HAL
open science

Accuracy of unidimensional and volumetric ultrasound measurements in predicting good pathological response to neoadjuvant chemotherapy in breast cancer patients

I. Gounaris, E. Provenzano, A. L. Vallier, L. Hiller, M. Iddawela, S. Hilborne, K. Taylor, P. Britton, H. M. Earl, R. Sinnatamby

To cite this version:

I. Gounaris, E. Provenzano, A. L. Vallier, L. Hiller, M. Iddawela, et al.. Accuracy of unidimensional and volumetric ultrasound measurements in predicting good pathological response to neoadjuvant chemotherapy in breast cancer patients. Breast Cancer Research and Treatment, 2011, 127 (2), pp.459-469. 10.1007/s10549-011-1454-x . hal-00626285

HAL Id: hal-00626285

https://hal.science/hal-00626285v1

Submitted on 25 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ACCURACY OF UNIDIMENSIONAL AND VOLUMETRIC ULTRASOUND MEASUREMENTS IN PREDICTING GOOD PATHOLOGICAL RESPONSE TO NEOADJUVANT CHEMOTHERAPY IN BREAST CANCER PATIENTS

I Gounaris[1], E Provenzano[2,3], A L Vallier[1,2], L Hiller[4], M Iddawela[1,2], S Hilborne[2], K Taylor[2], P Britton[2,3], H M Earl[1-3,5] and R Sinnatamby[2,3]

1 Oncology Centre, Cambridge University Hospitals NHS Foundation Trust, Addenbrookes Hospital, Cambridge, CB2 0QQ, UK

2 Cambridge Breast Unit, Cambridge University Hospitals NHS Foundation Trust, Addenbrookes Hospital, Cambridge, CB2 2QQ, UK

3. National Institute for Health Research, Cambridge Biomedical Research Centre

4 Warwick Clinical Trials Unit, University of Warwick, Coventry, CV4 7AL, UK

5 Department of Oncology, University of Cambridge, Cambridge, CB2 0QQ, UK

Corresponding author: Dr Ioannis Gounaris, Oncology Centre, Cambridge University Hospitals NHS Foundation Trust, Box 193, Addenbrookes Hospital, Hills road, Cambridge, CB2 0QQ, UK

Email: ioannis.gounaris@cancer.org.uk

Tel: +441223404626

ABSTRACT

BACKGROUND

Pathological complete response (pCR) is an important predictor of long term survival in patients with breast cancer receiving neoadjuvant chemotherapy. At present, the accuracy of traditional radiological assessments during treatment in predicting pCR is poor.

METHODS

Unidimensional and 3D volumetric ultrasound measurements prior to, after 4 cycles (mid-treatment), and at the end of 8 cycles (end-treatment) of chemotherapy were available from a subset of 55 patients enrolled in Neo-tAnGo, a National Cancer Research Network (NCRN) UK neoadjuvant chemotherapy breast cancer trial. Proportional changes in longest diameter (LD) and volume as well as absolute residual size thresholds were examined for their ability to predict pCR or pCR plus minimal residual disease (pCR/MRD). Sensitivity, specificity, positive (PPV) and negative predictive values (NPV) and likelihood ratios (LRs) were calculated. Receiver-operator characteristic (ROC) curves and logistic regression models were also constructed.

RESULTS

At mid-treatment, neither complete radiological response, nor proportional LD or volume changes were found predictive of final pCR. A small residual tumour volume ($\leq 1cm^3$ vs. $>1cm^3$) at mid-treatment however was associated with pCR/MRD (p=0.014). Sensitivity, specificity, PPV, NPV, LR+ and LR- values were 61%, 77%, 61%, 77%, 2.62 and 0.51 respectively. The area under the ROC curve was 0.689 (p=0.03). Volume $\leq 1cm^3$ at mid-treatment was found significant in a logistic regression (OR: 0.194, p=0.011). At end-treatment, no ultrasound measurements were found predictive of pCR or pCR/MRD.

CONCLUSIONS

Proportional tumour size changes (the basis of the RECIST criteria) were not found predictive of good pathological response, although residual volume $\leq 1cm^3$ at mid-treatment was found to be predictive of pCR/MRD. However, multiple volume and LD thresholds were examined and uncorrected p-values presented, increasing the possibility of type I errors. Replication in an independent dataset is required.

KEYWORDS

Breast cancer, neoadjuvant chemotherapy, volumetric ultrasound, pathological response

INTRODUCTION

Despite recent decreases in incidence [1,2], breast cancer remains the commonest cancer in women and the second most common cause of death due to cancer after lung cancer. In the United States it was estimated that more than 40000 women died of breast cancer in 2009 [3].

Neoadjuvant chemotherapy (NAC) is an established treatment option for breast cancer patients presenting with large primary tumours or locally advanced disease. Numerous randomised clinical trials as well as a meta-analysis [4] have confirmed the equivalence of neoadjuvant and adjuvant chemotherapy in terms of disease-free and overall survival. Pathological complete response (pCR – the complete absence of invasive tumour in both the breast and the axillary lymph nodes) has emerged as a powerful prognostic marker for overall and disease-free survival following neoadjuvant chemotherapy [5,6]. However, tumour response to chemotherapy is a continuum from frank resistance to pCR. Ogston et al [7] showed that the Miller-Payne classification system that is based on residual tumour cellularity is an independent predictor of survival in multivariate analysis. Symmans et al [8] also reported that the Residual Cancer Burden (RCB) Index that takes into account residual tumour size and cellularity as well as the number and size of nodal metastases is also predictive of outcome, with minimal residual disease carrying the same excellent prognosis as pCR.

Patients electing to undergo NAC are closely monitored. In most centres in Europe and North America the monitoring schedule includes clinical breast examination at each chemotherapy visit as well as imaging of the breast with mammography, ultrasound (US) and, in selected patients, MRI at diagnosis, mid-way through chemotherapy, and after completion of chemotherapy, prior to surgery. This enables timely identification of patients whose tumours do not respond or frankly progress during NAC, although the latter is a rare occurrence [9]. It is also important for final surgical planning, identifying or confirming those patients suitable for breast conserving surgery.

