Network Meta-analysis of Margin Threshold for Women With Ductal Carcinoma In Situ

Shi-Yi Wang, Haitao Chu, Tatyana Shamliyan, Hawre Jalal, Karen M. Kuntz, Robert L. Kane, Beth A. Virnig

Background Negative margins are associated with reduced risk of ipsilateral breast tumor recurrence (IBTR) for women with ductal carcinoma in situ (DCIS) treated with breast-conserving surgery (BCS). However, there is no consensus about the best minimum margin width.

Methods We searched the PubMed database for studies of DCIS published in English between January 1970 and July 2010 and examined the relationship between IBTR and margin status after BCS for DCIS. Women with DCIS were stratified into two groups, BCS with or without radiotherapy. We used frequentist and Bayesian approaches to estimate the odds ratios (OR) of IBTR for groups with negative margins and positive margins. We further examined specific margin thresholds using mixed treatment comparisons and meta-regression techniques. All statistical tests were two-sided.

Results We identified 21 studies published in 24 articles. A total of 1066 IBTR events occurred in 7564 patients, including BCS alone (565 IBTR events in 3098 patients) and BCS with radiotherapy (501 IBTR events in 4466 patients). Compared with positive margins, negative margins were associated with reduced risk of IBTR in patients with radiotherapy (OR = 0.46, 95% credible interval [CrI] = 0.35 to 0.59), and in patients without radiotherapy (OR = 0.34, 95% CrI = 0.24 to 0.47). Compared with patients with positive margins, the risk of IBTR for patients with negative margins was smaller (negative margin >0 mm, OR = 0.45, 95% CrI = 0.38 to 0.53; >2 mm, OR = 0.38, 95% CrI = 0.28 to 0.51; >5 mm, OR = 0.55, 95% CrI = 0.15 to 1.30; and >10 mm, OR = 0.17, 95% CrI = 0.12 to 0.24). Compared with a negative margin greater than 2 mm, a negative margin of at least 10 mm was associated with a lower risk of IBTR (OR = 0.46, 95% CrI = 0.29 to 0.69). We found a probability of .96 that a negative margin threshold greater than 10 mm is the best option compared with other margin thresholds.

Conclusions Negative surgical margins should be obtained for DCIS patients after BCS regardless of radiotherapy. Within cosmetic constraint, surgeons should attempt to achieve negative margins as wide as possible in their first attempt. More studies are needed to understand whether margin thresholds greater than 10 mm are warranted.

J Natl Cancer Inst 2012;104:507–516

For women with ductal carcinoma in situ (DCIS) who are treated with breast-conserving surgery (BCS), margin status is recognized as one of the most important predictors of local recurrence regardless of subsequent radiotherapy (RT) (1–3). A recent meta-analysis (4) showed that the risk estimate for ipsilateral breast tumor recurrence (IBTR) for DCIS patients with a positive margin was 2.25-fold higher compared with patients with negative margins (4). However, there is no consensus among physicians about what constitutes an adequate negative resection margin (5,6). For example, Blair et al. (5) surveyed surgeons with an interest in breast cancer surgery and found that 34% would accept 1 mm or less as the minimal acceptable negative margin for DCIS. Half (52%) considered a 2-mm negative margin as acceptable, and only 4% thought that a 10-mm negative margin was the minimum acceptable width (5). The dilemma exists because inadequate margins may lead to a higher recurrence rate, whereas wider resections may result in poor cosmetic outcomes with no substantial clinical benefits. Furthermore, surgeon attitudes about appropriate margin width also depend on whether or not the patient would receive RT (6). Although RT after BCS is the standard of care, previous research has shown that approximately half of the patients undergoing BCS did not receive RT (7). Indeed, consensus guidelines have suggested that low-risk women may be treated by BCS alone (8,9).

If it were ethical to randomly assign women with DCIS to different margin widths, the optimal margin width might be known within 10–15 years. However, because random assignment is not an ethical option, we must rely on evaluations from observational studies to determine the best margin threshold. Yet, systematic reviews are impeded in making direct comparisons to determine...
CONTENTS AND CAVEATS

Prior knowledge
The risk of ipsilateral breast tumor recurrence (IBTR) after breast-conserving surgery (BCS) for ductal carcinoma in situ is reduced if the margin around the tumor resection is free of cancer cells. However, what the optimal free or negative margin should be is unknown.

Study design
The effectiveness of different margin widths was examined in a network meta-analysis (multiple-treatments meta-analysis) of 21 studies that examined the relationship between the risk of IBTR and margin status in women treated with BCS with or without radiotherapy.

Contribution
Overall, wider negative margins were associated with a reduced risk of IBTR regardless of RT status. Moreover, compared with negative margins larger than 2 mm, negative margins of at least 10 mm were associated with a lower risk of IBTR, regardless of treatment with RT.

