Meta-analytic comparison of randomized and nonrandomized studies of breast cancer surgery

Janet P. Edwards, BSc, MD, MPH*

Elizabeth J. Kelly, MSc*

Yongtao Lin, MLIS, BA, BEd†

Taryn Lenders, MLIS, BScN†

William A. Ghali MD, MPH‡§

Andrew J. Graham MD, MSc, MHSc*§

From the *Department of Surgery, University of Calgary, the †Health Information Network Calgary, the ‡Department of Medicine and Centre for Health Policy Studies, and §Community Health Sciences, University of Calgary, Calgary, Alta.

Correspondence to:
A.J. Graham
Foothills Medical Centre
1403–29th St. NW
Calgary AB T2N 2T9
andrew.graham@albertahealthservices.ca

Background: Randomized controlled trials (RCTs) are thought to provide the most accurate estimation of "true" treatment effect. The relative quality of effect estimates derived from nonrandomized studies (nRCTs) remains unclear, particularly in surgery, where the obstacles to performing high-quality RCTs are compounded. We performed a meta-analysis of effect estimates of RCTs comparing surgical procedures for breast cancer relative to those of corresponding nRCTs.

Methods: English-language RCTs of breast cancer treatment in human patients published from 2003 to 2008 were identified in MEDLINE, EMBASE and Cochrane databases. We identified nRCTs using the National Library of Medicine's "related articles" function and reference lists. Two reviewers conducted all steps of study selection. We included studies comparing 2 surgical arms for the treatment of breast cancer. Information on treatment efficacy estimates, expressed as relative risk (RR) for outcomes of interest in both the RCTs and nRCTs was extracted.

Results: We identified 12 RCTs representing 10 topic/outcome combinations with comparable nRCTs. On visual inspection, 4 of 10 outcomes showed substantial differences in summary RR. The pooled RR estimates for RCTs versus nRCTs differed more than 2-fold in 2 of 10 outcomes and failed to demonstrate consistency of statistical differences in 3 of 10 cases. A statistically significant difference, as assessed by the z score, was not detected for any of the outcomes.

Conclusion: Randomized controlled trials comparing surgical procedures for breast cancer may demonstrate clinically relevant differences in effect estimates in 20%–40% of cases relative to those generated by nRCTs, depending on which metric is used.

Contexte : On estime que les essais randomisés et contrôlés (ERC) offrent l'estimation la plus juste de l'effet « réel » des traitements. La qualité relative de l'estimation de l'effet dérivée des études non randomisées (EnRC) reste indéterminée, particulièrement en ce qui concerne la chirurgie, où les obstacles à la réalisation d'ERC optimaux sont nombreux. Nous avons procédé à une méta-analyse des estimations de l'effet dérivées d'ERC qui comparaient des interventions chirurgicales pour le cancer du sein, aux estimations dérivées d'EnRC assortis.

Méthodes : Nous avons relevé les ERC sur le traitement du cancer du sein chez l'humain publiés en langue anglaise entre 2003 et 2008 à partir des bases de données MEDLINE, EMBASE et Cochrane. Nous avons relevé les EnRC à l'aide des listes bibliographiques et de la fonction « articles connexes » de la National Library of Medicine. Deux examinateurs se sont chargés de toutes les étapes de la sélection des études. Nous avons inclus les études qui comparaient 2 groupes soumis à une chirurgie pour le traitement du cancer du sein. Nous avons extrait l'information sur les estimations de l'efficacité des traitements, exprimées sous forme de risque relatif (RR) de survenue des paramètres retenus, tant dans les ERC que dans les EnRC.

Résultats : Nous avons relevé 12 ERC représentant 10 associations thématiques/paramétriques à des EnRC comparables. À l'examen visuel, 4 paramètres sur 10 étaient substantiellement différents au plan des RR sommaires. Les estimations des RR regroupés provenant des ERC par rapport aux EnRC ont différé de plus du double pour 2 paramètres sur 10 et les différences statistiques ne se sont pas maintenues dans 3 cas sur 10. Selon le score z, les paramètres n'ont présenté aucune différence statistiquement significative.

Conclusion : Les ERC qui comparent des interventions chirurgicales pour le cancer du sein peuvent révéler des différences cliniquement pertinentes pour ce qui est des estimations de l'effet dans 20 %–40 % des cas, par rapport aux estimations générées par les EnRC, la variation étant due aux échelles utilisées.

