Central pathology review with two-stage quality assurance for pathological response after neoadjuvant chemotherapy in the ARTemis Trial

Authors: Jeremy Thomas, Elena Provenzano, Louise Hiller, Janet Dunn, Clare Blenkinsop, Louise Grybowicz, Anne-Laure Vallier, Ioannis Gounaris, Jean Abraham, Luke Hughes-Davies, Karen McAdam, Stephen Chan, Rizvana Ahmad, Tamas Hickish, Stephen Houston, Daniel Rea, Carlos Caldas, John M. S. Bartlett, David A. Cameron, Larry Hayward, Helena Margaret Earl for the ARTemis Investigators Group

ABSTRACT

The ARTemis Trial tested standard neoadjuvant chemotherapy (NAC) ± Bevacizumab (Bev) in the treatment of HER2 negative early breast cancer. We compare data from central pathology review with report-review and also the reporting behaviour of the two central pathologists.

800 women with HER2-negative early invasive breast cancer were recruited. Response to NAC± Bev was assessed from local pathology reports for pathological complete response (pCR) in breast and axillary lymph nodes. Tissue sections from the original core biopsy and surgical excision were centrally reviewed by one of two trial pathologists blinded to the local pathology reports. Pathologists recorded the response to chemotherapy descriptively and also calculated residual cancer burden (RCB). 10% of cases were double-reported to compare the central pathologists' reporting behaviour.

Full sample retrieval was obtained for 681 of the 781 patients (87%) who underwent surgery within the trial and were evaluable for pCR. 483 (71%) were assessed by JT, and 198 (29%) were assessed by EP. RCB calculations were possible in 587/681 (86%) of the centrally reviewed patients, since 94/681 (14%) had positive sentinel nodes removed before NAC thus invalidating RCB scoring. Good concordance was found between the two pathologists for RCB classes within the 65-patient quality assurance exercise (kappa 0.63 (95%CI 0.57-0.69)). Similar results were obtained for the between-treatment arm comparison both from the report-review and the central pathology-review. For pCR, report-review was as good as central pathology review but for minimal residual disease, report-review overestimated the extent of residual disease.

In the ARTemis Trial central pathology review added little in the determination of pCR but had a role in evaluating low levels of residual cancer burden. Calculation of RCB proved to be a simple and reproducible method of quantifying response to NAC as demonstrated by performance comparison of the two pathologists.

Funding: Cancer Research UK (CRUK/08/037), Roche, Sanofi -Aventis.

Manuscript metrics

Abstract: 294 words

Paper: 3162 words excluding abstract, acknowledgements and references

Figures: 4

Tables: 4

The ARTemis trial is an open-label, randomised, phase 3 trial assessing the efficacy of neoadjuvant bevacizumab added to docetaxel followed by fluorouracil, epirubicin, and cyclophosphamide, for women with HER2-negative early breast cancer. Its primary endpoint was pathological complete response (pCR), defined as the absence of invasive disease in the breast and axillary lymph nodes. Initially, the two randomised arms of the trial were compared in terms of rates of pCR as determined by a two-reader blinded review of local pathology reports.[1] In addition a central pathology review and a large-scale two-stage pathology quality assurance exercise was undertaken. Thereby the accuracy of this commonly used primary endpoint in neoadjuvant chemotherapy breast cancer trials was assessed and compared with the two-reader report-review which until now has been the standard used by this group. [2] In addition the reliability of central specimen review has been investigated by independent double-reading of residual cancer burden (RCB) categories carried out by the two central pathologists in a subset of cases. This allows us to report on the comparison between assessment of local pathology report and central pathology review of original diagnostic material and also the reporting behaviour of the two reviewing pathologists. Although central pathological review has been carried out in studies reporting major centre results [3] as far as the authors are aware this is the first report of central pathology review of pCR with RCB scoring and category definition, carried out as part of a multi-centre large randomised phase 3 trial.

