

## BREAST IMPLANTS AND BREAST CANCER — REANALYSIS OF A LINKAGE STUDY

HEATHER BRYANT, M.D., AND PENNY BRASHER, PH.D.

**Abstract Background.** In 1992, Berkel and colleagues reported in the *Journal* the results of their study of the potential association of breast augmentation and breast cancer. The study reported that women who had breast augmentation had a significantly lower subsequent risk of breast cancer ( $P < 0.01$ ) than the general population, with a standardized incidence ratio of 0.48 overall. Assuming a 10-year induction period (that is, assuming that cancers found within 10 years of the augmentation might have been the result of a process begun before surgery and therefore should not be considered), the reported standardized incidence ratio was 0.16. Problems were later identified involving some of the study methods. This paper reports a second analysis of these data.

**Methods.** We used a data set from Alberta Health Care to identify eligible women with bilateral breast augmentation. Using a combination of deterministic and probabilistic methods, we linked this data set to the Alberta Cancer Registry to identify subsequent breast cancers that developed during the study period. Multiple estimates

IN 1992, Berkel and colleagues reported the results of a study examining the potential relation between breast implantation and the subsequent development of breast cancer.<sup>1</sup> After this work was published, a number of potential errors in the study methods were noted. This resulted in a review of the analysis by an external review team and the subsequent recommendation that a letter be written to the *Journal* pointing out some of the problems in the original work. As a result, a brief letter was published in January 1994,<sup>2</sup> outlining the areas of concern and noting that further analysis was under way. This paper reports on the results of the reanalysis.

The original study was carried out by linking two large data bases. The first, the files of Alberta Health Care, the provincially funded health insurance plan, contains information on all procedures billed to the plan; in Alberta there is nearly universal coverage of the population under this plan. The investigators had requested an abstract of these files to obtain data on every breast-augmentation procedure billed to Alberta Health Care over the enrollment period of the study (1973 to 1986). After the files were edited to remove men, women outside the target age group, and duplicate or other ineligible files, they were linked to the Alberta Cancer Registry — a population-based registry that records all new invasive cancers occurring in residents of the province. Its completeness for invasive cancers is believed to exceed 95 percent; the degree of completeness for the reporting of in situ cancer is not certain.

The data bases were linked in such a way as to re-

of standardized incidence ratios were calculated on the basis of differing study-eligibility dates, induction periods, and types of breast-cancer (invasive only or invasive plus in situ).

**Results.** The reanalysis found substantial differences in the numbers of person-years at risk, resulting in higher standardized incidence ratios than in the original analysis. The final ratios for all breast cancers, with October 1, 1973, used as the starting date of the study, were 0.76 (95 percent confidence interval, 0.55 to 1.02), 0.85 (95 percent confidence interval, 0.58 to 1.19), and 0.68 (95 percent confidence interval, 0.32 to 1.25) for induction periods of 0, 5, and 10 years, respectively. None of these standardized incidence ratios were significantly different from 1.

**Conclusions.** On the basis of this reanalysis, the incidence of breast cancer among the women who had breast augmentation could not be said to be either significantly higher or lower than that among the general population over the period during which this cohort was followed. (N Engl J Med 1995;332:1535-9.)

move cases in which the augmentation could be shown to have occurred after a diagnosis of breast cancer (and which were thus deemed ineligible for further follow-up), and then to ascertain the number of women with cosmetic implants in whom breast cancer subsequently developed. The actual number of breast-cancer cases identified in this manner was compared with the expected number, which was calculated by applying the age-specific breast-cancer rates in the province against the Alberta Health Care cohort's number of age-specific person-years at risk. The ratios that resulted — the standardized incidence ratios — were thus a comparison of observed and expected rates; standardized incidence ratios below 1 would indicate that fewer breast cancers were found than expected, and those above 1 would indicate a greater number of breast cancers among the implant cohort than among the general population.