Randomized controlled trials (RCTs) have been regarded as superior to the nonrandomized observational studies (nRCTs) because of the biases that can arise in nRCTs and the diverging results of nRCTs relative to RCTs. For example, laparoscopic repair of inguinal hernia gained popularity over open repair as a result of favourable results in nRCTs. Subsequently, numerous RCTs have been performed and assessed in a meta-analysis that concluded that the recurrence rate after laparoscopic inguinal hernia repair was no different than that for open repair. The RCTs also warned the surgical community of significantly higher rates of major visceral complications from a laparoscopic approach. Such unexpected findings of RCTs compared with nRCTs have led to a call for careful examination of surgical research, including increased use of RCTs.[1,2]

Studies by Concato and colleagues,[3] Benson and Hartz[4] and Ioannidis and colleagues[5] have renewed the debate regarding the role of RCTs relative to nRCTs. The first 2 groups reported that nRCTs provide similar results to RCTs and argue for an expanded role for nRCTs, particularly to exploit clinical databases. Meanwhile, Ioannidis and colleagues[5] reported reasonable correlation between RCTs and nRCTs in many instances, but also found a greater than 2-fold difference in odds ratios in 33% of topics examined. Surgical studies made up a small proportion of the topics compared. Shikata and colleagues,[6] using a similar methodology to Ioannidis and colleagues, reported on a comparison of effects in RCTs versus nRCTs of digestive surgery. The authors reported that one-quarter of nRCTs gave different results than RCTs. Similar to the studies by Concato and colleagues and Ioannidis and colleagues, Shikata and colleagues included only studies for which meta-analyses of RCTs were available, and they included comparisons between surgical procedures and nonsurgical treatments. Surgical RCTs, as a result of being infrequently conducted, may be somewhat under-represented in meta-analyses. Therefore, studies such as that of Shikata and colleagues may capture only a subset of surgical RCTs.

A study specifically designed to examine the effect estimates for surgical procedures alone in RCTs and nRCTs has, to our knowledge, not been performed. MacLeod[7] and Hall and colleagues[8] have highlighted methodologic issues, such as standardization of procedures, outcome assessment and sample size, which suggest that surgical trials may be different than nonsurgical trials. Concato and colleagues[3] noted that observational investigations of surgical operations may be more prone to selection bias. This raises the possibility that discrepancies of effect estimates between RCTs and nRCTs examining surgical versus nonsurgical interventions may not reflect potential discrepancies in studies comparing 2 surgical interventions.

Given that the paradigm for adopting new surgical procedures differs somewhat from that of pharmacologic therapies and that nRCTs continue to play an important role in informing decisions about adoption of surgical therapies,

there is an ongoing need to formally compare effect estimates from nRCTs with those from RCTs.

The purpose of our study was to determine whether the conclusions reached from an assessment of the body of knowledge obtained from nRCTs would be consistent with those obtained from RCTs, with specific focus on a surgically treated disease. Would decisions based on the collective evidence from nRCTs be consistent with that obtained from the "highest level of evidence," an RCT addressing that same topic? We performed a systematic review and meta-analysis to compare the effect estimates of RCTs with those of nRCTs in the domain of breast cancer surgery. This comparison will help us to better understand the relation between study design, the resulting effect estimate and the subsequent impact on evolving knowledge of surgical procedures.

Methods

Search for RCTs

Our first goal was to identify all existing English-language RCTs published between January 2003 and May 2008 that included 2 surgical arms for the treatment of breast cancer. We limited our search to this 5-year period with the hope of identifying RCTs for which there would likely be a large cohort of earlier nRCTs available for comparison. We identified RCTs through a search of MEDLINE, EMBASE and the Cochrane databases. In MEDLINE and EMBASE, we used a sensitive search strategy proposed by the Cochrane Collaboration for identifying RCTs. Where possible, the databases were searched by linking 2 broad content areas: breast cancer and surgical procedures. Medical Subject Headings (MeSH) were used in the MEDLINE search, including the terms "neoplasm," "breast neoplasms" and "surgical procedures, operative." The EMBASE search consisted primarily of a keyword search, but we used EMTREE terms, such as "surgical technique," "breast tumor" and "breast surgery." This strategy was modified for searches in the Cochrane database. Detailed search strategies are available on request.

Two reviewers (J.P.E., and either E.J.K. or A.J.G.) independently screened citations by title and abstract to identify studies for full-text review. To be included as eligible RCTs, studies had to be truly randomized and had to involve comparisons of 2 surgical procedures used for the treatment of breast cancer. Individual full text articles were then independently reviewed to determine eligibility. Disagreements were resolved by consensus. We then searched the reference lists of included articles and consulted a clinical expert to identify other relevant RCTs.

