Title of the article: Oncological Outcomes of Omitting Breast Surgery in Patients Achieving Complete Response After Neoadjuvant Therapy: A Systematic Review and Meta-Analysis

Abstract.
Aims: To evaluate the oncological outcomes of omitting breast surgery in breast cancer patients achieving pathological complete response (pCR) after neoadjuvant systemic treatment (NST), and to assess the influence of response assessment methods on treatment safety.
Materials & Methods: A systematic review and meta-analysis were conducted in accordance with the PRISMA 2020 guidelines. PubMed/MEDLINE, Scopus, Web of Science, Embase, Cochrane CENTRAL, and ScienceDirect were searched without date restrictions. Studies evaluating oncological outcomes in breast cancer patients with complete response after NST who did not undergo definitive breast surgery were included. RRs with 95% CIs were calculated using a Mantel–Haenszel random-effects model. Heterogeneity in response assessment methods was considered a priori.
Findings: Eight studies published 1986 - 2021 were included, comprising comparative cohorts, prospective phase II trials, a database study, and a case series. Comparative meta-analysis demonstrated no significant differences at 5 years between surgery omission and surgery groups for disease-free survival (RR 0.92, 95% CI 0.78–1.08), overall survival (RR 1.08, 95% CI 0.97–1.21), or metastasis-free survival (RR 1.08, 95% CI 0.97–1.20), with low heterogeneity (I^2 = 0%). Locoregional recurrence showed a non-significant trend favouring surgery. Contemporary studies using vacuum-assisted biopsy (VAB)-confirmed pCR reported 0% ipsilateral breast tumour recurrence, whereas older studies relying on clinical or imaging criteria reported higher recurrence rates.
Conclusions: Omission of breast surgery after NST may be feasible in highly selected patients when pCR is reliably confirmed, particularly using VAB-guided protocols. Nevertheless, this approach remains investigational pending results from ongoing randomized trials.

Key words: Breast Cancer; Neoadjuvant Chemotherapy; Pathological Complete Response; Breast Surgery Omission; Vacuum-Assisted Biopsy; Meta-Analysis.

Introduction

Breast cancer is the most commonly diagnosed malignancy among women worldwide and remains a leading cause of cancer-related mortality, with approximately 2.3 million new cases and 685,000 deaths reported globally in 2020 (1). Advances in systemic treatment, molecular classification, imaging, radiotherapy, and surgical techniques have substantially changed its management and improved patient outcomes over time (2,3). Contemporary breast cancer treatment is increasingly guided by tumour biology, particularly oestrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) status, rather than by anatomical characteristics alone (3).

Surgery remains a central component of curative treatment for localized breast cancer, either as the primary intervention or after neoadjuvant systemic therapy (3–5). Nevertheless, breast cancer surgery has progressively evolved toward less extensive procedures. Long-term randomized trials have demonstrated that breast-conserving surgery followed by radiotherapy provides survival outcomes comparable to those of mastectomy in appropriately selected patients (6,7). These findings established the principle that surgical morbidity may be reduced without compromising oncological outcomes when patient selection and complementary treatment are appropriate.

Neoadjuvant systemic therapy (NST) is now widely used, particularly for biologically aggressive and treatment-sensitive breast cancer subtypes (4,5,8). In addition to reducing tumour burden and facilitating breast conservation, NST provides an opportunity to evaluate tumour response in vivo. Pathological complete response (pCR), generally defined as the absence of residual invasive carcinoma in the breast and sampled axillary lymph nodes (ypT0/is ypN0), is associated with

favourable event-free and overall survival, particularly in patients with triple-negative breast cancer (TNBC) and HER2-positive disease (9,10). These subtypes also demonstrate higher probabilities of achieving pCR than hormone receptor-positive/HER2-negative breast cancer (9,10). The increasing pCR rates achieved with modern systemic regimens have therefore stimulated interest in further de-escalating locoregional treatment, including the potential omission of breast surgery (11–14).

However, the feasibility of a non-operative approach depends on the reliable identification of patients with true pCR. Clinical complete response and radiological complete response do not consistently exclude microscopic residual disease, and imaging alone may produce clinically relevant false-negative findings (11,13–15). Image-guided tumour-bed sampling, particularly vacuum-assisted biopsy (VAB) using multiple large-gauge samples and accurate clip targeting, has improved the ability to predict pCR (13–15). Nevertheless, diagnostic performance varies according to the sampling technique, tumour subtype, imaging response, and definition of complete response (13–15). Consequently, patients classified as complete responders in earlier studies based only on clinical examination or imaging may not be comparable with patients whose pCR is confirmed using contemporary biopsy protocols.

The omission of breast surgery also does not represent the omission of all locoregional treatment. Radiotherapy remains an essential component of contemporary non-operative protocols and is intended to control potential microscopic residual disease in the breast and regional lymphatic areas (11–14). Therefore, the oncological safety of surgery omission should be interpreted within an integrated treatment framework comprising effective systemic therapy, rigorous pathological response assessment, appropriate axillary management, radiotherapy, and intensive surveillance.

Although several retrospective cohorts and prospective feasibility studies have evaluated breast surgery omission after NST, the available evidence remains heterogeneous (11–14). Previous studies differ in their definitions of complete response, diagnostic methods, tumour subtypes, treatment periods, radiotherapy protocols, comparator groups, and duration of follow-up. These differences complicate the interpretation of survival and recurrence outcomes and have not been sufficiently integrated in previous quantitative analyses. Therefore, this systematic review and meta-analysis aimed to compare oncological outcomes between patients who omitted breast surgery and those who underwent surgery after achieving an apparent complete response to NST. It also examined how response-assessment methods, particularly VAB-based confirmation of pCR, influenced locoregional recurrence and the potential safety of a non-operative approach.

Material and Methods

Study Design

The purpose of this meta-analysis and systematic review was to compile all available information regarding the cancer risks associated with not performing breast reconstruction in patients who had a full response following neoadjuvant breast cancer treatment. Clinical complete response (cCR), radiological complete response (rCR), and pathological full response (pCR) are now all considered components of complete response due to the anticipated variation in response evaluation methodology among qualified studies. This diversity was known in advance to be a major contributor to clinical variability. The method of systematic review adhered to the PRISMA 2020 guidelines (16). The Mantel-Haenszel random-effects model was utilised in the meta-analysis that was conducted using RevMan 5.4.