Although routinely performed in clinical practice, the literature regarding the utility of mammography and US in assessing response to NAC and predicting pCR is surprisingly sparse and consists mostly of relatively small retrospective case series. Von Minckwitz et al [9] reported that the sensitivity of US after 2 cycles of chemotherapy for predicting eventual pCR was only 7.6% in the GeparTrio study. Chapgar et al [10]

noted that, in 189 patients treated with NAC at the MD Anderson Cancer Center (Houston, TX), the correlation of physical examination, mammography and US with pathological tumour size was weak to moderate with correlation coefficients of 0.42, 0.41 and 0.42 respectively. Small case series however, have reported much higher accuracy of US in predicting pCR [11,12] and there is increasing interest in employing advanced ultrasound-based technologies, for example assessing vascularity, to further improve performance [13,14].

The ability to accurately predict which patients are destined to achieve a good response to NAC early in the course of treatment would be a major advance in the field. It would greatly facilitate clinical trial designs of early switch to alternative cytotoxic or molecularly targeted agents (or even early recourse to surgery) in an attempt to improve pCR rates and ultimately survival from breast cancer.

In this study, utilising data collected prospectively in the context of a randomised, phase III trial of NAC (Neo-tAnGo), we attempted to correlate changes in either tumour longest diameter (LD) or volume as determined by conventional and 3D US with final pathology at surgery. We investigated the performance of proportional changes in tumour LD and volume in predicting good response to chemotherapy. We also examined the ability of absolute residual tumour size, in terms of either LD or volume, to predict the same.

METHODS

The Neo-tAnGo study

The Neo-tAnGo study is a randomized phase III study of NAC that enrolled 831 patients with high-risk early breast cancer from 57 UK centres between January 2005 and September 2007. In a 2-by-2 factorial design these patients were randomised to 4 cycles of epirubicin- cyclophosphamide (EC) followed by 4 cycles of paclitaxel with or without gemcitabine, or the reverse sequence (i.e. the taxane with or without gemcitabine first, followed by the anthracycline component). The primary endpoint was pCR and preliminary results have been reported [15].

Imaging Substudy

As per the main study protocol, all enrolled patients underwent tumour assessment, usually by mammography and/or US at baseline, after 4 cycles of chemotherapy (mid-

treatment) and after finishing chemotherapy, just prior to surgery (end-treatment). In addition, patients enrolled at Cambridge were offered 3D volumetric US and breast MRI at the same time-points (evaluation of MRI performance will be reported separately).

US data were acquired using a combination of conventional B-mode US (Aplio 80; Toshiba Medical Systems, Tochigiken, Japan) with a 12-MHz linear transducer, followed by a freehand 3D US data set, using a system developed in collaboration with our University's Department of Engineering. For the 3D acquisition, an optical position sensor (Polaris; Northern Digital, Waterloo, Ontario, Canada) was used, to continually track the orientation of 15 infrared-emitting diodes attached to the transducer. High spatial resolution US images (0.1 X 0.1mm pixel size) together with their spatial orientation data, were simultaneously recorded to an 800-MHz personal computer (World of Computers, Cambridge, England) equipped with research software (Stradwin; Department of Engineering, Cambridge University, Cambridge, England; freely available from: http://mi.eng.cam.ac.uk/~rwp/stradwin). This non-commercial research software was also used to analyze and display the 3D data and has been described in detail elsewhere [16,17]. All 3D US data acquisition was performed by 1 of 2 experienced Consultant radiologists. The 3D US dataset was reviewed at the personal computer, the tumour was manually segmented and a series of outlines representing the tumour was produced. Volume measurements were produced from the target outlines for each dataset using the Stradwin software [16,17].

Histopathology Review

Following surgery, all pathology reports were reviewed by an experienced breast histopathologist (EP) in conjunction with the study's chief investigator (HME). Each one separately assessed the response of the primary tumour using 5 categories: pCR, minimal residual disease (MRD, >90% response), chemotherapy response present but not gradable, no evidence of response or unable to comment. In case of disagreement, the case was reviewed jointly and final agreement reached by consensus. It should be noted that the categories of MRD and pCR that were used as outcome measures in this project correspond exactly to Grades 4 and 5 of the "Miller and Payne" system [7]. After the analysis of the imaging substudy data had been completed, an independent pathology review of all specimens was undertaken and the RCB Index scores were calculated for the patients enrolled at Cambridge. We elected not to repeat the analysis

using the RCB Index, as this is invalid in patients with a positive sentinel lymph node (SLN) biopsy [8], a fact that would further limit the sample size.

Substudy Endpoints

The primary objective of this substudy was to compare the performance of tumour LD and volume changes, as well as size of residual disease as measured by US in predicting good pathological response to chemotherapy. Two alternative definitions of good response were used: pCR only and pCR or MRD (pCR/MRD). pCR was defined as the absence of any invasive component at examination of both the primary tumour site and any axillary lymph nodes. Samples with *in situ* disease alone could still be classified as showing pCR if there was no associated invasive component. MRD was defined as evidence of >90% tumour necrosis at surgery, without any macroscopic residual axillary lymph node involvement.

In order to assess US performance, radiological response to chemotherapy was classified on the basis of either proportional change in tumour size or absolute residual tumour size. For LD proportional changes the following threshold values were used: 30% (corresponding to partial response by RECIST criteria), 50%, 75% and 100% (complete response [CR]) LD reduction. Proportional volume changes were examined at the following thresholds: 50%, 65% (corresponding to 30% LD change assuming a spherical volume), 75%, 90% and 100% (CR) reduction. The following absolute residual size thresholds were also used: residual LD less than 10, 20 and 30mm and residual volume less than 1, 2 and 3cm^3. For patients with multifocal tumours, US measurements were summed and therefore all the reported values are per patient and not per lesion.

Fifty-five patients had at least two available unidemensional and 3D US assessments. This study focused on the 48 patients that had available measurements at baseline and mid-treatment. We elected to do so as the utility of being able to predict good pathological response early is greater compared to just prior to surgery since, in the latter case, the tumour soon becomes available for detailed histopathological assessment. The outcomes for the 38 patients with available measurements and baseline and end-treatment are mentioned briefly.

Statistical Analysis

Correlation coefficients were calculated between unidimensional and volumetric measurements. Mean US LD and volume measurements at baseline and at the assessment time-point were compared with a paired t-test. The mean tumour diameters and volumes in patients with or without good response were compared using t-tests.