Search for nRCTs

We identified nRCTs using a number of strategies. For each identified RCT, we used the "related articles" function of the

National Library of Medicine database within PubMed to generate an extensive list of abstracts for subsequent screening. This "related article" search function identifies articles using a probabilistic content similarity algorithm based primarily on abstract text. This feature supports a qualitative approach to exploring large document collections.[9] Next, we performed a manual search of reference lists of previously identified RCTs. The 2 reviewers then independently screened reference citations by title and abstract to identify studies with comparable populations, interventions and comparators to the matched RCTs and to determine which ones would be eligible for full text review. Satisfaction of all of the following criteria led to inclusion of nRCTs: publication in English, nonrandomized study design and comparison of 2 groups that were comparable to those represented in the relevant RCT. We excluded nRCTs that used a historical control group. As the purpose of our study was to determine whether the conclusions reached from an assessment of the findings from nRCTs published before a corresponding RCT, we included nRCTs without a formal evaluation of study quality.

The outcomes of interest were determined after identifying matched groups of RCTs and nRCTs in a hierarchical manner. If mortality or recurrence were assessed in both the RCTs and nRCTs, these were employed as the outcome for analysis. If these data on mortality or recurrence were not available from both study types, we used objectively measured outcomes found in both study types. Finally, if objectively measured outcomes were not available, we used subjectively assessed or self-reported outcomes as a last resort.

Data extraction

As primary outcomes were not specified in all of the RCTs, outcomes were chosen for each RCT/nRCT comparison in an inclusive manner, extracting data for outcomes represented in both study types in which event rates permitted calculation of relative risk (RR). For each study, we extracted information on the year of publication, number of patients in each study arm, duration of follow-up and number of events in each arm of the outcome(s) of interest.

Statistical analysis

To compare the effect estimates of the RCTs and nRCTs for each outcome identified, the data from the RCTs and nRCTs were combined within study type grouping to generate pooled effect estimates. A summary RR for each study type and outcome was determined using a DerSimonian and Laird random effects model. Heterogeneity was assessed using the Q score and was considered to be significant at $p < 0.10$. Statistical analysis was performed using Stata software version 9.2 (StataCorp).

As there are no established metrics for comparison of RCTs with nRCTs, we employed both qualitative and quantitative approaches. We used a compilation of our 4 methods to draw conclusions from our data and did not prioritize any one over another. First, we performed a visual inspection of the results presented in a summarized forest plot format. An informal consensus process among all investigators was used to determine agreement or disagreement for each outcome. Second, we defined discrepancies based on differences in the relative magnitude of treatment. The RRs of nRCTs were deemed to be clinically significantly different from the RRs of the RCTs if they were at least double or less than half the RR of the RCTs. We selected this as the most clinically relevant comparison measure using the precedent found in the studies of Ioannidis and colleagues[5] and Shikata and colleagues.[6] Third, we evaluated whether there was a failure to demonstrate consistency of the statistical differences shown by the summary RRs. Agreement using this method implies that either both study types reflected no statistical difference in treatment effect or that there was a statistical difference shown and in the same direction for both study types. A similar method has been described by Barraclough and Govindan.[10] Finally, we performed a z score comparison of RR estimates from RCTs versus nRCTs. This measure is somewhat underpowered to detect clinically significant differences between pooled RRs involving small sample sizes, but was also reported in the studies by Ioannidis and colleagues[5] and Shikata and colleagues.[6]

RESULTS

Identification of breast cancer RCTs

Our database search strategy yielded a total of 4805 candidate abstracts. These were screened by 2 reviewers to identify 15 RCTs satisfying our inclusion criteria (Fig. 1).

Identification of corresponding nRCTs

A varying number of candidate nRCTs were identified (ranging from 173 for outcome 1 to 4315 for outcome 2; Table 1).[11–48] The candidate articles were then evaluated by 2 independent reviewers to identify nRCTs addressing each of the topics represented by the previously identified RCTs and to determine whether inclusion criteria were met. The yield of this multistep search is described in Table 1. More detailed flow charts describing article selection for nRCTs for each topic are available on request.