Eligibility Criteria

Research studies were considered for inclusion if they fulfilled all the following requirements: (1) patients with breast cancer who had histological confirmation and had received neoadjuvant chemotherapy or neoadjuvant systemic therapy (NAC) or pCR (complete response) without undergoing definitive breast surgery were included in the study; (3) the studies reported at least one pre-specified oncological outcome, such as DFS, OS, MFS, LRR, ipsilateral breast tumour recurrence

(IBTR), or distant recurrence, or (4) there was enough quantitative data to enable data extraction. Eligible studies included both comparison and single-arm designs. Included papers included case series, clinical trials, cohort studies (both retrospective and prospective), and clinical trials with a minimum of five patients in the no-surgery group. Breast cancer subtype, country of origin, and publication date were not criteria. Supplementary Appendix S1 contains all of the exclusion criteria.

Search Strategy and Study Selection

CENTRAL, Web of Science, ScienceDirect, Scopus, Embase, PubMed/MEDLINE, and Scopus were among the databases that were researched thoroughly. The date range that could be searched was not limited. Researcher searched for breast cancer, neoadjuvant therapy, full response, and surgery omission using a mix of free-text keywords and MeSH phrases (Supplementary Appendix S2). An additional search was conducted using ClinicalTrials.gov to find finished or continuing trials, citation monitoring of important publications, and examine each review article and study that was included in the review by hand. After two reviewers looked over each study's abstract and title, we read each study's full text and compared it to our criteria. Researcher talked things out and came to an understanding. Researcher used a PRISMA flow diagram to record the steps of the selecting process.

Data Extraction and Quality Assessment

In order to independently retrieve the data, two reviewers employed a standardised form.. Patient and tumour demographics; treatment details (neoadjuvant regimen, radiotherapy protocol, axillary management); response assessment (definition and method), study details (author, year of publication, nation, research methodology, enrolment time, sample size, length of follow-up), and oncological outcomes (event counts, denominators, and time points) were documented. Recognising that this is an estimate since Kaplan-Meier rates account for censoring, absolute event counts were calculated using the stated rates and sample sizes when only proportional survival outcomes were reported. The occurrence of events was given precedence when their reported counts were accessible.

The Newcastle-Ottawa Scale (NOS) for comparative observational studies was used to evaluate the methodological quality of the included research. The NOS ranges from 0 to 9 stars and measures selection, comparability, and outcome ascertainment. Study quality was rated as high if it was 7-9, moderate if it was 4-6, and low if it was 0-3. Researcher used narrative analysis to assess the potential for bias in single-arm research.

Definitions

Systemic treatments such as cytotoxic chemotherapy, anti-HER2 targeted medicines, immunotherapy, and endocrine therapy are all part of neoadjuvant systemic therapy (NST), which is given before definitive surgical treatment. A subtype of cytotoxic chemotherapy is known as neoadjuvant chemotherapy (NAC). No signs of residual invasive cancer were found during the histopathological evaluation of breast and axillary lymph node samples taken following surgical excision or image-guided vacuum-assisted biopsy (VAB). This finding is referred to as a pathology complete response (pCR). A clinical complete response (cCR) occurs when a physical examination finds no signs of disease. “Radiological complete remission” (rCR) occurs when no illness is detectable on imaging modalities.

The following cancer outcome measures were utilised: DFS, OS, MFS, locoregional recurrence (LRR), ipsilateral breast tumour recurrence (IBTR), where treatment was administered, and distant recurrence (CR), where cancer has spread to other locations.

Statistical Analysis

Quantitative meta-analysis was performed using RevMan 5.4 with a Mantel-Haenszel random-effects model to account for anticipated methodological and clinical heterogeneity. Effect sizes were expressed as risk ratios (RRs) with 95% confidence intervals (CIs). For survival outcomes, events were defined as surviving patients, whereas for recurrence outcomes, events represented patients experiencing recurrence. Statistical significance was set at $P < 0.05$. Heterogeneity was assessed using the I^2 statistic, Cochran's Q test ($P < 0.10$ considered significant), and Tau^2. When substantial heterogeneity ($I^2 > 50\%$) was identified, potential sources were explored through forest plot inspection and predefined clinical variables. Subgroup analysis or meta-regression was not performed when studies per outcome were ≤3, and publication bias was not assessed when fewer than 10 studies were available per outcome

Outcomes and Criteria for Quantitative Pooling

Primary outcomes were DFS and OS at 5 and 10 years. Secondary outcomes included MFS at 5 years, LRR, and distant recurrence. Research was encompassed in quantitative meta-analysis if they provided binary outcome data for both a no-surgery group and a surgery comparator group at comparable time points. Studies not meeting these criteria were synthesized narratively.

Sensitivity Analysis

Pre-specified sensitivity analyses included: (1) excluding research that is likely to be biased; (2) exclusion of outlying studies contributing disproportionately to heterogeneity; and (3) comparison of random-effects versus fixed-effects models.

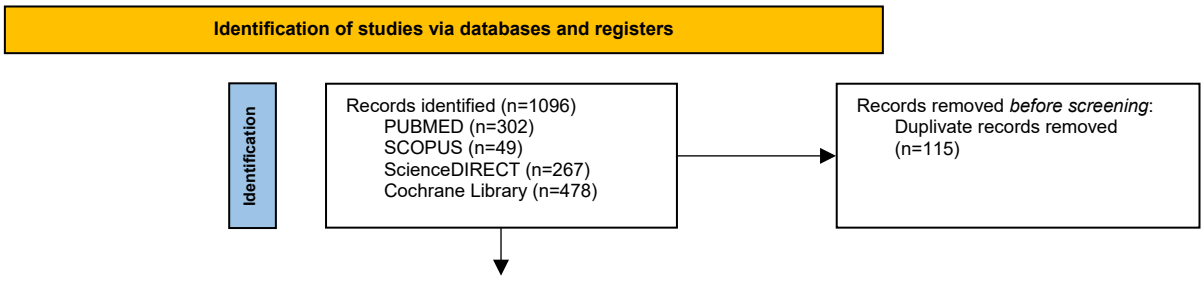
Ethical Considerations

Being a systematic review of already published data and not involving direct interaction with human subjects, this study did not require ethical approval or informed permission. The research was conducted in accordance with the principles outlined in the Declaration of Helsinki.