PATIENTS AND TREATMENTS

Between May 2009 and January 2013, the ARTemis trial recruited 800 women ≥18 years old with newly diagnosed HER2-negative early invasive breast cancer (radiological tumour size >20 mm, with or without axillary involvement). Patients with

inflammatory cancer, T4 tumours with direct extension to the chest wall or skin, and ipsilateral supraclavicular lymph node involvement were also eligible with any size of primary tumour. Full eligibility criteria details have been described in detail elsewhere.[1] Patients were randomised from 66 UK sites and assigned, via a central computerised minimisation procedure, to three cycles of docetaxel (100 mg/m² once every 21 days) followed by three cycles of fluorouracil (500 mg/m²), epirubicin (100 mg/m²), and cyclophosphamide (500 mg/m²) once every 21 days (D-FEC), without or with four cycles of bevacizumab (15 mg/kg) (Bev+D-FEC). 781 patients (98% of the randomised 800) underwent surgery following their neo-adjuvant treatment and could be assessed, via local pathology reports, for the primary endpoint of absence of invasive breast cancer in the breast and axillary lymph nodes.

METHODS

Diagnostic and surgical excision histopathology slides were requested from the relevant participating sites for all of the 781 evaluable patients. All retrieved cases underwent central independent review, blinded to the local histopathology report, by an experienced breast histopathologist with a special interest in neoadjuvant clinical trials (JT & EP) between June 2011 and March 2016. The reviewing pathologist was not the same as the pathologist who had previously assessed the slides locally, and/or would have access to the histopathology results at their hospital. Any missing slides or additional relevant operations (e.g. sentinel lymph node biopsy) were re-requested as necessary. The variables recorded were maximum invasive tumour size in two dimensions, whole tumour size (including DCIS) in two dimensions, post treatment tumour grade, presence of lymphovascular invasion, presence and nature of in situ disease, percentage tumour cellularity, percentage cellularity that is in situ

disease, total number of lymph nodes, number of positive lymph nodes and size of largest nodal metastasis.

As well as assessing the validity of the findings from the blinded review of local pathology reports, an inter-pathologist reproducibility exercise was also undertaken. For this, a randomly chosen 10% of patients had samples reviewed by both pathologists for determination of levels of agreement between central review findings. To simplify this exercise variables were restricted to those required to calculate the RCB score: invasive size (length and width), percent tumour cellularity, percent of tumour that is DCIS, size of largest nodal metastasis and number of positive nodes.[4] The 10% sample was randomly chosen, whilst ensuring a representative RCB class split, as recorded by the first pathologist's reviews.

The results of the central pathology review were also compared with the outcome results as determined by the central review of the original histopathology reports from the source laboratory. In particular, we compared rates of pCR and minimal residual disease (MRD) as determined by formal assessment of the RCB and by interpretation of the histopathology. The local pathologists were not given any formal reporting guidelines specifically for this trial.

STATISTICAL METHODS

Agreement between the two pathologists' RCB classes, and also between central review and local reports in determination of pCR, was undertaken using the kappa statistic. Agreement between the two pathologists in terms of RCB scores and its six components were scrutinised using Bland-Altman plots and assessed using overall

concordance correlation coefficient (CCC) [5] Comparison of patient characteristics between groups was undertaken using chi-squared tests with continuity correction where appropriate. Logistic regression was used to assess the effect of randomised treatment arm on pCR rates, after adjustment for stratification factors.

RESULTS

A total of 22,916 slides from 727 patients were reviewed. Full sample retrieval was obtained for 681 (87%) of the 781 ARTemis patients who underwent surgery within the trial and were evaluable for the primary endpoint of pCR. 483/681 patients (71%) were assessed by JT, and 198/681 patients (29%) by EP. The maximum number of slides per patient was 164; median 29 slides. 94/681 patients (14%) had a positive pre-chemotherapy sentinel lymph node biopsy (SLNB) thus invalidating the calculation of an RCB score at surgery. RCB scores and classes were thus calculated on the remaining 587 patients (75% of the 781). Patient characteristics of the 587 patients with assessable RCB appeared representative of the trial sample as a whole (Table 1).