Among the 15 RCTs identified, 3 studies[49–51] were excluded from further analysis because no relevant nRCTs were identified. For the comparable RCTs and nRCTs, a range of 1–3 matched outcomes were identified, resulting in 10 outcomes available for comparison (Fig. 1). The clinical topics addressed by these RCT/nRCT combinations
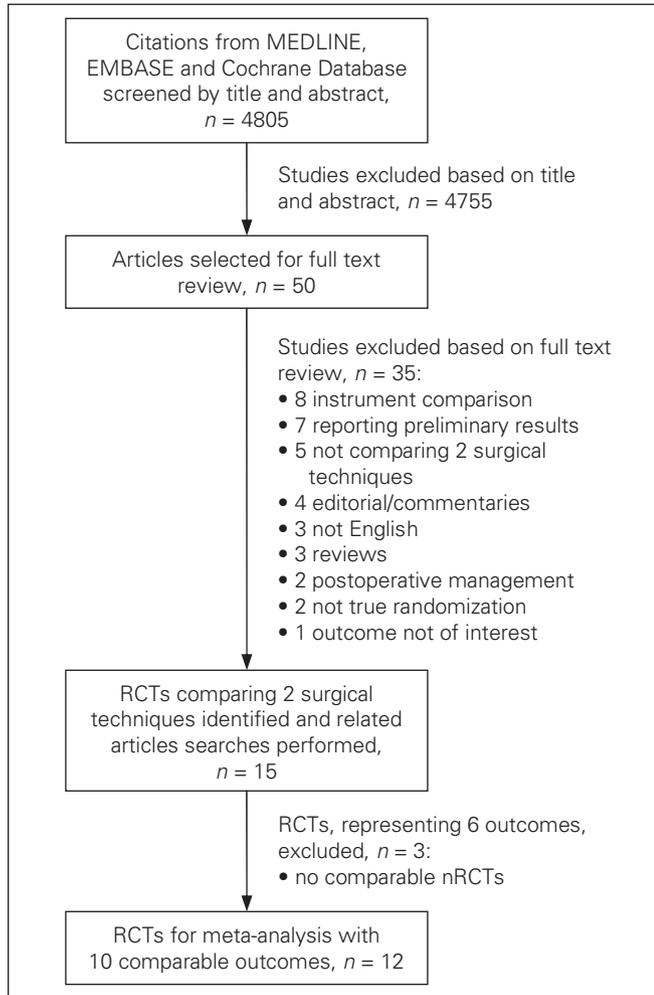
Fig. 1. Identification of randomized controlled trials (RCTs) and related nonrandomized trials (nRCTs) comparing 2 surgical interventions in breast cancer.

included some of the major controversies in breast cancer surgery of recent years, such as breast conserving surgery versus modified radical mastectomy and sentinel lymph node biopsy versus axillary lymph node dissection. Other topics included whether or not preserving the intercostobrachial nerve had substantial impact on sensory outcomes and whether there was a significant difference in lymph node harvest and complication rate with preservation of the pectoralis minor muscle.

Comparative meta-analysis of RCTs and nRCTs by topic

Random effects meta-analysis was performed for each of the 10 outcomes to produce pooled RR estimates for RCTs and nRCTs. Detailed forest plots presenting these individual topic/outcome analyses are presented in Appendix 1, available at cma.ca/cjs. Summarized results showing pooled RCT and nRCT RR estimates for all topics and outcomes are presented in Figure 2.

The summary RRs for each outcome were evaluated for concordance between the nRCTs and RCTs using our 4 methods of comparison (Table 2). On visual inspection of the RR estimates in Figure 2, 6 of 10 comparisons appeared to have generally close agreement, whereas 4 had apparent discrepancies (outcomes 2, 3, 6 and 8). If the formal definition of disagreement is a greater than or equal to 2-fold difference in RR, only 2 of 10 outcomes met the definition for substantial disagreement (outcomes 2 and 3). When failure to demonstrate consistency of the statistical differences was considered, 3 of 10 topics showed disagreement in the directionality or presence of treatment effect (outcomes 3, 5, 6). The z score resulted in no statistically significant differences in treatment effect size.

Table 1. Topics of randomized controlled trials in breast cancer surgery and the number of related nonrandomized studies					
Outcome comparison, topic	RCT study	No. patients in RCT	No. related article abstracts	No. full text nRCTs screened	No. nRCTs (patients)
ICB nerve dissection v. preservation					
1. Sensory deficit[11,12]	Freeman et al.[11] Torresan et al.[12]	158	173	7	4 (410)[13–16]
ALND v. SLNB					
2. Death[17–19]	Mansel et al.[17]	2 167	4 351	18	1 (298)[30]
3. Local recurrence[17–19]	Veronesi et al.[18] Zavagno et al.[19]				4 (4 481)[20–23]
4. Numbness[24,25]	Fleissig et al.[24]	1 127	391	44	1 (370)[26]
5. Swelling[24,25]	Purushotham et al.[25]				1 (370)[26]
6. Reduced arm mobility[24,25]					1 (370)[26]
Mastectomy v. breast-conserving therapy					
7. Death[27–30]	Veronesi et al.[27]	2 561	3 047	43	10 (7 544)[31–38,47,48]
8. Local recurrence[27–30]	Arriagada et al.[28] Poggi et al.[29] Fisher et al.[30]				12 (12 537)[31–42]
Pectoral minor dissection v. preservation					
9. No. lymph nodes removed[43]	Freitas-Júnior et al.[43]	426	343	24	2 (789)[44,45]
10. Complications[43]					1 (101)[46]
ALND = axillary lymph node dissection; ICB = intercostobrachial; nRCT = nonrandomized study; RCT = randomized controlled trial; SLNB = sentinel lymph node biopsy.					

