DOI:10.1093/jnci/djs528
Advance Access publication January 7, 2013

REVIEW

Meta-Analysis of Magnetic Resonance Imaging in Detecting Residual Breast Cancer After Neoadjuvant Therapy

Michael L. Marinovich, Nehmat Houssami, Petra Macaskill, Francesco Sardanelli, Les Irwig, Eleftherios P. Mamounas, Gunter von Minckwitz, Meagan E. Brennan, Stefano Ciatto†

Manuscript received September 3, 2012; revised November 5, 2012; accepted November 6, 2012.

Correspondence to: Michael Luke Marinovich, MPH, Screening and Test Evaluation Program, Sydney School of Public Health, A27, Edward Ford Building, University of Sydney, NSW 2006, Australia (e-mail: luke.marinovich@sydney.edu.au).

Background It has been proposed that magnetic resonance imaging (MRI) be used to guide breast cancer surgery by differentiating residual tumor from pathologic complete response (pCR) after neoadjuvant chemotherapy. This meta-analysis examines MRI accuracy in detecting residual tumor, investigates variables potentially affecting MRI performance, and compares MRI with other tests.

Methods A systematic literature search was undertaken. Hierarchical summary receiver operating characteristic (HSROC) models were used to estimate (relative) diagnostic odds ratios ([R]DORs). Summary sensitivity (correct identification of residual tumor), specificity (correct identification of pCR), and areas under the SROC curves (AUCs) were derived. All statistical tests were two-sided.

Results Forty-four studies (2050 patients) were included. The overall AUC of MRI was 0.88. Accuracy was lower for "standard" pCR definitions (referent category) than "less clearly described" (RDOR = 2.41, 95% confidence interval [CI] = 1.11 to 5.23) or "near-pCR" definitions (RDOR = 2.60, 95% CI = 0.73 to 9.24; $P = .03$.) Corresponding AUCs were 0.83, 0.90, and 0.91. Specificity was higher when negative MRI was defined as contrast enhancement less than or equal to normal tissue (0.83, 95% CI = 0.64 to 0.93) vs no enhancement (0.54, 95% CI = 0.39 to 0.69; $P = .02$), with comparable sensitivity (0.83, 95% CI = 0.69 to 0.91; vs 0.87, 95% CI = 0.80 to 0.92; $P = .45$). MRI had higher accuracy than mammography ($P = .02$); there was only weak evidence that MRI had higher accuracy than clinical examination ($P = .10$). No difference in MRI and ultrasound accuracy was found ($P = .15$).

Conclusions MRI accurately detects residual tumor after neoadjuvant chemotherapy. Accuracy was lower when pCR was more rigorously defined, and specificity was lower when test negativity thresholds were more stringent; these definitions require standardization. MRI is more accurate than mammography; however, studies comparing MRI and ultrasound are required.

J Natl Cancer Inst;2013;105:321–333

Neoadjuvant chemotherapy (NAC) has a well-established role in the management of breast cancer (1–4). For women with operable disease at presentation, the primary aim of NAC is the achievement of pathologic complete response (pCR) prior to surgery (5,6), which has been shown to confer improvements in long-term disease-free and overall survival relative to cases in which residual invasive tumor remains after NAC (7,8). Accurate ascertainment of whether pCR has been achieved or, conversely, accurate detection of the presence of residual tumor is needed to inform surgical planning (9). Currently, assessment of the presence or absence of residual tumor after NAC informs the extent of subsequent surgery; however, the avoidance of surgery remains a future goal for patients in whom an absence of residual tumor can be accurately detected (10).

Various breast imaging modalities have been used to detect whether residual malignancy is present or absent after NAC, of which magnetic resonance imaging (MRI) has been increasingly used and recommended in recent years (9,11). In this systematic review, we examine the evidence on the ability of MRI to identify whether residual malignancy is present or whether pCR has been achieved at completion of NAC, report estimates of MRI accuracy and comparative accuracy, and investigate variables potentially affecting MRI accuracy in the NAC setting.

Methods

Identification of Studies

A systematic search of the biomedical literature up to February 2011 was undertaken to identify studies assessing the accuracy of MRI after NAC in differentiating the presence of residual tumor from the absence of disease (ie, pCR). MEDLINE and EMBASE

were searched through EMBASE.com; PREMEDLINE, Database of Abstracts of Reviews of Effects, Heath Technology Assessment, and Cochrane databases were searched through Ovid. Search terms were selected to link MRI with breast cancer and response to NAC. Keywords and medical subject headings included "breast cancer," "nuclear magnetic resonance imaging," "MRI," "neoadjuvant," and "response." The full search strategy is available in Supplementary Appendix A (available online). Reference lists were also searched, and content experts were consulted to identify additional studies.

Review of Studies and Eligibility Criteria

All abstracts were screened for eligibility by one author (M. L. Marinovich), and a sample of 10% was assessed independently by a second author (N. Houssami) to ensure consistent application of the eligibility criteria. Eligible studies were required to have enrolled patients with newly diagnosed breast cancer undergoing NAC, with MRI undertaken after NAC to detect the presence of residual tumor before surgery. Studies must have the counts required to estimate sensitivity (the proportion of patients with residual tumor correctly classified by MRI as having disease present) and specificity (the proportion of cases with pCR in whom MRI declared an absence of residual tumor) or sufficient data to allow a 2×2 table to be extracted. Pathologic response based on surgical excision was the reference standard, but studies were not excluded if alternative reference standards were used in a minority of patients. Where studies presented comparisons with alternative assessment methods (ultrasound, clinical examination, mammography), estimates of accuracy were also extracted or derived for these tests. Studies in which MRI was undertaken only during NAC and studies that enrolled fewer than 10 patients were ineligible. One study (12) was identified a priori as having used a fixed MRI contrast dose, rather than dosage per unit of body weight, and was therefore excluded.

Potentially eligible citations were reviewed in full to determine eligibility (M. L. Marinovich or N. Houssami). The screening and inclusion process is summarized in Supplementary Appendix B (PRISMA flowchart) (available online).

Data Extraction

Data that was related to test accuracy, study design, patient characteristics, tumors, treatment, technical details of MRI, comparator tests, and the reference standard were extracted independently by two authors (M. L. Marinovich, and either S. Ciatto, M.E. Brennan, or F. Sardanelli). Study-level definitions of pathologic response and MRI thresholds for the absence of residual tumor were categorized according to the criteria in Supplementary Appendix C (available online). Quality appraisal was undertaken using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) checklist (modified for application to studies of residual tumor detection in this setting) (13,14). Disagreements were resolved by discussion and consensus, with arbitration by a third author (N. Houssami) when required.

Statistical Analysis

Descriptive analyses were conducted using coupled forest plots and scatter plots of study-specific estimates of sensitivity and specificity. Because studies varied in the criterion used to define a positive test result, the Rutter and Gatsonis hierarchical summary receiver operating characteristic (HSROC) model (15) was used to model MRI accuracy in terms of the diagnostic odds ratio (DOR): [sensitivity/(1–sensitivity)]/[(1–specificity)/specificity]. The DOR is the ratio of the odds of MRI being positive when residual tumor is truly present relative to the odds of MRI being positive when pCR has been achieved. A DOR of 1 means that the test does not discriminate between patients with and without residual tumor; higher values indicate better test performance (16). The HSROC model takes into account uncertainty in estimates of sensitivity and specificity within studies, as well as additional unexplained variation (heterogeneity) between studies, by the inclusion of random study effects for test accuracy and threshold (a function of the underlying test positivity rate). A shape parameter, fitted in the model as a fixed effect, allows for asymmetry in the SROC curve (ie, variation in accuracy by test threshold). For a symmetrical SROC curve, the estimated $\log_e$(DOR) is constant across thresholds. Supplementary Appendix D (available online) shows a detailed specification of the model.

HSROC models were fitted using PROC NLMIXED in SAS version 9.2 (SAS Institute, Cary, NC) (17). The distribution of the random effects for accuracy and threshold was checked for each model to ensure that normality assumptions were met. For models in which the variation in accuracy between studies was observed to be negligible, test accuracy was modeled as a fixed effect. Further detail of the model-fitting strategy is described by Macaskill (18).

Covariables were added to the HSROC model to assess whether the shape or position (accuracy) of the SROC curve(s) was associated with differences in patient, test, treatment, and study characteristics. Covariables were age (median of <50 years vs >50 years); histology (proportion that was invasive ductal carcinoma); stage (proportion that was stage I/II); human epidermal growth factor receptor 2 and estrogen receptor status (proportion that was receptor positive); chemotherapy type (anthracycline-based, anthracycline/taxane-based, anthracyclines/taxanes alone or combined, other); surgery type (proportion that was mastectomy); time from MRI to surgery (mean); midpoint of study enrollment period; definition of pCR (see Supplementary Appendix C, available online); prevalence of pCR; comparative vs noncomparative study design; prospective vs retrospective design; and consecutive vs nonconsecutive patient enrollment. Each covariable was modeled separately, and their contributions to the model were assessed by the likelihood ratio test (19). Where subgroups of studies used equivalent MRI contrast enhancement thresholds (ie, no contrast enhancement vs enhancement less than or equal to normal breast tissue) to define a negative result for residual tumor, summary estimates of sensitivity and specificity were derived for these thresholds. Ninety-five percent confidence intervals (CIs) for the expected sensitivity and specificity and the t statistics and corresponding P values for differences between MRI thresholds were derived using the ESTIMATE command in PROC NLMIXED (18).