INTER-RATER REPRODUCIBILITY OF PATHOLOGISTS

65 patients were double reviewed by JT and EP. The 65 patients were representative of the 587 sample as a whole in terms of patient characteristics (Table 1) and the random sampling technique determined that they were also representative in terms of RCB class as recorded by the first pathologist's central review.

RCB class

The two pathologists showed very similar reporting profiles for RCB class (observed frequencies of RCB 0:1:2:3 being 14:9:32:10 for pathologist 1 and 13:9:34:9 for pathologist 2) (Table 2). In 52/65 (80%) of patients there was agreement on RCB class, and in 13/65 (20%) where there was disagreement none were more than one RCB class different. A good level of agreement was observed over all RCB classes (Kappa 0.70 (95%CI 0.55-0.84)) (Figure 1). No differences were found between patient groups where JT and EP agreed on RCB class (n=52) or disagreed (n=13), in terms of randomised treatment arm or stratification variables (age, ER status, tumour size, clinical involvement of axillary nodes, locally advanced/inflammatory disease data not shown).

RCB score

For the 13 patients where there was disagreement in RCB class, the majority of disagreements were due to the two pathologists' RCB scores falling just either side of the published RCB score cut-points of 1.36 and 3.28 (Figure 2). There was good overall concordance in RCB score (CCC 0.75 (95%CI 0.40-0.91)), with the average discrepancy in RCB score being of the magnitude 0.245 (IQR 0.135-0.501, range 0.085-1.840).

Components of the RCB score

Focusing on the 13 patients where the two pathologists differed in RCB class assignment, the greatest inter-rater variability was in the assessment of percentage of DCIS within the tumour (CCC -0.04 (95%CI -0.30-0.21)) and, to a lesser extent, in the assessment of invasive size [CCC 0.20 (95%CI -0.12–0.47) for width and CCC 0.35 (95%CI -0.11-0.68) for length] and percent of tumour cellularity (CCC 0.30

(95%CI -0.05-0.59)). The strongest agreement was observed in identification of number of positive nodes (CCC 0.95 (95%CI 0.85-0.98)) followed by size of the largest nodal metastasis (CCC 0.74 (95%CI 0.37-0.91)).

Sources of discrepancy

Seven cases where there was a disagreement in RCB class due to substantial differences in size measurement, cellularity or nodal status were reviewed again with joint discussion by the two pathologists. Sources of discrepancy included interpretation of multiple tumour foci as one lesion or multiple lesions, measurement of lesion size from single slides or estimating total number of slides, inclusion of pre-treatment SLN metastases in the RCB calculation, errors in measurement, and interpretation of degenerate cells in post treatment lymph nodes as metastasis or not.

CENTRAL REVIEW OF PATHOLOGY SPECIMENS VS REVIEW OF LOCAL PATHOLOGY REPORTS: INTER-METHOD RELIABILITY

Both methods determined similar levels of pCR in the 587 patients where both assessment results were available; 121 (21%) with RCB class 0 from central pathology review and 119 (20%) reported as pCR from local pathology report (Table 3). A good level of agreement was observed between the two methods' findings when grouped as the 3 levels of RCB 0 (pCR) vs RCB 1 (MRD) vs RCB 2/3 (Moderate/extensive disease) (kappa 0.63 (95%CI 0.57-0.69) (Figure 3). However, for 6 patients, the level of disagreement was by more than one category (1 patient with pCR from the report-review but RCB Category 2 from specimen review, and 5

patients with Moderate/extensive disease from the report-review but with RCB category 0 from the specimen review.)

Slides for 5 of the 6 cases were available for second review by one of the pathologists (EP). Sources of discrepancy included not receiving all the tumour slides for review (2 cases) and interpretation of residual tumour as DCIS or invasive disease (1 case). In one case, the second review agreed with the histopathology report (residual tumour) rather than the central review (pCR). In another case called pCR on central review, the discrepancy appears to be due to inconsistency in the original report in calling tumour cells in the node viable and non-viable; both central reviewers thought this represented an area of necrosis.