Discussion

The importance of evaluating innovative surgical procedures has become increasingly recognized,[52] with the role of RCTs compared with nRCTs under debate.[3–6] The present study demonstrates that the effect estimates comparing 2 surgical procedures for breast cancer in RCTs and corresponding nRCTs showed clinically important differences in 20%–40% of cases. The proportion of clinically important differences varied depending on which of our metrics was used.

Our study extends and expands on the work of Concato and colleagues,[3] Benson and Hartz,[4] Ioannidis and colleagues[5] and Shikata and colleagues[6] as, to our knowledge, it is the first to explicitly examine comparisons of 2 surgical procedures rather than medical versus surgical treatment and to compare RCT versus nRCT studies not previously included in a meta-analysis of breast cancer surgery. Interestingly, our results are similar to previous published results. Ioannidis and colleagues,[5] who examined primarily medical topics included in a previously published meta-analysis, found a greater than 2-fold difference in odds ratios (OR) in 29% of outcomes. This result is comparable to our finding of disagreement for 2 of 10 (20%) comparisons in breast cancer surgery. Shikata and colleagues,[6] who reported on digestive surgery topics previously studied in meta-analyses, found a greater than 2-fold difference in effect estimates for 7 of 16 (44%) of topics.

One of the challenges in this type of study is that an accepted definition of agreement of effect estimates between RCTs and nRCTs does not exist.[53] We chose to consider several measures, including visual inspection, magnitude of difference in effect estimates, failure to demonstrate consistency of the statistical differences, and the z score for statistical differences.

Visual inspection of results, although qualitative, is an important first step in data analysis. The results of our study demonstrated discrepancies in 4 of 10 outcomes. A qualitative analysis such as this may lack interobserver reliability, but it does suggest divergent effect estimates between RCTs and nRCTs in a substantial proportion of comparisons.

Examining the results of failure to demonstrate consistency of the statistical differences has an intuitive appeal as it examines whether nRCTs comparing surgical procedures

Table 2. Comparison of statistical and magnitude of effect discrepancies between randomized controlled trial and nonrandomized study results

Discrepancy definition	All outcomes, n = 10	Topic/outcome comparison no.
Visual assessment of directionality of effect		
nRCTs overestimate treatment effect	3	2, 3, 6
nRCTs underestimate treatment effect	1	8
Agreement	6	1, 4, 5, 7, 9, 10
Magnitude of effect, ≥ 2-fold difference in RR	2	2, 3
Disagreement of statistical significance	3	3, 5, 6
z score, absolute value > 1.96	0	NA

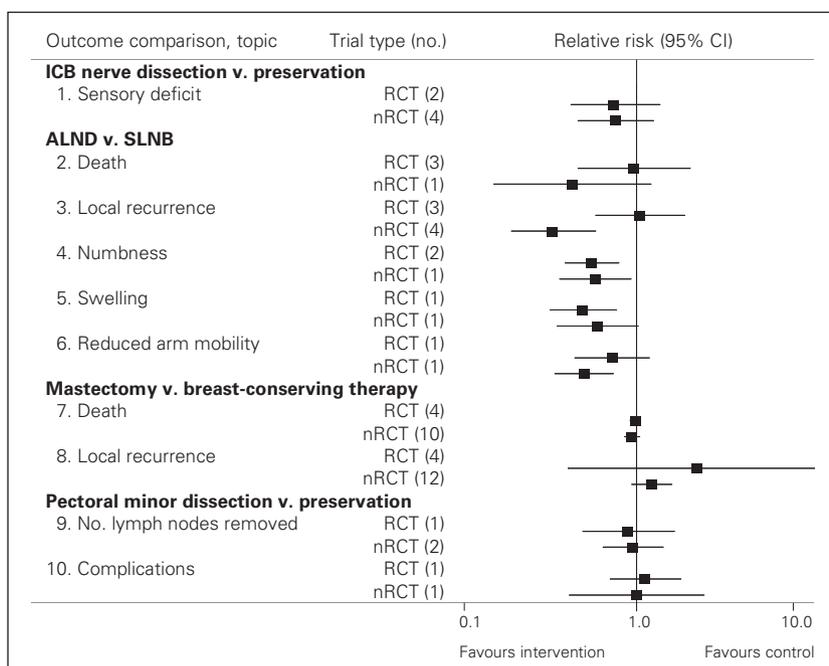NA = not applicable; nRCTs = nonrandomized study; RR = relative risk.