The HSROC model was also used to compare the test performance of MRI relative to ultrasound, clinical examination, and mammography for subgroups of studies in which MRI and at least one comparator test were evaluated in the same patients (or in patient groups that substantially overlapped). Test type was included as a covariable, with separate models used for each comparison.

Where there was no evidence of asymmetry in the estimated SROC curves (assessed by the likelihood ratio test), the shape

parameter was set to zero, and the relative DOR (RDOR) was used to compare accuracy for levels of the covariable. Ninety-five percent confidence intervals for RDORs were derived from the asymptotic standard error of the estimate reported by PROC NLMIXED and assuming a *t* distribution, as described previously (18). The area under the curve (AUC) for each fitted SROC was computed by the method described by Walter (20) to provide a global measure of accuracy or using numerical integration when curves were asymmetric. The fitted curves were displayed graphically, superimposed on a scatter-plot in ROC space of study-specific estimates of (sensitivity, 1-specificity) pairs. Plotted curves were restricted to the range of data points.

Differences in QUADAS items between studies were tested using χ^2 or Fisher exact tests, as appropriate. All tests of statistical significance were two-sided; the level chosen for statistical significance was .05.

Results

Study Characteristics

A total of 2107 citations were identified. Forty-four studies (21–64) were eligible for inclusion in our meta-analysis, reporting data on 2949 patients (n = 2967 cancers) undergoing MRI and/or comparator tests; MRI data were reported for 2050 patients (n = 2068 cancers). Studies enrolled patients between 1990 and 2008 (median midpoint of recruitment = year 2001) and included a median of 36 patients in the analysis of MRI accuracy (range = 14–208). Characteristics of included studies are summarized in Table 1 and Supplementary Appendices C (pCR definitions and MRI thresholds) and E (MRI technical characteristics) (available online).

Patients enrolled in included studies had predominantly stage II and III cancer, and the majority had invasive ductal carcinoma (see Table 1). NAC was primarily anthracycline-taxane based, either sequential or in combination. Trastuzumab was used in 11 studies (range of patients within studies = 1.5%–62.5%). Radiotherapy was given before surgery in two studies (54,64). For studies that specified the type of surgery undertaken, a majority of patients underwent breast conservation. Study quality appraisal is summarized in Supplementary Appendix F (available online).

MRI Details

The majority of studies used dynamic contrast-enhanced MRI (86.4%) with a 1.5-T magnet (77.3%). Dedicated bilateral breast coils were used in all studies in which the coil type was reported. All studies that provided detail on contrast employed gadolinium-based materials, most commonly gadopentetate dimeglumine (50.0%), typically at the standard dosage of 0.1 mmol/kg body weight (61.4%) (see Supplementary Appendix E, available online). Ten studies (22.7%) (27,29–32,37,38,42,46,50) considered MRI to be negative (absence of residual tumor) when there was an absence of contrast enhancement; a further six studies (13.6%) (23,24,26,40,49,57) defined MRI negativity as contrast enhancement less than or equal to normal breast tissue. The remaining studies either did not report MRI negativity in terms of the degree of contrast enhancement (n = 20, 45.4%) (21,22,25,33,34,36,39,41,43–45,48,52,54,56,58,59,61–63) or did not specify a threshold (n = 8, 18.2%) (28,35,47,51,53,55,64) (see Supplementary Appendix C, available online).

Reference Standard

Pathology from surgical excision was the reference standard for all patients in all but two studies; test results were verified by localization biopsy in a small number (6.2%) of patients in one study (36) and by follow-up (41.2%) in another (64).

Definitions of reference standard positivity (presence of residual tumor) and negativity (pCR) varied across studies (see Supplementary Appendix C, available online). Twenty studies (21,23–25,27,30–33,36,37,41,43,44,48,52,56,57,59,62) (45.5%) defined pCR as the absence of invasive cancer on pathological examination, with or without the presence of ductal carcinoma in situ (DCIS; ie, residual DCIS was considered negative). Nine of these studies (27,30,32,36,37,41,44,56,57) provided data that allowed DCIS to be classified as either positive or negative for residual disease; primary analyses classified DCIS as negative on the reference standard, consistent with the Miller–Payne grading system (65), and the effect of classifying residual DCIS as positive in these studies was explored in sensitivity analyses. In four additional studies (9.1%) (34,45,50,53), pCR was defined as the absence of any residual invasive cancer or DCIS (ie, residual DCIS was considered positive). In 12 studies (27.3%) (26,38–40,46,47,49,51,58,60,61,63), nonspecific definitions that did not describe whether residual DCIS was considered positive or negative (eg, pCR was defined simply as the absence of residual disease/malignancy or the measurement of residua being zero) were employed. Four studies (9.1%) (28,42,54,55) allowed reference standard negativity to include small clusters of microscopic invasive cells or similar definitions of minimal residual disease ("near-pCR"). A further four studies (9.1%) (22,29,35,64) did not define reference standard positivity and negativity.

pCR Rates

Study-specific pCR rates ranged between 2.6% and 54.9%, with a median of 16.0%. The rates are presented in Supplementary Appendix G (available online), stratified by response definition.

Accuracy of MRI

Study-specific estimates of MRI sensitivity and specificity are presented in Figure 1. Median sensitivity across studies was 0.92 (interquartile range [IQR] = 0.85–0.97), and median specificity was 0.60 (IQR = 0.39–0.96). Table 2 reports the overall and covariable-specific modeled estimates of MRI accuracy (derived from separate models for each covariable); in all but one of these models (midpoint of patient enrollment = year 2000 or earlier vs year 2001 or later), the shape parameter was not statistically significant (ie, SROC curves were symmetrical). Overall, the AUC for MRI based on all 44 studies was 0.88; the SROC curve for all studies is shown in Figure 2.

MRI accuracy (DOR) differed according to the applied study-level definition of pCR (P = .03). Figure 3 displays SROC curves stratified by pCR definition. Accuracy was lowest in studies that permitted residual DCIS in the definition of pCR; relative to this referent group, accuracy was higher in studies that excluded residual DCIS from the definition of pCR (RDOR = 1.31, 95% CI = 0.33 to 5.20), applied a nonspecific definition of reference standard positivity/negativity (RDOR = 2.41, 95% CI = 1.11 to 5.23), or used a near-pCR definition (RDOR = 2.60, 95% CI = 0.73 to 9.24). Relatively few studies excluded DCIS from the pCR definition (n = 4) or used a near-pCR outcome (n = 4); hence confidence

Table 1. Summary of cohort, tumor, treatment and reference standard characteristics of included studies*