Radiological and pathological responses at the thresholds defined were compared with Fisher's exact test or χ^2 as appropriate. For thresholds significant at this first analysis, sensitivities, specificities, positive and negative predictive values, likelihood ratios as well as accuracy were calculated from the tables. To further examine the trade-offs between sensitivity and specificity, receiver-operating characteristic (ROC) curves were constructed and the corresponding areas under the curve (AUC) were calculated and compared using a modification of Wilcoxon's rank sum test.

Univariate logistic regression models were constructed including age, tumour grade, hormone and HER2 receptor status, clinical T stage, nodal status, chemotherapy sequence, proportional tumour size change thresholds and residual size thresholds at mid-treatment. Threshold values were entered in the model one at a time. The outcome variables were pCR and pCR/MRD and covariates with $P<0.10$ at univariate analysis were entered into multivariate models. In an unplanned *post hoc* analysis, the investigations were repeated for the HER2 positive and HER2 negative patients separately.

As this was an exploratory study using a convenience sample, formal sample power calculations were not done. All analyses were conducted using SPSS version 17.0 unless otherwise specified. Significance was determined by $p<0.05$ and no corrections for multiple testing were performed. The Toronto Centre for Evidence-Based Medicine online calculator was used for calculating confidence intervals for sensitivity, specificity, positive and negative predictive values and likelihood ratios (available from: http://ktclearinghouse.ca/cebm/practise/ca/calculators/statscalc). The freely available MedCalc program (www.medcalc.be) was used to compare different AUC curves.

RESULTS

Patients

Forty-eight women had both conventional and 3D US measurements of their primary tumours at baseline and mid-treatment (Table 1).

This was a cohort of young patients (median age 46.5 years; 30 [62.5%] being <50 years old) with high risk breast cancer. Twelve (25%) patients had multifocal disease at presentation and 21 (44%) presented with T3 or T4 primaries. SLN biopsy was not mandated by the trial protocol and was not in routine clinical use at the start of the trial; only 12 (25%) of the patients underwent an SLN biopsy procedure. There were relatively high rates of oestrogen receptor negativity (31%) and HER2 positivity (42%). Five (10%) patients were triple negative (i.e. negative for ER, PR and HER2).

Treatment received and final pathology

Nineteen (40%) women received 4 cycles of EC followed by paclitaxel with or without gemcitabine, whereas 29 (60%) received the reverse sequence. As paclitaxel was administered in a dose dense fashion (every two weeks), the women receiving taxane first had their mid-treatment US scan after 8 weeks of chemotherapy, while the women in the anthracycline-first arm underwent the same evaluation after 12 weeks. All 48 women successfully completed 8 cycles of NAC and underwent surgery (11 had breast conserving surgery and 37 a mastectomy). The 20 women with HER2 positive tumours also received adjuvant trastuzumab for 1 year.

Eleven (23%) of patients achieved pCR (Table 2) with 38% achieving pCR/MRD. The mean residual invasive tumour size was 18.83mm (SE 3.22mm), while the combined mean residual invasive and *in situ* component was 27.53mm (SE 3.54mm).

Baseline and mid-treatment ultrasound and clinical examination measurements

Tumour sizes by US and clinical examination, using callipers, at baseline and mid-treatment are presented in Table 3. US measurements of LD at baseline were on average 6.27mm smaller than the corresponding clinical measurements (95% CI 1.37-11.17mm, $p=0.013$). At mid-treatment, as assessed by US, the tumours had significantly decreased in size on both uni-dimensional and volumetric criteria (both $p<0.001$). As would be expected, the LD and volumetric measurements were significantly correlated both at baseline and mid-treatment ($r=0.74$ at baseline and $r=0.57$ at mid-treatment, both $p<0.001$). On a per-patient, proportional basis, tumour LD changes ranged from +19% to -100% (radiological CR) between baseline and mid-treatment. The corresponding volume changes were +118% to -100%. The mean LD and volume changes were 30.2% (S.E. 5.5%) and 51% (S.E. 7.5%) respectively.

Ultrasound measurements and pathological response

The 11 patients with pCR and the 37 patients without pCR had similar mean tumour LD and volume at mid-treatment (Table 4). Mean LD was 27.7mm and 23.92mm (p=0.572) and mean volumes were 2.95cm^3 and 2.49cm^3 respectively (p=0.691). Similarly, at mid-treatment, LD decreased by 33.4% in patients with eventual pCR and by 29.2% in patients without pCR (p=0.754). The corresponding volume changes were 62.3% and 47.5% (p=0.415). There is therefore, no evidence in this sample that tumours that have shrunk proportionately more at mid-treatment, as measured by US, are more likely to show pCR. The results were essentially unchanged when combined pCR/MRD (18 patients) was considered as the response variable (Table 4).

Using a threshold of 30% reduction in LD that would be consistent with the RECIST response criteria, 27 (56%) of the 48 women had a radiological response. Six (22%) of these women eventually showed a pCR, an outcome shared with 5 (24%) of the 21 women with less than 30% reduction in tumour LD at mid-treatment (p=1). Using thresholds of 50% and 75% LD reduction did not improve US performance (Table 5). Even among the 3 patients that had a radiological CR (rCR) at mid-treatment, only 1 (33%) showed eventual pCR, whereas 10 (22%) of the 45 patients with less than rCR did so (p=0.551).

As far as volume changes are concerned, 30 (62.5%) women had a >50% reduction in tumour volume, of which 7 (23%) showed a pCR, a proportion not statistically different from the 4 (22%) of 18 women with less than 50% volume reduction that eventually showed a pCR (p=1). Using a threshold of 65% volume reduction, which corresponds to a 30% LD reduction if a spherical volume is assumed, the corresponding proportions were 6/24 (25%) and 5/24 (21%) respectively (p=1). Thresholds of 75%, 90% and even 100% (rCR) did not improve ultrasound performance (Table 5).

The ability of absolute residual tumour size rather than proportional change at mid-treatment to predict response was also examined. None of the thresholds examined could separate patients with or without pCR in a statistically significant way (Table 5).

The same proportional change and absolute size thresholds were analysed for the outcome of pCR/MRD. No statistically significant findings emerged when examining the various proportional change thresholds' ability to separate the patients with good response to chemotherapy (pCR/MRD) from the rest (Table 6). However, examining residual tumour size thresholds (Table 6) revealed that patients with residual tumour volume ≤1cm^3 at mid-treatment were more likely to show pCR/MRD than those with

larger residual volumes. Eleven (61%) of the 18 patients with residual volume ≤1cm^3 had pCR/MRD compared with 7 (23%) of the 30 patients with larger residual tumour volumes (uncorrected p=0.014).