Implications
Within cosmetic constraints, the widest possible free margins should be achieved in BCS. More studies are needed to determine if margins wider than 10 mm are warranted.

Limitations
Most of the data was from observational studies, and thus subject to selection bias because RT might not be randomly allocated. The categories for defining different free margin widths were somewhat arbitrary because of inconsistent reporting. The model assumed a linear risk of recurrence over time, but this may not be the case. Data on wider negative margins are limited and studies with and without RT were combined.

From the Editors
the best margin threshold because original studies are heterogeneous both in definitions and reporting. Not only do studies define positive margins in a range of ways (from tumor touching the ink to free margin width <1 mm) but also, specific margin thresholds are subject to authors’ discretion and to inconsistent reporting. In the absence of collecting individual patient data from these observational studies, we must depend on indirect comparisons of different margin widths. Viewing each margin group as a competing strategy, we can objectively assess the effect of different margin status. For example, using “positive margin” groups as a common comparator, an indirect estimate of the benefit of “margin threshold = 2 mm” over “margin threshold = 0 mm” can be obtained by comparing “margin threshold = 2 mm” vs “positive margin” and “margin threshold = 0 mm” vs “positive margin.”

We therefore performed a meta-analysis based on a comprehensive literature review (10,11). We included clinical trials that examined BCS plus RT and BCS alone for the treatment of DCIS. Using traditional (frequentist) and Bayesian approaches, we first describe pairwise direct comparisons between groups with negative and positive margins. We further examined specific margin thresholds using both Bayesian hierarchical models and frequentist nonlinear-random-effects models for indirect comparisons (12–14).

Methods

Literature Identification
We searched the PubMed database for studies of DCIS published in English between January 1970 and July 2010. We describe elsewhere the details of the search algorithm, study selection, and assessment of risk of bias (11). We searched Medical Subject Headings (MESH), titles, and abstracts for the terms Ductal Carcinoma In Situ, DCIS, noninfiltrating intraductal carcinoma, carcinoma in situ, intraductal carcinoma, ductal carcinoma in situ of the breast, localized breast cancer, and stage 0 breast cancer. Study quality was assessed using the framework recommended in the manual of comparative effectiveness reviews (http://effectivehealthcare.ahrq.gov/repFiles/2007_10DraftMethodsGuide.pdf) and was categorized into three groups: well designed (low risk of bias), fair (susceptible to some bias), and poor (high risk of bias).

We included studies that investigated the association between eligible outcomes and patient characteristics, tumor characteristics, and/or treatment strategies. We selected randomized controlled trials and observational studies with sample sizes more than 100 case patients of DCIS. We recorded name of first author, journal and year of publication, country of origin, years and sources of patient enrollment, patient and tumor characteristics (such as median age, follow-up time, percentage of comedonecrosis, tumor size, and pathological grade), treatment received, sample size per arm, and outcomes. We abstracted rates of the outcomes in a variety of subgroups, such as age, margin status, comedonecrosis, tumor size, and pathological grade. When the original studies performed multivariable analyses, we also recorded relative measure of the association between patient/tumor characteristics and outcomes. For the purpose of this review, we included trials regardless of treatments, as long as IBTR was reported separately for women treated with BCS alone and BCS plus RT. Different arms from the same trial were treated as separate studies. Furthermore, we restricted studies to those that provided subgroup data based on different margin widths and excluded studies that did not report margin status. When results for the same patient group were reported more than once, we included only the report with the longest follow-up. If we found more than one study from a particular institution, we used the results from the largest sample size or the latest published articles. We calculated the number of recurrences from endpoint percentages if we could not locate an exact estimate in the published studies. Methods used for assessing risk of bias and summary of characteristics of included studies are given in Supplementary Tables 1–3 (available online).

Two investigators (S.-Y.W. and T.A.S.) independently extracted data and discussed any discrepancies with a third investigator (B.A.V.). We cross-checked all of the data we entered into calculations against the original articles. We did not contact the original investigators.

Statistical Analyses
To synthesize evidence comprehensively, we applied both frequentist and Bayesian approaches. For direct pairwise comparisons between positive and negative margins, we analyzed trials using either tumor touching the ink (margin = 0 mm) or margin less than 1 mm as the positive margin groups. We used data on 0-mm margins when data...
Different margin thresholds were treated as different treatments. We used binomial likelihood to model the probability of IBTR within each treatment (threshold status) arm. In each trial, we defined a study-specific baseline effect using log odds of IBTR of the control (positive margin) group, taking RT status and the definition of positive margin into account. The median follow-up time, a study-level covariate, was also included to assess how it affected the log odds of each margin group. We then modeled the effect of treatment (log odds ratio) for each threshold group. For each treatment (threshold), we estimated the treatment-specific effect from the mean intervention effect compared with the positive margin control. We further derived comparisons among threshold = 2 mm, 5 mm, and 10 mm from differences between their specific effects. Mean and Bayesian 95% credible intervals (CrI) for treatment effects were estimated and expressed as odds ratios for presentation. The models with a smaller value of deviance information criterion were selected.