Results

Study Selection and Characteristics

Eight studies published between 2003 and 2024 met the pre-specified inclusion criteria (Table 1). These included four retrospective comparative cohorts (17–20), one national database analysis (21), two prospective single-arm phase 2 trials (22,23), and one retrospective case series (24). The studies spanned three countries (UK, France, and US) and treatment periods from 1986 to 2021, with median follow-up durations of 26.4 to 144 months. A critical methodological distinction separated the studies into two generations: older cohorts (Ring, Daveau, Apte) relied on clinical and/or radiological response assessment to define complete response, whereas more recent studies (Kuerer, Johnson, Tasoulis) employed vacuum-assisted biopsy (VAB) for direct pathological confirmation of pCR. This variation in response assessment represents a major source of clinical heterogeneity and is considered in the interpretation of all subsequent analyses. Detailed individual study characteristics are provided in Supplementary Results S5.



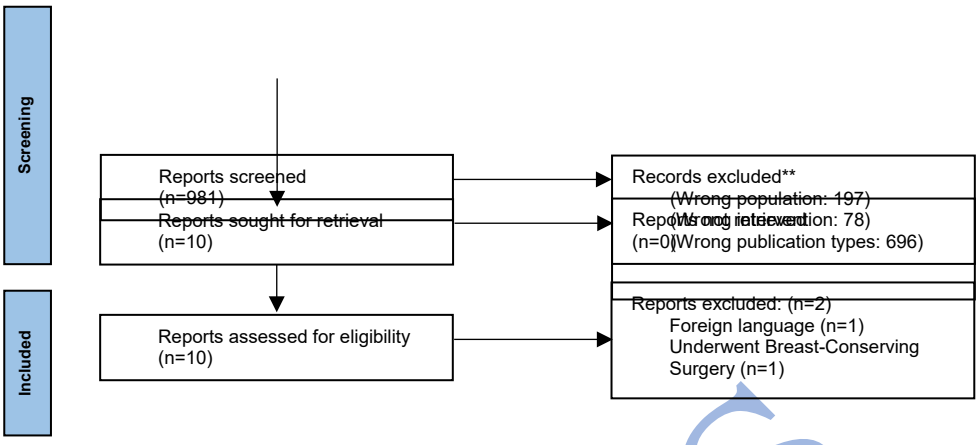


Figure 1. PRISMA 2020 flow diagram of study selection.

Table 1. Characteristics of Included Studies

Study	Year	Country	Design	N (NS)	N (Surg)	F/U (mo)	CR Def.	Key Outcomes
Ring	2003	UK	Retro. cohort	69	67	87	cCR	DFS, OS, MFS, LRR, DR
Clouth	2007	UK	Retro. cohort	20	81	33.5	pCR	MFS
Daveau	2011	France	Retro. cohort	100	65	144	cCR	DFS, OS, MFS, LRR, DR
Apte	2021	UK	Retro. cohort	28	93	138	rCR/pCR	DFS, OS, LRR, DR
Özkurt	2019	USA	Retro. (NCDB)	45*	3,938	37	cCR	OS
Johnson	2023	USA	Prosp. phase 2	6	—	44.3	pCR (VACB)	IBTR
Kuerer	2022	USA	Prosp. phase 2	31	—	26.4	pCR (VACB)	IBTR, OS
Tasoulis	2024	UK	Retro. series	7	—	67	pCR (VAB)	LRR, DFS, OS

*NS = no surgery; Surg = surgery comparator; F/U = follow-up; mo = months; CR Def. = complete response definition; cCR = clinical complete response; rCR = radiological complete response; pCR = pathological complete response; VACB = vacuum-assisted core biopsy; VAB = vacuum-assisted biopsy. *45 of 350 patients without surgery achieved cCR. — = single-arm study.*

Data Handling and Analytical Decisions

Several study-specific methodological decisions were required during data extraction to ensure analytical integrity. Johnson et al. (22) and Kuerer et al. (23) reported data from the same multicentre phase 2 trial (NCT02945579); to avoid double-counting, primary quantitative data were drawn from Kuerer et al. (the definitive expansion phase, n=50), with Johnson et al. cited for supplementary longer-term follow-up of the feasibility cohort. For Apte et al. (20), survival outcomes (OS, DFS at 10 years) used evaluable denominators (n=112) after exclusion of 9 non-breast-cancer deaths, whereas recurrence outcomes used total enrolled denominators (n=121), consistent with the original publication. Apte et al.'s wide local excision (WLE) and mastectomy subgroups were combined into a single "surgery" arm (n=93 enrolled / n=85 evaluable). For Ring et al. (17), the "metastatic, total"

figures were used for distant recurrence, which may overlap with locoregional events. Clouth et al. (18) was grouped by pCR status rather than surgical management, as metastasis outcomes were reported in this stratification; consequently, only MFS data were pooled from this study. Outcomes were extracted at 5-year and 10-year time points where available. A complete tabulation of all data handling decisions with justifications is provided in Supplementary Results Table S1.

Meta-Analysis Results

Despite variations in complete response definitions (clinical, radiological, or pathological), meta-analysis was considered appropriate because all studies evaluated the same clinical question: oncological outcomes after omission of breast surgery following apparent complete response to neoadjuvant therapy. Variability in response assessment was recognised a priori as a source of clinical heterogeneity and addressed using a random-effects model, sensitivity analyses, and narrative interpretation. Quantitative synthesis included four comparative studies (Ring, Clouth, Daveau, and Apte) reporting binary outcome data for both no-surgery and surgery groups.

