-percent level, a one-half-of-1-percentage-point difference in survival would translate to an odds ratio of 49.5/50.5 divided by 50/50—or .98, indicating that the odds of survival were **98 percent** as good with one therapy as with the other. But when survival rates are about 90 percent, a one-half-of-1-percentage-point difference translates to an odds ratio of 89.5/10.5 divided by 90/10—or .94, indicating that the odds are **94 percent** as good with one therapy as with the alternative. In this sense, for the patient population and outcome criterion examined in this report, odds ratios may seem exaggerated relative to the absolute size of the difference in survival rates.

However, referring again to table 5, the results shown for quintile 3 do **not** meet our criteria for use of the term "similar" because the observed (nonsignificant) difference between the survival rates is greater than 1.5 percentage points. According to our criteria, this nonsignificant pattern should be regarded as inconclusive.

A Further Check on Medical Practice Results

The propensity-score adjustments were intended to minimize selection bias on measured variables, such as tumor size and demographic characteristics. However, noncancer-related life-threatening illnesses or conditions, such as serious heart disease, were not measured in the SEER data and therefore could not be included in the propensity score. Such illnesses or conditions might at once influence treatment selection and limit 5-year survival—and could represent a form of selection bias not accounted for by the propensity scores.[25]

SEER data does, however, include codes for cause of death. Therefore, it was possible to check for selection bias on illnesses and conditions not related to cancer in the following way: We performed a sensitivity analysis in which we reproduced table 5 **omitting** patients who were coded as having died of illnesses and conditions unrelated to cancer within the 5-year interval. As indicated in table 7, with those patients omitted, the difference in survival following breast-conservation therapy and mastectomy is, on average, again within 1.5 percentage points of zero, and it is not statistically significant.

[25]For example, if women with serious heart disease are not selected for therapy that includes radiation in the chest area, patients selected for mastectomy would, as a result, be less likely to survive for 5 years than those selected for breast conservation—regardless of the effectiveness of their cancer treatment.

Table 7: Treatment Effect for SEER Cases, by Quintile, Omitting Patients Who Died of Causes Unrelated to Cancer[a]

Quintile[b]	Treatment	Number of node-negative patients	Rate of survival versus cancer death	Standard error[c]
1	Breast conservation	54	88.8%	4.3%
	Mastectomy	966	90.5%	0.9%
2	Breast conservation	102	86.0%	3.4%
	Mastectomy	917	87.7%	1.0%
3	Breast conservation	184	89.4%	2.3%
	Mastectomy	841	91.4%	1.0%
4	Breast conservation	279	92.0%	1.6%
	Mastectomy	742	91.5%	1.0%
5	Breast conservation	453	90.7%	1.3%
	Mastectomy	589	90.7%	1.1%
Adjusted across quintiles	Breast conservation	1,072	89.4%[d]	1.3%
	Mastectomy	4,055	90.4%[d]	0.5%

[a]Patients dying of causes other than cancer or of unknown or unrecorded causes within 5 years of diagnosis are omitted from this table. All those included either survived 5 years or are known to have died from cancer.

[b]Quintile based on propensity score.

[c]Standard errors calculated as specified in Mosteller and Tukey (1977). The difference of 1.0 percentage point favoring mastectomy (that is, 89.4 percent - 90.4 percent = –1.0 percent) is **not** significant at the .05 level.

[d]The odds ratio for these survival percentages is .90; the ratio is defined as the odds of surviving (to not surviving) following breast-conservation therapy to the odds of surviving (to not surviving) following mastectomy.

At the same time, however, the breast-conservation and mastectomy survival rates within each of the first three quintiles fall short of our criteria for similarity; specifically, although the differences between the breast-conservation and mastectomy survival rates for these quintiles are not statistically significant, each is slightly larger than 1.5 percentage points. According to our criteria, the separate results for quintiles 1 through 3 are inconclusive. Yet when results for these quintiles are considered together—and compared to the results for quintiles 4 and 5—there are two potential implications: (1) breast-conservation therapy may not have been quite as effective as mastectomy for some of the patients who were less likely to receive it—such as those who resided in "low-lumpectomy" areas (in which breast-conservation therapy was relatively uncommon); and (2) breast conservation has been at least as effective as mastectomy for those who were most likely to receive it.

There are various possible explanations for this nonsignificant pattern, based on the different components of the propensity score.[26] (See appendix I.) However, at the present time, exploratory analyses would be difficult, at best, because within the rather homogeneous group of patients examined in this report, there is a relatively small number of breast-conservation patients (1,072) and only about one-third of them (340) fall into quintiles 1 through 3.[27]

Cross Design Comparisons and Strength of the Evidence

Step 3 consists of cross design comparisons and a consideration of the evidence. An informal comparison of the summary results for step 1 and step 2 suggests that the average treatment effect estimated in the statistical analysis of selected SEER cases is similar to the effects observed in the single-center and multicenter randomized studies. The more precise comparisons in tables 8 and 9 show that, quantitatively, this is indeed the case.[28] But do these data constitute sufficient evidence to conclude that the effectiveness of breast-conservation therapy in day-to-day medical practice really is, at least on average, similar to its effectiveness in randomized studies?

To address this issue, we considered (1) the potential differences distinguishing the SEER analysis from single-center and multicenter randomized studies (including the potential for hidden selection bias in the SEER analysis) and (2) the impact that these potential differences might have on the treatment effects we observed. We then used an additional type of cross design comparison as a validity check.

[26]One possibility is that breast conservation was relatively new in 1983. Thus, in "low lumpectomy" areas, there would have been few surgeons experienced in this approach—and the effectiveness of breast conservation (at least during the time frame examined here) may have been lessened as a result.

[27]This is because, by definition, the patients in quintiles 1 through 3 are less likely than others to receive breast-conservation therapy.

[28]The U.S.-NSABP was omitted from tables 8 and 9 because the results of NCI's multicenter audit of that study had not been issued.

Table 8: Comparison of Treatment Effects Based on Survival Differences[a]

Cross design comparison	Survival rate for breast conservation minus survival rate for mastectomy: SEER difference[b]	Survival rate for breast conservation minus survival rate for mastectomy: Randomized studies' difference[c]	Comparison of differences: SEER difference minus randomized studies' difference
SEER cases versus **single-center** studies	–0.6%	0.0%	–0.6%
SEER cases versus **multicenter studies** (omitting U.S.-NSABP)	–1%[d]	0%[d]	–1%[d]
SEER cases versus single-center and multicenter studies (omitting U.S.-NSABP)	–1%[e]	0%[e]	–1%[e]

[a]Node-negative patients only.

[b]From table 6.

[c]From table 3.

[d]The negative 1-percent figure for the difference between breast-conservation and mastectomy survival rates for the SEER data is rounded from –0.6 percent; the 0 percent figure for the difference between breast-conservation and mastectomy survival rates in multicenter randomized studies is rounded from 0.2 percent; comparison of SEER data versus multicenter studies (omitting U.S.-NSABP) was calculated as (- 0.6%) - 0.4% = –1.0%, which rounds to –1 percent. These figures were rounded because for one multicenter study, the only available data were rounded to the nearest whole percent.

[e]The negative 1-percent figure for the difference between breast-conservation and mastectomy survival rates for the SEER data is rounded from –0.6 percent; the 0 percent figure for the difference between breast-conservation and mastectomy survival rates in randomized studies is rounded from 0.2 percent; comparison of SEER data versus multicenter studies was calculated as (–0.6%) - 0.2 = –0.8%, which rounds to –1 percent. These figures were rounded because for one multicenter study, the only available data were rounded to the nearest whole percent.