In each Markov chain Monte Carlo sampling, we ranked as best the threshold option with the lowest absolute odds. The probability that each treatment was best was derived from the percentage of best ranking across all simulations. To monitor convergence, we ran three multiple chains from different initial parameter starting values and visually examined trace plots and sample autocorrelation plots. We further assessed convergence of Markov chains using the Gelman and Rubin convergence statistic (19). Based on the visual examination of trace plots and the Gelman and Rubin convergence diagnosis, the mixing of Markov chain Monte Carlo tended to be good, that is, Markov chains tended to converge within 1000 iterations. To be conservative, we used a burn-in period consisting of 5000 iterations to ensure that sample draws from the posterior distribution were unaffected by initial parameter starting values. An additional 50000 random draws were used for posterior summaries to minimize Monte Carlo simulation error. To ensure that overall effects were dominated by data from the trials and not influenced by choice of initial distribution, we selected vague prior distributions, including normal densities (mean 0, variance 10000) for means and an inverse gamma distribution with (a, b) = (0.001, 0.001) for variance of the logarithm of the odds ratios. We performed sensitivity analyses varying the vague priors with an inverse gamma distribution (0.1, 0.1) for the variance. We also allowed the prior distributions of log odds ratios to have a double exponential distribution or a Student’s t distribution in the sensitivity analyses (20).

We applied the same approach, except for the probability of the best threshold option using a frequentist nonlinear mixed-effects model. The predicted proportion of IBTR was estimated by the model controlling threshold and treatment status. Details of the data and methods in the mixed treatment comparisons are given in Supplementary Table 4 and Supplementary Methods (available online).

We conducted meta-analyses using Stata version 11 (StataCorp, College Station, TX) for frequentist pairwise comparisons, SAS version 9.2 (SAS Institute Inc, Cary, NC) for frequentist mixed treatment comparisons, and WinBUGS version 1.4.3 for the Bayesian approach. We reported 95% confidence intervals or credible intervals. All statistical tests were two-sided.
Results

Eligible Studies
The 3581 scrutinized reports yielded 21 eligible studies published in 24 articles (1,21–43) (Figure 1). Seventeen studies examined BCS with RT for DCIS, and 15 studies examined BCS without RT (Table 1). The median follow-up time for these studies ranged from 43 to 132 months. In total, 1066 IBTR events occurred in 7564 patients, including BCS alone (565 IBTR events in 3098 patients) and BCS with RT (501 IBTR events in 4466 patients). We excluded from the pooled analyses two studies with no positive margin group for comparison (28,35). In addition, we excluded three studies (29,42,43) that overlapped with a multicenter trial of BCS plus RT (24). We included the larger study because it had a larger sample size than the three studies combined. We confirmed that this decision did not affect our results.

Comparisons Between Positive Margin and Negative Margin
Consistent with the literature, patients with negative margins were less likely to experience IBTR than patients with a positive margin regardless of subsequent RT, with the frequentist and Bayesian models yielding almost identical odds ratios (IBTR for DCIS patients receiving BCS plus RT with negative margins vs those with positive margins, frequentist model: OR = 0.45, 95% CI = 0.36 to 0.57, P < .001; Bayesian hierarchical model: OR = 0.46, 95% CrI = 0.35 to 0.59, Figure 2; IBTR for DCIS patients with negative margins receiving BCS alone vs those with positive margins, frequentist model: OR = 0.34, 95% CI = 0.25 to 0.46, P < .001; Bayesian hierarchical model: OR = 0.34, 95% CrI = 0.24 to 0.47, Figure 3). The magnitude of the estimates was not affected by controlling for length of follow-up or definition of positive margin. When combining BCS with and without RT, the unadjusted odds ratio for IBTR in patients with negative margins in the frequentist model was identical to the Bayesian estimate after controlling for RT status, length of follow-up, and definition of free margin (OR = 0.40, 95% CI = 0.33 to 0.48, P < .001). Although the odds ratio in the BCS-alone group seemed smaller than that in the BCS plus RT group, meta-regression analysis found that the confidence interval for the RT status includes 0 (OR = 0.10, 95% CI = −0.26 to 0.45), indicating that RT makes no substantial contribution to margin effects. The Bayesian model reached the same conclusion.