Table 2. Summary of Pooled Meta-Analysis Results (Mantel–Haenszel Random-Effects)

Outcome	k	NS Events/N	Surg Events/N	Pooled RR (95% CI)	I^2	P	Interpretation
DFS – 5 yr	2	107/169	90/132	0.92 [0.78–1.08]	0%	0.30	No significant difference
DFS – 10 yr	3	108/196	118/217	1.10 [0.66–1.86]	90%	0.71	High heterogeneity; unreliable
MFS – 5 yr	3	149/189	159/213	1.08 [0.97–1.20]	0%	0.16	No significant difference
OS – 5 yr	2	143/169	103/132	1.08 [0.97–1.21]	0%	0.16	No significant difference
OS – 10 yr	3	150/196	139/217	1.23 [0.89–1.69]	86%	0.20	High heterogeneity; unreliable
LRR	3	39/197	28/225	1.52 [0.72–3.20]	44%	0.27	Trend toward higher LRR in NS
Distant recurrence	3	53/197	71/225	0.83 [0.45–1.52]	62%	0.55	No significant difference

RR = risk ratio; CI = confidence interval; k = number of studies; NS = no surgery. For survival outcomes, RR > 1.0 favours no surgery; for recurrence outcomes, RR > 1.0 indicates higher recurrence in no surgery group. Forest plots are provided in Supplementary Results.

Five-Year Outcomes

No significant differences were observed between surgery and no-surgery groups across all five-year outcomes, with no statistical heterogeneity (I^2 = 0%). Pooled analyses demonstrated comparable disease-free survival (RR 0.92, 95% CI 0.78–1.08), overall survival (RR 1.08, 95% CI 0.97–1.21), and metastasis-free survival (RR 1.08, 95% CI 0.97–1.20). Consistent findings across studies from different countries and treatment eras suggest that omission of surgery does not appear to compromise short-to-medium-term survival in patients achieving complete response after neoadjuvant therapy.

Study or Subgroup	No Surgery Events	No Surgery Total	Surgery Events	Surgery Total	Weight	Risk Ratio M-H, Random, 95% CI	Year	Risk Ratio M-H, Random, 95% CI
Ring et al., 2003	42	69	43	67	39.0 %	0.95 [0.73, 1.23]	2003	
Daveau et al., 2011	65	100	47	65	61.0 %	0.90 [0.73, 1.11]	2011	

Total (95% CI)		**169**		**132**	**100%**	**0.92 [0.78, 1.08]**	
Total events	107		90				

Heterogeneity: Tau² = 0.00; Chi² = 0.10, df = 1 (P = 0.75); I² = 0%
Test for overall effect: Z = 1.03 (P = 0.31)

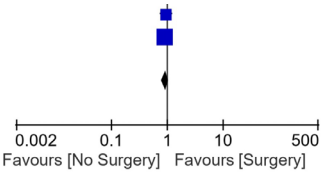


Figure 2. Forest plot of five-year DFS comparing no surgery versus surgery groups.

Study or Subgroup	No Surgery		Surgery			Risk Ratio		Risk Ratio
	Events	Total	Events	Total	Weight	M-H, Random, 95% CI	Year	M-H, Random, 95% CI
Ring et al., 2003	52	69	47	67	26.6%	1.07 [0.87, 1.32]	2003	
Clouth et al., 2007	18	20	61	81	30.7%	1.20 [0.99, 1.45]	2007	
Daveau et al., 2011	79	100	51	65	42.8%	1.01 [0.86, 1.18]	2011	
Total (95% CI)		**189**		**213**	**100%**	**1.08 [0.97, 1.20]**		
Total events	149		159					

Heterogeneity: Tau² = 0.00; Chi² = 1.85, df = 2 (P = 0.40); I² = 0%
Test for overall effect: Z = 1.41 (P = 0.16)

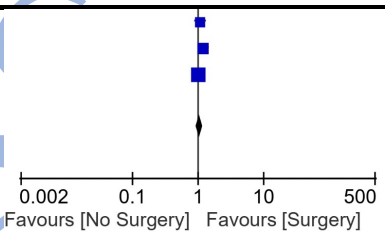


Figure 3. Forest plot of five-year MFS comparing no surgery versus surgery groups.

Study or Subgroup	No Surgery Events	Total	Surgery Events	Total	Weight	Risk Ratio M-H, Random, 95% CI	Year	Risk Ratio M-H, Random, 95% CI
Ring et al., 2003	52	69	50	67	31.3%	1.01 [0.83, 1.23]	2003	
Daveau et al., 2011	91	100	53	65	68.7%	1.12 [0.98, 1.27]	2011	
Total (95% CI)		**169**		**132**	**100.0%**	**1.08 [0.97, 1.21]**		
Total events	143							

Heterogeneity: Tau² = 0.00; Chi² = 0.74, df = 1 (P = 0.39); I² = 0%
Test for overall effect: Z = 1.42 (P = 0.16)

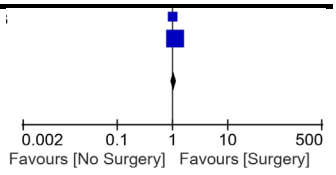
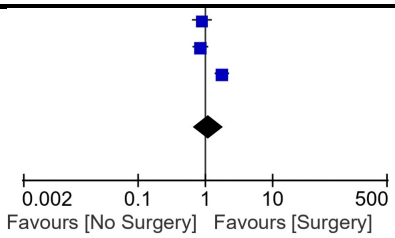


Figure 4. Forest plot of five-year OS comparing no surgery versus surgery groups.

Ten-Year Outcomes

In contrast to the relatively stable findings at five years, estimates of ten-year DFS and OS were statistically unreliable due to very high heterogeneity (I^2 = 90% and 86%, respectively). Pooled 10-year DFS showed an RR of 1.10 (95% CI: 0.66–1.86, P = 0.71; Figure 5) and 10-year OS an RR of 1.23 (95% CI: 0.89–1.69, P = 0.20; Figure 6). The heterogeneity was largely driven by the study by Apte et al. (20), which reported markedly better 10-year outcomes in the non-surgical group (DFS 88.9% vs. 50.6%; OS 92.6% vs. 56.5%). However, the Apte surgery group included a mastectomy cohort with substantially worse baseline tumour characteristics (clinical T3 stage, invasive lobular predominance), introducing considerable confounding by indication that likely explains the observed disparity. When the Apte study was excluded in sensitivity analysis, the pooled 10-year DFS RR was approximately 0.86 (95% CI: 0.70–1.06, I^2 = 0%) for 10-year DFS, a result consistent with the five-year findings and suggesting that the underlying treatment effect may be similar at both time points when appropriately comparable populations are analyzed (Supplementary Results S6). Comparison of fixed-effects versus random-effects models yielded similar point estimates for five-year outcomes with narrower confidence intervals under the fixed-effects model, further supports the stability of these findings.