Table 9: Comparison of Treatment Effects Based on Odds Ratios[a]

Cross design comparison	Difference in odds ratios[b]	
	Estimate of cross design difference	Significance of difference
Odds ratio from the SEER analysis minus the odds ratio for		
Single-center randomized studies	.94-1.00 = –0.06	Not significant[c]
Multicenter randomized studies (omitting U.S.-NSABP)	.94-1.05 = –0.11	Not significant[d]
Single-center and multicenter studies (omitting U.S.- NSABP)	.94-1.03 = –0.09	Not significant[e]

[a]Node-negative patients only. In this table, the odds ratio representing the effect of breast-conservation therapy relative to mastectomy for the SEER data is compared to the corresponding odds ratios calculated for single-center and multicenter studies.

[b]The odds ratio for SEER patients is from table 6. The odds ratios for the single-center and multicenter randomized studies are from table 3.

[c]Significance test performed at the .05 level. Standard error of the difference is .33.

[d]Significance test performed at the .05 level. Standard error of the difference is .23.

[e]Significance test performed at the .05 level. Standard error of the difference is .20.

Differences Across Study Designs

Three potential cross design differences could affect comparisons of the treatment effect estimated for the SEER medical practice data to the treatment effects observed in single-center and multicenter randomized studies. These are

- potential differences in actual treatment effectiveness (SEER versus single-center and multicenter studies),
- potential differences in patients (again, SEER versus single-center and multicenter studies), which might be related to differences in treatment effectiveness, and
- lack of randomization in the SEER data versus randomization in the single-center and multicenter studies—which could lead to differences in the estimates of treatment effectiveness.

Each of these potential differences could affect the comparison of treatment effects (SEER versus single-center and multicenter studies) in the following ways:

- If there are real differences in treatment effectiveness (for example, if breast-conservation therapy is less effective than mastectomy in

day-to-day medical practice), this would affect the comparison of effects—SEER versus randomized studies. (This is, in fact, the hypothesis we have sought to test.)

- If there are differences in patients—again SEER cases versus randomized studies—this also could affect the comparison of effects, but only if breast-conservation therapy is, in fact, more or less effective for the particular kinds of patients who were included in the SEER analysis than for the kinds of patients included in the randomized studies.
- And, the lack of randomization in the SEER data could affect our estimate of the treatment effect in day-to-day medical practice **if** (1) the kinds of SEER patients who were selected for one treatment had better prognoses than those selected for the other treatment—and (2) this was not corrected as part of our analysis.[29]

In the foregoing analyses, our intent was to test for whether the effectiveness of breast-conservation therapy relative to mastectomy was indeed the same in day-to-day medical practice as in single-center and multicenter randomized studies. In comparing the effect of nominally identical treatments across designs, our goal was to identify **the first type of difference** listed above. We therefore attempted to minimize the influence of each of the other two potential differences.

With respect to differences in patients, we selected SEER patients that were at least roughly comparable to those treated in the randomized studies. With respect to selection bias, the fact that we began with a homogeneous group of SEER patients (node-negative, tumors 4 cm or less, age 70 or younger) argues against substantial amounts of bias.[30] We used the propensity-score method to minimize bias on tumor size and on other measured variables. We also conducted a sensitivity analysis to check for selection bias on life-threatening diseases or conditions other than cancer (for example, heart disease)—and found none. Nevertheless, we realize that despite such efforts, some patient differences or some degree of hidden selection bias can persist.

The similarity of the average treatment effect observed for the SEER medical practice data and the effects observed for the randomized studies (that is, the results shown in tables 8 and 9) argue that none of the potential differences listed above had a major impact. The most parsimonious interpretation of the data presented in tables 8 and 9 is that

[29]That is, selection bias can distort a treatment effect only if an **unmeasured** variable is related to **both** treatment selection and likelihood of survival.

[30]Refer to table 4 and to appendix I for more complete descriptions of the SEER cases examined here.

breast-conservation therapy is, on average, similarly effective to mastectomy in day-to-day medical practice.

Logically, however, it is also possible that if two of three potential cross design differences occurred simultaneously, they could "balance each other out" to produce a false impression of similar treatment effects across designs. Of particular relevance is the **possibility** that hidden selection bias in the SEER data analysis (specifically, a hidden bias toward selecting better-prognosis patients for breast conservation) could "counterbalance" treatment differences (specifically, less effective breast-conservation therapy in medical practice)—and thus create an impression of similar treatment effects across study designs.

Combined-Treatments Survival Rates

We reasoned that an additional indication of the relative effectiveness of treatments across designs would be afforded by a comparison of (1) the **combined-treatments** survival rate for the SEER analysis to (2) the corresponding rates for single-center and multicenter studies. Logically, the SEER combined-treatments survival rate is not affected by internal selection bias. Thus, if the SEER rate proved to be similar to the corresponding rates in single-center and multicenter studies, this would point to minimal differences both in patients and in treatment effectiveness across the designs.[31] In short, similar combined-treatments survival rates for the selected SEER cases and for a set of randomized studies would support the conclusion of similar overall effectiveness of breast conservation in day-to-day medical practice and in the randomized studies.

In contrast, if the SEER combined-treatments survival rate proved to be **different from** the corresponding rates for randomized studies, a number of interpretations would be possible—including a difference in patients as well as a difference in the general quality of treatments being given.

In comparing combined-treatments survival rates across studies, it is necessary to take account of any differences in tumor size—specifically, any differences between the tumor sizes of the selected SEER patients and the patients in the single-center and multicenter randomized studies. This

[31]The alternative to accepting this explanation of a similarity in combined-treatments survival rates for SEER and for randomized studies would be to argue that a particular combination of patient differences (specifically, **patients with worse prognoses in the randomized studies**) and treatment differences (worse results for breast conservation in the SEER data) had produced the similar combined-treatments survival rates. This explanation does **not** seem plausible to us in the current instance because it seems unlikely that physicians would refer patients with worse prognoses to a randomized study that included a **less extensive treatment**.

is because tumor size is related to patient survival. As previously noted, four of the six randomized studies had a tumor-size limit of 4 cm, whereas two studies had a limit of 2 cm; the roughly comparable set of SEER patients had a tumor-size limit of 4 cm.

The comparison of combined-treatments survival rates is easiest to make for SEER data versus **multicenter studies**. This is because all multicenter studies had, in effect, the same tumor-size limit (4 cm) and the SEER cases selected for our analyses were also subjected to the 4-cm limit. Therefore, in this section, we separately discuss (1) the comparison of the SEER combined-treatments survival rate to the combined-treatments survival rate for multicenter studies and (2) the corresponding comparison for SEER and single-center studies.

Table 10 (first row) shows the combined-treatments 5-year survival rate for the full set of SEER cases used in the foregoing analyses; this rate—86.9 percent (or 87 percent, rounded)—is appropriate for comparison to the **multicenter** studies. As shown in the bottom row of table 11, the difference in rates is only 1 percentage point and is not significant. The most parsimonious explanation of this result is that, at least on average and with respect to 5-year survival, there are (1) no substantial differences between the patients in our SEER analysis and the patients in the multicenter studies and (2) no large difference between the effectiveness of breast-conservation therapy or mastectomy across the two types of analyses.[32]

[32]Small or even moderate hidden differences in the effectiveness of one of the treatments (but not the other) could only be detected in a much more sensitive analysis.