Posterior inferences were not substantially changed by sensitivity analyses varying the prior distributions assuming variance from an inverse gamma prior distribution (0.1, 0.1) or by the prior distribution of log odds ratios assuming a double exponential distribution or a Student’s t distribution. The mean odds ratios for BCS regardless of RT were identical in these sensitivity analyses (BCS plus RT: mean OR = 0.46, all three sensitivity analyses; BCS alone, assuming variance from an inverse gamma prior distribution [0.1, 0.1]: mean OR = 0.34; assuming log odds ratios from a double exponential distribution or a Student’s t distribution: OR = 0.35). Our estimates of ORs and 95% CrIs were not greatly affected by our a priori choices.

Comparisons Among Different Negative Margin Width
Results of odds ratio of IBTR and probability that each margin status is the best were assessed for each margin threshold (Table 2). Using the Bayesian hierarchical model, the 10-mm threshold had the lowest odds ratio (patients with positive margins as a reference group) of IBTR (OR = 0.17, 95% CrI = 0.12 to 0.24) compared with the odds ratio of IBTR for 0-mm threshold (OR = 0.45, 95% CrI = 0.38 to 0.53), 2-mm threshold (OR = 0.38, 95% CrI = 0.28 to 0.51), and 5-mm threshold (OR = 0.55, 95% CrI = 0.15 to 1.30). Bayesian analysis determined that the probability of being the best option (compared with other margin widths) was .04 for a 5-mm

Figure 1. Flow chart for trial selection for multiple-treatments meta-analysis of margin thresholds for ductal carcinoma in situ. BCS = breast-conserving surgery; RT = radiotherapy.
threshold and .96 for a 10-mm threshold for all patients. The probability of a greater than 2-mm margin being the best strategy was almost 0. There is evidence that a negative margin larger than 10 mm leads to lower recurrence rates compared with a negative margin larger than 2 mm. Comparisons of the 10-mm threshold group with the 2-mm threshold group showed meaningful difference in the risk of IBTR (OR = 0.46; 95% CI = 0.27 to 0.66, P < .001). Notably, the coefficient for margin definitions include 0, suggesting no substantial effect on the margin thresholds.

The findings from the frequentist nonlinear mixed-effects models (Table 3) were similar to the Bayesian model results (Table 2). The odds of IBTR for the 10-mm threshold group was statistically significantly lower than that for the 2-mm threshold group (OR = 0.46, 95% CI = 0.27 to 0.66, P < .001). Notably, the coefficient for median follow-up time was positive, indicating that this study covariate might confound the association between margin widths and IBTR and thus should be included in the adjustment.