Two standardized incidence ratios were reported in the original study. The first included all breast cancers identified in the study, including those detected soon after the implantation procedure, and was reported to be 0.48. The second allowed for an induction period of 10 years. Induction periods are intended to account for the time between the onset of any exposure and a resultant carcinogenic effect, followed by the lag between the first carcinogenic change and the growth of that change into a clinically detectable cancer. Because of these time lags, one could assume that a breast cancer detected shortly after an implantation procedure was not, in fact, the result of the procedure itself but was the result of a process under way before the surgery. The standardized incidence ratio with the assumption of a 10-year induction period was said in the original study to be 0.16. In summary, it was reported that there was a significantly lower rate

From the Division of Epidemiology, Prevention and Screening, Alberta Cancer Board, 1040 7th Ave. S.W., Suite 120, Calgary, AB T2P 3G9, Canada, where reprint requests should be addressed to Dr. Bryant.

of breast cancer among the patients with implants than in the general population.

We pointed out several potential problems in this analysis in the letter to the editor of the *Journal* published last year.<sup>2</sup> First, there was a consistent overcalculation of about half a year for every woman in the cohort because dates were rounded in the original study. Second, there was no adjustment for women who moved from the province after their implantations; if breast cancer subsequently developed in these women, the Alberta Cancer Registry would not be routinely notified of it. In addition, the linkage was based on a limited number of variables, some of which could have changed during the study period. The Alberta Health Care number (AHC number) which is used by the provincial health insurance plan to identify patients and is also routinely collected in the registry, did not, at the time of the study, remain unchanged over a person's lifetime but could change with marriage, divorce, or other changes in status. Thus, there is a potential for underlinkage in this type of study. The reanalysis addressed these and other problems, and resulted in standardized incidence ratios for induction periods of both 0 and 10 years that were not significantly different from 1.

## METHODS

### Description of Data Bases

As in the earlier study, the women with breast implants were identified from Alberta Health Care data on payment claims for breast implantation. The files available from Alberta Health Care for this reanalysis included the AHC number, registrant's surname (the registrant was most frequently the male "head of household"), the initials (and rarely the first name) of the woman undergoing the breast augmentation, the date of the implantation procedure, the fee code (Q66 or Q67, taken to refer to bilateral or unilateral implantation, respectively), sex, and the starting and ending dates of coverage under any given AHC number. The starting and ending dates of coverage were not present in the files analyzed for the original study.

Because AHC numbers may have changed, the data files used for this reanalysis (but not the first analysis) also included, where possible, information on multiple AHC numbers for each person's record. Alberta Health Care had used linkage procedures to construct these patient histories, although the accuracy and completeness of this process is not known.

The second data base used in this analysis was the Alberta Cancer Registry, as previously described. In the original study, all patients with first primary breast cancers diagnosed from 1973 through 1990 were selected; a similar time period was used throughout most of this reanalysis, although an additional follow-up year (1991) has been added for brief comment. The reanalysis also distinguished between *in situ* and invasive breast cancers as coded in the Alberta Cancer Registry. Data routinely available in the cancer registry include the woman's full name, AHC number, date of birth, and, in many cases, previous surnames if applicable. If a woman's AHC number is known to change during treatment or follow-up, it is updated in the registry, and the original number is not saved; there is no active notification to the cancer registry of changes in AHC numbers after treatment.

### Cohort Definition

People were excluded from the cohort because they were male, received their implants outside the study period (as defined further below), or were less than 20 or more than 64 years of age at their most recent birthday. Only first implantations were included; subse-

quent procedures were excluded. Women were also excluded if the first implantation noted was unilateral. This was done to minimize the chance of including in the inception cohort women who had actually had reconstructive surgery after breast cancer, and who thus would have a risk of breast cancer during the subsequent follow-up period that would be different from that of the normal population. Finally, women who had surgery during a gap in their Alberta Health Care coverage of a year or more were also excluded from the cohort to limit the possibility that nonresidents of Alberta might be included.