Fig. 2. Relative risk (RR) estimates from randomized controlled trials (RCTs) versus nonrandomized trials (nRCTs) for each topic. The number of studies contributing to each RR is in brackets beside the study type. ALND = axillary lymph node dissection; CI = confidence interval; ICB = intercostobrachial; SLNB = sentinel lymph node biopsy.

are identifying beneficial, harmful or ineffective treatments in the same manner as RCTs. Our study demonstrated discrepancies in 3 of 10 outcome comparisons. Although not previously used as a measure of agreement, clear divergence in effect estimates with regards to benefit or harm could impact treatment recommendations.

The final measure of agreement we used was a 2-fold or greater difference in RR. This measure of agreement has been used previously in studies similar to ours and has clear face validity. Our results demonstrated that in 20% of outcomes the effect estimate of nRCTs compared with RCTs differed by more than 2 times. We believe that a discrepancy in effect estimates of this magnitude could influence surgical practice and clinical practice guidelines in a divergent manner.

Limitations

An important caveat to our study, and to the preceding studies by Ioannidis and colleagues[5] and Shikata and colleagues,[6] is that we present pooled RRs and head-to-head comparisons of RCT and nRCT findings in the presence of heterogeneity across studies. For some outcomes, there was significant heterogeneity of results across studies (see Appendix 1, Figs. S1–S10). We found significant heterogeneity across studies for 2 of 5 RCT meta-analyses that included more than 1 RCT for breast cancer surgery and for 3 of 5 nRCT meta-analyses that we performed. Meanwhile, the study by Ioannidis and colleagues[5] revealed significant heterogeneity in 23% of the RCT meta-analyses and 41% of the nRCT meta-analyses. Shikata and colleagues[6] noted a similar estimate of heterogeneity (41%) in observational studies. The use of random-effects models partially mitigates this concern. Although it was not the intention of our study to make inferences with regard to efficacy of therapies, the pooling of studies with significant heterogeneity could still be criticized. Such pooling is not standard in rigorous meta-analysis owing to the risk of pooling unequal confounding variables and, thus, is a limitation of our approach.

A potential further limitation to our study was the inclusion of RCTs published only in English and within a recent 5-year time period. A similar time and language limitation was applied to our process of identifying comparable nRCTs. The recent 5-year period was selected with the hopes of including higher-quality RCTs, which more likely would have been published recently. The more recent trials were also felt to have a larger number of related nRCTs. The selection of nRCTs involved manual searches and using the PubMed filter for related articles. This strategy may have missed studies that a more exhaustive search strategy would not have missed.

Despite these limitations, our study expands on and extends the findings of previous studies because, to our knowledge, we focused on a surgical area not previously assessed by such studies. We extend the work of Shikata and colleagues,[6] which included only studies that were previously included in meta-analyses. Most surgical studies have not been included in meta-analyses owing to the small numbers of studies being performed in each topic area, thus our study methods were more inclusive in this respect. We also included a broader definition of agreement/disagreement, focusing on factors important to surgeons in clinical decision-making. We also perhaps reach a different global conclusion than Benson and Hartz,[4] whose global conclusions hinted that nRCTs approximate RCTs much of the time and that RCTs may thus not always be needed. Our study findings in the domain of breast cancer surgery underline that effect estimates may differ by more than 2-fold in at least 20% of RCT versus nRCT comparisons.

Conclusion

Although nRCTs often produce comparable effect estimates to RCTs, in 20% of cases the results will differ. As we cannot predict when an nRCT will be misleading, the surgical community should strive to conduct more RCTs of surgical therapies wherever feasible.

References

1. McCormack K, Scott NW, Go PM, et al. Collaboration the EU Hernia Trialists. Laparoscopic techniques versus open techniques for inguinal hernia repair. *Cochrane Database Syst Rev* 2008;1:CD001785 10.1002/14651858.

2. Horton R. Surgical research or comic opera: questions, but few answers. *Lancet* 1996;347:984-5.

3. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 2000;342:1887-92.

4. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000;342:1878-86.

5. Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001;286:821-30.

6. Shikata S, Nakayama T, Noguchi Y, et al. Comparison of effects in

randomized controlled trials with observational studies in digestive surgery. *Ann Surg* 2006;244:668-76.