### Table 1. Characteristics of eligible studies*

<table>
<thead>
<tr>
<th>Author (ref)</th>
<th>Publication year</th>
<th>Study duration</th>
<th>Patient source</th>
<th>Median follow-up time, mo</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bijker (21)</td>
<td>2006</td>
<td>1986–1996</td>
<td>The European Organization for Research and Treatment of Cancer (EORTC)</td>
<td>126</td>
<td>BCS, BCSRT</td>
</tr>
<tr>
<td>Solin (24)</td>
<td>2005</td>
<td>1973–1995</td>
<td>10 institutions in four countries in North America and Europe</td>
<td>102</td>
<td>BCSRT</td>
</tr>
<tr>
<td>MacAusland (25)</td>
<td>2007</td>
<td>1987–2004</td>
<td>Four institutions (Women and Infant’s Hospital, Rhode Island Hospital, St Elizabeth’s Medical Center, and Tufts-New England Medical Center)</td>
<td>55.2</td>
<td>BCS</td>
</tr>
<tr>
<td>Neuschatz (1)</td>
<td>2001</td>
<td>1986–1997</td>
<td>Breast Health Center at New England Medical Center</td>
<td>54</td>
<td>BCSRT</td>
</tr>
<tr>
<td>Chan (27)</td>
<td>2001</td>
<td>1978–1997</td>
<td>Breast Unit of the University Hospital of South Manchester</td>
<td>47</td>
<td>BCS, BCSRT</td>
</tr>
<tr>
<td>Wong (28)</td>
<td>2006</td>
<td>1995–2002</td>
<td>Dana-Farber/Harvard Cancer Center</td>
<td>43</td>
<td>BCS</td>
</tr>
<tr>
<td>Ben-David (29)</td>
<td>2007</td>
<td>1985–2002</td>
<td>Department of Radiation Oncology at the University of Michigan</td>
<td>74.4</td>
<td>BCSRT</td>
</tr>
<tr>
<td>Cataliotti (30)</td>
<td>1992</td>
<td>1968–1990</td>
<td>Department of Surgery and Radiotherapy of the University and General Hospital of Caraghi in Florence</td>
<td>94</td>
<td>BCS, BCSRT</td>
</tr>
<tr>
<td>Turaka (31)</td>
<td>2009</td>
<td>1998–2007</td>
<td>Fox Chase Cancer Center</td>
<td>81.6</td>
<td>BCSRT</td>
</tr>
<tr>
<td>Rudloff (32)</td>
<td>2010</td>
<td>1991–NS</td>
<td>Memorial Sloan-Kettering Cancer Center, New York</td>
<td>132</td>
<td>BCS, BCSRT</td>
</tr>
<tr>
<td>Ringberg (33)</td>
<td>2000</td>
<td>1987–1991</td>
<td>Population based Regional Tumor Registry in Lund</td>
<td>63</td>
<td>BCS</td>
</tr>
<tr>
<td>Tunon-de-Lara (34)</td>
<td>2001</td>
<td>1971–1995</td>
<td>Regional Cancer Center in Bordeaux</td>
<td>86</td>
<td>BCS, BCSRT</td>
</tr>
<tr>
<td>Hughes (35)</td>
<td>2009</td>
<td>1997–NS</td>
<td>The Eastern Cooperative Oncology Group and North Central Cancer Treatment Group</td>
<td>75.6</td>
<td>BCS</td>
</tr>
<tr>
<td>Varnek (36)</td>
<td>1995</td>
<td>1980–1993</td>
<td>University Medical Center and Tucson Medical Center in Tucson, AZ</td>
<td>43</td>
<td>BCSRT</td>
</tr>
<tr>
<td>Sahoo (37)</td>
<td>2005</td>
<td>1986–2000</td>
<td>University of Chicago</td>
<td>63</td>
<td>BCSRT</td>
</tr>
<tr>
<td>Nakamura (38)</td>
<td>2002</td>
<td>1972–2002</td>
<td>University of Southern California</td>
<td>105</td>
<td>BCSRT</td>
</tr>
<tr>
<td>MacDonald (39)</td>
<td>2005</td>
<td>1972–2004</td>
<td>University of California</td>
<td>57</td>
<td>BCS</td>
</tr>
<tr>
<td>Fish (40)</td>
<td>1998</td>
<td>1979–1997</td>
<td>University of Toronto</td>
<td>60</td>
<td>BCS</td>
</tr>
<tr>
<td>Kestin (41)</td>
<td>2000</td>
<td>1980–1993</td>
<td>William Beaumont Hospital, Royal Oak, MI</td>
<td>84</td>
<td>BCSRT</td>
</tr>
<tr>
<td>Goldstein (42)</td>
<td>2000</td>
<td>1980–1993</td>
<td>William Beaumont Hospital, Royal Oak, MI</td>
<td>84</td>
<td>BCSRT</td>
</tr>
<tr>
<td>Rodrigues (43)</td>
<td>2002</td>
<td>1973–2000</td>
<td>Yale University School of Medicine</td>
<td>98.4</td>
<td>BCSRT</td>
</tr>
</tbody>
</table>

* BCS = breast-conserving surgery alone; BCSRT = breast-conserving surgery plus radiotherapy; NS = the end of the study period was not specified.
same institution and yielding more accurate results. For example, RT. Second, we carefully ascertained patient sources in each study, receive RT (approximately half the patients undergoing BCS did not not restricted to clinical trials with treatment of BCS plus RT. In these two approaches, we explored the effect of different margin thresholds in DCIS. Previous literature (restricted analysis to those non-duplicated but eligible studies – 52) found that the odds ratio of IBTR was 0.44 (95% CI = 0.35 to 0.56). In contrast, Dunne et al. (47) reported. Our choice of control group—negative or positive margins—had virtually no impact on this finding. Third, the approach to assessment of specific margin thresholds by Dunne et al. is methodologically unconvincing. Although Dunne et al. (47) stratified the data by margin widths, they pooled results from studies and did not use studies as a stratification factor. By summing the raw data to compute the odds ratios, the analysis of Dunne et al. (47) might introduce confounding and lead to incorrect answers (53,54). In contrast, we used a network meta-analysis framework to assess specific margin thresholds. Our approach ensures that each study serves as its own control, which minimizes potential confounding and avoids a potential problem known as Simpson paradox, a paradox in which the result present in different groups is reversed when the groups are combined (54).

Fourth and most important, these results differ from the conclusion of Dunne et al. (47) with respect to the optimal width of negative surgical margin. They concluded that a negative margin threshold of 2 mm seemed equal to a larger margin for women with DCIS receiving BCS with RT, although there was a trend suggesting that a 5-mm free margin was better than a 2-mm free results are consistent with Wang et al. (4). Using data mostly from multivariable adjustment including both treatments, Wang et al. (4) found that the odds ratio of IBTR was 0.44 (95% CI = 0.35 to 0.56). In contrast, Dunne et al. (47) estimated an odds ratio of 0.36 (95% CI = 0.27 to 0.47) in patients with negative margins. Their findings might be overestimated because they included multiple studies from the same institution that drew on overlapping patient sources. When we replicated the Dunne et al. study (47) but restricted analysis to those non-duplicated but eligible studies (48–52), the risk of IBTR was similar to our findings (frequentist approach, OR = 0.44, 95% CI = 0.34 to 0.56), and importantly, higher than the odds ratio of 0.36 that Dunne et al. (47) reported. Our choice of control group—negative or positive margins—had virtually no impact on this finding. Third, the approach to assessment of specific margin thresholds by Dunne et al. is methodologically unconvincing. Although Dunne et al. (47) stratified the data by margin widths, they pooled results from studies and did not use studies as a stratification factor. By summing the raw data to compute the odds ratios, the analysis of Dunne et al. (47) might introduce confounding and lead to incorrect answers (53,54). In contrast, we used a network meta-analysis framework to assess specific margin thresholds. Our approach ensures that each study serves as its own control, which minimizes potential confounding and avoids a potential problem known as Simpson paradox, a paradox in which the result present in different groups is reversed when the groups are combined (54).