Binary classification tests, likelihood ratios and overall accuracy of the 1cm^3 residual volume threshold at mid-treatment

Residual tumour volume ≤1cm^3 had a sensitivity of 61% (95% C. 39-80%), specificity of 77% (95% CI 59-88%) and overall accuracy of 71% for the outcome of pCR/MRD. PPV and NPV were also good at 61% (95% CI 39-80%) and 77% (95% CI 59-88%) respectively. The positive likelihood ratio (LR+) was 2.62 (95% CI 1.24-5.52) and the negative likelihood ratio (LR-) was 0.51 (95% CI 0.28-0.94). A ROC curve was constructed, using residual tumour volume ≤1cm^3 as a binary cut-off point. The AUC value was 0.689 (95% CI 0.528-0.849, p=0.03) implying discrimination significantly different from chance (Fig. 1).

Logistic Regression modelling

Univariate binary logistic regression models, using pCR/MRD as the outcome variable, with age, tumour grade, hormone and HER2 receptor status, clinical T stage, nodal status, chemotherapy sequence, proportional tumour size change thresholds and residual size thresholds at mid-treatment as covariates were constructed. The following three variables were significant at p≤0.1: residual volume at mid-treatment ≤1cm^3, grade and age (Table 7). The odds ratio (OR) for pCR/MRD was 0.194 for patients with residual volume >1cm^3 compared to ≤1cm^3 (95% CI 0.054-0.69, p=0.011). When all three variables were forced into a multivariate model (Table 8), none retained statistical significance. Residual volume ≤1cm^3 however, still trended towards statistical significance (OR 0.279, 95% CI 0.072-1.082, p=0.065).

Subgroup analyses

The 20 (42%) patients with HER2-positive tumours had higher pCR rates than the rest (30% vs. 18%) but the difference was not statistically significant (p=0.488). Similarly, 9 (45%) of 20 HER2-positive patients achieved pCR/MRD compared to 9 (32%) of 28 HER2-negative patients (p=0.385). When proportional US threshold changes were tested for their ability to predict pCR and pCR/MRD, separately in the HER2-positive and HER2-negative patients, none reached statistical significance. However, in HER2-positive patients, tumour volume ≤1cm^3 was an even more powerful predictor of pCR.

Five (71%) of 7 HER2-positive patients with residual tumour volume ≤1cm^3 achieved pCR compared to 1 (8%) of 13 with larger tumours (p=0.007). When the outcome of pCR/MRD was examined, residual volume ≤1cm^3 for HER2-positive patients retained statistical significance, with 6 (86%) of 7 patients having pCR/MRD compared to 3 (23%) of 13 with larger tumours (p=0.017). Sensitivity, specificity, PPV, NPV, LR+, LR- and overall accuracy were calculated for the outcome of pCR in HER2-positive patients with residual tumour volume ≤1cm^3 at mid-treatment. The results were 83%, 86%, 71%, 92%, 5.83, 0.19 and 85% respectively. The AUC of the ROC curve was 0.845 (Fig. 2), with discrimination statistically significantly different from chance (p=0.017). The 1cm^3 threshold however, was not significant in HER2-negative patients.

Measurements at the end of chemotherapy

Thirty-eight women had available measurements at baseline and end-treatment (of which 31 [82%] had also assessments at mid-treatment and form part of the main analysis). The correlation between the US measurement of residual tumour LD at end-treatment and the size of residual invasive tumour at pathological assessment was weak to moderate, although statistically significant (r=0.359, p=0.027). The correlation was only marginally improved when the size of residual invasive plus *in situ* tumour was considered (r=0.381, p=0.018). None of the proportional or absolute size thresholds were statistically significant in this cohort (data not shown).

DISCUSSION

The main findings of our analysis can be summarised as follows: i) proportional LD and volume changes showed very poor ability in predicting pCR or pCR/MRD in the current series; ii) irrespective of their initial size, tumours with volumes ≤1cm^3 at mid-treatment, were more likely to show eventual pCR/MRD compared to tumours >1cm^3; iii) in a *post hoc* unplanned subgroup analysis, this observation seemed to be limited to HER2 positive tumours

The use of the traditional thresholds of 30% and 100% reduction in LD, equivalent to partial (PR) and complete (rCR) radiological response by RECIST criteria, could not separate patients into good pathological responders or not, in a statistically significant way. Similarly, proportional volume thresholds did not show an advantage. This is to an extent expected, as there was strong correlation between LD and volume measurements at each time point (r=0.57-0.848), implying that most tumours had a reasonably regular

shape, limiting the scope for dramatic differences between the unidimensional and volumetric approaches.

In our series, only 1 of 3 patients with rCR at mid-treatment showed eventual pCR and a further 1 showed MRD. Other authors have also reported poor concordance between US-assessed rCR and pCR [11]. As all sonograms were performed by experienced breast radiologists, one must consider the possibility that US is inherently limited in predicting pCR. Breast tumours in patients receiving NAC do not just melt away. They tend to fragment [18,19] and sometimes tumour necrosis is accompanied by a stromal reaction [20]. Furthermore, US cannot distinguish between *in situ* and invasive components. The consequence of these is that US will often show a residual mass in cases were pathological examination reveals pCR or MRD. The opposite can also be true as in cases of tumour necrosis without stromal reaction, ultrasound will show rCR whereas pathological examination may reveal residual viable tumour. The correlation coefficient r=0.359 between end-treatment US-assessed LD and pathology in our series similar to the value r=0.29 reported by Fiorentino et al [21] but lower than other reports in the literature (Table 9). Of note, the correlation only marginally improved (r=0.381) when residual *in situ* carcinoma was taken into account.

We tested various other proportional LD and volume thresholds for their ability to predict pCR or pCR/MRD. A similar approach in breast cancer patients receiving NAC has been reported by Akazawa et al [22]. These authors tested various 2D US as well as 2D and 3D MRI thresholds at end-treatment for their ability to predict recurrence in a series of 51 patients. They found that proportional size change thresholds of 50%, 60%, 70%, 75% and 80% as assessed by 2D US were not predictive of disease recurrence. However, >75% volume reduction as assessed by MRI was predictive with an uncorrected p=0.03. Similarly to the situation with our series, multiple performed comparisons increase the study's vulnerability to type I errors.