Variable	Number providing data		Median estimate	IQR	Range
	Studies	Patients			
Cohort characteristics					
No., all tests	44	2949 (2967 cancers)	37	26–56	14–869
No., MRI only†	44	2050 (2068 cancers)	36	25–52	14–208
Recruitment midpoint year	35	2574	2001	2000–2005	1992–2007
Age, mean or median, years	37	2664	49.0	47.0–51.4	42.0–56.0
Menopausal status					
Premenopausal	11	744	59.3%	49.5%–68.8%	42.6%–75.4%
Peri-/postmenopausal	11	205	40.7%	31.2%–50.5%	24.6%–57.4%
Tumor characteristics					
Clinical size, mean or median, cm†	13	1342	4.9	4.3–6.2	2.4–8.2
T stage†					
T1	16	40	0.0%	0.0%–2.3%	0.0%–60.0%
T2	14	487	50.1%	39.4%–60.5%	5.0%–84.9%
T3	14	303	31.1%	24.4%–37.5%	12.3%–47.9%
T4	15	156	13.2%	2.7%–21.0%	0.0%–57.5%
Tx	15	4	0.0%	0.0%–0.0%	0.0%–3.0%
Stage					
I	23	5	0.0%	0.0%–0.0%	0.0%–7.0%
II	21	442	56.8%	42.1%–66.7%	0.0%–86.7%
III	20	278	35.4%	30.9%–55.0%	13.3%–94.1%
IV	22	24	0.0%	0.0%–0.0%	0.0%–27.5%
Histology†					
IDC	34	1930	81.0%	69.5%–88.6%	46.7%–100.0%
ILC or IDC/ILC	34	353	11.5%	5.9%–18.8%	0.0%–33.3%
Other	34	159	2.8%	0.0%–6.9%	0.0%–16.1%
Unknown or NR	34	50	0.0%	0.0%–0.0%	0.0%–27.5%
Nodal status					
Positive	21	1234	67.4%	51.1%–73.3%	38.4%–93.8%
Negative	21	892	32.6%	22.2%–43.8%	6.2%–61.6%
Unknown or NR	21	56	0.0%	0.0%–3.1%	0.0%–15.6%
Grade					
I	11	64	11.5%	1.9%–17.8%	0.0%–35.6%
II	11	239	43.8%	29.7%–50.0%	28.9%–56.2%
I and II combined	12	785	52.9%	45.8%–61.7%	43.2%–71.2%
III	12	499	33.7%	25.2%–39.5%	23.1%–55.1%
Unknown or NR	12	177	13.3%	6.1%–16.6%	0.0%–25.0%
ER†					
Positive	22	1195	57.4%	46.3%–62.7%	29.2%–73.0%
Negative	20	691	40.7%	30.6%–48.4%	2.5%–70.8%
Unknown or NR	18	200	1.7%	0.0%–11.5%	0.0%–32.5%
PR†					
Positive	13	259	37.8%	31.4%–50.0%	6.8%–64.4%
Negative	12	316	46.1%	28.8%–62.8%	10.0%–70.8%
Unknown or NR	12	107	3.6%	0.0%–26.2%	0.0%–66.1%
HER2†					
Positive	15	266	32.3%	23.1%–48.6%	0.0%–62.5%
Negative	15	555	62.5%	38.5%–75.6%	4.4%–100.0%
Unknown or NR	15	45	0.0%	0.0%–7.1%	0.0%–60.0%
Treatment					
NAC regimen					
Anthracycline-based	39	25	7.2%	0.0%–80.0%	0.0%–100.0%
Antracycline-taxane-based	40	1928	63.9%	5.9%–100.0%	0.0%–100.0%
Other	39	12	0.0%	0.0%–2.6%	0.0%–100.0%
Studies using trastuzumab with NAC	11‡				
Trastuzumab used	10	123	22.9%	5.6%–48.6%	1.5%–62.5%
Trastuzumab not used	10	586	77.1%	51.4%–94.4%	37.5%–98.5%
Type of surgery†					
BCS	31	1230	46.2%	19.0%–59.4%	0.0%–100.0%
Mastectomy	31	932	53.8%	34.4%–76.2%	0.0%–100.0%
No surgery	32	9	0.0%	0.0%–0.0%	0.0%–41.2%

(Table continues)

Table 1 (Continued).

Variable	Number providing data: Studies	Number providing data: Patients	Median estimate	IQR	Range
Reference standard					
Type of reference standard†					
Pathology	44	2958	100.0%	100.0%–100.0%	58.8%–100.0%
Other	44	9	0.0%	0.0%–0.0%	0.0%–41.2%
Time from MRI to Sx, mean or median/estimate, days	20	1581	24	7.5–28	1–45
Prevalence of pCR, patients with MRI only†	44	2068	16.0%	11.1%–26.4%	2.6%–54.9%

* BCS = breast conserving surgery; DCIS = ductal carcinoma in situ; ER = estrogen receptor; HER2 = human epidermal growth factor receptor 2; IDC = invasive ductal carcinoma; ILC = invasive lobular carcinoma; IQR = interquartile range; MRI = magnetic resonance imaging; NAC = neoadjuvant chemotherapy; NR = not reported; pCR = pathologic complete response; PR = progesterone receptor; Sx = surgery.

† Values based on the number of cancers.

‡ Used in 11 studies, but figures based on 10 studies for which the proportion of patients who received trastuzumab was reported.

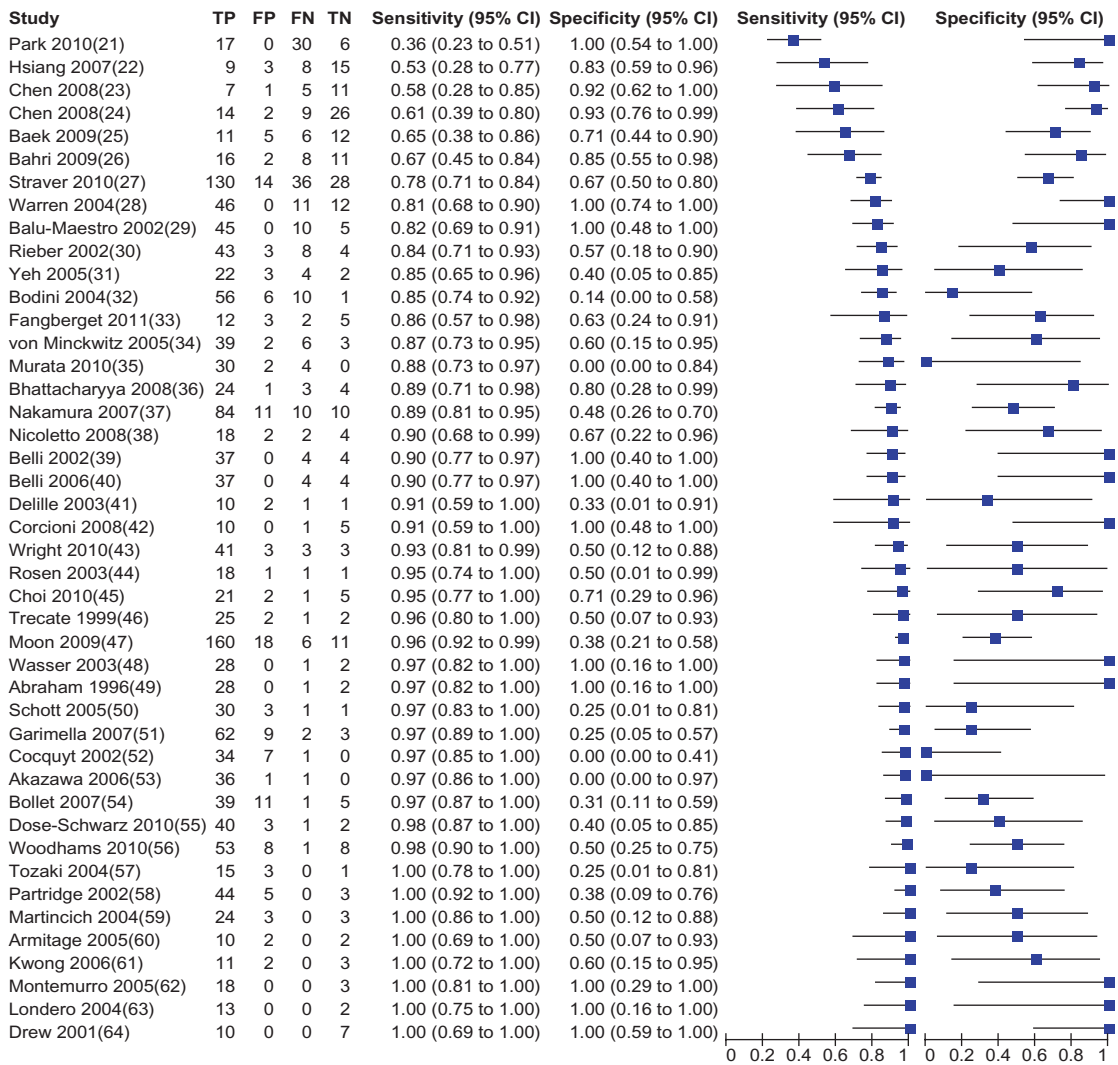

Figure 1. Forest plot of study-specific estimates of magnetic resonance imaging (MRI) sensitivity and specificity. The **black squares** and **horizontal lines** represent the estimate and 95% confidence interval (CI) for each study. The overlap between cohorts reported in the two studies by Chen et al. (23,24) is 14 patients. The two Belli et al. (39,40) studies report different study cohorts; there were no overlapping patients. FN = false negative; FP = false positive; TN = true negative; TP = true positive.

intervals around RDORs for these definitions are relatively wide. AUCs for pCR definitions were 0.83 for the absence of invasive disease, with or without DCIS; 0.86 for the absence of invasive disease and DCIS; 0.90 for nonspecific definitions; and 0.91 for near-pCR.

The midpoint of patient enrollment to each study (year 2000 or earlier vs year 2001 or later) was also associated with MRI accuracy. However, different SROC curve shapes were observed for earlier and later studies (P = .01) (see Supplementary Appendix H, available online); it is therefore not possible to report single DORs (or a RDOR) as summary measures of accuracy. Earlier studies reported consistently high sensitivity across the range of specificity values, whereas a trade-off between sensitivity and specificity at different thresholds was evident in later studies. Studies with a midpoint of patient enrollment of year 2000 or earlier reported higher overall accuracy than those with a midpoint of year 2001 or later (AUC = 0.92 vs 0.83). Comparison of QUADAS items between earlier vs later studies suggested no major differences in study quality between levels of this covariable (P > .05 for all QUADAS items). No statistical evidence was found for associations between MRI accuracy and other variables related to study design, patient characteristics, and treatment characteristics (Table 2).

Sensitivity analyses, in which modelling was repeated with changed pCR definitions in nine studies (see "Reference Standard"), resulted in similar parameter estimates to those in the primary analysis. Sensitivity analyses were also undertaken to exclude one study (23) with a patient cohort that overlapped with a second study (24); additional analyses excluded two studies that did not use a reference standard of pathologic examination in all patients (36,64). Exclusion of these studies did not substantially affect parameter estimates.