Table 10: Combined-Treatments Survival Rates for Three SEER Comparison Groups

Comparison group	Number in sample	Estimate of 5-year survival	Confidence interval
All selected SEER cases (tumor size limit: 4 cm[a])	5,326	86.9%	86.0% to 87.8%
Subset of selected SEER cases (tumor size limit: 2 cm[b])	3,588	89.9%	88.9% to 90.9%
Weighted composite[c]	[c]	89.4%[c]	88.6% to 90.3%

[a]The 4-cm limit directly corresponds to the effective limit for the multicenter studies and for one single-center study (U.S.-NCI); see table I.3 in appendix I.

[b]The 2-cm limit directly corresponds to the limit for two of the three single-center studies; see table I.3 in appendix I.

[c]The weighted composite estimate, which directly corresponds to the tumor-size limits for the three single-center studies taken together, is a weighted combination of the other two estimates. The weights were chosen according to the relative sizes of the U.S.-NCI study (4-cm limit) and the Milan and French studies combined (2-cm limit). Specifically, the estimate for the full set of selected SEER cases (top row) was given a weight of 16 percent (reflecting the fact that the effective n for node-negative patients in the U.S.-NCI study is 120—see table I.2 of appendix I); the estimate for the subset of patients with a tumor sized 2 cm or less (middle row) was given a weight of 84 percent (reflecting the fact that the n's for node-negative patients in the Milan and French studies are 520 and 121). Specifically, the weighted average was calculated as: 16 percent times 86.9 percent, **plus** 84 percent times 89.9 percent.

Table 11: Cross Design Comparison of Combined-Treatments Survival Rates

	Difference in 5-year survival rates	
Cross design comparison	Estimate of the cross design difference	Confidence interval[a]
Rate for appropriate SEER comparison group **minus** rate for		
Single-center randomized studies	89.4% - 93.7% = –4.3%	- 5.6% to - 3.0%[b]
Multicenter randomized studies (omitting U.S.-NSABP)	87% - 88% = –1%[c]	- 3% to 1%[d]

[a]The 95-percent confidence interval is based on the standard error of the difference between survival estimates, which was calculated by taking the square root of the sum of the estimated variances of the two survival estimates.

[b]Because the 95% confidence interval does not overlap 0 (the point of equivalence), the difference is significant at the .05 level.

[c]Because results for one multicenter study could not be obtained to the nearest tenth of 1 percent, results were rounded to the nearest whole percent. (In this instance, 86.9% was rounded to 87% and 88.0% was rounded to 88%; the difference of 1 percentage point is the same regardless of whether rounding takes place before or after the subtraction.)

[d]Because this confidence interval does overlap 0 (the point of equivalence), the difference is not significant.

The comparisons are more complex for SEER versus the single-center studies because two of the three studies had a 2-cm limit. The SEER weighted composite estimate in the last row of table 10 (89.4 percent) combines (1) the survival estimate for the full set of selected SEER cases (4-cm limit) with (2) the survival estimate for the subset of cases defined with a 2-cm limit. (See table 10, note c.) This survival estimate is appropriate for comparison to the single-center studies' estimate (93.7 percent). From table 11, it is clear that with breast-conservation and mastectomy patients taken together, the 5-year survival rate for patients in **single-center** randomized studies is higher than the rate for the corresponding SEER estimate—by a difference of 4.3 percentage points, which is statistically significant. The meaning of this finding is unclear. It could be explained by the argument that implementations of treatments in single-center studies are generally better than implementations in multicenter studies or day-to-day medical practice (which seems to be logical). But it could also be explained by hidden selection of patients with better prognoses for the single-center studies.

Summary and Conclusions

In this report, we examined the relative effectiveness of breast-conservation therapy and mastectomy for patients treated in three contexts: single-center randomized studies, multicenter randomized studies, and day-to-day medical practice. In each context, the summary data indicated that 5-year survival was similar following the two alternative treatments. The best outcomes for both treatments occurred in the single-center studies; however, outcomes for the SEER medical practice patients were comparable to outcomes in the multicenter studies.

We recognize that database analyses are vulnerable to hidden selection bias. But we believe such bias is likely to be minimal in the SEER analyses presented here because (1) a homogeneous group of patients was examined, (2) careful adjustments were made for differences in tumor size and demographic characteristics (using the propensity-score method), and (3) a check for possible selection bias on life-threatening factors unrelated to cancer (such as heart disease) reaffirmed our initial conclusion. In addition, the fact that the combined-treatments survival rate was similar in multicenter studies and in the SEER data points to similar levels of treatment effectiveness across these two designs.

We caution that this analysis does not prove the absence of selection bias in the SEER analysis—and that these results are limited to the patient population, treatments, and outcome that we were able to examine

empirically. Nevertheless, virtually all the evidence that we were able to examine pointed toward the similarity of patient survival following breast-conservation and mastectomy—in day-to-day medical practice as well as in the randomized studies. Only one caveat was suggested by the results of our analyses: A minority of breast-conservation patients—the kinds of patients for whom breast-conservation therapy was relatively **unlikely** to be used (based on factors such as residence in areas where breast-conservation is relatively uncommon) but who nevertheless did receive it—may have achieved slightly better results with mastectomy. The observed difference, however, was not statistically significant.

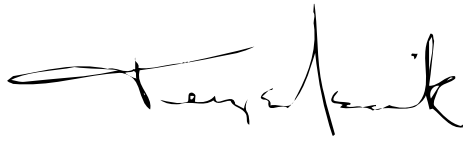
Agency Comments

This report does not examine agency programs; thus, we did not request agency comments. However, we obtained reviewer comments from staff at the National Cancer Institute and the Agency for Health Care Policy and Research; from a number of university-based researchers with expertise in statistics, research methods, or breast cancer; and from investigators in charge of each of the randomized studies. (See appendix II.)

We will be sending copies of this report to the Director of the National Cancer Institute and to other interested parties. We will also make copies available upon request.

If you have any questions, please call me at (202) 512-2900, or call Robert L. York, Director of Program Evaluation in Human Services Areas, at (202) 512-5885 or Judith A. Droitcour, Assistant Director, at (202) 512-5885. Major contributors to this report are listed in appendix III.

Sincerely yours,



Terry E. Hedrick
Assistant Comptroller General

Contents

Abbreviations

DBCG	Danish Breast Cancer Cooperative Group
EORTC	European Organization for Research and Treatment of Cancer
IGR	Institut Gustave-Roussy
NCI	National Cancer Institute
NIH	National Institutes of Health
SEER	Surveillance, Epidemiology and End Results database
U.S.-NSABP	National Surgical Adjuvant Breast Project

Technical Appendix

Confidence Intervals

Some of the tables in this report present 95-percent confidence intervals in addition to point estimates. These intervals reflect the fact that estimates of the parameter in question (for example, the odds ratio) might fluctuate because of random variation in the data.[1] If the 95-percent confidence interval for an odds ratio includes 1 (the point of equivalent odds), there is **no** statistically significant difference (at the .05 level) between the odds of survival following breast-conservation therapy and the odds of survival following mastectomy. Similarly, if the 95-percent confidence interval for a difference in percentages includes 0 (the point of equivalence), there is no statistically significant difference between the percentages being compared (at the .05 level). A statistically significant difference is one that is not likely to have occurred by chance alone. The utility of confidence intervals and significance tests is not limited to randomly selected samples. (See Winch and Campbell, 1969.)