Discussion

The results of this study indicate that wider margins minimize the risk of IBTR and should be a priority for clinicians making surgical plans. The risk of IBTR was lower with a negative margin larger than 10 mm than with a negative margin larger than 2 mm, regardless of RT status.

Network meta-analyses, also known as “mixed treatment comparisons meta-analyses” or “multiple-treatments meta-analyses,” reflect the network of comparisons when three or more treatments are proposed for the same disease (45). Mixed treatment comparisons offer the advantage of facilitating simultaneous inference regarding all treatments, enabling users to select the best treatment (12). Meta-regression aims to explain the difference between intervention effects through one or more characteristics of the studies involved. For example, meta-regression uses study-level covariates to handle heterogeneity (46). Previous literature (45) has demonstrated the danger of naive interpretations of indirect comparisons when network meta-analyses do not account for potential confounders (45). Taking advantage of the combined strengths of these two approaches, we explored the effect of different margin thresholds in DCIS.

This analysis improves on the study conducted by Dunne et al. (47) in four ways. First, unlike Dunne et al. (47), our study was not restricted to clinical trials with treatment of BCS plus RT. In practice, approximately half the patients undergoing BCS did not receive RT (7). Therefore, the results of our study may apply to women receiving BCS alone, as well as those receiving BCS plus RT. Second, we carefully ascertained patient sources in each study, possibly reducing bias by avoiding overlapping inputs from the same institution and yielding more accurate results. For example, in comparing positive and negative margin status, the point estimate of the odds ratio was 0.46 if receiving BCS with RT. These
margin. Prior literature indicates that indirect comparisons that discard within-trial comparisons are liable to bias (53,56); in contrast, our mixed treatment comparisons seem more appropriate. To make our result more comparable to Dunne et al. (47), we analyzed the model including data only from the BCS plus RT subgroup. The odds ratios for IBTR for patients with thresholds of 0, 2, and 10 mm were similar to the results shown in Table 2. The probability that a threshold of 5 or 10 mm was the best option was 0.97 in all patients receiving BCS plus RT. In contrast, the probability that a negative margin greater than 2 mm was the best option was only 0.02. We found consistent evidence that the larger the margin the lower the IBTR, whether restricted to the BCS plus RT group or not.

Our finding that free margins of at least 10 mm decrease the risk of IBTR has important clinical implications. In clinical practice, BCS strives for a balance between wider free margins and cosmetic outcomes. However, these findings provide a strong rationale for surgeons to reconsider the seeming dilemma presented by wider negative margins vs better cosmetic outcomes. Although the findings of Dunne et al. (47) suggest that surgeons should target only a 2-mm free margin, the results of this study suggest that 2 mm may not be an adequate threshold and that surgeons performing first BCS should, under cosmetic constraints, strive to achieve free margins as wide as possible. By attaining wider negative margins with first surgeries, surgeons could improve quality of care and minimize IBTR, as well as decrease the risk of re-excision surgery, resulting in better long-term outcomes without further cosmetic compromise.

Decisions about target margins for first surgeries should be tailored to patients’ prognostic factors, cosmetic outcomes, and patient preferences. For example, if a patient accepts about a 20% probability of recurrence, one of her options is BCS plus RT (Table 3). In this case, her surgeon should target the widest possible negative margins within cosmetic constraints. If her choice is BCS alone, the surgeon should attempt to achieve margin width greater than 0 mm; however, the final margin status involves even greater uncertainty and cosmetic outcomes could be compromised. In contrast, if the patient cannot tolerate a 5% probability of recurrence, she should receive BCS plus RT, with an aim of final negative margins wider than 10 mm (Table 3). She must undergo both treatments, not one or the other, because neither RT alone nor margins wider than 10 mm alone could fulfill her expectation of a less than 5% probability of recurrence. It is worth noting that the additional benefits from a wider margin are quite small for women receiving BCS plus RT; The probability of IBTR differs by 1.7% between margin greater than 0 mm and margin greater than 2 mm and 5% between margin greater than 2 mm and margin greater than 10 mm (Table 3). Because absolute reduction in IBTR is greater for women receiving BCS alone, surgeons should target a wider margin for this high-risk group. We acknowledge that these results may not apply to all patient subgroups, because risk of IBTR is affected by other prognostic factors, such as age, grade,