Threshold-Specific Sensitivity and Specificity of MRI

Ten studies used a complete absence of MRI enhancement to identify pCR (threshold 1), and six studies used contrast enhancement equal to or less than normal breast tissue to define a negative MRI result (threshold 2). The summary estimate of specificity was higher for threshold 2 (0.83, 95% CI = 0.64 to 0.93) than for threshold 1 (0.54, 95% CI = 0.39 to 0.69; P = .02), with comparable pooled sensitivity (0.83, 95% CI = 0.69 to 0.91; vs 0.87, 95% CI = 0.80 to 0.92; P = .45). Summary estimates are displayed in Supplementary Appendix I (available online).

Comparisons of the Accuracy of MRI and Other Tests

Table 2 presents comparisons between MRI (referent group) and comparator tests, based on subgroups of studies undertaking MRI and clinical examination [11 studies (31,32,34,43,44,50,52,54,56,58,64)], ultrasound [10 studies (31,34,36,42,45,50,54,55,62,63)], or mammography [7 studies (34,36,43,50,54,55,64)]. There was evidence that mammography had lower accuracy than MRI (RDOR = 0.27, 95% CI = 0.07 to 1.02; P = .02) (AUC = 0.89 vs 0.95; SROC presented in Figure 4). The analysis was repeated after removing one potentially influential study (36), and the statistically significant difference in accuracy between MRI and mammography remained (RDOR = 0.36, 95% CI = 0.08 to 1.60; P = .04), with comparable AUCs (0.88 vs 0.94).

There was only weak evidence that clinical examination had lower accuracy than MRI (RDOR = 0.53, 95% CI = 0.22 to 1.28; P = .10; AUC = 0.83 vs 0.89; SROC presented in Figure 5). Accuracy favored MRI in four studies; in the remaining seven studies, MRI was observed to have higher sensitivity but lower specificity than clinical examination. The lower accuracy observed for ultrasound compared with MRI was not statistically significant (RDOR = 0.54, 95% CI = 0.20 to 1.44; P = .15; AUC 0.90 vs 0.93; SROC presented in Figure 6). Differences in sensitivity were generally small in all 10 studies comparing the tests; in three of four studies with relatively larger differences in specificity, this difference favored MRI.

Discussion

In the neoadjuvant setting, accurate information on whether residual malignancy is present or whether pCR has been achieved assists in guiding surgical management of breast cancer. We modeled the accuracy of breast MRI, when performed preoperatively after NAC, through evidence synthesis from 44 studies (MRI data for 2068 cancers). Studies generally showed high sensitivity (correct detection of residual tumor), with evidence of heterogeneity in the estimates of specificity (correct identification of pCR) (Figure 1). Our meta-analysis showed that the capability of MRI for differentiating the presence of residual malignancy from pCR had an overall AUC of 0.88 and that overall accuracy differed according to definition of pCR and study timeframe.

Our meta-analysis adds substantially to earlier work (66), not only by including a greater number of studies but also by addressing comparative accuracy of MRI and other tests. In addition, we extensively explored study-level covariables, which allowed us to identify new associations and to provide methodologically appropriate estimates of sensitivity and specificity according to MRI positivity thresholds. In this meta-analysis, the median pCR rate was 16.0%, and, although this varied across 44 studies (range = 2.6%–54.9%) and an earlier review based on fewer studies suggested MRI accuracy was associated with rates of pCR (66), there was no statistically significant association between pCR rate and MRI accuracy in our models. However, the accuracy of MRI differed according to pCR definition (P = .03; see Figure 3). Relative to a referent group of studies using a clearly described "standard" definition (no invasive tumor, with or without the presence of residual DCIS) the accuracy of MRI was higher in studies using pCR definitions that were not clearly described (RDOR = 2.41, 95% CI = 1.11 to 5.23). Underlying methodological problems within studies may be associated with a poorly defined outcome definition, contributing to an overestimation of the accuracy of MRI relative to studies that employed clearly described standardized definitions of pCR.

Compared with a standard pCR definition, RDORs for studies that defined pCR as an absence of both invasive tumor and DCIS or as near-pCR were 1.31 (95% CI = 0.33 to 5.20) and 2.60 (95% CI = 0.73 to 9.24), respectively. Wide confidence intervals around these estimates reflect relatively few studies using the latter definitions (n = 4 for each definition); however, an increase in accuracy when residual DCIS is excluded vs included in the pCR definition is consistent with previous studies that reported lower MRI sensitivity in detecting DCIS relative to invasive cancer (67). Similarly, MRI has been observed to have limitations in detecting scattered, microscopic tumor foci after NAC (11,68); the estimated RDOR for near-pCR relative to a standard pCR definition may reflect fewer false

Table 2. Univariate models of magnetic resonance imaging (MRI) accuracy and comparisons of the accuracy of MRI and clinical examination, ultrasound, and mammography*

Covariable examined in univariate models of MRI accuracy	Number of studies in model	DOR (95% CI)	RDOR (95% CI)	*P*†	AUC for MRI
Base model of MRI accuracy, no covariables	44	17.89 (11.45 to 27.95)	—	—	0.88
Study characteristics					
pCR definition‡	40				
No invasive, DCIS may be present	20	10.59 (6.17 to 18.16)	1.00 (referent)	.03	0.83
No invasive, no DCIS	4	13.87 (3.95 to 48.67)	1.31 (0.33 to 5.20)		0.86
No residual disease, not further specified	12	25.47 (12.72 to 51.00)	2.41 (1.11 to 5.23)		0.90
Near-pCR, residual invasive cells	4	27.52 (8.83 to 85.82)	2.60 (0.73 to 9.24)		0.91
Midpoint of study enrolment period§	35				
Year 2000 and earlier	12	—	—	—	0.92
Year 2001 and later	23	—	—		0.83
Study-specific prevalence of pCR, continuous covariable	44	16.68 (10.50 to 26.50)‖	0.96 (0.73 to 1.28)¶	.75	0.87‖
Comparative study	44				
No, study reports MRI accuracy only	19	14.60 (8.65 to 24.66)	1.00 (referent)	.22	0.86
Yes, study reports MRI and other test	25	24.00 (12.34 to 46.66)	1.64 (0.74 to 3.65)		0.90
Design	44				
Prospective	14	14.77 (7.07 to 30.86)	1.00 (referent)	.82	0.86
Retrospective	19	19.47 (10.05 to 37.73)	1.32 (0.51 to 3.41)		0.88
Unknown	11	19.41 (8.86 to 42.54)	1.31 (0.46 to 3.75)		0.88
Patient enrollment	44				
Consecutive clinically defined cohort	6	18.47 (5.98 to 57.07)	1.00 (referent)	.64	0.88
Consecutive patients who had the test	12	11.77 (5.64 to 24.54)	0.64 (0.17 to 2.41)		0.84
Nonconsecutive	20	21.09 (11.64 to 38.22)	1.14 (0.32 to 4.02)		0.84
Unknown	6	19.23 (6.60 to 56.08)	1.04 (0.23 to 4.80)		0.89
Patient characteristics					
Median age, years	37				
<50	22	22.78 (11.96 to 43.41)	1.00 (referent)	.22	0.90
>50	15	13.31 (6.62 to 26.76)	0.58 (0.24 to 1.45)		0.85
Histology, proportion IDC relative to other invasive types, continuous covariable	34	17.24 (10.18 to 29.20)‖	0.97 (0.60 to 1.58)¶	1.00	0.88‖
Stage, I or II, %, continuous covariable	21	29.85 (11.49 to 77.56)‖	0.79 (0.55 to 1.13)¶	.17	0.91‖
HER2+, %	15				
<30%	7	11.54 (3.35 to 39.71)	1.00 (referent)	.75	0.84
>30%	8	13.09 (1.31 to 131.08)	1.13 (0.04 to 32.07)		0.85
ER+, %, continuous covariable	22	14.96 (9.56 to 23.41)‖	1.21 (0.82 to 1.78)¶	.32	0.86‖
Treatment characteristics					
Chemotherapy type	44				
Anthracycline-based	7	10.14 (3.51 to 29.32)	1.00 (referent)	.66	0.83
Anthracycline/taxane-based	16	20.44 (10.17 to 41.08)	2.02 (0.58 to 6.97)		0.89
Anthracyclines, taxanes alone or combined	17	18.50 (9.92 to 34.52)	1.83 (0.55 to 6.04)		0.88
Other	4	18.70 (4.19 to 83.42)	1.84 (0.30 to 11.49)		0.88
Proportion mastectomy, relative to BCS/no surgery	31				
<50%	14	19.95 (9.34 to 42.61)	1.00 (referent)	.75	0.89
>50%	17	22.68 (10.75 to 47.86)	1.14 (0.41 to 3.19)		0.90
Time from MRI to surgery, continuous covariable	20	17.11 (9.52 to 30.74)‖	0.99 (0.94 to 1.03)#	.53	0.87‖
Comparative accuracy studies					
MRI	11	19.73 (4.00 to 97.30)	1.00 (referent)	.10	0.89
Clinical examination		10.45 (4.30 to 25.37)	0.53 (0.22 to 1.28)		0.83
MRI	10	42.94 (7.08 to 260.55)	1.00 (referent)	.15	0.93
Ultrasound		23.25 (9.61 to 56.21)	0.54 (0.20 to 1.44)		0.90
MRI	7	73.74 (8.18 to 665.01)	1.00 (referent)	.02	0.95
Mammography		20.12 (7.18 to 56.42)	0.27 (0.07 to 1.02)		0.89

* AUC = area under the receiver operating characteristics curve; BCS = breast conserving surgery; CI = confidence interval; DCIS = ductal carcinoma in situ; ER = estrogen receptor; HER2 = human epidermal growth factor receptor 2; IDC = invasive ductal carcinoma; MRI = magnetic resonance imaging; pCR = pathologic complete response; RDOR = relative diagnostic odds ratio; — = not applicable.