The unique approach of our study was to examine absolute residual size thresholds, hypothesising that tumours that had shrunk below a given size would be more likely to show pCR or pCR/MRD, irrespective of their initial size or the proportional size change. Adopting this approach, it was found that residual volume $\leq 1cm^3$ at mid-treatment was predictive of pCR/MRD. The AUC of the ROC curve was 0.689, indicating discriminatory ability significantly better than chance alone (p=0.03). Choi et al [23] reported that US-assessed rCR had an AUC of 0.68 in predicting pCR (p>0.05)

whereas Keune et al [24] found it to be 0.71 (p-value not given). The AUC for the 1cm^3 volume threshold compares favourably to those, with the added advantage that it is determined at the midway point rather than the end of chemotherapy. However, the multiple (26) hypothesis tests that were involved at arriving at this threshold raise the possibility of Type I statistical error. Therefore, independent validation is a larger dataset is needed. Subgroup analysis showed that the threshold of volume ≤1cm^3 at mid-treatment was particularly informative for HER2-positive tumours for both pCR and pCR/MRD. Again, this may reflect a chance finding, as calculations involved multiple hypothesis tests and the sample sizes were very small (20 HER2-positive and 28 HER2-negative patients). It should also be kept in mind that, although pathological responses such as MRD that fall short of pCR may confer a similarly good prognosis, further validation of this outcome measure is required.

US during NAC has a limited ability to predict pCR and, despite some positive findings, volumetric ultrasound in the current study did not result in marked improvements in performance. The limitations of clinical examination [25] and mammography [10] have also been extensively discussed in the literature. Breast MRI [11,22] appears to be more accurate than US, but not enough to enable it to be used to predict pCR with certainty. For reasons already discussed (tumour fragmentation, stromal reactions, presence of *in situ* residual carcinoma), it may be that no anatomy-based imaging technique will be able to result in highly accurate predictions.

In conclusion, this study adds to the evidence that proportional size changes, either unidimensional or volumetric, as assessed by US cannot discriminate between patients with and without good pathological responses. However, residual volume ≤1cm^3 at mid-treatment could potentially predict which patients are destined to exhibit good pathological response 8 to 12 weeks before the end of chemotherapy, particularly in the HER2-positive population. Validation in different datasets is now required and, if successful, would facilitate trial designs that assign patients to different treatment arms at mid-treatment based on predicted pathological responses.

ACKNOWLEDGEMENTS

We would like to thank Drs Graham Treece, Richard Prager and Andrew Gee at the Department of Engineering, University of Cambridge who set up the 3D US imaging system and the whole Medical Imaging Group at the department for loaning of the

equipment and providing technical support. We would also like to thank Miss Natalie Hicks for assistance with data input.

EP, PB, HME and RS are supported with research funding from the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre (BMRC).

CONFLICTS OF INTEREST

The authors declare no relevant conflicts of interest

REFERENCES

1. Li SI, Daling JR (2007) Changes in breast cancer incidence rates in the United States by histologic subtype and race/ethnicity, 1995 to 2004. Cancer Epidemiol Biomarkers Prev 16:2773-80.
2. Ravdin P, Cronin K, Howlander N, Berg C, Chlebowski R, Feuer E et al (2007) The decrease in breast-cancer incidence in 2003 in the United States. New Engl J Med 356:1670-4.
3. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ (2009) Cancer Statistics 2009. CA Cancer J Clin 59:225-49.
4. Mauri D, Pavlidis N, Ioannidis J (2005) Neoadjuvant versus adjuvant systemic treatment in breast cancer: a meta-analysis. J Natl Cancer Inst 97:188-94.
5. Sataloff DM, Mason BA, Prestipino AJ, Seinige UL, Lieber CP, Baloch Z (1995) Pathologic response to induction chemotherapy in locally advanced carcinoma of the breast: a determinant of outcome. J Am Coll Surg 180:297-306.
6. Kuerer HM, Newman LA, Smith TL, Ames FC, Hunt KK, Dhingra K et al (1999) Clinical course of breast cancer patients with complete pathologic primary tumour and axillary lymph node response to doxorubicin-based neoadjuvant chemotherapy. J Clin Oncol 17:460-9.
7. Ogston KN, Miller ID, Payne S, Hutcheon AW, Sarkar TK, Smith I et al (2003) A new histological grading system to assess response of breast cancers to primary chemotherapy: prognostic significance and survival. Breast 12:320-7.

8. Symmans WF, Peintinger F, Hatzis C, Rajan R, Kuerer H, Valero V et al (2007) Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. J Clin Oncol 25:4414-22.
9. von Minckwitz G, Kummel S, Vogel P, Hanusch C, Eidtmann H, Hilfrich J et al (2008) Intensified neoadjuvant chemotherapy in early-responding breast cancer: Phase III randomised GeparTrio study. J Natl Cancer Inst 100:552-62.
10. Chapgar AB, Middleton LP, Sahin AA, Dempsey P, Buzdar AU, Mirza AN et al (2006) Accuracy of physical examination, mammography and ultrasonography in predicting residual pathologic tumour size in patients treated with neoadjuvant chemotherapy. Ann Surg 243:257-64.
11. Kanazawa T, Akashi-Tanaka S, Iwamoto E, Takasugi M, Shien T, Kinoshita T et al (2005) Diagnosis of complete response to neoadjuvant chemotherapy using diagnostic imaging in primary breast cancer patients. Breast J 11:311-6.
12. Schott AF, Roubidoux MA, Helvie MA, Hayes DF, Kleer CG, Newman LA et al (2005) Clinical and radiologic assessments to predict breast cancer pathologic complete response to neoadjuvant chemotherapy. Breast Cancer Res Treat 92:231-8.
13. Zhu Q, Tannenbaum S, Hedge P, Kane M, Xu C, Kurtzman SH (2008) Noninvasive monitoring of breast cancer during neoadjuvant chemotherapy using optical tomography with ultrasound localization. Neoplasia 10:1028-40.
14. Singh G, Kumar P, Parshad R, Seith A, Thulkar S, Hosten N (2010) Role of color doppler indices in predicting disease-free survival of breast cancer patients during neoadjuvant chemotherapy. Eur J Radiol 75:e158-62.
15. Earl HM, Vallier A, Hiller L, Fenwick N, Iddawela M, Hughes-Davies L et al (2009) Neo-tAnGo: a neoadjuvant randomized phase III trial of epirubicin/cyclophosphamide and paclitaxel+/-gemcitabine in the treatment of women with high-risk early breast cancer (EBC): first report of the primary endpoint, pathological complete response (pCR). J Clin Oncol 27:s522.
16. Prager R, Gee A, Treece G, Berman L (2002) Freehand 3D ultrasound without voxels: volume measurement and visualisation using the Stradx system. Ultrasonics 40:109-15.
17. Treece G, Gee A, Prager R, Cash CJ, Berman L (2003) High definition freehand 3D ultrasound. Ultrasound in Medicine and Biology 29:529-46.