Definition of Terms

In comparing patient survival rates—for example, in comparing the 5-year survival rate for breast-conservation patients to the corresponding rate for mastectomy patients—we termed the two rates "similar" when

- the observed difference between rates was less than 1.5 percentage points (absolute value), **and**
- that difference in rates was not statistically significant.[2] (See table I.1.)

Table I.1: Labeling Survival Rates as Similar, Higher, or Lower: Criteria for Difference in Rates

	Size-of-difference criterion	
Statistical criterion	**<1.5 percentage points[a]**	≥1.5 percentage points[a]
Not significant	"Similar" survival rates	Nonsignificant pattern (inconclusive owing to a lack of significance)
Significant	Precise estimate of a small difference[b]	"Higher" or "lower" survival rates

[a]Absolute value.

[b]Requires very large samples.

[1]Strictly speaking, the meaning of the confidence interval is as follows: Conceptualizing repeated randomized studies in which investigators followed the same procedures and constructed the same kind of interval, 95 percent of the time that interval would include the "true value."

[2]The difference between two survival rates would not be statistically significant if the 95-percent confidence interval surrounding that difference overlapped zero (the point of equivalence).

When a comparison of survival rates showed a difference of 1.5 percentage points or larger—**and** that difference was statistically significant—we used the terms "higher" and "lower."

When survival rates differed by 1.5 percentage points or more—but statistical significance was not attained—we termed the result a nonsignificant pattern. (A nonsignificant pattern is considered **inconclusive** because of the lack of statistical significance. See table I.1.)

This approach recognizes that a high degree of statistical power is required to detect significant differences as small as 1.5 percentage points.[3] Without a high degree of statistical power, we believe it would be inappropriate to term results "similar" merely because of a failure to find a significant difference.

With respect to the remaining possibility depicted in table I.1—a difference of less than 1.5 percentage points that **is** statistically significant—we note that this would **not occur** except where extremely large samples allowed very precise estimates. Were any findings to fall into this category, the conclusion would be that a real, although relatively small, difference does exist—and has been estimated very precisely.[4]

A size-of-difference criterion (cutting point) was used because of the relative imprecision of the estimates, given the existing studies and data. We wished to choose a cutting point that, in our judgment, would represent a difference in survival rates that could reasonably be considered "similar." Thus, we rejected potential cutting points that seemed too high (such as 5 percentage points) because we believed most patients would not consider survival rates that differed by that amount to be similar. In this context, a criterion of 1 percentage point or less versus a larger difference initially seemed reasonable. We chose 1.5 as the specific cutting point (that is, a difference of less than 1.5 percentage points versus 1.5 or greater) because it was possible to obtain most, though not all, survival estimates rounded to the nearest tenth of a percent.

Finally, while we believe 1.5 percentage points is a reasonable cutting point for purposes of defining "similar" levels of survival in this study, we

[3]For example, if the true survival rates following two alternative treatments were 85.5 percent and 87.0 percent, a power of .90 to detect this 1.5-percentage-point difference at the 95-percent confidence level would require nearly **11,000** patients in **each** treatment group. In the area of breast-conservation therapy and mastectomy, the samples in the randomized studies—and in the database analysis presented here—fall short of this number.

[4]In this report, no results fell into this category.

recognize that it is, to some extent, arbitrary. We do not mean to imply that this figure represents the point at which a particular physician or patient would distinguish between a "meaningful difference" and an irrelevant one. We are also cognizant of the fact that, for every 10,000 patients who receive a treatment characterized by even a 1-percentage-point lower survival rate than an available alternative treatment, there would be 100 deaths that could have been avoided by choosing the other treatment—provided that the observed 1-percentage-point difference is, in fact, a real difference and not merely the result of random variation.

In this report, we have avoided use of the term "equivalent" to describe the survival rates observed for breast-conservation and mastectomy patients. A technical reason for this is that to claim "equivalent" survival following the two treatments would require the confidence interval surrounding the difference to be **so small** that it could be entirely enclosed by a prespecified interval—specifically, one defined such that all values within it would be justifiable as clinical equivalence. (See Fleiss, 1992.) That is, not only would we have to justify a difference of 1.5 percentage points as clinically equivalent, but both the upper and lower bounds of the confidence interval surrounding our estimate of the difference would have to be within 1.5 percentage points of zero. This degree of precision would only be possible with very large samples.

Randomized Studies

This section (1) describes our methods of combining randomized studies' results, including the use of "effective n's" and rounding rules; (2) describes the patients included in the six randomized studies that met our criteria; and (3) briefly discusses the two English studies that were omitted from our analyses because they did not meet our treatment criteria.

Combining Randomized Studies' Results

We conducted the meta-analysis of six randomized studies primarily to produce information that could be compared to the separate statistical analysis of selected cases from the SEER database. We began our work for the meta-analysis of randomized studies' results by calculating for each randomized study an odds ratio for 5-year survival (because the outcome criterion for the SEER analysis was 5-year survival). Then we tested for the homogeneity of the odds ratios and, because no significant heterogeneity was found, combined them in a common odds ratio. Specifically, we used the Mantel-Haenszel (1959) method and the STAT XACT program produced

by Cytel Software of Cambridge, Massachusetts. STAT XACT uses the Breslow-Day (1980) method of testing for homogeneity of odds ratios. The confidence intervals surrounding the odds ratios were also calculated using the STAT XACT program and are based on the variance estimation method of Robins, Breslow, and Greenland (1986).

Effective N's

Three of the six randomized studies—the Milan study, the French study conducted at the Institut Gustave-Roussy, and the U.S.-NSABP—had both (1) started long enough ago that, except for patients lost to follow-up, all had been followed for 5 years and (2) calculated **recent** estimates of 5-year survival for node-negative patients. Thus, for these three studies, estimates of 5-year survival were based on 5 or more years of follow-up for all or almost all patients.

For the other three studies (U.S.-NCI, Danish, and EORTC), the 5-year survival estimates were actuarial and included a more substantial number of patients who had not been followed for 5 years.[5] To treat these actuarial estimates appropriately in our meta-analyses, we developed the following approach:

- obtain the standard errors of the actuarial estimates (that is, standard errors that take account of how long each patient has been followed up);[6]
- calculate the "effective n" associated with each actuarial estimate, according to the formula shown by Cutler and Ederer (1958);[7]
- multiply the actuarial estimate of 5-year survival by the effective n—thus obtaining the **effective number** who survived (and, by subtraction, died) in each treatment group of each study; and
- use these "effective n's" in calculating the common odds ratio for the meta-analysis.[8]

Effective n's for the three studies were calculated as shown in table I.2.

[5]These actuarial estimates of 5-year survival include patients followed for less than 5 years, with appropriate calculations that maximize the utility of the available data.

[6]In the three studies for which we derived "effective n's," the estimated standard errors of the actuarial estimates either were available in the published literature or we obtained them from investigators.

[7]This formula is simply: standard error of the actuarial estimate (calculated to take account of how long each patient has been followed up) = the square root of (p*q divided by the effective n). Here, p refers to the actuarial estimate of the proportion surviving; q = 1 — p. Substituting the figures for the standard error, p and q, one solves for the effective n.

[8]An expert in survival analysis (Dr. John Wong of the New England Medical Center) agreed that such an approach would be appropriate.