ARTemis PRIMARY ENDPOINT RESULTS

The ARTemis trial's primary endpoint was previously reported using the local pathology report-reviews on 781 patients and showed significantly more Bev+D-FEC patients achieving a pCR compared with D-FEC patients: 22% (95%CI 18–27) of 388 Bev+D-FEC patients compared with 17% (95%CI 13–21) of 393 D-FEC patients (adjusted p=0.03) (Table 4A).[1] Using the RCB categories from the central pathology specimen review the results remained the same: 25% (95%CI 20–30) of 290 Bev+D-FEC patients achieved an RCB 0 compared with 16% (95%CI 12–21) of 297 D-FEC patients (adjusted p=0.02) (Table 4B).

Likewise previously, using local pathology report-reviews, pCR rates had been found to differ significantly across both ER status (ER negative 38% [95% CI 32–45],

weakly positive 41% [29–53], strongly positive 7% [5–9]; p<0.0001), and tumour grade (grade 1/2 7% [4–11], grade 3 29% [25–34]; p<0.0001). Using the central pathology specimen review, similar results were found for rates of RCB 0; ER negative 39% [95% CI 32–46], weakly positive 35% [23–48], strongly positive 7% [5–11] (p<0.0001) and grade 1/2 7% [4–12], grade 3 31% [26–37] (p<0.0001).

DISCUSSION

This review focused on the presence or absence of pCR in the excision specimen including the presence of residual DCIS. Local pathologists were not given reporting proformas or guidelines for assessment of response which have been shown to aid concordance between pathologists in clinical trials. [6] Because the reviewing pathologists were assessing the original sections in the overwhelming majority of cases analytical issues do not impinge on this central review although differences in practice among different local laboratories would necessitate caution in drawing any comparison between centres.

In this review the pathologists were blinded to the macroscopic description and therefore had to reconstruct the tumour bed dimensions from the slides as best they could. Normally a pathologist would record a block map to aid reconstruction of the tumour area when viewing the slides, and this was highlighted as being of particular importance for accurate assessment of response in the recent BIG-NABCG working group recommendations.[7] In some cases the tumour bed was present on megaslides and this made assessment much easier. The assessment of tumour bed size is often not straightforward following neoadjuvant chemotherapy because the tumour is poorly defined macroscopically, and it can be difficult to determine the

tumour boundaries histologically. Tumour cellularity can be very heterogeneous, and is also difficult to assess in spite of the availability of online guidance tools [7] [8] and there is inconsistency amongst pathologists in these assessments.[9] Agreement about pCR should however be good and will only usually cause difficulty if small residues of tumour cells are overlooked, or if there is difficulty in interpreting in situ from invasive disease. Although in this study the best level of agreement in the reviewing pathologists' cross-over study was of numbers of lymph nodes this also is not always easy to determine without the macroscopic description. In some cases the local pathologist had written on the slide to state the number of nodes present. The concordance between the two pathologists is better than recorded in a recent review of consistency of reporting of RCB and replicates the finding that the reporting of the lymph node component of the score is more reproducible. [9] Given the limitations of this study detailed above with lack of access to source reports and block descriptions, the RCB is shown to be a very robust system for quantifying residual disease in the clinical trial context.

The central pathology review was immensely labour-intensive. The maximum number of slides submitted for a single case was 164 (median 29 per case). At best it was only possible to review three or four cases per hour. Not only was it time-consuming for the pathologists but it also placed a burden on local pathology departments retrieving slides and a considerable logistic burden for the Trials Office. One must question whether the exercise was worth the effort given that there was no clinically significant numerical change in the end results. However, one cannot generalise about central pathology review. In some trials central review is used at the outset to confirm eligibility whether this be Her-2 or ER status for example or a

particular tumour type e.g. triple negative breast cancer. In the ALTTO Trial both Her-2 and ER status were changed following central re-testing in 5 – 15% of cases.[10] An important distinction must be drawn here between re-testing, potentially using different reagents and conditions, and the review of original diagnostic material. In ARTemis we accepted a patient's eligibility as reported but reviewed critically the endpoint which was very specifically pathological.