7. McLeod RS. Issues in surgical randomized controlled trials. *World J Surg* 1999;23:1210-4.

8. Hall JC, Mills B, Nguyen N, et al. Methodologic standards in surgical trials. *Surgery* 1996;119:466-72.

9. Lin J, Wilbur WJ. PubMed related articles: a probabilistic topic-based model for content similarity. *BMC Bioinformatics* 2007;8:423.

10. Barraclough H, Govindan R. Biostatistics primer: what a clinician ought to know: subgroup analyses. *J Thorac Oncol* 2010;5:741-6.

11. Freeman SR, Washington SJ, Pritchard T, et al. Long term results of a randomised prospective study of preservation of the intercosto-brachial nerve. *Eur J Surg Oncol* 2003;29:213-5.

12. Torresan RZ, Cabello C, Conde DM, et al. Impact of the preservation of the intercostobrachial nerve in axillary lymphadenectomy due to breast cancer. *Breast J* 2003;9:389-92.

13. Nishiki M, Yamane M, Amano K, et al. Evaluation of intercosto-brachial nerve preservation in modified radical mastectomy for breast cancer — preliminary report on sensory disturbance test results. *Hiroshima J Med Sci* 1984;33:559-62.

14. Paredes JP, Puente JL, Potel J. Variations in sensitivity after sectioning the intercostobrachial nerve. *Am J Surg* 1990;160:525-8.

15. Taylor KO. Morbidity associated with axillary surgery for breast cancer. *ANZ J Surg* 2004;74:314-7.

16. Temple WJ, Ketcham AS. Preservation of the intercostobrachial nerve during axillary dissection for breast cancer. *Am J Surg* 1985;150:585-8.

17. Mansel RE, Fallowfield L, Kissin M, et al. Randomized multicenter trial of sentinel node biopsy versus standard axillary treatment in operable breast cancer: the ALMANAC Trial. *J Natl Cancer Inst* 2006;98:599-609.

18. Veronesi U, Paganelli G, Viale G, et al. Sentinel-lymph-node biopsy as a staging procedure in breast cancer: update of a randomised controlled study. *Lancet Oncol* 2006;7:983-90.

19. Zavagno G, De Salvo GL, Scalco G, et al. A randomized clinical trial on sentinel lymph node biopsy versus axillary lymph node dissection in breast cancer: results of the Sentinella/GIVOM trial. *Ann Surg* 2008;247:207-13.

20. Haid A, Knauer M, Koberle-Wuhrer R, et al. Medium-term follow-up data after sentinel node biopsy alone for breast cancer. *Eur J Surg Oncol* 2006;32:1180-5.

21. Giuliano AE, Haigh PI, Brennan MB, et al. Prospective observational study of sentinel lymphadenectomy without further axillary dissection in patients with sentinel node-negative breast cancer. *J Clin Oncol* 2000;18:2553-9.

22. Naik AM, Fey J, Gemignani M, et al. The risk of axillary relapse after sentinel lymph node biopsy for breast cancer is comparable with that of axillary lymph node dissection: a follow-up study of 4008 procedures. *Ann Surg* 2004;240:462-8; discussion 468-71.

23. Takei H, Suemasu K, Kurosumi M, et al. Recurrence after sentinel lymph node biopsy with or without axillary lymph node dissection in patients with breast cancer. *Breast Cancer* 2007;14:16-24.

24. Fleissig A, Fallowfield LJ, Langridge CI, et al. Post-operative arm morbidity and quality of life. Results of the ALMANAC randomised trial comparing sentinel node biopsy with standard axillary treatment in the management of patients with early breast cancer. *Breast Cancer Res Treat* 2006;95:279-93.

25. Purushotham AD, Upponi S, Klevesath MB, et al. Morbidity after sentinel lymph node biopsy in primary breast cancer: results from a randomized controlled trial. *J Clin Oncol* 2005;23:4312-21.

26. Husen M, Paaschburg B, Flyger HL. Two-step axillary operation increases risk of arm morbidity in breast cancer patients. *Breast* 2006;15:620-8.

27. Veronesi U, Cascinelli N, Mariani L, et al. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. *N Engl J Med* 2002;347:1227-32.

28. Arriagada R, Le MG, Guinebretiere JM, et al. Late local recurrences in a randomised trial comparing conservative treatment with total mastectomy in early breast cancer patients. *Ann Oncol* 2003;14:1617-22.

29. Poggi MM, Danforth DN, Sciuto LC, et al. Eighteen-year results in the treatment of early breast carcinoma with mastectomy versus breast conservation therapy: the National Cancer Institute Randomized Trial. *Cancer* 2003;98:697-702.

30. Fisher B, Anderson S, Bryant J, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. *N Engl J Med* 2002;347:1233-41.