18. El-Didi MH, Moneer MM, Khaled HM, Makarem S (2000) Pathological assessment of the response of locally advanced breast cancer to neoadjuvant chemotherapy and its implications for surgical management. Surg Today 30:249-54.
19. Huber S, Medl M, Vesely M, Czembirek H, Zuna I, Delorme S (2000) Ultrasonographic tissue characterization in monitoring tumour response to neoadjuvant chemotherapy in locally advanced breast cancer (work in progress). J Ultrasound Med 19:677-86.
20. Pinder SE, Provenzano E, Earl E, Ellis IO (2007) Laboratory handling and histology reporting of breast specimens from patients who have received neoadjuvant chemotherapy. Histopathol 50:409-17.
21. Fiorentino C, Berruti A, Bottini A, Bodini M, Brizzi MP, Brunelli A et al (2001) Accuracy of mammography and echography versus clinical palpation in the assessment of response to primary chemotherapy in breast cancer patients with operable disease. Breast Cancer Res Treat 69:143-51.
22. Akazawa K, Tamaki Y, Taguchi T, Tanji Y, Miyoshi Y, Kim SJ et al (2008) Potential of reduction in total tumour volume measured with 3-D MRI as a prognostic factor for locally–advanced breast cancer patients treated with primary chemotherapy. Breast J 14:523-31.
23. Choi JH, Lim HI, Lee SK, Kim WW, Kim SM, Cho E et al (2010) 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 102:392-7.
24. Keune J, Jeffe D, Schootman M, Hoffman A, Gillanders W, Aft R (2010) Accuracy of ultrasonography and mammography in predicting pathologic response after neoadjuvant chemotherapy for breast cancer. Am J Surg 199:477-84.
25. Bear HD, Anderson S, Smith R, Geyer C, Mamounas EP, Fisher B et al (2003) Sequential preoperative or postoperative docetaxel added to preoperative doxorubicin plus cyclophosphamide for operable breast cancer: National Surgical Adjuvant Breast and Bowel Project Protocol B-27. J Clin Oncol 21:4165-74.
26. Akazawa K, Tamaki Y, Taguchi T, Tanji Y, Miyoshi Y, Kim SJ et al (2006) Preoperative evaluation of residual tumour extent by three-dimensional

magnetic resonance imaging in breast cancer patients treated with neoadjuvant chemotherapy. Breast J 12:130-7.

27. Balu-Maestro C, Chapellier C, Bleuse A, Chanalet I, Chauvel C, Largillier R (2002) Imaging in evaluation of response to neoadjuvant breast cancer treatment - benefits of MRI Breast Cancer Res Treat 72:145-52.
28. Herrada J, Iyer R, Atkinson EN, Sneige N, Buzdar A, Hortobagyi G (1997) Relative value of physical examination, mammography, and breast sonography in evaluating the size of the primary tumour and regional lymph node metastases in women receiving neoadjuvant chemotherapy for locally advanced breast carcinoma. Clin Cancer Res 3:1565-9.
29. Londero V, Bazzocchi M, Del Frate C, Puglisi F, Di Loreto C, Francescutti G, Zuiani C (2004) Locally advanced breast cancer: comparison of mammography, sonography and MR imaging in evaluation of residual disease in women receiving neoadjuvant chemotherapy. Eur Radiol 14:1371-9.
30. Peintinger F, Kuerer H, Anderson K, Boughey J, Meric-Bernstam F, Singletary SE et al (2006) Accuracy of the combination of mammography and sonography in predicting tumor response in breast cancer patients after neoadjuvant chemotherapy. Ann Surg Oncol 13:1443-9.
31. Segara D, Krop IE, Garber JE, Winer E, Harris L, Bellon JR et al (2007) Does MRI predict pathologic tumor response in women with breast cancer undergoing preoperative chemotherapy? J Surg Oncol 96:474-80.
32. Singh S, Pradhan S, Shukla RC, Ansari MA, Kumar A (2005) Color Doppler ultrasound as an objective assessment tool for chemotherapeutic response in advanced breast cancer. Breast Cancer 12:45-51.
33. Forouhi P, Walsh JS, Anderson TJ, Chetty U (1994) Ultrasonography as a method of measuring breast tumour size and monitoring response to primary systemic treatment. Br J Surg 81:223-5.
34. Akashi-Tanaka S, Fukutomi T, Watanabe T, Katsumata N, Nanasawa T, Matsuo K et al (2001) Accuracy of contrast-enhanced computed tomography in the prediction of residual breast cancer after neo-adjuvant chemotherapy. Int J Cancer 96:66-73.