† *P* value from two-sided likelihood-ratio test.

‡ pCR (absence of residual tumor) was considered to be negative on the reference standard.

§ Curves are not symmetric and do not have the same shape (see figure in Supplementary Appendix H, available online).

‖ Estimate at the median value of the continuous covariable (prevalence of pCR = 16.0%; proportion IDC [excluding histology unknown or not reported] = 84.2%; proportion of stage I/II = 56.8%; proportion ER+ = 57.4%; time from MRI to surgery = 24 days).

¶ Expected RDOR per 10% increase in covariable.

Expected RDOR per 1-day increase in time from MRI to surgery.

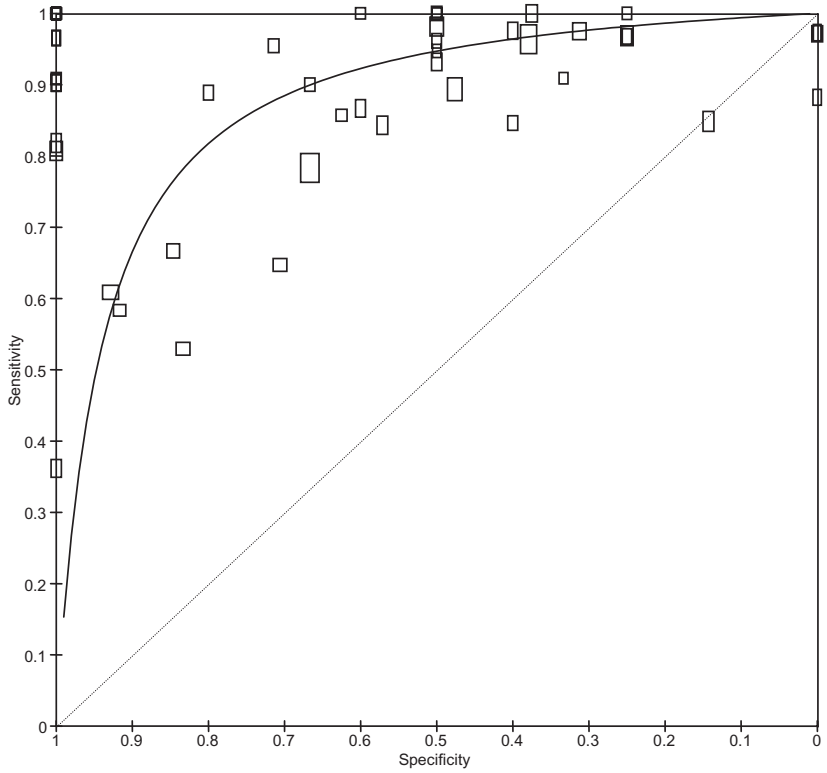

Figure 2. Summary receiver operating characteristics curve for magnetic resonance imaging from all included studies. The **black squares** represent estimates of sensitivity and specificity for each study. The **relative size** of the black square represents study sample size.

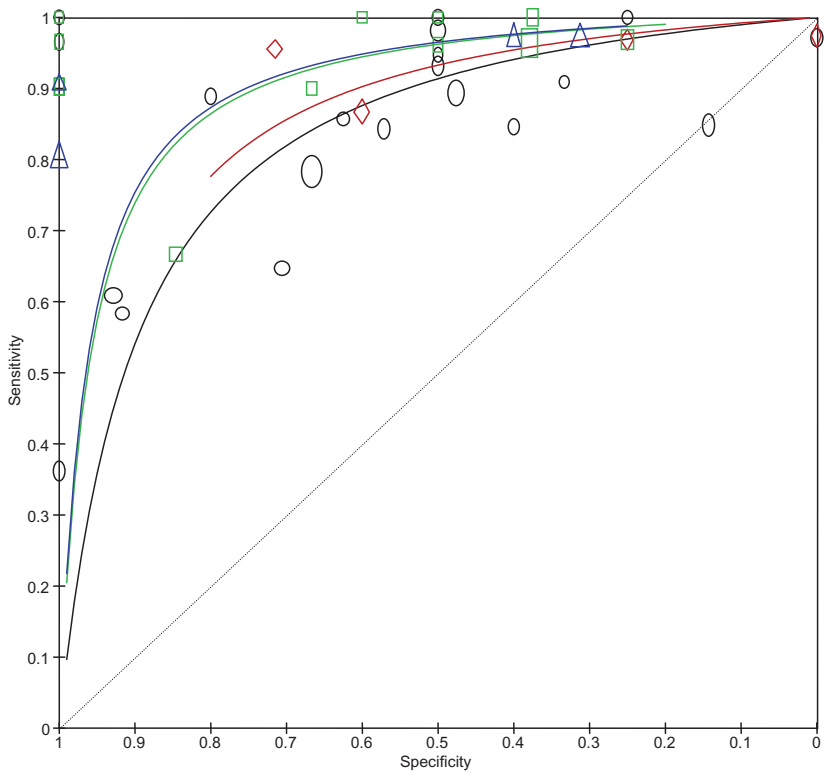

Figure 3. Summary receiver operating characteristics curves for magnetic resonance imaging stratified by definition of pathologic complete response (pCR). The **black circles** and **black curve** represent studies that defined pathologic response as the absence of invasive cancer, with or without the presence of ductal carcinoma in situ (DCIS). The **red diamonds** and **red curve** represent studies that defined pathologic response as the absence of invasive cancer and DCIS. The **green squares** and **green curve** represent studies that defined pathologic response as no residual disease (not further specified). The **blue triangles** and **blue curve** represent studies that defined pathologic response as near-pCR (minimal residual disease). The **relative size** of the symbols represents study sample size.

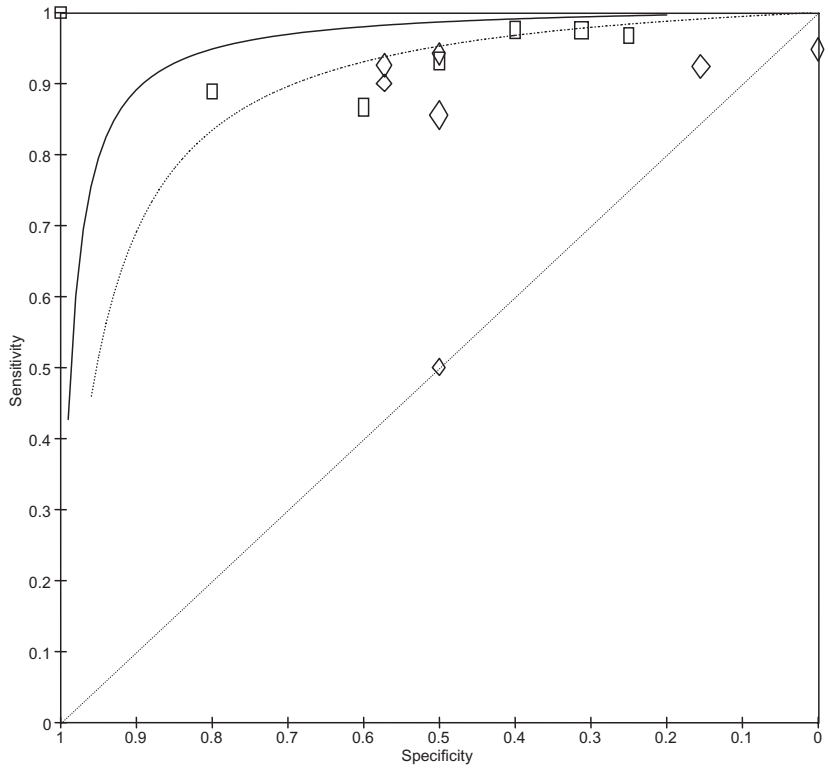

Figure 4. Summary receiver operating characteristics curves for magnetic resonance imaging (MRI) vs mammography. The **black squares** and **solid curve** represent MRI. The **black diamonds** and **dashed curve** represent mammography. The **relative size** of the symbols represents study sample size.