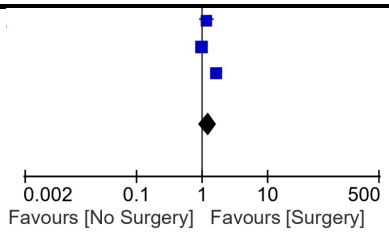
Study or Subgroup	**No Surgery**		**Surgery**			**Risk Ratio**		**Risk Ratio**
	Events	**Total**	**Events**	**Total**	**Weight**	**M-H, Random, 95% CI**	**Year**	**M-H, Random, 95% CI**
Ring et al., 2003	32	69	35	67	32.0%	0.89 [0.63, 1.25]	2003	
Daveau et al., 2011	52	100	40	65	33.8%	0.84 [0.65, 1.11]	2011	
Apte et al., 2021	24	27	43	85	34.2%	1.76 [1.37, 2.25]	2021	
Total (95% CI)		**196**		**217**	**100%**	**1.10 [0.66, 1.86]**		
Total events	108		118					



Heterogeneity: Tau^2 = 0.19; Chi^2 = 20.17, df = 2 (P < 0.0001); I^2 = 90%
Test for overall effect: Z = 0.37 (P = 0.71)

Figure 5. Forest plot of ten-year DFS comparing no surgery versus surgery groups.

Study or Subgroup	**No Surgery**		**Surgery**			**Risk Ratio**		**Risk Ratio**
	Events	**Total**	**Events**	**Total**	**Weight**	**M-H, Random, 95% CI**	**Year**	**M-H, Random, 95% CI**
Ring et al., 2003	48	69	40	67	31.4%	1.17 [0.91, 1.50]	2003	
Daveau et al., 2011	77	100	51	65	35.4%	0.98 [0.83, 1.16]	2011	
Apte et al., 2021	25	27	48	85	33.2%	1.64 [1.32, 2.03]	2021	
Total (95% CI)		**196**		**217**	**100%**	**1.23 [0.89, 1.69]**		
Total events	150		139					



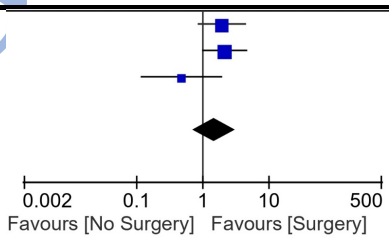
Heterogeneity: Tau² = 0.07; Chi² = 13.81, df = 2 (P = 0.0001); I² = 86%
Test for overall effect: Z = 1.27 (P = 0.20)

Figure 6. Forest plot of ten-year OS comparing no surgery versus surgery groups.

Locoregional Recurrence and Distant Recurrence

Locoregional recurrence (LRR) showed a non-significant trend toward higher recurrence in the no-surgery group (RR 1.52, 95% CI 0.72–3.20; I^2 = 44%). Moderate heterogeneity likely reflected differences in response assessment methods and patient selection across studies. No significant difference was observed for distant recurrence (RR 0.83, 95% CI 0.45–1.52), although substantial heterogeneity was present (I^2 = 62%). The wide confidence interval and variability between studies indicate limited precision for this endpoint.

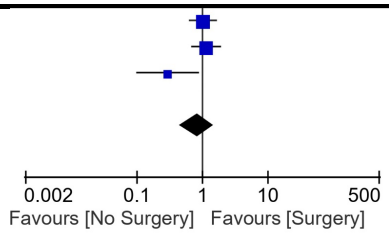
Study or Subgroup	No Surgery Events	Total	Surgery Events	Total	Weight	Risk Ratio M-H, Random, 95% CI	Year	Risk Ratio M-H, Random, 95% CI
Ring et al., 2003	14	69	7	67	36.8%	1.94 [0.84, 4.51]	2003	
Daveau et al., 2011	23	100	7	65	41.2%	2.14 [0.97, 4.69]	2011	
Apte et al., 2021	2	28	14	93	20.2%	0.47 [0.11, 1.96]	2021	
Total (95% CI)		**197**		**225**				
Total events	39		28					



Heterogeneity: Tau² = 0.19; Chi² = 3.55, df = 2 (P = 0.17); I² = 44%
Test for overall effect: Z = 1.10 (P = 0.27)

Figure 7. Forest plot of locoregional recurrence without metastasis comparing no surgery versus surgery groups.

Study or Subgroup	No Surgery Events	Total	Surgery Events	Total	Weight	Risk Ratio M-H, Random, 95% CI	Year	Risk Ratio M-H, Random, 95% CI
Ring et al., 2003	22	69	31	67	41.0%	1.02 [0.62, 1.67]	2003	
Daveau et al., 2011	28	100	16	65	39.4%	1.14 [0.67, 1.93]	2011	
Apte et al., 2021	3	28	34	93	19.6%	0.29 [0.10, 0.88]	2021	
Total (95% CI)		**197**		**225**	**100%**			
Total events	53		71					



Heterogeneity: Tau² = 0.17; Chi² = 5.21, df = 2 (P = 0.07); I² = 62%
Test for overall effect: Z = 0.59 (P = 0.55)

Figure 8. Forest plot of distant recurrence comparing no surgery versus surgery groups.

Narrative Synthesis: Single-Arm and Non-Comparable Studies

Four studies were not amenable to quantitative pooling but offered important complementary insights and are therefore summarised narratively. Özkurt et al. (21) provided the largest real-world dataset, derived from the NCDB. Among 93,417 patients who received NAC, 350 did not undergo surgery, including 45 patients who achieved clinical complete response (cCR). Five-year OS among these 45 patients was 96.8%, comparable to 92.5% observed among 3,938 patients with surgically confirmed pCR (P = 0.15). Treatment response appeared to be a stronger determinant of outcome than surgical intervention itself, as patients without cCR had significantly lower five-year overall survival (69.8%; P = 0.012). Smaller tumour size (T1–T2, OR 6.56) and node-positive disease (OR 5.02) independently predicted cCR. Despite limitations related to database analyses, these findings suggest that treatment response may have greater prognostic significance than surgery alone.