FIGURE CAPTIONS:

Figure 1. AUC under the ROC curve for residual volume ≤1cm^3 and outcome of pCR/MRD

Figure 2. AUC under the ROC curve for residual volume ≤1cm^3 in HER2 positive patients and outcome of pCR

Table 1. Characteristics of patients with baseline and mid-treatment ultrasound measurements available

Characteristic	Patients n=48 (%)
Age	
Median	46.5
Range	31-64
<50	30 (62.5%)
≥50	18 (37.5%)
Side	
Left	22 (46%)
Right	26 (54%)
Histology	
Ductal	46 (96%)
Lobular	2 (4%)
Number of lesions	
1	36 (75%)
2	9 (19%)
3	3 (6%)
Clinical T stage	
T1	2 (4%)
T2	25 (52%)
T3	17 (35%)
T4 (inflammatory)	4 (9%)
Nodal involvement	
Clinically involved – CB positive	13 (27%)
Clinically negative – No CB	23 (48%)
SLNB positive	4 (9%)
SLNB negative	8 (17%)
Grade	
1	0
2	19 (40%)
3	29 (60%)
ER status	
Positive	33 (69%)
Negative	15 (31%)
HER2 status	
Positive	20 (42%)

Negative	28 (58%)
Triple Negative	5 (10%)

CB: Core biopsy SLNB: Sentinel Lymph Node Biopsy; ER: oestrogen receptor

Table 2. Histopathological findings

Characteristic	Patients (%)
Pathological assessment	
pCR	11(23%)
MRD	7(15%)
Evidence of response	21(44%)
Response cannot be classified	9(19%)
Residual tumour size	
Invasive only	18.83mm (SE 3.22)
Invasive + in situ	27.53mm (SE 3.54)

pCR: pathological complete response; MRD: Minimal Residual Disease

Table 3. Mean baseline and mid-treatment measurements by ultrasound and clinical examination

	Baseline (SE)	**Mid-Treatment (SE)**
Clinical examination (LD)	42.83mm (3.19)	Not available[a]
Ultrasound LD	36.76mm (SE 2.34)	24.79mm (SE 2.11)
Ultrasound volume	6.43cm^3 (SE 1.08)	2.6cm^3 (SE 0.49)

a Clinical examination results not consistently documented

LD: Longest diameter

Table 4. Ultrasound measurements at mid-treatment between responders and non-responders

	Responders Mean	**Non-Responders Mean**	**Difference**	**95% CI of the difference**	**P value**
	Response = pCR				
	n=11	n=37			
LD (mm)	27.7	23.92	3.78	-10.34 -17.89	0.572
Volume (cm^3)	2.95	2.49	0.47	-1.88 - 2.81	0.691
% LD change	-33.4%	-29.2%	-4.2%	-30.8 – 22.5	0.754
% volume change	-62.3%	-47.5%	-14.8%	-50.9 – 21.4	0.415
	Response = pCR plus MRD				
	n=18	n=30			
LD (mm)	24.02	25.25	-1.23	-10.07 – 7.62	0.781
Volume (cm^3)	2.09	2.9	-0.82	-2.84 – 1.21	0.422
% LD change	-31.8%	-29.3%	-2.5%	-25.7 – 20.7	0.829
% volume change	-59.7%	-45.7%	-14.1%	-45.4 – 17.3	0.371

pCR: pathological complete response; LD: Longest Diameter; MRD: Minimal Residual Disease

Table 5. Thresholds of ultrasound-assessed proportional changes and residual tumour sizes at mid-treatment, split by eventual pCR

	pCR	No pCR	%pCR	%no pCR	p value
LD reduction					
<30%	5	16	24%	76%	>0.99
≥30%	6	21	22%	78%	
<50%	8	26	24%	76%	>0.99
≥50%	3	11	21%	79%	
<75%	10	35	22%	78%	0.551
≥75%	1	2	33%	67%	
<100%	10	35	22%	78%	0.551
100%	1	2	33%	67%	
Volume reduction					
<50%	4	14	22%	78%	>0.99
≥50%	7	23	23%	77%	
<65%	5	19	21%	79%	>0.99
≥65%	6	18	25%	75%	
<75%	6	24	20%	80%	0.724
≥75%	5	13	28%	72%	
<90%	9	33	21%	79%	0.609
≥90%	2	4	33%	67%	
<100%	10	35	22%	78%	0.551
100%	1	2	33%	67%	
Residual LD					
≤10mm	2	4	33%	67%	0.609
>10mm	9	33	21%	79%	

≤20mm	5	16	24%	76%	>0.99
>20mm	6	21	22%	78%	
≤30mm	7	28	20%	80%	0.458
>30mm	4	9	31%	69%	
Residual volume					
≤1cm^3	6	12	33%	67%	0.288
>1cm^3	5	25	17%	83%	
≤2cm^3	6	23	21%	79%	0.732
>2cm^3	5	14	26%	74%	
≤3cm^3	6	27	18%	82%	0.283
>3cm^3	5	10	33%	67%	

pCR: pathological complete response; LD: longest diameter

Table 6. Thresholds of ultrasound-assessed proportional changes and residual tumour sizes at mid-treatment, split by eventual pCR or MRD

	pCR/ MRD	No pCR /MRD	%pCR /MRD	%no pCR /MRD	p value
LD reduction					
<30%	7	14	33%	67%	0.765
≥30%	11	16	41%	59%	
<50%	13	21	38%	62%	>0.99
≥50%	5	9	36%	64%	
<75%	16	29	36%	64%	0.547
≥75%	2	1	67%	33%	
<100%	16	29	36%	64%	0.547
100%	2	1	67%	33%	
Volume reduction					
<50%	7	11	39%	61%	>0.99
≥50%	11	19	37%	63%	
<65%	8	16	33%	67%	0.766
≥65%	10	14	42%	58%	
<75%	11	19	37%	63%	>0.99
≥75%	7	11	39%	61%	
<90%	15	27	36%	64%	0.658
≥90%	3	3	50%	50%	
<100%	16	29	36%	64%	0.547
100%	2	1	67%	33%	
Residual LD					
≤10mm	3	3	50%	50%	0.658
>10mm	15	27	36%	64%	

≤20mm	9	12	43%	57%	0.558
>20mm	9	18	33%	67%	
≤30mm	13	22	37%	63%	>0.99
>30mm	5	8	38%	62%	
Residual volume					
≤1cm^3	11	7	61%	39%	**0.014**
>1cm^3	7	23	23%	77%	
≤2cm^3	13	16	45%	55%	0.235
>2cm^3	5	14	26%	74%	
≤3cm^3	13	20	39%	61%	0.757
>3cm^3	5	10	33%	67%	

pCR: pathological complete response; MRD: minimal residual disease; LD: longest diameter; vol: volume

Table 7. Univariate logistic regression for prediction of pCR or MRD at mid-treatment