Table I.2: Effective N's for Three Randomized Studies' 5-Year Survival Estimates

Study and treatment	Proportion surviving	Standard error	Actual n	Effective n[a]
U.S.-NCI				
Breast conservation	.939	.030	74	64
Mastectomy	.947	.030	67	56
Danish				
Breast conservation	.874	.020	289	275
Mastectomy	.859	.022	288	250
EORTC				
Breast conservation	.890	.021	238	222
Mastectomy	.900	.019	237	237[b]

[a]Rounded to the nearest whole number (patient).

[b]Actual n. For the EORTC mastectomy group, the effective n was larger than the actual n, apparently because of rounding; we therefore used the actual n of 237. (The EORTC employed a 2-to-1 randomization initially and later adjusted probabilities to achieve equal numbers in each treatment group.)

Rounding Rules

The most precise estimates available were used. The combined-studies (and combined-treatments) survival estimates were, where possible, rounded to the nearest tenth of a percentage point. Because estimates for one multicenter randomized study (EORTC) were only available to the nearest full percentage point, summary figures involving that study were rounded to the nearest full percentage point—to avoid implying greater precision than was possible. Odds ratios were calculated using the most precise figures possible; however, in preparing the data from each randomized study, the number of patients who died within 5 years and the number who survived were calculated from reported percentages and then rounded to the nearest patient (whole number). Odds ratios, which were based on the rounded numbers of patients, were themselves rounded at the second decimal place. Differences between reported survival rates were calculated using the most precise figures possible—and then rounded for presentation in tables. Slight differences in results may have occurred because of rounding procedures and the use of "effective n's" (described above).

Description of Patient Characteristics

This description of the characteristics of patients who participated in randomized studies is based on published eligibility requirements as well as informal requirements identified through data on the kinds of patients that were actually included (which we obtained, as needed, by calling

investigators). Briefly, all patients in the six randomized studies had invasive breast cancer.[9] As shown in table I.3, almost all patients were age 70 or younger and had tumors of 4 cm or less.[10] Two of the three single-center studies admitted only patients with tumors of 2 cm or less. Most randomized studies had numerous eligibility requirements in addition to the age and tumor-size limits. For the U.S. studies, these were as follows:

- **U.S.-NCI**. Tumor confined to breast and axillary nodes, no advanced local disease, no inflammatory carcinoma, no multiple masses or bilateral cancer, no Paget's disease, no prior cancer.
- **U.S.-NSABP**. No fixation to underlying muscle or chest wall, no clinical evidence of skin involvement or distant metastases, no multiple masses (unless all but one proved benign), no prior cancer.

Table I.3: Characteristics of Node-Negative Patients in Randomized Studies

Study	Years of patient enrollment	Age limit	Tumor-size limit
Single-center			
U.S.-NCI[a]	1979 to 1987	None stated; 10 patients aged 71 or older	5 cm; only **6** patients with tumors 4.01 cm to 5 cm
Milan	1973 to 1980	70 years	2 cm
French	1972 to 1980	70 years	2 cm
Multicenter			
Danish	1983 to 1989	69 years	In the group on which estimates are based, only **9** patients had tumors larger than 4 cm
EORTC[b]	1980 to 1986	70 years	"Not too large for good cosmesis;" only **8** patients had tumors larger than 4 cm
U.S.-NSABP	1976 to 1984	70 years	4 cm

[a]Patients in U.S.-NCI—the sole single-center study to include patients with tumors larger than 2 cm—comprise only about 16 percent of all node-negative patients in the three single-center studies; thus, single-center studies are dominated by patients with small tumors (2 cm or smaller).

[b]EORTC provided us with the information that for eight patients in their study, the diameter of the tumor was pathologically determined to be greater than 4 cm.

[9]Invasive cancer is "a stage of cancer in which cancer cells have spread to healthy tissue adjacent to the tumor" (Altman and Sarg, 1992, p. 143).

[10]The Danish study separately reported results for a high-risk group of patients who are not included here because they are generally outside the scope of this report. (Mostly, the high-risk patients were node-positive or their tumor sizes were larger than 5 cm.)

With respect to type of breast cancer (histology), the U.S.-NCI randomized study further noted that almost all patients had infiltrating duct carcinoma. The Milan study also reported that a majority of patients had this type of cancer.

The Excluded English Studies

Two English studies (Atkins et al., 1972; Hayward, 1981; Hayward and Caleffi, 1987) did not meet our treatment criteria because they did not include nodal dissection as part of the breast-conservation therapy that they provided.[11] The two studies are unique in several ways and are therefore briefly discussed in this appendix.

- First, treatments given in the two English studies differed from treatments given in other randomized studies. As mentioned above, the 1961 and 1971 English studies did not perform nodal dissection on breast-conservation patients. In addition, they have been criticized for providing inadequate radiation (Harris et al., 1983).
- Second, patient survival rates appeared to be considerably lower than in the six studies that met our criteria. This suggests that patients in the English studies may have had poorer prognoses or been subjected to poorer treatment implementations, or both.
- Third, the two English studies were conducted earlier than the other studies. They began in 1961 and 1971, and the 1971 study used the same procedures as the 1961 study. The six studies in our analysis were begun between 1972 and 1983.
- Fourth, in the two English studies, the overall pattern indicated that lumpectomy was less effective than mastectomy. In the first English study, it was clear—early on—that clinically node-positive patients who received lumpectomy showed lower survival rates than those who received mastectomy. Therefore, only clinically node-negative patients were included in the second English study; however, the clinically node-negative breast-conservation patients in the second English study showed lower 5-year survival than corresponding mastectomy patients. And when the **10-year** follow-up was completed for the **first** study, the clinically node-negative patients in that study also showed a pattern of higher survival with mastectomy than with lumpectomy.

Although the English studies did not qualify for our analyses, we believe they are noteworthy in that they caution that there is at least some question as to whether breast-conservation therapy and mastectomy

[11]Although these studies identified "clinically node-negative" patients (indeed, the second English included only clinically node-negative patients), it was not possible to separate out those who would have **tested** node-negative.

produce comparable survival results when treatment implementations are poorer or when patients have poorer prognoses.

SEER Cases Included in Our Analysis

SEER began recording the type of surgery that breast cancer patients received for the cohort diagnosed in 1983. At the time we performed the analyses reported here, SEER follow-up was available through 1990. We therefore selected patients diagnosed from 1983 through 1985—all of whom could be followed for 5 years. The number of positive nodes was not recorded for these diagnostic cohorts. Because the number of positive nodes is a key prognostic factor for early-stage node-positive patients—and may also be associated with selection of surgery—we believe it is necessary for a statistical analysis aimed at minimizing selection bias among node-positive patients. Data on longer term survival and on node-positive patients are provided by randomized studies. As more SEER data become available, SEER analyses that cover node-positive patients and longer term survival will be possible.

The SEER analyses presented in this report are based on 5,326 breast cancer patients. This dataset was formed by accessing the SEER database for 1983 to 1985 diagnoses and selecting patients who met the following criteria:

- no previous diagnosis with another cancer;
- type of treatment, disease-related, and demographic characteristics known;[12]
- patient followed for 5-years or longer;
- node-negative invasive breast cancer that had not spread beyond the breast (no chest wall involvement, no skin involvement, no attachment to the pectoral muscle);
- tumor 4 cm or smaller;
- type of cancer: infiltrating duct carcinoma or adenocarcinoma (NOS);
- type of treatment: if breast-conservation therapy, lumpectomy with nodal dissection plus radiation; if mastectomy, no "outlier" treatments (that is, no subcutaneous mastectomy, no mastectomy without nodal dissection, no radical mastectomy, no mastectomy plus radiation);[13] and
- age 70 or younger.[14]

[12]No autopsy-only or death-certificate-only cases were included.