31. Fan J, Wang L, Wang XJ, et al. Breast conservative therapy in east part of China: a retrospective cohort study. *J Cancer Res Clin Oncol* 2006;132:573-8.

32. Hermann RE, Esselstyn CB Jr, Grundfest-Broniatowski S, et al. Partial mastectomy without radiation is adequate treatment for patients with stages 0 and I carcinoma of the breast. *Surg Gynecol Obstet* 1993;177:247-53.

33. Kindermann G, Genz T. A comparison between the results of simple mastectomy and tumorectomy for breast cancer: the problem of local recurrence. *Arch Gynecol* 1985;237:67-73.

34. Noguchi M, Yagasaki R, Kawahara F, et al. Breast conserving treatment versus modified radical mastectomy in Japanese patients with operable breast cancer. *Int Surg* 1997;82:289-94.

35. Rauschecker HF, Sauerbrei W, Gatzemeier W, et al. Eight-year results of a prospective non-randomised study on therapy of small breast cancer. The German Breast Cancer Study Group (GBSG). *Eur J Cancer* 1998;34:315-23.

36. Rissanen PM. A comparison of conservative and radical surgery combined with radiotherapy in the treatment of stage I carcinoma of the breast. *Br J Radiol* 1969;42:423-6.

37. Romsdahl MM, Montague ED, Ames FC, et al. Conservation surgery and irradiation as treatment for early breast cancer. *Arch Surg* 1983;118:521-8.

38. Voogd AC, Nab HW, Crommelin MA, et al. Comparison of breast-conserving therapy with mastectomy for treatment of early breast cancer in community hospitals. *Eur J Surg Oncol* 1996;22:13-6.

39. Devitt JE. The influence of conservative and radical surgery on the survival of patients with breast cancer. *Can Med Assoc J* 1962;87:906-10.

40. Halverson KJ, Taylor ME, Perez CA, et al. Survival following breast-conserving surgery and irradiation or modified radical mastectomy in patients with invasive breast cancers with a maximum diameter of 1 cm. *Mo Med* 1993;90:759-63.

41. Geiger AM, Thwin SS, Lash TL, et al. Recurrences and second primary breast cancers in older women with initial early-stage disease. *Cancer* 2007;109:966-74.

42. Matthews RH, McNeese MD, Montague ED, et al. Prognostic implications of age in breast cancer patients treated with tumorectomy and irradiation or with mastectomy. *Int J Radiat Oncol Biol Phys* 1988;14:659-63.

43. Freitas-Júnior R, Oliveira EL, Pereira RJ, et al. Modified radical mastectomy sparing one or both pectoral muscles in the treatment of breast cancer: intra and postoperative complications. *Sao Paulo Med J* 2006;124:130-4.

44. Markandoo P, Smith P, Chaudary MA, et al. Preservation of pectoralis minor in axillary clearance for breast cancer. *Br J Surg* 1998;85:1547-8.

45. Muscolino G, Leo E, Sacchini V, et al. Resectable breast cancer: axillary dissection sparing pectoralis muscles and nerves. *Eur J Surg Oncol* 1988;14:429-33.

46. Hoefer RA Jr, DuBois JJ, Ostrow LB, et al. Wound complications following modified radical mastectomy: an analysis of perioperative factors. *J Am Osteopath Assoc* 1990;90:47-53.

47. Arriagada R, Le MG, Contesso G, et al. Predictive factors for local recurrence in 2006 patients with surgically resected small breast cancer. *Ann Oncol* 2002;13:1404-13.

48. Dunn MJ, Tartter PI. Comparison of mastectomy and lumpectomy for medial carcinoma of the breast. *Surg Gynecol Obstet* 1991;173:13-6.

49. Krag DN, Anderson SJ, Julian TB, et al. Technical outcomes of sentinel-lymph-node resection and conventional axillary-lymph-node dissection in patients with clinically node-negative breast cancer: results from the NSABP B-32 randomised phase III trial. *Lancet Oncol* 2007;8:881-8.

50. Lucci A, McCall LM, Beitsch PD, et al. Surgical complications associated with sentinel lymph node dissection (SLND) plus axillary lymph node dissection compared with SLND alone in the American College of Surgeons Oncology Group Trial Z0011. *J Clin Oncol* 2007;25:3657-63.

51. Dalberg K, Johansson H, Signomklao T, et al. A randomised study of axillary drainage and pectoral fascia preservation after mastectomy for breast cancer. *Eur J Surg Oncol* 2004;30:602-9.

52. McCulloch P, Altman DG, Campbell WB, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet* 2009;374:1105-12.

53. MacLehose RR, Reeves BC, Harvey IM, et al. A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies. *Health Technol Assess* 2000;4:1-154.