	Odds Ratio	**95% C.I.**	**P-value**
Age (<50 vs. ≥50)	**0.291**	**0.085-0.998**	**0.05**
Grade (2 vs. 3)	**0.286**	**0.076-1.071**	**0.063**
ER (negative vs. positive)	0.769	0.214-2.768	0.688
HER2 (negative vs. positive)	0.579	0.177-1.894	0.366
Sequence (anthracycline vs. paclitaxel first)	0.654	0.194-2.209	0.494
Clinical T-stage (T1/2 vs. T3/4)	0.667	0.205-2.165	0.5
Node (negative vs. positive)	0.654	0.194-2.209	0.494
≤1cm^3 residual vol (no vs. yes)	**0.194**	**0.054-0.69**	**0.011**

pCR: pathological complete response; MRD: minimal residual disease; LD: longest diameter; vol: volume; ER: oestrogen receptor

Table 8. Multivariate logistic regression for prediction of pCR or MRD at mid-treatment

	Odds Ratio	**95% C.I.**	**P-value**
Age (<50 vs. ≥50)	0.431	0.111-1.663	0.222
Grade (2 vs. 3)	0.427	0.103-1.77	0.241
≤1cm^3 residual vol (no vs. yes)	0.279	0.072-1.082	0.065
Constant	2.852		0.095

pCR: pathological complete response; MRD: minimal residual disease; vol: volume

Table 9. Correlation between ultrasound and pathological longest tumour diameter

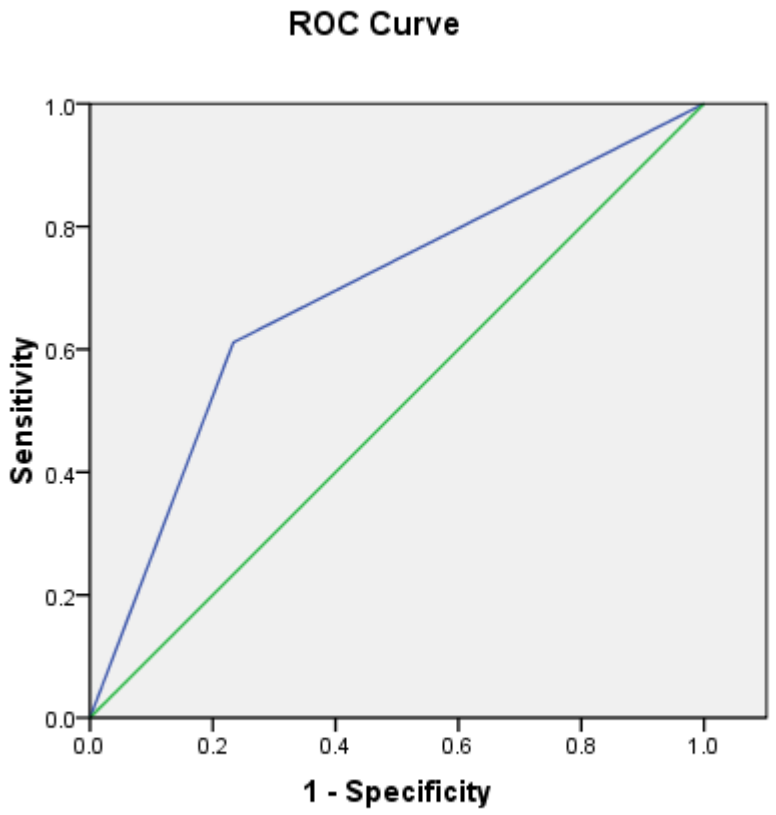
Study	Number of patients	Year of Publication	Correlation US-pathology[c]	Ultrasound Accuracy (limit)[c]
Akazawa et al [26]	38	2006	0.48	64.9% (2cm)
Balu-Maestro et al [27]	51	2002		43%
Chapgar et al [10]	189	2006	0.42	75% (1cm)
Fiorentino et al [21]	141	2001	0.29	
Herrada et al [28]	100	1997	0.60	
Keune et al [24]	104	2010		59.6% (1cm)
Londero et al [29]	16	2004	0.57	
Peintinger et al [30][a]	162	2006	0.66	82.7% (1cm)
Segara et al [31]	51	2007	0.61	66% (1cm)
Singh et al [32]	25	2005	0.69-0.81[b]	
Forouhi et al [33]	35	1994	0.94	
Akashi-Tanaka et al [34]	57	2001	0.56	

a Combined assessment of mammography and ultrasound

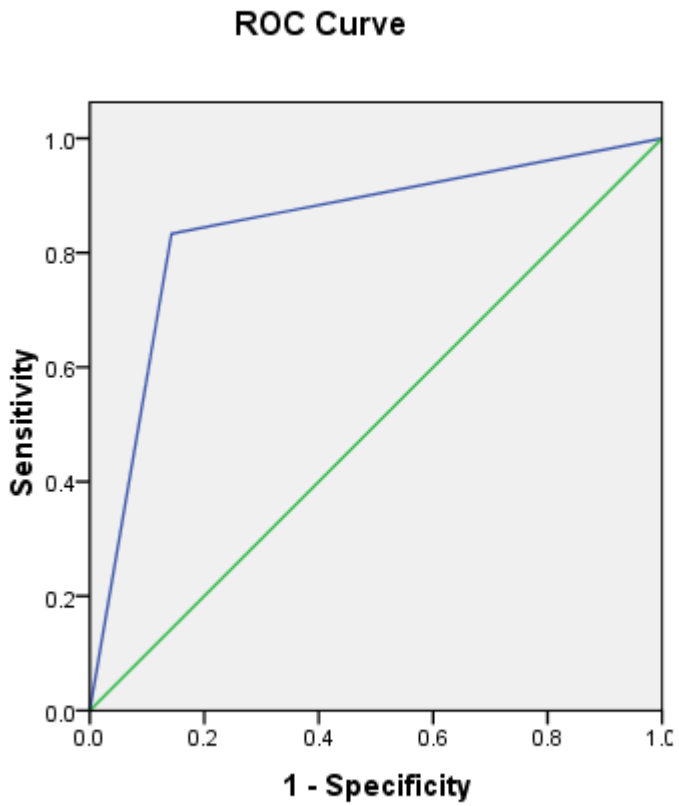
b Colour Doppler used – various indexes with different correlation coefficients reported

c The correlation column refers to correlation coefficient r as reported in the original publication. The ultrasound accuracy column refers to the proportion of tumours in which the difference between ultrasound-measured and pathological longest diameter was within the stated limit.

US: ultrasound



Diagonal segments are produced by ties.



Diagonal segments are produced by ties.