[13]Cases were also excluded if there was no breast surgery.

[14]Native American patients were excluded because of their very small numbers.

In the resulting dataset, which included 5,326 patients, about 20 percent of patients received breast-conservation therapy; the remaining 80 percent received mastectomy.

Preliminary analyses on a broader set of SEER patients included those that had been lost to follow-up before the requisite 5 years had elapsed following diagnosis (6.2 percent had been lost to follow-up). In these analyses, the patients who were followed for at least 5 years and those who were not, proved to be virtually identical with respect to both tumor size (the main prognostic factor for node-negative patients) and type of surgery. Specifically,

- Patients not followed had an average tumor size of 2 cm, as did those followed for all 5 years.
- Seventeen percent of the followed patients received breast-conservation therapy (as opposed to mastectomy), as did 17 percent of those lost to follow-up.

Derivation of Propensity Scores and Creation of Quintile Subclasses

To derive the propensity scores, we entered patient characteristics into a logistic regression model predicting selection for breast-conservation therapy. The six patient characteristics entered were

- year in which patient was diagnosed (time),
- geographic area of residence (place),[15]
- size of the patient's tumor,
- patient's age at diagnosis,
- marital status, and
- race or ethnicity.[16]

Because the ultimate objective of the propensity-score analysis was to enhance equivalence of the two SEER treatment groups on all measured variables, **all six** variables were included in the final model. Five of the six variables did prove to significantly affect a patient's probability of

[15]Specifically, this variable consists of the five states and four metropolitan areas that are covered by the SEER database.

[16]Other variables—type of breast cancer and extension of the cancer to the skin or pectoral muscle—would be relevant for broader patient populations but not for the analysis presented here.

receiving breast-conservation therapy.[17] The model also included one significant interaction term—the interaction of geographic area with diagnostic year. (See table I.4.)

As expected, patients with smaller tumors were more likely to receive breast-conservation therapy than patients with larger tumors. However, the other patient characteristics determining selection for breast-conservation therapy **argued against a unidimensional** selection process in which patients with better prognoses are consistently selected for breast-conservation therapy. Notably,

- Patients under 40 had relatively high odds of receiving breast-conservation therapy, although there is some evidence that they may have less favorable prognoses than middle-aged patients (de la Rochefordiere et al., 1993). [18]
- Asian women had lower odds than others of receiving breast-conservation therapy, although they may have somewhat better prognoses than other breast cancer patients.

The propensity scores (probabilities of breast-conservation therapy obtained using the model in table I.4) for the SEER patients examined here ranged from .01 to .69. The propensity scores were used to create five quintiles, as suggested by Rosenbaum and Rubin (1984). The first quintile consists of patients who were least likely to receive breast-conservation therapy, whereas the fifth quintile consists of those who were most likely to receive it.

[17]All six variables were included in the model because our goal was to eliminate even nonsignificant differences between the two groups, to the extent possible. Only marital status proved to be insignificant.

[18]The relationship that we observed between age and selection for breast-conservation therapy had been previously reported (Swanson et al., 1992).

Table I.4: Logistic Regression Model Predicting Selection for Breast-Conservation Therapy[a]

Characteristic	Estimated coefficient	Standard error	Coefficient/ standard error
DODY 1983-85[b]	0.9369	.2277	4.1146
Age group			
Under 40	1.1340	.1219	9.3027
40-49	0.7346	.0995	7.3829
50-59	0.2999	.0919	3.2633
60-70[c]	0		
Tumor size (cm)	–0.3695	.0436	–8.4748
Registry			
San Francisco-Oakland	2.3453	.4071	5.7610
Connecticut	1.1574	.4339	2.6674
Metropolitan Detroit	1.1439	.4192	2.7288
Hawaii	2.1083	.5549	3.7994
Iowa	0.1116	.4799	0.2325
New Mexico	0.3911	.6220	0.6288
Seattle-Puget Sound	2.6214	.4029	6.5063
Utah	1.4254	.4761	2.9939
Metropolitan Atlanta[c]	0		
Race or ethnicity			
White	0.1106	.2554	0.4330
Black	0.0318	.2980	0.1067
Asian	–0.7860	.3726	–2.1095
Hispanic[c]	0		
Marital status			
Never married	0.1879	.1760	1.0676
Married	0.2152	.1254	1.7161
Divorced or separated	0.1484	.1655	0.8967
Widowed	0		
Interaction: DODY and registry			
San Francisco-Oakland	–0.6872	.2462	–2.7912
Connecticut	–0.4541	.2672	–1.6995
Metropolitan Detroit	0.0747	.2531	0.2951
Hawaii	–1.1878	.3657	–3.2480
Iowa	–0.0904	.2878	–0.3141
New Mexico	0.0825	.3805	0.2168
Seattle-Puget Sound	–0.7078	.2444	–2.8961
Utah	–0.4198	.3050	–1.3764
Metropolitan Atlanta[c]	0		

(continued)

Characteristic	Estimated coefficient	Standard error	Coefficient/ standard error
Constant	–3.4311	.4827	–7.1081
–2 log likelihood = 4794.498			
Comparison with constant-only model			
Chi-square = 646.975 with 27 df			

[a]Selection for breast-conservation therapy (versus mastectomy) is predicted using date of diagnostic year (DODY, 1983 to 1985), SEER registry (geographic location), and patient characteristics. Breast-conservation therapy was coded 1 and mastectomy was coded 0.

[b]The years 1983, 1984, and 1985 were coded 0, 1, 2.

[c]Reference category.

As intended, the propensity-score quintiles differentiated between patient subgroups; that is, major differences **across** the quintiles were apparent. Notably, **half** (51 percent) of quintile 1 patients (low probability of breast-conservation therapy) had tumors larger than 2 cm; whereas only **14** percent of quintile 5 had tumors of that size.[19] With respect to geographic area, 70 percent of quintile 1 patients were from Iowa, metropolitan Detroit, or metropolitan Atlanta; by contrast, 73 percent of quintile 5 patients were from the San Francisco-Oakland or the Seattle-Puget Sound registries. Only 6 percent of quintile 1 patients were diagnosed in 1985, compared to 66 percent of quintile 5.