Prospective studies by Kuerer et al. (23) and Johnson et al. (22) provide the highest-quality evidence available for the omission strategy. Out of fifty patients with TNBC or HER2-positive cancer, thirteen (62% of the total) who opted out of surgery and received whole-breast radiation with tumour bed boost achieved pCR in a multicenter phase 2 trial. Upon completion of a median 26.4-month follow-up, there was a total of one hundred percent overall survival (OS), and there was zero ipsilateral breast tumour recurrence (IBTR). 71% of TNBC cases, 55% of HER2-positive diseases, and 81% of HER2-positive/hormone receptor-negative cancers had subtype-stratified pCR confirmation rates. The earlier feasibility cohort reported by Johnson et al. (n=6, median follow-up 44.3 months) similarly demonstrated 0% IBTR, providing preliminary evidence of durability. Notably, Kuerer et al. also demonstrated the limitations of imaging alone: among 17 patients with radiological complete response on imaging, 3 cases (18%) were determined to have VACB-retained invasive illness, underscoring the necessity of tissue-based confirmation.

The longest median follow-up among modern VAB-guided trials, Tasoulis et al. (24) detailed results in seven patients with VAB-confirmed pCR who were treated non-surgically. No locoregional recurrence, distant recurrence, or death was observed during follow-up. Notably, this cohort included patients with tumours up to 70 mm and multifocal disease, providing preliminary evidence that the omission approach may not necessarily be limited to the smallest tumours if pCR is rigorously confirmed.

Risk of Bias and Heterogeneity

NOS quality assessment scores ranged from 5 to 7 out of 9, indicating moderate-to-high methodological quality. Ring et al. and Daveau et al. achieved the highest scores (7/9), while Clouth et al. and Apte et al. were rated moderate. The main limitation across studies was selection bias, as patients managed without surgery were non-randomly selected and often represented favourable responders. In addition, none of the studies adjusted for important prognostic factors such as tumour subtype, grade, or baseline stage, limiting control of confounding.

A consistent pattern was observed in heterogeneity analyses: five-year endpoints showed excellent consistency across studies (I^2 = 0% for DFS, OS, and MFS), while ten-year outcomes exhibited substantial heterogeneity (I^2 = 86–90%), driven primarily by the Apte et al. study as described above. Because all of the studies were observational, there is a chance of bias, the samples were too small to draw firm conclusions, and the results were inconsistent over the long run, the GRADE criteria would place the overall certainty of evidence at a very low level for all outcomes. Table S2 of the Supplementary Results and Appendix S6 provide the results of all sensitivity analyses, as well as narrative risk of bias assessments for single-arm trials and detailed NOS scoring.

Evidence-Based Candidate Profile for Breast Surgery Omission

An important clinical question arising from this review is the identification of patients most appropriate for surgery omission. This section integrates data from all eight included studies to provide an evidence-based framework for patient selection (Table 3). The candidate criteria described below are derived from the observed characteristics of cohorts in which omission was

associated with favourable outcomes, particularly in contemporary studies reporting zero recurrence.

Tumour Molecular Subtype

Tumour molecular subtype emerged as a key determinant of eligibility for surgery omission because of its strong association with pCR. TNBC and HER2-positive tumours showed the highest pCR rates, reaching 71% and 55%, respectively, with HER2-positive/hormone receptor-negative disease achieving up to 81% in the Kuerer et al. study. Modern VAB-guided prospective studies exclusively enrolled TNBC and HER2-positive patients and consistently reported 0% ipsilateral breast tumour recurrence. In contrast, hormone receptor-positive/HER2-negative tumours demonstrated low pCR rates and a higher risk of late relapse, leading to their exclusion from contemporary omission trials.

Clinical Stage and Histology

Early-stage unicentric tumours (cT1–2, ideally ≤4 cm) are preferred candidates for the omission approach. Smaller tumour size showed promise as a standalone predictor of achieving cCR in the Özkurt et al. (21) study (T1–T2: OR 6.56, $P < 0.001$). Initial cN0 or cN1 disease is considered acceptable provided that nodal conversion to ycN0 is confirmed post-NAC by sentinel lymph node biopsy (SLNB) or targeted axillary dissection (TAD). Regarding histological type, invasive ductal carcinoma (IDC) is the optimal histology; invasive lobular carcinoma (ILC) should generally be excluded due to its characteristically diffuse growth pattern, which renders imaging assessment unreliable, and its lower pCR rates compared with IDC. High tumour grade (Grade 3) is associated with higher pCR probability and was a consistent feature across all successful omission cases in the included studies.

Response Assessment and pCR Confirmation

The pCR confirmation method emerged as the most critical determinant of oncological safety across all included studies. A stark contrast was observed between studies employing different assessment strategies: those using clinical and/or imaging response criteria alone reported LRR rates of 7–31% (Ring: 21%, Daveau: 31%, Clouth: 12.5%), whereas studies using VAB-confirmed pCR reported 0% IBTR across all patients (Kuerer: 0/31 at 26.4 months; Johnson: 0/6 at 44.3 months; Tasoulis: 0/7 at 67 months). Kuerer et al. Demonstrated that radiological complete response alone was insufficient for omission decisions: VACB revealed persistent invasive disease in three out of seventeen patients (18%) whose complete remission was confirmed by imaging. The current standard for pCR confirmation requires multimodality imaging, including MRI, combined with VAB using a ≥9-gauge needle, a minimum of 12 cores, and marker clip targeting. When these technical criteria are strictly applied, false-negative rates can be reduced to 0–5%.