The original study included procedures with the fee code Q66 or Q67 that were performed between January 1, 1973, and December 31, 1986. Our reanalysis indicated that these study dates, in fact, introduced many ineligible people into the cohort. Until at least October 1973, and possibly as late as October 1974, these fee codes indicated different surgical procedures (ganglion and lymphoedema treatment, respectively). Because arguments supporting the use of either October 1973 or October 1974 as the starting date for the study could be made, we performed the analyses using both starting dates.

Finally, women who were found in subsequent linkage procedures (see below) to have had a breast cancer diagnosed before the first recorded bilateral implantation were removed from the cohort.

### Definitions of Study Starting and End Points

The starting point for entry into the implant cohort was the date on which the first known bilateral augmentation was performed. End points were chosen to indicate either the primary end point of interest (diagnosis of breast cancer) or the date on which further follow-up for breast-cancer incidence ended. Follow-up could end as a result of death (from any cause) or of leaving the province, and dates from the Alberta Health Care file were used to determine these events. Similarly, women who had a lapse of more than one year in Alberta Health Care coverage after the implantation and were subsequently reentered in the Alberta Health Care file did not contribute person-years at risk to the analyses during the time outside Alberta; this guarded against the possibility that an unreported cancer was diagnosed outside the province during that period.

### Linkage Procedures

The study was carried out with a combination of deterministic and probabilistic linkage methods. Briefly, deterministic linkage searches for files that match on the basis of variables as predetermined by the program; probabilistic linkage uses weights derived from the frequency distributions of the identifying information to produce a score for potential matches.<sup>3</sup> In the first deterministic matching, exact matches of AHC numbers were sought. Subsequent deterministic runs searched for matches for various combinations of birth date and name. Finally, the LinkPro program<sup>4</sup> was used to determine probabilistic weightings of the available matching elements. Both methods converted surnames into phonetic codes to minimize the possibility that matches might be missed because of spelling errors in surnames.

All files that matched on the basis of AHC numbers were checked to ascertain whether they had been found to be true matches in the earlier study. Any additional matches, as well as a sample of the Alberta Health Care matches reported in the earlier study, were sent for chart review to determine whether implantations had been recorded.

For the other deterministic and probabilistic linkages, the printouts were inspected visually first. Those with an obvious reason for an imperfect match (commonly, the same phonetic code for a surname but a clearly different actual surname) were not followed further. Files with a reasonably high probability of matching were sent in sequential batches for review by the staff of the Alberta Cancer Registry.

### Calculation of Number of Person-Years at Risk and Expected Number of Breast Cancers

Using Alberta Cancer Registry records for first invasive breast cancers for the years 1972 through 1991 and population-denominator data from Statistics Canada,<sup>5</sup> we calculated age- and calendar-specific incidence rates with the use of three-year moving averages for five-

year age groups. For each person in the cohort, contributions of time at risk specific for age group and calendar year were calculated. Finally, the incidence rates calculated for the Alberta population were applied to the number of person-years at risk in the cohort to derive the expected number of cancers. More detail is presented in the Appendix.

The standardized incidence ratio was calculated as the ratio of the observed to the expected number of breast-cancer cases. Ninety-five percent confidence intervals were calculated with a Poisson distribution of breast-cancer cases assumed.<sup>6</sup> Since some of the strata included very few cases, Mantel-Haenszel ratios of incidence risk and corresponding confidence intervals<sup>7</sup> were also calculated for comparison.

## RESULTS

### Definition of the Implant Cohort

The Alberta Health Care data set included information on 12,569 persons. The final numbers of eligible cohort members who had implants, determined for the two possible study starting dates according to the eligibility criteria noted above, are shown in Table 1.

### Availability of Matching Data

Of the 10,900 women eligible for the study, approximately 50 percent appeared to have a single AHC number and registrant surname throughout the study period. For the remaining half there was evidence of a change in one or both variables. This could represent an underestimation of the true frequency of changes of name and AHC number, since the degree of successful linkage of multiple numbers and names in the Alberta Health Care files also has an effect on these proportions. Other analyses of the data indicated that this linkage was less successful for women with earlier implantation dates, suggesting that the number of matching variables for these earlier implantations is not complete. In addition, 10.1 percent of the Alberta Health Care files had incomplete birth-date information.