Within each propensity-score quintile, we checked the breast-conservation therapy and mastectomy groups for equivalence on all six variables. No major differences were found; two relatively minor differences were adjusted for, as follows:

- First, with respect to tumor size, within four of five quintiles, a slightly higher proportion of mastectomy patients than breast-conservation patients had tumors larger than 2 cm. For example, within quintile 5, 15 percent of mastectomy patients had tumors larger than 2 cm, as compared to 12 percent of breast-conservation patients. We therefore adjusted results **within each quintile** so that the patients with larger

[19]By definition, no SEER patient in the group examined here had a tumor larger than 4 cm.

tumors would contribute equally to the mastectomy survival estimate and to the breast-conservation survival estimate for that quintile.[20]

- Second, with respect to year of diagnosis, within quintile 5 there was a significant difference between mastectomy patients and breast-conservation patients: 64 percent of the mastectomy patients in quintile 5 had been diagnosed in 1985 as compared to 70 percent of breast-conservation patients. Although year of diagnosis is not generally associated with differences in patient survival, we took the precaution of adjusting results for quintile 5 so that patients diagnosed in 1985 would contribute equally to that quintile's breast-conservation survival estimate and its mastectomy survival estimate (as would patients diagnosed in 1984 and 1983).[21]

Using the quintiles together with the additional adjustments ensures that the comparison between survival rates following breast-conservation therapy and mastectomy is based on patient groups that were adjusted to be as "equivalent" as possible on all relevant measured variables.[22]

[20]Specifically, two separate tumor-size groups were defined: (1) patients with tumors 2 cm or smaller and (2) patients with tumors 2.1 cm to 4.0 cm. **Within** each quintile, we divided the mastectomy patients into these two tumor-size groups; we then divided the breast- conservation patients into these two groups. Five-year survival was calculated for each quintile-by-treatment-by-tumor-size group. Finally, within each quintile, we calculated a weighted average survival rate for mastectomy patients and for breast-conservation patients. Specifically, within each quintile, the relative sizes of the two tumor-size groups were determined with both treatment groups combined; these figures were then used as weights in calculating the separate weighted average survival rate for mastectomy patients and the rate for breast-conservation patients in each quintile.

[21]Specifically, we defined six subgroups, based on crossing the three diagnostic years with the two tumor-size groups. Within quintile 5, we divided the mastectomy patients into these six subgroups. We then divided the breast-conservation patients into the six subgroups. Finally, we calculated a weighted average survival rate for mastectomy patients and for breast-conservation patients, using weighting procedures analogous to those described in the previous footnote. (In other words, the relative sizes of the six subgroups in quintile 5 were determined with both treatment groups combined; these figures were then used as weights in calculating the separate weighted average survival rates for mastectomy patients in quintile 5 and breast-conservation patients in quintile 5.)

[22]As a final check **within** each quintile, we compared patients receiving breast-conservation therapy to patients receiving mastectomy with respect to their **average** tumor size—separately for each of the tumor-size-by-treatment subgroups (and for quintile 5, for each tumor-size-by-treatment-by-diagnostic year subgroup). For **every** subgroup, the average tumor size for breast-conservation patients and mastectomy patients proved to be virtually identical.

List of Experts

The experts listed here commented on one or more drafts of the report or advised us on the methods used in our analyses, or both. We are grateful for the gracious contributions of all these individuals.

Knud West Andersen, Ph.D., Danish Breast Cancer Cooperative Group, Copenhagen, Denmark

Rodrigo Arriagada, M.D., Institut Gustave-Roussy, Villajuif, France

John Bailar, M.D., Ph.D., Department of Epidemiology and Biostatistics, McGill University

M. Blichert-Toft, M.D., Danish Breast Cancer Cooperative Group, Copenhagen, Denmark

Robert F. Boruch, Ph.D., Department of Statistics and Graduate School of Education, University of Pennsylvania

David S. Cordray, Ph.D., Vanderbilt Institute for Public Policy Studies, Vanderbilt University

Ian S. Fentiman, M.D., Guy's Hospital, London, England

Victor Hasselblad, Ph.D., Center for Health Policy, Duke University

Susan Love, M.D., Breast Center, University of California at Los Angeles Medical Center

Claire Maklan, Ph.D., Agency for Health Care Policy and Research, Department of Health and Human Services

Charles Manski, Ph.D., Department of Economics, The University of Wisconsin

Carol Redmond, Ph.D., National Surgical Adjuvant Breast Project, University of Pittsburgh

Donald B. Rubin, Ph.D., Department of Statistics, Harvard University

Seth Steinberg, Ph.D., Biostatistics and Data Management Section, National Cancer Institute

J.A. van Dongen, M.D., European Organization for Research and Treatment of Cancer, Amsterdam, The Netherlands

Umberto Veronesi, M.D., Istituto Nazionale per lo Studio e la Cura dei Tumori, Milan, Italy

Paul Wortman, Ph.D., Department of Psychology, State University of New York at Stony Brook

Major Contributors to This Report

Program Evaluation and Methodology Division

Principal Contributors	Judith A. Droitcour, Assistant Director and Project Manager Eric Larson
Principal Advisers	Eleanor Chelimsky Richard L. Linster George Silberman Michele Orza Richard Weston Donald Keller
Referencer	Venkareddy Chennareddy

Technical Library

Carol Johnson

Bibliography

Altman, Roberta, and Michael Sarg. The Cancer Dictionary. New York: Facts on File, 1992.

Andersen, Knud West. Personal communication. Danish Breast Cancer Cooperative Group, Aug. 10, 1993.

Atkins, Sir Hedley, et al. "Treatment of Early Breast Cancer: A Report after Ten Years of a Clinical Trial," British Medical Journal, 2:423-29, 1972.

Blichert-Toft, M., et al. "Danish Randomized Trial Comparing Breast Conservation Therapy with Mastectomy: Six Years of Life-Table Analysis," Journal of the National Cancer Institute Monographs, 11:19-25, 1992.

Blichert-Toft, M., et al. "A Danish Randomized Trial Comparing Breast-Preserving Therapy with Mastectomy in Mammary Carcinoma: Preliminary Results," Acta Oncologica, 27(Fasc. 6a):671-77, 1988.

Breslow, N.E., and N.E. Day. Statistical Methods in Cancer Research: Volume 1 - The Analysis of Case Control Studies. Lyon, France: International Agency for Research on Cancer (Pub. No. 32), 1980.

Byar, David P. "Why Data Bases Should Not Replace Randomized Clinical Trials," Biometrics, 36:337-42, 1980.

Cutler, Sidney J., and Fred Ederer. "Maximum Utilization of the Life Table Method in Analyzing Survival," Journal of Chronic Diseases, 8:699-712, 1958.

de la Rochefordiere, Anne, et al. "Age as a Prognostic Factor in Premenopausal Breast Carcinoma," Lancet, 341:1039-43, 1993.

Dickersin, K., and J. Berlin. "Meta-Analysis: State-of-the-Science," Epidemiologic Reviews, 14:154-76, 1992.

Droitcour, Judith A., George Silberman, and Eleanor Chelimsky. "Cross Design Synthesis: A New Form of Meta-analysis for Combining Results from Randomized Clinical Trials and Medical-Practice Databases," International Journal of Technology Assessment in Health Care, 9(3):440-49, 1993.

Ellenberg, Susan S. "Meta-Analysis: The Quantitative Approach to Research Review," Seminars in Oncology, 15:472-81, 1988.

Fleiss, Joseph L. "General Design Issues in Efficacy, Equivalency, and Superiority Trials," Journal of Periodontal Research (special issue), 27:306-13, 1992.

Freidlin, Boris. Personal communication. EMMES Corp., June 23, 1994.

GAO. (See U.S. General Accounting Office.)

Hankey, Benjamin F., et al. "Overview." In Barry A. Miller et al., Cancer Statistics Review, 1973-1989 (NIH Pub. No. 92-2789). Bethesda, Md.: National Institutes of Health, 1992.

Harris, Jay R., et al. "Conservative Surgery and Radiotherapy for Early Breast Cancer," Cancer, 66 (Sept. 15 Supp.):1427-38, 1990.

Harris, Jay R., Samuel Hellman, and William Silen (eds.). Conservative Management of Breast Cancer: New Surgical and Radiotherapeutic Techniques. Philadelphia: J.B. Lippincott, 1983.