Radiotherapy, Axillary Management, and Surveillance

A tumour bed increase in conjunction with whole-breast radiation was an absolute need, non-negotiable component of all omission protocols across the eight included studies. Modern protocols delivered 40–50 Gy in 15–25 fractions with a tumour bed boost of 10–16 Gy. For initially cN+ patients, regional nodal irradiation and surgical axillary confirmation (SLNB/TAD) demonstrating ycN0 status are required before committing to breast surgery omission. Patients managed without surgery require intensive, structured surveillance, imaging every six months for the initial two years, followed by yearly multimodality imaging (mammography, ultrasound, and MRI) for a minimum of five to ten years. Importantly, in the study by Kuerer et al. (23), up to 29% of no-surgery patients in the trial required biopsy for false-positive imaging findings during follow-up, all of which proved benign,

highlighting the considerable surveillance burden inherent to this approach and the need for both patient compliance and adequate institutional resources to support long-term monitoring.

Table 3. Synthesized Evidence-Based Candidate Profile for Breast Surgery Omission

Criterion	Favourable (Supports Omission)	Unfavourable (Caution/Contraindication)	Key Evidence Source
Tumour subtype	TNBC; HER2-positive (especially HER2+/HR−)	HR+/HER2− (luminal); ILC	Kuerer 2022; Johnson 2023; Tasoulis 2024; Özkurt 2019
Initial T-stage	cT1–T2 (≤4 cm); unicentric disease	cT4; inflammatory breast cancer	Kuerer 2022; Johnson 2023; Özkurt 2019
Initial N-stage	cN0 or cN1 (if post-NST ycN0 confirmed by SLNB/TAD)	Persistent cN+ after NST	Kuerer 2022; Tasoulis 2024; Daveau 2011
Histological type	Invasive ductal carcinoma (IDC)	Invasive lobular carcinoma (ILC)	Kuerer 2022; Apte 2021
Tumour grade	Grade 3 (high grade; chemosensitive)	Grade 1–2 in HR+ tumours	Apte 2021; Tasoulis 2024
pCR confirmation	VAB (≥9-gauge, ≥12 cores, marker clip targeting)	Clinical/imaging only; core biopsy <12 samples	Kuerer 2022; Johnson 2023; Tasoulis 2024
Imaging modality	MRI + mammography + USS; rCR on all modalities	Mammography/USS only; discordant imaging	Kuerer 2022; Ring 2003
NST regimen	TCHP (HER2+); AC+T ± immunotherapy (TNBC)	CMF; anthracycline-only; suboptimal regimen	Kuerer 2022; Clouth 2007
Radiotherapy	Whole-breast RT + tumour bed boost (mandatory)	Refusal of radiotherapy (absolute contraindication)	All 8 studies
Axillary management	cN0: regional nodal RT. cN1→ycN0: SLNB/TAD confirming clearance	Persistent nodal positivity post-NST	Kuerer 2022; Tasoulis 2024; Daveau 2011
Surveillance capacity	Access to high-quality multimodality imaging; intensive follow-up	Poor compliance; limited imaging access	Kuerer 2022; Johnson 2023
Patient age	≥40 years	<40 years (higher recurrence risk)	Daveau 2011; Kuerer 2022

TNBC = triple-negative breast cancer; HER2+ = HER2-positive; HR = hormone receptor; ILC = invasive lobular carcinoma; IDC = invasive ductal carcinoma; VAB = vacuum-assisted biopsy; VACB = vacuum-assisted core biopsy; MRI = magnetic resonance imaging; USS = ultrasound; RT = radiotherapy; NST = neoadjuvant systemic therapy; SLNB = sentinel lymph node biopsy; TAD = targeted axillary dissection; TCHP = docetaxel-carboplatin-trastuzumab-pertuzumab; AC+T = anthracycline-cyclophosphamide followed by taxane.

Overall Clinical Interpretation

Current evidence suggests comparable five-year outcomes between surgery and no-surgery groups, supported by low heterogeneity ($I^2 = 0\%$) and consistent findings from the NCDB analysis. However, locoregional recurrence showed a clinically relevant trend toward higher risk in older studies relying on clinical or imaging response assessment, whereas contemporary studies using VAB-confirmed pCR reported 0% ipsilateral breast tumour recurrence up to 67 months. These findings suggest that

accurate pCR confirmation, particularly with VAB, may substantially reduce locoregional recurrence risk.

The three studies reporting zero locoregional recurrence, Kuerer (23), Johnson (22), and Tasoulis (24), all shared five essential features: (1) restriction of eligibility to chemosensitive subtypes (TNBC and HER2-positive); (2) VAB-confirmed pCR with rigorous sampling protocols; (3) mandatory whole-breast radiotherapy with tumour bed boost; (4) surgical axillary assessment for initially node-positive patients; and (5) intensive multimodality surveillance. By contrast, historical studies reporting LRR rates of 7–31% uniformly lacked one or more of these features—most critically, biopsy-confirmed pCR. Taken together, these observations suggest that omitting breast surgery may be feasible and oncologically safe in a highly selected population defined by these five criteria, although the current evidence base remains insufficient to establish this approach as the standard of care. Ongoing large-scale randomized controlled trials (NRG-BR005, MICRA, NOSTRA, ASTARTE, RESPONDER) are expected to provide the definitive evidence needed to determine whether non-operative management can be integrated into standard medical treatment.

Discussion

This systematic review and meta-analysis suggest that the omission of breast surgery after an apparent complete response to neoadjuvant systemic therapy may be feasible in carefully selected patients. However, the findings should be interpreted as evidence supporting further investigation rather than as confirmation that surgery can be routinely omitted. The current evidence is derived predominantly from retrospective cohorts and small prospective feasibility studies, with important differences in patient selection, response-assessment methods, systemic treatment, radiotherapy protocols, and duration of follow-up.

The apparent similarity in intermediate-term survival between surgical and non-surgical groups is consistent with the established prognostic importance of pathological complete response. Previous pooled analyses have shown that patients who achieve pCR after neoadjuvant therapy generally have more favourable event-free and overall survival than those with residual disease, particularly in TNBC and HER2-positive breast cancer (9,25). This suggests that tumour sensitivity to systemic therapy and the resulting biological response may substantially influence prognosis. The NCDB analysis by Özkurt et al. similarly indicated that treatment response was an important determinant of survival among patients who did not undergo surgery (21).