Review of pathology reports by the two principal investigators was made more difficult by a lack of standardisation of how local reports were written – not all units use easy-to-read synoptic reports. Moreover the majority of standardised reports are designed for the adjuvant setting without specific fields for the additional variables that need to be recorded post neoadjuvant therapy, such as tumour cellularity and fibrosis in lymph nodes. Standardisation of routine reporting in clinical practice for neoadjuvant cases has been addressed recently by an international working group which should make this easier when designing future clinical trials.[11] It is possible that should such standardisation be adopted, a measure of response such as RCB could be calculated locally. Also, although there is some evidence that pathologists are better at assessing chemotherapy response by reading pathology reports than are practicing clinicians[12] this was not borne out by our data. It was evident on review of some of the cases where there were discrepancies between report-review and central review that this was due to missing slides. In this trial the two pathologists were involved in the resolution of disputed MRDs on report-review and were probably helpful in that area but MRD was not an end point of the trial. Furthermore one must urge caution in trying to make direct comparisons between RCB and more descriptive approaches to assessment of response to NAC

particularly in equating RCB1 with MRD. RCB1 is strongly dependent on low tumour cellularity while MRD as determined by the report review, where there was often no information on comparison of pre and post treatment tumour cellularity, was heavily influenced by residual tumour size and these are not always the same.

The literature on central pathology review of clinical trials is limited. The NSABP requires central pathology review for its randomised clinical trials and central reviewers are trained to operate with 90% concordance on pathological features compared with 65% concordance between local and central reporting in the NSABP B-18 trial for example. [13] Recently reported central review of bone marrow fibrosis showed a concordance of 58% between central and local reporting whereas a central panel of three reviewing pathologists achieved consistency of 88% for all three pathologists and 98% for two.[14] However we were unable to find any reports of central pathology review in the context of neoadjuvant breast cancer trials.

We chose RCB as our method of measuring chemotherapy response primarily because it gave a numerical score which proved particularly convenient when it came to the cross-over study between the two pathologists. Its principal shortcoming is the lack of comparison with the baseline core biopsy but from a clinical point of view the tumour burden following chemotherapy is a sensible feature to measure and has been shown to correlate well with outcomes at 10 years follow up.[15] One of the important aspects of ARTemis is the future programme of translational research and that has required sections from core biopsies, excised tumours and nodes to be marked up for future tissue sampling. A pathologist would certainly be required to support that aspect of a future trial. The central review process described here also

provides great confidence in the recorded ARTemis endpoints, thus supporting subsequent translational work aimed at understanding the determinants of individual tumour response and the correlations of that with long term outcomes. Recently it has been shown that combining RCB with Ki67 measurements further increases the predictive power of this tool. [16] Also there is a growing interest in either post-neo-adjuvant studies, or allowing patients to enter other studies post-NACT which means that where low volumes of residual disease are permitted in such studies, perhaps caution is needed about relying on local reporting – whereas for pCR or bulk residual disease one can probably rely on local reporting.

CONCLUSION

Central pathology review of the ARTemis trial has allowed a direct comparison with report-review and has shown that when the primary end point of the trial pCR, is compared, the two methods are equally effective. Central pathology review has a place in the assessment of minimal residual disease but if that is not an agreed pre-specified trial end point there is little extra value in doing this. Learning from the experience of ARTemis, future neoadjuvant clinical trials could be improved by training in the routine calculation of RCB. Also, standardised routine reporting using report templates would greatly assist in report-review.[17] Such training might provide more robust reporting of RCB categories, facilitating future clinical management, when current and planned trials of adjuvant treatment in patients not achieving a pCR to neoadjuvant therapy come to fruition.