Table 2. Multiple treatment comparisons among groups with different margin threshold in breast-conserving surgery (BCS) with or without radiotherapy*

<table>
<thead>
<tr>
<th>Margin threshold and treatment</th>
<th>With RT adjustment only</th>
<th>With covariate adjustment†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequentist</td>
<td>Bayesian</td>
</tr>
<tr>
<td></td>
<td>Mean OR (95% CI)</td>
<td>Probability of best option</td>
</tr>
<tr>
<td>T = 0 mm†</td>
<td>0.45 (0.38 to 0.53)§</td>
<td>0</td>
</tr>
<tr>
<td>T = 2 mm†</td>
<td>0.37 (0.26 to 0.49)§</td>
<td>0.38 (0.27 to 0.51)</td>
</tr>
<tr>
<td>T = 5 mm†</td>
<td>0.46 (0 to 0.94)§</td>
<td>0.49 (0.13 to 1.17)</td>
</tr>
<tr>
<td>T = 10 mm†</td>
<td>0.17 (0.11 to 0.24)§</td>
<td>0.18 (0.12 to 0.25)</td>
</tr>
<tr>
<td>T = 10 vs 2 mm</td>
<td>0.47 (0.27 to 0.67)§</td>
<td>0.47 (0.30 to 0.71)</td>
</tr>
<tr>
<td>BCS plus radiotherapy vs BCS alone</td>
<td>0.48 (0.33 to 0.63)‡</td>
<td>0.49 (0.34 to 0.69)</td>
</tr>
</tbody>
</table>

* CI = confidence interval; CrI = credible interval; OR = odds ratio; RT = radiotherapy; T = threshold. Analyses from the frequentist nonlinear mixed-effects model and the Bayesian hierarchical model. All statistical tests were two-sided.
† Adjusted with radiotherapy and length of follow-up;
‡ Patients with positive margin are the reference group.
§ P < .001.
|| P = .06.

Table 3. Predicted probabilities of IBTR stratified by margin threshold and treatment*

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Positive margin, mean (95% CI)</th>
<th>0 mm, mean (95% CI)</th>
<th>2 mm, mean (95% CI)</th>
<th>5 mm, mean (95% CI)</th>
<th>10 mm, mean (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCS plus RT</td>
<td>20% (16 to 24), N = 698</td>
<td>10% (8 to 13), N = 2057</td>
<td>9% (6 to 11), N = 742</td>
<td>11% (1 to 20), N = 23</td>
<td>4% (3 to 6), N = 86</td>
</tr>
<tr>
<td>BCS alone</td>
<td>35% (29 to 41), N = 423</td>
<td>20% (16 to 23), N = 1262</td>
<td>17% (12 to 22), N = 163</td>
<td>20% (3 to 36), N = 10</td>
<td>9% (5 to 12), N = 421</td>
</tr>
</tbody>
</table>

* BCS = breast-conserving surgery; CI = confidence interval; IBTR = ipsilateral breast tumor recurrence; RT = radiotherapy. The predicted probabilities of IBTR were estimated by the frequentist nonlinear mixed-effects model controlling for threshold and treatment status. All statistical tests were two-sided.
and tumor size. Nonetheless, our results should facilitate clinical decision making by helping patients and their physicians choose among reasonable options.

We emphasize the importance of wider margins, but we equally underscore that RT should not be omitted except for the low-risk group. First, RT cannot eliminate the influence of margin effect because the risks of IBTR continue decreasing with wider free margins. This suggests that RT cannot be relied upon to mitigate the negative impact of positive margins. Alternatively, RT is able to profoundly decrease IBTR regardless of margin status. Given that BCS is subject to cosmetic constraint and not all surgeries can guarantee 10-mm free margins, RT should always be considered the top priority. Because margins and treatments are the modifiable prognostic factors, we highlight that RT should complement (and not be supplanted by) the targeting of wider free margins to minimize IBTR.