Hayward, John L. "The Guy's Hospital Trials on Breast Conservation." In Jay R. Harris, Samuel Hellman, and William Silen (eds.), Conservative Management of Breast Cancer: New Surgical and Radiotherapeutic Techniques. Philadelphia: J.B. Lippincott, 1983, pp. 77-90.

Hayward, John, and Maira Caleffi. "The Significance of Local Control in the Primary Treatment of Breast Cancer," Archives of Surgery, 122:1244-47, 1987.

Kahn, Harold A., and Christopher T. Sempos. Statistical Methods in Epidemiology. New York: Oxford University Press, 1989.

Lee-Feldstein, Anna, Hoda Anton-Culver, and Paul J. Feldstein. "Treatment Differences and Other Prognostic Factors Related to Breast Cancer Survival: Delivery Systems and Medical Outcomes," Journal of the American Medical Association, 271(15):1163-68, 1994.

Lichter, Allen S., et al. "Mastectomy Versus Breast-Conserving Therapy in the Treatment of Stage I and II Carcinoma of the Breast: A Randomized Trial at the National Cancer Institute," Journal of Clinical Oncology, 10(6):976-83, 1992.

Louis, Thomas A., Harvey V. Fineberg, and Frederick Mosteller. "Findings for Public Health From Meta-Analyses," Annual Review of Public Health, 6:1-20, 1985.

Mignolet, Francoise. Personal communication. Brussels: EORTC Data Center, Oct. 12, 1994.

Mosteller, Frederick, and John W. Tukey. Data Analysis and Regression. Reading, Mass.: Addison-Wesley, 1977.

National Institutes of Health. "Early-Stage Breast Cancer—NIH Consensus Conference," Journal of the American Medical Association, 265(3):391-95, 1991.

Office of Technology Assessment. (See U.S. Congress, Office of Technology Assessment.)

Percy, Constance, Valerie Van Holten, and Calum Muir. International Classification of Diseases for Oncology, 2nd ed. Geneva: World Health Organization, 1990.

Robins, J., N. Breslow, and S. Greenland. "Estimators of the Mantel-Haenszel Variance Consistent in Both Sparse Data and Large-Strata Limiting Models," Biometrics, 42:311-23, 1986.

Rosenbaum, Paul R., and Donald B. Rubin. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score," Journal of the American Statistical Association, 79(387):516-24, 1984.

Rubin, Donald B., and N. Thomas. "Characterizing the Effect of Matching Using Linear Propensity Score Methods with Normal Distributions," Biometrika, 79(4):797-809, 1992.

Sacks, Nigel P.M., and M. Baum. "Primary Management of Carcinoma of the Breast," Lancet, 342:1402-08, 1993.

Sarrazin, Daniele. Personal communication. Institut Gustave-Roussy, Aug. 6, 1993.

Sarrazin, Daniele, et al. "Ten-year Results of a Randomized Trial Comparing a Conservative Treatment to Mastectomy in Early Breast Cancer," Radiotherapy and Oncology, 14:177-84, 1989.

Sarrazin, Daniele, et al. "Conservative Treatment Versus Mastectomy in Breast Cancer Tumors with Macroscopic Diameter of 20 Millimeters or Less: The Experience of the Institut Gustave-Roussy," Cancer, 53:1209-13, 1984.

Sarrazin, Daniele, et al. "Conservative Treatment Versus Mastectomy in T1 or Small T2 Breast Cancer—A Randomized Clinical Trial." In Jay R. Harris, Samuel Hellman, and William Silen (eds.), Conservative Management of Breast Cancer: New Surgical and Radiotherapeutic Techniques. Philadelphia: J.B. Lippincott, 1983, pp. 101-11.

Stablein, D.M. Personal communication. EMMES Corp., June 10, 1994a.

Stablein, D.M. A Reanalysis of NSABP Protocol B06: Final Report. Potomac, Md.: EMMES Corp., 1994b.

Steinberg, Seth M. Personal communication. National Cancer Institute, July 13, 1993.

Straus, K., et al. "Results of the National Cancer Institute Early Breast Cancer Trial," Journal of the National Cancer Institute Monographs, 11:27-32, 1992.

Swanson, G. Marie, et al. "Trends in Conserving Treatment of Invasive Carcinoma of the Breast in Females," Surgery, Gynecology & Obstetrics, 171:465-71, 1990.

U.S. Congress, Office of Technology Assessment. Identifying Health Technologies That Work: Searching for Evidence, OTA-H-608. Washington, D.C.: U.S. Government Printing Office, 1994.

U.S. General Accounting Office. Cross Design Synthesis: A New Strategy for Medical Effectiveness Research (GAO/PEMD-92-18). Washington, D.C.: U.S. General Accounting Office, 1992.

van Dongen, J.A., et al. "Factors Influencing Local Relapse and Survival and Results of Salvage Treatment after Breast-Conserving Therapy in Operable Breast Cancer: EORTC Trial 10801, Breast Conservation Compared With Mastectomy in TNM Stage I and II Breast Cancer," European Journal of Cancer, 28A(4-5):801-05, 1992a.

van Dongen, J.A., et al. "Randomized Clinical Trial to Assess the Value of Breast-Conserving Therapy in Stage I and II Breast Cancer, EORTC 10801 Trial," Journal of the National Cancer Institute Monographs, 11:15-18, 1992b.

Veronesi, Umberto. Personal communication. Istituto Europeo di Oncologia, June 21, 1994.

Veronesi, Umberto. "Local Control and Survival in Early Breast Cancer: The Milan Trial," International Journal of Radiation: Oncology—Biology—Physics, 12:717-20, 1986a.

Veronesi, Umberto, et al. "Comparison of Halsted Mastectomy with Quadrantectomy, Axillary Dissection, and Radiotherapy in Early Breast Cancer: Long-Term Results," European Journal of Cancer and Clinical Oncology, 22(9):1085-89, 1986b.

Veronesi, Umberto, et al. "Comparing Radical Mastectomy with Quadrantectomy, Axillary Dissection, and Radiotherapy in Patients with Small Cancers of the Breast," New England Journal of Medicine, 305(1):6-11, 1981.

Winch, Robert, and Donald Campbell. "Proof? No. Evidence? Yes. The Significance of Significance Tests," American Sociologist, May 1969:140-43.

Winchester, David P., and James D. Cox. "Standards for Breast-Conservation Treatment," CA-A Cancer Journal for Clinicians, 42(3):134-59, 1992.

Woolf, B. "On Estimating the Relation Between Blood Group and Disease," Annals of Human Genetics, 19:251-53, 1955.

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are $2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

Orders by mail:

U.S. General Accounting Office
P.O. Box 6015
Gaithersburg, MD 20884-6015

or visit:

Room 1100
700 4th St. NW (corner of 4th and G Sts. NW)
U.S. General Accounting Office
Washington, DC

Orders may also be placed by calling (202) 512-6000 or by using fax number (301) 258-4066.

Each day, GAO issues a list of newly available reports and testimony. To receive facsimile copies of the daily list or any list from the past 30 days, please call (301) 258-4097 using a touchtone phone. A recorded menu will provide information on how to obtain these lists.

United States
General Accounting Office
Washington, D.C. 20548-0001

Bulk Mail
Postage & Fees Paid
GAO
Permit No. G100

Official Business
Penalty for Private Use $300

Address Correction Requested