Nevertheless, pCR should not be interpreted as equivalent to cure. Circulating and disseminated tumour cells may persist after an apparently complete local response and may contribute to subsequent distant recurrence (26,27). Moreover, the absence of a surgical specimen limits the ability to confirm that all residual invasive and in situ disease has been eradicated. These considerations are particularly important for interpreting long-term outcomes because the available evidence remains limited by heterogeneous comparator groups, incomplete pathological verification, and differences in treatment eras. Therefore, the absence of an apparent survival disadvantage should not be interpreted as proof of oncological equivalence.

The principal concern associated with surgery omission is locoregional control. Earlier studies that classified response using clinical examination or imaging alone generally reported less favourable local outcomes than contemporary studies using biopsy-confirmed pCR (17–19). Clinical and radiological complete responses do not reliably exclude microscopic residual disease because imaging accuracy varies according to tumour subtype, lesion pattern, treatment response, and imaging modality (13,15,28). Consequently, some patients classified as complete responders in earlier studies may still have harboured residual invasive carcinoma or ductal carcinoma in situ.

Contemporary studies have attempted to overcome this limitation by performing image-guided tumour-bed sampling. Vacuu m-assisted biopsy enables more extensive tissue acquisition than conventional core-needle biopsy and can improve the identification of residual disease when performed with appropriate clip targeting, large-gauge devices, and multiple representative samples

(15,28,29). MRI-guided biopsy may provide additional diagnostic information in selected patients (30). However, findings from the MICRA trial demonstrate that minimally invasive biopsy can still miss residual disease that is subsequently detected in surgical specimens (31). The safety of a non-operative strategy therefore depends not simply on whether a biopsy is performed, but on the quality of tumour-bed localization, sampling adequacy, radiological–pathological concordance, and multidisciplinary review.

Radiotherapy is another essential component of the non-operative approach. In the studies included in this review, surgery omission did not mean omission of all locoregional treatment. Patients generally continued to receive whole-breast irradiation, frequently combined with a tumour-bed boost. Evidence from breast-conserving treatment has established that radiotherapy substantially reduces the risk of local recurrence, while a tumour-bed boost provides additional local control in appropriately selected patients (32–34). In a non-operative protocol, radiotherapy functions together with systemic therapy and rigorous pathological assessment to control possible microscopic residual disease. It should therefore not be regarded as a direct or independent replacement for surgery.

Tumour subtype is also central to patient selection. Contemporary omission studies have primarily included patients with TNBC or HER2-positive disease because these subtypes are more likely to achieve pCR following modern systemic therapy and demonstrate a stronger association between pCR and favourable outcomes (9,10,22–24). In contrast, hormone receptor-positive/HER2-negative tumours have lower pCR rates and may recur after prolonged intervals, making short- or intermediate-term follow-up insufficient to establish safety (35,36). Invasive lobular carcinoma also warrants particular caution because its diffuse growth pattern may reduce the accuracy of imaging and biopsy-based response assessment, while its pathological response rate is generally lower than that of invasive ductal carcinoma (37). Patients with inflammatory breast cancer or locally extensive disease should not currently be considered candidates for surgical de-escalation outside clinical trials because of their elevated locoregional risk (38,39).

Axillary disease must be evaluated separately from response within the breast. Patients who initially present with node-positive disease require reliable post-treatment assessment of the axilla. Sentinel lymph node biopsy or targeted axillary dissection may be necessary to confirm nodal clearance before breast surgery omission is considered. Persistent nodal disease would indicate incomplete locoregional response and would substantially weaken the rationale for a non-operative pathway. Decisions regarding regional nodal irradiation should likewise be individualized according to the initial nodal stage and post-treatment findings.

The findings of this review indicate that omission should be considered only within a tightly controlled multidisciplinary framework. Potential candidates are patients with biologically treatment-sensitive tumours, a favourable clinical and imaging response, adequately sampled and biopsy-confirmed pCR, appropriate axillary assessment, acceptance of radiotherapy, and access to structured long-term surveillance. Even among carefully selected patients, surveillance may result in additional imaging and biopsies for suspicious findings, as observed in prospective omission studies (22,23). These diagnostic procedures and the anxiety associated with intensive monitoring should be considered when evaluating the overall benefits and burdens of treatment de-escalation.

The possible advantages of avoiding surgery include reduced perioperative morbidity, preservation of breast anatomy, faster recovery, and potentially improved quality of life. However, these outcomes were not consistently measured in the included studies and could not be quantitatively synthesized. Therefore, the assumption that omission necessarily improves patient-reported outcomes remains unproven. Future trials should evaluate not only recurrence and survival but also breast symptoms, body image, anxiety, surveillance burden, treatment satisfaction, healthcare utilization, and cost-effectiveness.

Several limitations affect the interpretation of this review. Most comparative evidence was observational and therefore susceptible to selection bias and confounding by indication. Patients managed without surgery were often exceptional responders with more favourable tumour biology, whereas surgical comparator groups were not always clinically equivalent. Important prognostic characteristics, including tumour subtype, stage, grade, systemic regimen, and response-assessment method, were not consistently adjusted for. The studies also differed in their definitions of complete response, biopsy techniques, radiotherapy protocols, outcome definitions, and follow-up periods. The small number of eligible studies and events limited the ability to conduct subgroup analyses, meta-regression, publication-bias assessment, and definitive evaluation of long-term outcomes. Accordingly, the certainty of the available evidence remains very low, and the pooled findings should be considered hypothesis-generating.

Future prospective randomized studies should use standardized definitions of pCR, clearly specified image-guided biopsy protocols, uniform axillary assessment, contemporary subtype-specific systemic therapy, consistent radiotherapy regimens, and sufficiently long follow-up. Patient-reported outcomes and the consequences of intensive surveillance should also be incorporated. Until such evidence becomes available, breast surgery omission should remain investigational and should be undertaken only within clinical trials or rigorously monitored prospective protocols.

Conclusions

Breast surgery omission after neoadjuvant systemic therapy appears potentially feasible in highly selected patients with treatment-sensitive breast cancer when pathological complete response is rigorously confirmed and radiotherapy and structured surveillance are provided. However, the current evidence has very low certainty and does not establish oncological equivalence; therefore, this approach remains investigational and is not recommended as the standard of care.