Because of the changes in registrant surnames, the presence of a birth name on the breast-cancer registry, if applicable, was an important factor in the likelihood of matching. There was a birth name present in the files for approximately 80 percent of the breast-cancer cases between 1973 and 1990. Only 69 percent of the files registered between 1973 and 1975 recorded a birth name, and it was recorded even less frequently in the years before the study starting date.

### Estimation of Number of Person-Years at Risk

Using the methods described previously and the October 1973 starting date, we calculated the number of person-years at risk for the entire cohort to be 89,219 years, without consideration of an induction period. Similar calculations of person-years at risk were made to include induction periods of various durations (1, 5, and 10 years), and with October 1974 as the starting date. The distribution of eligible cohort members according to year of procedure and the frequency of various numbers of person-years at risk are shown in Table 2. Approximately one quarter of the women con-

Table 1. Reasons for Exclusion from the Implant Cohort, According to the Starting Date of the Study.

REASON FOR EXCLUSION	STARTING DATE	
	Oct. 1, 1973	Oct. 1, 1974
<b>Total cohort</b>	12,569	12,569
Implantation outside study period	563	1,089
Unilateral implant	631	625
Coverage inactive for 1-yr gap around time of implantation	46	39
Male sex	67	45
Under 20 yr old	349	332
Over 64 yr old	13	8
<b>Total eligible before linkage</b>	10,900	10,431
Bilateral implantation after breast cancer identified through linkage	65	63
<b>Final cohort</b>	10,835	10,368

tributed less than 5 person-years at risk, and just over a third contributed more than 10.

### Linkages Identified

In the original study, linkage was used to identify women who had had breast cancers before their implantations. In the present study, women with unilateral implants were removed from the cohort first. An additional 65 women were removed from the cohort because their bilateral implantations were postcancer reconstructions.

The primary purpose of the linkage was to identify the women whose breast cancers developed after breast augmentation. Using the methods described earlier, we identified a total of 45 matches involving breast cancer diagnosed between October 1, 1973, and December 31, 1990, of which 5 were in situ cases (including 1 representing an Alberta Health Care match unconfirmed by chart review). When we used the 1974 starting date, we identified 39 postimplantation breast cancers, of which 3 were in situ.

### Standardized Incidence Ratios

In the original study, standardized incidence ratios were used as the end point of interest. In Table 3, the standardized incidence ratios and their 95 percent con-

Table 2. Dates of Implantations and Number of Person-Years at Risk for the Implant Cohort.\*

CHARACTERISTIC	NO. OF WOMEN	% OF TOTAL
Date of implantation		
1973-1975	1230	11.4
1976-1978	2186	20.2
1979-1981	2876	26.5
1982-1984	2799	25.8
1985-1986	1744	16.1
Person-years at risk		
<5	2654	24.5
5-9.99	4362	40.3
10-14.99	3029	28.0
≥15	790	7.3

\*Cohort includes 10,835 women, as identified in Table 1.

Table 3. Standardized Incidence Ratios and 95 Percent Confidence Intervals, According to Induction Period and Cohort Starting Date.

INDUCTION PERIOD	TOTAL BREAST CANCERS		INVASIVE BREAST CANCERS ONLY	
	1973 COHORT	1974 COHORT	1973 COHORT	1974 COHORT
None	0.76 (0.55–1.02)	0.72 (0.51–0.99)	0.72 (0.51–0.98)	0.71 (0.50–0.98)
1 yr	0.81 (0.59–1.08)	0.77 (0.55–1.05)	0.76 (0.55–1.04)	0.76 (0.53–1.05)
5 yr	0.85 (0.58–1.19)	0.81 (0.53–1.18)	0.82 (0.30–1.15)	0.80 (0.52–1.18)
10 yr	0.68 (0.32–1.25)	0.51 (0.19–1.12)	0.66 (0.30–1.25)	0.56 (0.21–1.22)

confidence intervals are shown both for total (including in situ) breast cancers and for invasive breast cancers only, for the four chosen induction periods. The two possible starting dates are presented separately. Neither of the standardized incidence ratios comparable with those in the original study is significantly different from 1.