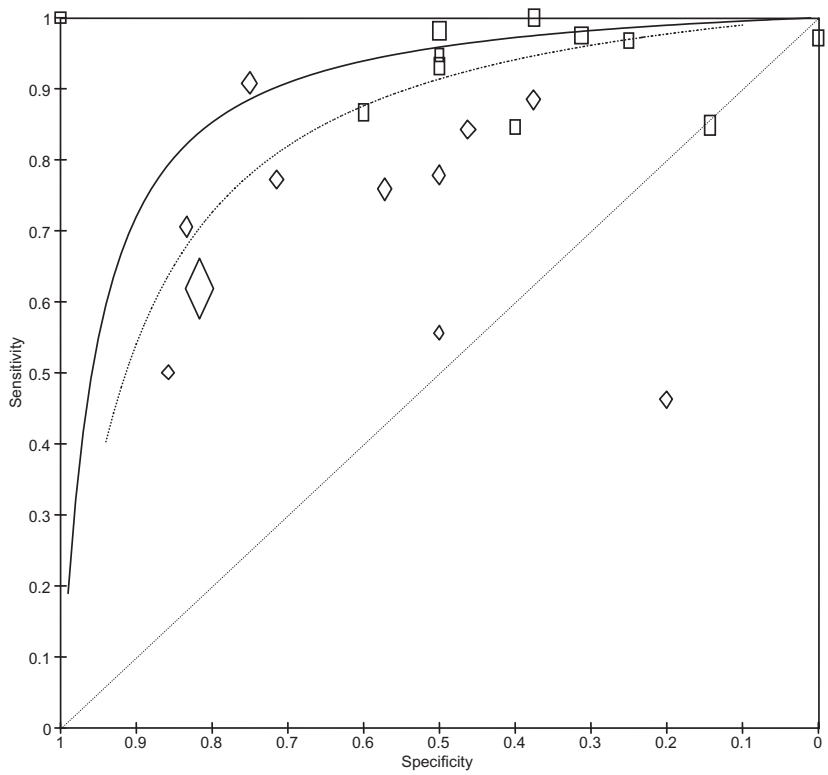

Figure 5. Summary receiver operating characteristics curves for magnetic resonance imaging (MRI) vs clinical examination. The **black squares** and **solid curve** represent MRI. The **black diamonds** and **dashed curve** represent clinical examination. The **relative size** of the symbols represents study sample size.

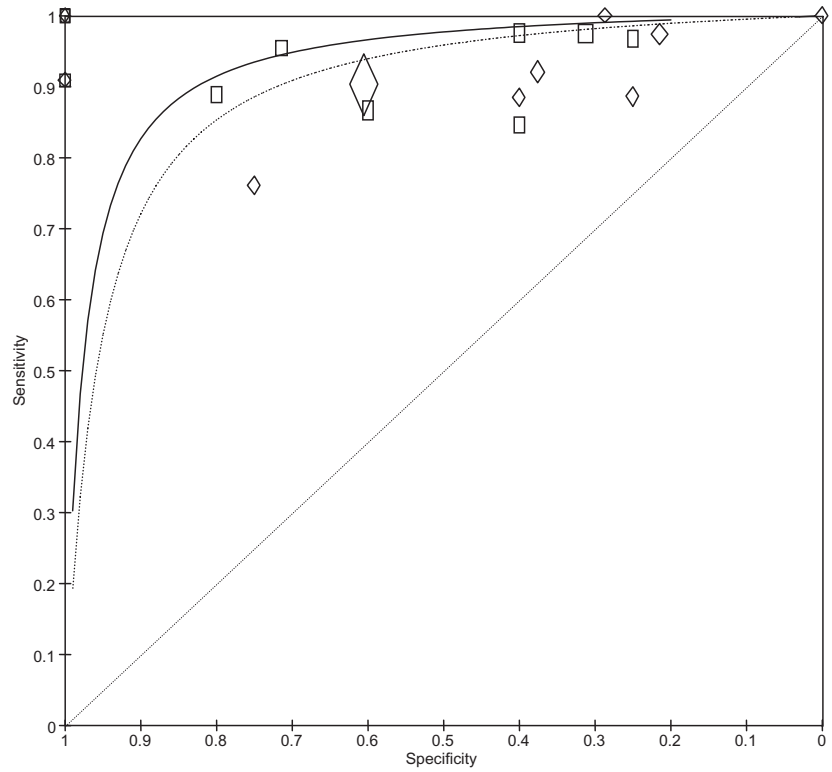

Figure 6. Summary receiver operating characteristics curves for magnetic resonance imaging (MRI) vs ultrasound. The **black squares** and **solid curve** represent MRI. The **black diamonds** and **dashed curve** represent ultrasound. The **relative size** of the symbols represents study sample size.

negatives and a consequent increase in true negative MRI results when a near-pCR definition is used. Given that near-pCR may plausibly overestimate accuracy relative to standard pCR definitions and given the impact of residual malignancy on prognosis (69), the use of near-pCR as an outcome in the preoperative, post-NAC setting is not recommended. This analysis highlights the importance of standardizing pCR definitions in the NAC setting (70).

One of the strengths of our work is the characterization (Table 1) and evaluation in analysis (Table 2) of a large number of covariables related to study quality, patient characteristics, tumor characteristics, MRI, and treatment. We found that studies with a midpoint of patient enrollment of year 2000 or earlier reported higher overall AUC (AUC = 0.92) than those with a midpoint of year 2001 or later (AUC = 0.83), although SROC curve shapes differed for earlier and later studies (P = .01). Examination of the curves indicated that earlier studies had consistently high sensitivity across the range of specificity values, whereas a trade-off between sensitivity and specificity at different thresholds was evident in later studies. With the evolution of MRI technology over time, it may appear counterintuitive that relatively lower accuracy was observed in more recent studies; however, a meta-analysis of preoperative MRI (71) also suggested similar findings. No clear differences in study quality between timeframes were observed to account for the above finding; however, it is possible that earlier studies may have involved radiologists with MRI expertise and that later studies involved readers with less MRI-dedicated expertise, reflecting broader adoption of MRI in breast imaging practice. It may also be possible that readers in more recent studies adjusted the implicit threshold used to define MRI positivity/negativity in response to the relatively lower specificity reported in many earlier studies.

In our subgroup analysis of studies that reported contrast enhancement thresholds applied to declare a positive or negative test, there was no statistically significant difference between thresholds for summary estimates of MRI sensitivity (ie, correct detection of residual malignancy) (0.83 vs 0.87; P = .45). However, MRI specificity (ie, correct identification of pCR) was statistically significantly greater when contrast enhancement equal to or less than normal breast tissue was considered negative for residual tumor compared with a complete absence of contrast uptake (0.83 vs 0.54; P = .02), which reflects the likelihood that enhancement caused by inflammatory or reactive changes post-NAC may be considered false positive for residual malignancy using the latter threshold. These findings raise concerns about MRI potentially underestimating the effect of NAC in achieving pCR where a stringent threshold is applied for defining the absence of residual malignancy; however, when pCR is identified by contrast enhancement equal to or less than normal breast tissue, the relatively higher specificity of MRI may allow better planning of breast conserving surgery. Standardization of MRI interpretation criteria/thresholds in this clinical setting is required.

Our analysis showed that the accuracy of MRI was statistically significantly higher than that of mammography (P = .02). There was only weak evidence suggesting that MRI also had greater accuracy than clinical examination (P = .10). Differences in summary accuracy estimates for MRI and ultrasound were not statistically significant (P = .15). These subgroup analyses were based on fewer studies because they were limited to studies that directly compared tests and, therefore, have relatively reduced power to detect differences in test accuracy. This may account for the lack of statistical differences between MRI and ultrasound accuracy in subgroup analysis; however,

the findings may also represent true similarity in accuracy for MRI and ultrasound. We were unable to compare the accuracy of MRI with a combination of ultrasound and clinical examination because of a lack of studies that presented these data. The high relative cost of MRI, combined with potential advantages of clinical examination and ultrasound in terms of accessibility, suggest that a combination of the latter may be a reasonable alternative testing strategy to MRI in preoperative assessment after NAC. We recommend that future research aim to compare the combined accuracy of ultrasound and clinical examination with that of MRI in the NAC setting.

This study has some limitations. The reporting of information related to methodological quality was highly variable between studies and individual QUADAS items (see Supplementary Appendix F, available online), and some studies did not adequately describe MRI technical details (see Supplementary Appendix E, available online). Investigators should be encouraged to fully describe study methodology, MRI technology, and technique to allow the risk of bias and the generalizability of study findings to be assessed. Furthermore, relatively recent improvements in MRI technology that may be expected to potentially improve accuracy (eg, multichannel coils; ≥3-T magnets; contrast materials with high relaxivity; additional sequences allowing for diffusion weighted imaging) were underrepresented in studies included in this analysis. The effect of these developments on the accuracy of MRI should be the subject of further study.