Frequentist and Bayesian approaches use different software and frameworks. For example, we used the STATA procedure “metan” and the SAS procedure NLMIXED in the frequentist approach and WinBUGS in the Bayesian approach. Despite different frameworks, these approaches can be considered complementary. When there is lack of prior knowledge for the parameters of interest and weak prior distributions are assumed, inferences obtained by Bayesian and frequentist methods generally agree. When suitable and informative prior distributions can be constructed to incorporate knowledge of model parameters, the Bayesian approach is particularly attractive (57–59). For example, if an informative prior distribution for the log odds of IBTR for the reference group is used, one can obtain narrower credible intervals for the odds ratios. In addition, once a posterior distribution is obtained by Bayesian methods, one can directly provide information such as the posterior probability that a treatment is best, which is difficult to obtain using the frequentist approach. In this report, comparisons of groups of negative and positive margins between the two approaches did not differ substantially, but minor differences in the magnitude of the point estimate emerged in indirect comparisons. For example, the frequentist method but not the Bayesian approach indicated that the median follow-up time might confound the association between margin widths and IBTR. Because statistical conclusions might differ between frequentist and Bayesian methods, our analyses should be more comprehensive for using both. For example, we reported the results from both models with and without the follow-up time covariate (Table 2). Furthermore, although a fixed-effects model generally produced more compelling results, it does not allow the true effects to differ across studies and is therefore not justified (60). Most meta-analysts currently apply a random-effects model, which often produces more conservative results than those of fixed-effect models, to accommodate differences among studies (54).

Although earlier studies often reported ipsilateral DCIS and invasive recurrence separately, only four studies reported detailed information for each treatment and margin subgroup (27, 34, 37, 40). The evidence was insufficient to conclude whether margin size had a greater effect on DCIS or invasive ductal carcinoma recurrence. Because the risk of invasive recurrence but not that of DCIS recurrence influences patient survival, future studies should specifically focus on ipsilateral invasive recurrence events. Additionally, although age at diagnosis has been shown previously to be an important predictor of IBTR (61, 62), we did not perform meta-regression using age as a covariate because of inconsistent reporting and limited median age variation.

This study also had some limitations. First, most of our data came from observational studies. As a result, the analyses might be vulnerable to selection bias to the extent that RT might be allocated nonrandomly. However, because it is unethical to randomly assign patients into different margin status groups, we stratified studies by RT status; thus, the impact of this bias should be minimal. Second, our meta-analysis is based on available published results; thus, publication bias cannot be dismissed. Moreover, because the inclusion criteria incorporated a more general perspective on the incidence, treatment, and outcomes of DCIS, instead of margin status alone, some of the odds ratio point estimates in the results contradicted our expectations. However, our results were robust irrespective of analytical methods and sensitivity analyses, even when we included small trials. Third, the categories for defining different free margin widths were somewhat arbitrary. We based our four groups of negative margins on how they were typically reported by earlier studies that specified different margins. We evaluated the best margin threshold by treating each margin width as an intervention. We acknowledge the limitation of excluding studies that did not report margin status. However, our estimates based on this more restrictive group of studies seemed reasonable. For example, our predicted 10% probability of IBTR at 5 years for margins greater than 10 mm is consistent with the results from a prospective trial (35). Ideally, to investigate the best threshold, data should be categorized as margins less than 0 mm, between 0 and 2 mm, between 2 and 5 mm, between 5 and 10 mm, and larger than 10 mm. We acknowledge that in our analysis the “threshold = 0 mm” group may include the “threshold = 2 mm” group, which also encompasses the wider groups. Although our categorization may bias toward small threshold groups, assuming that wider margins have better outcomes, we still found a 10-mm threshold to be better than a 2-mm threshold. Fourth, recurrence risk may not be constant over time (63). Our model, which assumed that the log odds ratio of IBTR of a specific free margin width compared with those of a positive margin have a linear relationship over time, might not fully control for heterogeneity. Finally, data on wider negative margins are limited. Whereas most studies reported negative margins and positive margins, few reported outcomes for wider margins compared with narrow free margins. Our analyses were based on data combining BCS with and without RT together with the assumption that the OR of IBTR from a specific margin width is similar regardless of RT status. Although the results from our meta-regression of positive vs negative margins tend to support our assumption, this issue has not been completely validated. Data were limited for margins larger than 2 mm, and further studies are necessary to confirm our findings.

In conclusion, surgical margins negative for DCIS should be obtained after BCS regardless of RT. A 10-mm free margin appears to be superior to a 2-mm free margin with respect to risk of IBTR. Within cosmetic constraint, surgeons should target free margins as wide as possible in their first attempt. Our results, departing from those of a previous review, emphasize the importance
of using mixed treatment comparisons for evaluating health-care interventions and informing medical decisions.

References


**Funding**


**Notes**

The authors declare no conflict of interest. The authors of this report are responsible for its content. Statements in this article should not be construed as an endorsement by the Agency for Healthcare Research and Quality or the US Department of Health and Human Services. The authors thank Bradley Carlin and Francois Sainfort for their valuable comments.

**Affiliations of authors:** Division of Health Policy and Management, University of Minnesota School of Public Health, Minneapolis, MN (S-YW, TS, HJ, KMK, RLK, BAV); Division of Biostatistics, University of Minnesota School of Public Health, Minneapolis, MN (HC); Minnesota Evidence-based Practice Center, Minneapolis, MN (TS, RLK).