Although standardized incidence ratios were calculated so as to allow comparison of these results with those in the original paper, incidence rate ratios were also calculated with the use of Mantel-Haenszel procedures, as described previously. These resulted in estimates that differed by an absolute value of less than 3 percent from the standardized incidence ratios presented above.

Finally, all of the above analyses used the same study ending date as the original study — that is, December 31, 1990. Since the 1992 publication of that study, data have become available on the 1991 breast-cancer cases. This updating resulted in the identification of nine more breast cancers. The standardized incidence ratios based on total breast cancers and the 1973 cohort starting date were 0.79, 0.83, 0.87, and 0.78 for induction periods of 0, 1, 5, and 10 years, respectively. All the confidence intervals for these estimates included 1.

### DISCUSSION

The original study identified 11,670 women for inclusion. Only 10,835 women (or 10,368, for the cohort starting on October 1, 1974) were left in the implant cohort in the present study, representing a reduction of about 7 percent in cohort size. The major factors contributing to this difference were the exclusion of women with unilateral implants and the redefinition of the cohort starting and ending dates to coincide with Alberta Health Care coding changes.

There is a greater discrepancy between the two analyses in the numbers of person-years at risk used to calculate breast-cancer risk. The original study derived an estimate of 124,494 person-years at risk, whereas the maximal estimate in this analysis was 89,219. This discrepancy can be ascribed to three major sources. First, the reduced cohort would obviously have fewer person-years at risk; the effect would be greater than the actual reduction in the number of cohort members, since a large proportion of the women removed from the cohort were excluded early in the study, thus taking with them a disproportionately high number of years of follow-up. Second, out-migration was not accounted for in

the original study, whereas this analysis used the ending date for Alberta Health Care coverage to address this occurrence. The actual out-migration rate indicated by the Alberta Health Care files of the eligible cohort was 16.5 percent; this is comparable with estimates of expected out-migration that we can derive by extrapolating from the

available Statistics Canada information for most of the relevant cohort years. Finally, the consistent overestimation of approximately half a year of follow-up per woman in the original study accounted for an additional reduction in the number of person-years at risk of almost 5 percent.

The number of person-years at risk for the 10-year induction period was also overestimated in the original work. This was the result of including all the person-years at risk for any woman whose implantation date was more than 10 years before the study ending date. However, calculations based on an induction period assume that a woman is not at risk for implant-related cancers that develop during this time; neither the cancers nor the person-years at risk that accumulate during this time are included. The discrepancies between the original study and this reanalysis were therefore greater in the standardized incidence ratios calculated for the 10-year induction period (0.16 vs. 0.68) than for the overall estimate with no induction period (0.48 vs. 0.76).

Despite the presence of more AHC numbers in the files and the combination of deterministic and probabilistic linkage methods, the overall number of postimplantation cancers identified was 45; in the original study, 47 linkages were identified and 6 were discarded, for a total of 41 postimplantation cancers. The matches in this study represent the 41 from the original study plus 4 that were identified through the use of additional identifying data available in the updated files. None of the six cases discarded from the original study were included in the reanalysis. Five of the six discarded cases involved implantation dates earlier than October 1973 (in fact, it was the investigation of these cases that led to the discovery of the need for a new study starting date), and the sixth did not meet the criteria for a true match used in this analysis.