In summary, our meta-analysis has shown good overall accuracy for MRI, although accuracy estimates varied with the definition of pCR, which highlights the importance of standardizing pCR definitions. Subgroup analysis also suggests that MRI may be more likely to be false positive for residual malignancy (thereby falsely underestimating the effect of NAC in achieving pCR) in studies that defined absence of residual malignancy on MRI as contrast enhancement less than or equal to that of normal breast tissue, rather than an absence of enhancement. In comparative studies, MRI was more accurate than mammography, but no differences in accuracy were observed between MRI and other less technically complex and costly tests (ultrasound, clinical examination). However, relatively few studies reported direct comparisons between MRI and other tests, and the comparative accuracy of MRI and combined ultrasound and clinical examination warrants further investigation in well-designed clinical trials.

References

1. Makhoul I, Kiwan E. Neoadjuvant systemic treatment of breast cancer. *J Surg Oncol*. 2011;103(4):348–357.
2. Fisher B, Bryant J, Wolmark N, et al. Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. *J Clin Oncol*. 1998;16(8):2672–2685.
3. Von Minckwitz G, Costa SD, Raab G, et al. Dose-dense doxorubicin, docetaxel, and granulocyte colony-stimulating factor support with or without tamoxifen as preoperative therapy in patients with operable carcinoma of the breast: a randomized, controlled, open phase IIb study. *J Clin Oncol*. 2001;19(15):3506–3515.
4. Von Minckwitz G, Kummel S, Vogel P, et al. Intensified neoadjuvant chemotherapy in early-responding breast cancer: phase III randomized GeparTrio study. *J Natl Cancer Inst*. 2008;100(8):552–562.
5. Kaufmann M, Hortobagyi GN, Goldhirsch A, et al. Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: an update. *J Clin Oncol*. 2006;24(12):1940–1949.
6. Kaufmann M, Von Minckwitz G, Bear HD, et al. Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: new perspectives 2006. *Ann Oncol*. 2007;18(12):1927–1934.
7. Chollet P, Amat S, Cure H, et al. Prognostic significance of a complete pathological response after induction chemotherapy in operable breast cancer. *Br J Cancer*. 2002;86(7):1041–1046.
8. Montagna E, Bagnardi V, Rotmensz N, et al. Pathological complete response after preoperative systemic therapy and outcome: relevance of clinical and biologic baseline features. *Breast Cancer Res Treat*. 2010;124(3):689–699.
9. Sardanelli F, Boetes C, Borisch B, et al. Magnetic resonance imaging of the breast: recommendations from the EUSOMA working group. *Eur J Cancer*. 2010;46(8):1296–1316.
10. Heys SD, Chaturvedi S. Primary chemotherapy in breast cancer: the beginning of the end or the end of the beginning for the surgical oncologist? *World J Surg Oncol*. 2003;1.
11. Morrow M, Waters J, Morris E. MRI for breast cancer screening, diagnosis, and treatment. *Lancet*. 2011;378(9805):1804–1811.
12. DeMartini WB, Lehman CD, Peacock S, Russell MT. Computer-aided detection applied to breast MRI: assessment of CAD-generated enhancement and tumor sizes in breast cancers before and after neoadjuvant chemotherapy. *Academic Radiol*. 2005;12(7):806–814.
13. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3.
14. Whiting PF, Weswood ME, Rutjes AWS, Reitsma JB, Bossuyt PNM, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol*. 2006;6.
15. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. 2001;20(19):2865–2884.
16. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol*. 2003;56(11):1129–1135.
17. SAS Institute. SAS/STAT User's Guide. Version 8. Cary, NC: SAS Institute; 1999.
18. Macaskill P. Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *J Clin Epidemiol*. 2004;57(9):925–932.
19. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Version 1.0. The Cochrane Collaboration; 2010:1–61.
20. Walter SD. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Stat Med*. 2002;21(9):1237–1256.
21. Park SH, Moon WK, Cho N, et al. Diffusion-weighted MR imaging: pretreatment prediction of response to neoadjuvant chemotherapy in patients with breast cancer. *Radiology*. 2010;257(1):56–63.
22. Hsiang DJ, Yamamoto M, Mehta RS, et al. Predicting nodal status using dynamic contrast-enhanced magnetic resonance imaging in patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy with and without sequential trastuzumab. *Arch Surg*. 2007;142(9):855–860.
23. Chen JH, Mehta RS, Nalcioglu O, Su MY. Inflammatory breast cancer after neoadjuvant chemotherapy: can magnetic resonance imaging precisely diagnose the final pathological response? *Ann Surg Oncol*. 2008;15(12):3609–3613.
24. Chen JH, Feig B, Agrawal G, et al. MRI evaluation of pathologically complete response and residual tumors in breast cancer after neoadjuvant chemotherapy. *Cancer*. 2008;112(1):17–26.
25. Baek HM, Chen JH, Nie K, et al. Predicting pathologic response to neoadjuvant chemotherapy in breast cancer by using MR imaging and quantitative 1H MR spectroscopy. *Radiology*. 2009;251(3):653–662.
26. Bahri S, Chen JH, Mehta RS, et al. Residual breast cancer diagnosed by MRI in patients receiving neoadjuvant chemotherapy with and without bevacizumab. *Ann Surg Oncol*. 2009;16(6):1619–1628.
27. Straver ME, Loo CE, Rutgers EJT, et al. MRI-model to guide the surgical treatment in breast cancer patients after neoadjuvant chemotherapy. *Ann Surg*. 2010;251(4):701–707.

28. Warren RML, Bobrow LG, Earl HM, et al. Can breast MRI help in the management of women with breast cancer treated by neoadjuvant chemotherapy? *Br J Cancer*. 2004;90(7):1349–1360.
29. Balu-Maestro C, Chapellier C, Bleuse A, Chanalet I, Chauvel C, Largillier R. Imaging in evaluation of response to neoadjuvant breast cancer treatment benefits of MRI. *Breast Cancer Res Treat*. 2002;72(2):145–152.
30. Rieber A, Brambs HJ, Gabelmann A, Heilmann V, Kreienberg R, Kuhn T. Breast MRI for monitoring response of primary breast cancer to neoadjuvant chemotherapy. *Eur Radiol*. 2002;12(7):1711–1719.
31. Yeh E, Slanetz P, Kopans DB, et al. Prospective comparison of mammography, sonography, and MRI in patients undergoing neoadjuvant chemotherapy for palpable breast cancer. *Am J Roentgenol*. 2005;184(3):868–877.
32. Bodini M, Berruti A, Bottini A, et al. Magnetic resonance imaging in comparison to clinical palpation in assessing the response of breast cancer to epirubicin primary chemotherapy. *Breast Cancer Res Treat*. 2004;85(3):211–218.
33. Fangberget A, Nilsen LB, Hole KH, et al. Neoadjuvant chemotherapy in breast cancer-response evaluation and prediction of response to treatment using dynamic contrast-enhanced and diffusion-weighted MR imaging. *Eur Radiol*. 2011;21(6):1188–1199.
34. Von Minckwitz G, Raab G, Caputo A, et al. Doxorubicin with cyclophosphamide followed by docetaxel every 21 days compared with doxorubicin and docetaxel every 14 days as preoperative treatment in operable breast cancer: The GEPARDUO study of the German Breast Group. *J Clin Oncol*. 2005;23(12):2676–2685.
35. Murata Y, Kubota K, Hamada N, et al. Diffusion-weighted magnetic resonance imaging for assessment after neoadjuvant chemotherapy in breast cancer, based on morphological concepts. *Oncol Lett*. 2010;1(2):293–298.
36. Bhattacharyya M, Ryan D, Carpenter R, Vinnicombe S, Gallagher CJ. Using MRI to plan breast-conserving surgery following neoadjuvant chemotherapy for early breast cancer. *Br J Cancer*. 2008;98(2):289–293.
37. Nakamura S, Ishiyama M, Tsunoda-Shimizu H. Magnetic resonance mammography has limited ability to estimate pathological complete remission after primary chemotherapy or radiofrequency ablation therapy. *Breast Cancer*. 2007;14(2):123–130.
38. Nicoletto MO, Nitti D, Pescarini L, et al. Correlation between magnetic resonance imaging and histopathological tumor response after neoadjuvant chemotherapy in breast cancer. *Tumori*. 2008;94(4):481–488.
39. Belli P, Romani M, Costantini M, et al. Role of magnetic resonance imaging in the pre and postchemotherapy evaluation in locally advanced breast carcinoma. *Rays*. 2002;27(4):279–290.
40. Belli P, Costantini M, Malaspina C, Magistrelli A, LaTorre G, Bonomo L. MRI accuracy in residual disease evaluation in breast cancer patients treated with neoadjuvant chemotherapy. *Clin Radiol*. 2006;61(11):946–953.
41. Delille JP, Slanetz PJ, Yeh ED, Halpern EF, Kopans DB, Garrido L. Invasive ductal breast carcinoma response to neoadjuvant chemotherapy: Noninvasive monitoring with functional MR imaging—pilot study. *Radiology*. 2003;228(1):63–69.
42. Corcioni B, Santilli L, Quercia S, et al. Contrast-enhanced US and MRI for assessing the response of breast cancer to neoadjuvant chemotherapy. *J Ultrasound*. 2008;11(4):143–150.
43. Wright FC, Zubovits J, Gardner S, et al. Optimal assessment of residual disease after neo-adjuvant therapy for locally advanced and inflammatory breast cancer—clinical examination, mammography, or magnetic resonance imaging? *J Surg Oncol*. 2010;101(7):604–610.
44. Rosen EL, Blackwell KL, Baker JA, et al. Accuracy of MRI in the detection of residual breast cancer after neoadjuvant chemotherapy. *Am J Roentgenol*. 2003;181(5):1275–1282.
45. Choi JH, Lim HI, Lee SK, et al. The role of PET CT to evaluate the response to neoadjuvant chemotherapy in advanced breast cancer: comparison with ultrasonography and magnetic resonance imaging. *J Surg Oncol*. 2010;102(5):392–397.
46. Trecate G, Ceglia E, Stabile F, et al. Locally advanced breast cancer treated with primary chemotherapy: comparison between magnetic resonance imaging and pathologic evaluation of residual disease. *Tumori*. 1999;85(4):220–228.
47. Moon HG, Han W, Lee JW, et al. Age and HER2 expression status affect MRI accuracy in predicting residual tumor extent after neo-adjuvant systemic treatment. *Ann Oncol*. 2009;20(4):636–641.
48. Wasser K, Sinn HP, Fink C, et al. Accuracy of tumor size measurement in breast cancer using MRI is influenced by histological regression induced by neoadjuvant chemotherapy. *Eur Radiol*. 2003;13(6):1213–1223.
49. Abraham DC, Jones RC, Jones SE, et al. Evaluation of neoadjuvant chemotherapeutic response of locally advanced breast cancer by magnetic resonance imaging. *Cancer*. 1996;78(1):91–100.
50. Schott AF, Roubidoux MA, Helvie MA, et al. Clinical and radiologic assessments to predict breast cancer pathologic complete response to neoadjuvant chemotherapy. *Breast Cancer Res Treat*. 2005;92(3):231–238.
51. Garimella V, Qutob O, Fox JN, et al. Recurrence rates after DCE-MRI image guided planning for breast-conserving surgery following neoadjuvant chemotherapy for locally advanced breast cancer patients. *Eur J Surg Oncol*. 2007;33(2):157–161
52. Cocquyt VF, Villeirs GM, Blondeel P, et al. Assessment of response to preoperative chemotherapy in patients with stage II and III breast cancer: the value of MRI. *Breast*. 2002;11(4):306–315.
53. Akazawa K, Tamaki Y, Taguchi T, et al. Preoperative evaluation of residual tumor extent by three-dimensional magnetic resonance imaging in breast cancer patients treated with neoadjuvant chemotherapy. *Breast J*. 2006;12(2):130–137.
54. Bollet MA, Thibault F, Bouillon K, et al. Role of dynamic magnetic resonance imaging in the evaluation of tumor response to preoperative concurrent radiochemotherapy for large breast cancers: a prospective phase II study. *Int J Radiat Oncol Biol Phys*. 2007;69(1):13–18.
55. Dose-Schwarz J, Tiling R, Avril-Sassen S, et al. Assessment of residual tumour by FDG-PET: conventional imaging and clinical examination following primary chemotherapy of large and locally advanced breast cancer. *Br J Cancer*. 2010;102(1):35–41.
56. Woodhams R, Kakita S, Hata H, et al. Identification of residual breast carcinoma following neoadjuvant chemotherapy: Diffusion-weighted imaging—comparison with contrast-enhanced MR imaging and pathologic findings. *Radiology*. 2010;254(2):357–366.
57. Tozaki M, Uno S, Kobayashi T, et al. Histologic breast cancer extent after neoadjuvant chemotherapy: comparison with multidetector-row CT and dynamic MRI. *Radiat Med*. 2004;22(4):246–253.
58. Partridge SC, Gibbs JE, Lu Y, Esserman LJ, Sudilovsky D, Hylton NM. Accuracy of MR imaging for revealing residual breast cancer in patients who have undergone neoadjuvant chemotherapy. *Am J Roentgenol*. 2002;179(5):1193–1199.
59. Martincich L, Montemurro F, De Rosa G, et al. Monitoring response to primary chemotherapy in breast cancer using dynamic contrast-enhanced magnetic resonance imaging. *Breast Cancer Res Treat*. 2004;83(1):67–76.
60. Armitage P, Behrenbruch C, Brady M, Moore N. Extracting and visualizing physiological parameters using dynamic contrast-enhanced magnetic resonance imaging of the breast. *Med Image Anal*. 2005;9(4):315–329.
61. Kwong MS, Chung GG, Horvath LJ, et al. Postchemotherapy MRI overestimates residual disease compared with histopathology in responders to neoadjuvant therapy for locally advanced breast cancer. *Cancer J*. 2006;12(3):212–221.
62. Montemurro F, Martincich L, De Rosa G, et al. Dynamic contrast-enhanced MRI and sonography in patients receiving primary chemotherapy for breast cancer. *Eur Radiol*. 2005;15(6):1224–1233.
63. Londero V, Bazzocchi M, Del Frate C, et al. Locally advanced breast cancer: comparison of mammography, sonography and MR imaging in evaluation of residual disease in women receiving neoadjuvant chemotherapy. *Eur Radiol*. 2004;14(8):1371–1379.
64. Drew PJ, Kerin MJ, Mahapatra T, et al. Evaluation of response to neoadjuvant chemoradiotherapy for locally advanced breast cancer with dynamic contrast-enhanced MRI of the breast. *Eur J Surg Oncol* 2001;27(7):617–620.
65. Ogston KN, Miller ID, Payne S, et al. A new histological grading system to assess response of breast cancers to primary chemotherapy: prognostic significance and survival. *Breast*. 2003;12(5):320–327.
66. Yuan Y, Chen XS, Liu SY, Shen KW. Accuracy of MRI in prediction of pathologic complete remission in breast cancer after preoperative therapy: a meta-analysis. *Am J Roentgenol*. 2010;195(1):260–268.
67. Berg WA, Madsen KS, Schilling K, et al. Breast cancer: comparative effectiveness of positron emission mammography and MR imaging in presurgical planning for the ipsilateral breast. *Radiology*. 2010;258(1):59–72.

68. Turnbull LW. Dynamic contrast-enhanced MRI in the diagnosis and management of breast cancer. *NMR Biomed*. 2009;22(1):28–39.
69. Von Minckwitz G, Kaufmann M, Kummel S, et al. Definition and impact of pathological complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. Results from the German neoadjuvant trials meta-analysis. *J Clin Oncol*. 2012;30 (15):1796–1804.
70. Marinovich ML, Sardanelli F, Ciatto S, et al. Early prediction of pathologic response to neoadjuvant therapy in breast cancer: systematic review of the accuracy of MRI. *Breast*. 2012;21(5):669–677.
71. Houssami N, Ciatto S, Macaskill P, et al. Accuracy and surgical impact of magnetic resonance imaging in breast cancer staging: systematic review and meta-analysis in detection of multifocal and multicentric cancer. *J Clin Oncol*. 2008;26(19):3248–3258.

Funding

This work was partly funded by the National Health and Medical Research Council (NHMRC) (program grant 633003 to the Screening & Test Evaluation Program). MLM was supported by an NHMRC postgraduate scholarship.

Notes

M. L. Marinovich conceived and co-ordinated the study, conducted the literature searches and review of studies, performed the data extraction and statistical analysis, and drafted the manuscript. N. Houssami conceived the study, advised on literature searches and study eligibility, contributed to resolution of data extraction, advised on clinical aspects and data interpretation, and contributed to drafting the manuscript. P. Macaskill conceived the statistical methods used, advised on data analysis and interpretation, and contributed to drafting the manuscript. F. Sardanelli contributed to data extraction, advised on MRI technical issues and clinical aspects, and contributed to drafting the manuscript. L. Irwig advised on methodological aspects and data interpretation and contributed to drafting the manuscript. E. P. Mamounas advised on clinical aspects and contributed to drafting the manuscript. G. von Minckwitz advised on clinical aspects and contributed to drafting the manuscript. M. E. Brennan contributed to data extraction and to drafting the manuscript. S. Ciatto contributed to data extraction, advised on clinical aspects and contributed to drafting the manuscript. All authors read and approved the final manuscript.

The funder had no role in the design of the study; the collection, analysis and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication.

Affiliations of authors: Screening and Test Evaluation Program (STEP), Sydney School of Public Health, University of Sydney, Sydney, Australia (MLM, NH, PM, LI, MEB); Dipartimento di Scienze Biomediche per la Salute, Università degli Studi di Milano, Milan, Italy (FS); Unità di Radiologia, IRCCS Policlinico San Donato, Milan, Italy (FS); Aultman Health Foundation, Canton, OH (EPM); German Breast Group & Universitäts-Frauenklinik Frankfurt, Neu-Isenburg, Germany (GvM); U.O. Senologia Clinica e Screening Mammografico, Dipartimento di Radiodiagnostica, APSS, Trento, Italy (SC). †Died May 2012.