We believe it probable that underlinkage still exists, for several reasons. Information identifying multiple AHC numbers is scarce among women who received implants in the early years of the study. These are the study years most likely to have contributed to the number of breast cancers in the cohort, since the women would potentially have been followed for more person-years at risk and a higher proportion of them would be moving into the high-risk years by the ending date of the study. Information identifying alternative surnames is also not always available in the Alberta Cancer Registry. Thus, a woman who married and changed both

surname and AHC number could have slipped through the linkage procedure.

Although we used several linkage steps to limit this possibility, the very few matching variables available made this unwieldy. It was deemed unproductive, for example, to search every potential link that involved matching only first initials and the month and year of birth, although it is possible that there were some true linkages in that group.

As expected, the accumulation of person-years at risk and the increased ages of the women resulted in our finding more breast cancers in the later study years; seven breast cancers, or 16 percent of the total, were identified in 1990, and an additional nine breast cancers were found in 1991. The additional cases resulted in standardized incidence ratios ranging from 0.78 to 0.87, which were higher than those for the shorter follow-up period. This would point to the need for future periodic follow-up of this cohort.

Although the original study found women with implants to have a significantly lower risk of breast cancer than the general population, our reanalysis does not support this conclusion. The major problems in the original analysis each moved the estimate of the standardized incidence ratio downward; their combined effect resulted in estimates that were substantially lower than the current data would support. Despite the fact that all of the 20 standardized incidence ratios in this reanalysis (including those with the 1991 end date) are below 1, only 3 confidence intervals — all of which ignore an induction period — marginally exclude 1, and there is still some concern about possible remaining underlinkage.

In neither analysis have efforts been made to control for possible self-selection of women into the implant group, which could have produced biases if their risk of breast cancer had been different from that of the general population before implantation. Furthermore, be-

cause the use of polyurethane-covered implants is believed to be very limited in Alberta, we can say nothing about the relation of this type of implant to the risk of breast cancer. It would appear, however, that for the study women with implants, over the period included in this analysis, the apparent risk of breast cancer cannot be said to be either higher or significantly lower than that of the general population.

We are indebted to Drs. Stephan Lanes, Sander Greenland, Leslie Roos, Martha Fair, and James Till for their helpful comments, and to Donna Turner and Andrew McMillan for technical assistance.

## APPENDIX

### *Calculation of the Number of Person-Years at Risk and the Expected Number of Breast Cancers*

The Alberta Cancer Registry was used to identify first breast cancers diagnosed in Alberta women during the period 1972 to 1991. Age- and calendar-specific incidence rates,  $i_{ij}$ , were calculated with the use of three-year moving averages for five-year age groups:  $i = 20-24, \dots, 85-89, 90+$ , and  $j = 1973, \dots, 1990$ . For each person in the cohort ( $k = 1, \dots, n$ ), time-at-risk contributions specific for age group and calendar year were calculated as  $py_{ijk}$ . The expected number of cancers was determined as

$$E = \sum_{i=20-24}^{90+} \sum_{j=1973}^{1990} \sum_{k=1}^n i_{ij} (py_{ijk}).$$

## REFERENCES

1. Berkel H, Birdsall DC, Jenkins H. Breast augmentation: a risk factor for breast cancer? *N Engl J Med* 1992;326:1649-53.
2. Bryant H, Brasher PMA, van de Sande JH, Turc J-M. Review of methods in "Breast augmentation: a risk factor for breast cancer?" *N Engl J Med* 1994; 330:293.
3. Howe GR, Lindsay J. A generalized iterative record linkage computer system for use in medical follow-up studies. *Comput Biomed Res* 1981;14:327-40.
4. Wajda A. LinkPro user's manual, version 1. Winnipeg, Man.: InfoSoft, 1992.
5. Electronic Data Dissemination Division. Statistics Canada. Ottawa, Ont.: STATCAN, CANSIM disc, 1993.
6. Bailar JC III, Ederer F. Significance factors for the ratio of a Poisson variable to its expectation. *Biometrics* 1964;20:639-43.
7. Greenland S, Robins JM. Estimation of a common effect parameter from sparse follow-up data. *Biometrics* 1985;41:55-68.