Acknowledgments

The trial was funded by Cancer Research UK (CRUK/08/037 for trial coordination), Roche (provision of free study drug bevacizumab and unrestricted educational grant for trial coordination), and Sanofi–Aventis (unrestricted educational grant for trial coordination). We acknowledge the 116 investigators and their teams from 66 participating UK centres who entered patients into the ARTemis trial. Our gratitude also goes to the 800 women who kindly participated in our study. We thank the data and safety monitoring committee members: I Craig Henderson (University of California, San Francisco, CA, USA), Luca Gianni (Ospedale San Raffaele, Milan, Italy), Mark F Brady (Gynecologic Oncology Group Statistical & Data Center, Roswell Park Cancer Institute, Buffalo, NY, USA), and Xavier Pivot (Centre Hospitalier Regional Universitaire et Institut Regional Federatif de Cancerologie, Besancon, France). We thank the trials unit staff: phase 3 coordination at Warwick Clinical Trials Unit, Coventry, UK; translational coordination at University of Cambridge, Addenbrooke's Hospital, Cambridge, UK and University of Edinburgh Department of Oncology (BioBank); and statistical analysis at Warwick Clinical Trials Unit, UK.

Legends for Tables and Figures

Table 1: Patient characteristics

Table 2: RCB categories for the 65 patients, by the two pathologists

Table 3: Levels of residual cancer at surgery, from the two assessment methods for the 781 patients

Table 4A: Treatment arm comparison using local pathology report review data (n=781 patients) [1]

Table 4B: Treatment arm comparison using central pathology specimen review data (n=587 patients)

Figure 1: Level of agreement across two pathologists' rating of RCB class

Figures 2A and 2B: Inter-rater reliability of pathologists' RCB scores, where there is disagreement in RCB class (n=13 patients)

2A – Pathologist 1 v Pathologist 2 RCB Scores

2B – Average of the two pathologist's RCB Scores

Figure 3: Level of agreement across the two methods of review

REFERENCES

[1] Earl HM; Hiller L, Dunn JA; et al. Efficacy of neoadjuvant bevacizumab added to docetaxel followed by fluorouracil, epirubicin and cyclophosphamide, for women with HER2 negative early breast cancer (ARTemis): an open-label randomised phase 3 trial. Lancet Oncology 2015; **16**: 656-66. doi: 10.1016/S1470-2045(15)70137-3. Epub 2015 May 11

[2] Earl HM, Vallier A-L, Hillier L et al. Effects of the addition of gemcitabine, and paclitaxel-first sequencing, in neoadjuvant sequential epirubicin, cyclophosphamide, and paclitaxel for women with high-risk early breast cancer (Neo-tAnGo): an open-label, 2×2 factorial randomised phase 3 trial. Lancet Oncology 2014; **15**: 201-12

[3] Chavez-Mac Gregor M, Hubbard R, Meric-Bernstam F, et al Residual cancer burden (RCB) in breast cancer patients treated with taxane- and anthracycline-based neoadjuvant chemotherapy: The effect of race. Journal of Clinical Oncology, 2010; **28**, No 15 suppl: 607

[4] Symmans WF, Peintinger F, Hatzis C et al. Measurement of Residual Breast Cancer Burden to Predict Survival After Neoadjuvant Chemotherapy JCO J Clin Oncol 2007; **25**:4414-4422 DOI: 10.1200/JCO.2007.10.6823

[5] Lin L I-K. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics. 45(1) Mar. 1989, pp. 255-268. DOI: 10.2307/2532051. Stable URL: http://www.jstor.org/stable/2532051.

[6] Fisher ER, Constantino J. Quality assurance of pathology in clinical trials: The National Surgical Adjuvant Breast and Bowel Project Experience. Cancer; 1994: **74:** 2638-41.

[7] https://www.mdanderson.org/education-and-research/resources-for-professionals/clinical-tools-and-resources/clinical-calculators/calculators-rcb-pathology-protocol2.pdf (Accessed 19th August 2016)

[8] RCB Calculator. http://www3.mdanderson.org/app/medcalc/index.cfm?pagename=jsconvert3 Accessed 8th September 2016.

[9] Peintinger F, Sinn B, Hatzis C et al. Reproducibility of residual cancer burden for prognostic assessment of breast cancer after neo-adjuvant chemotherapy. Modern Pathology 2015; **28**: 913-20.

[10] McCullough AE, Dell'Orto P, Reinholz MM, et al. Central Pathology review of Her-2 and ER in early breast cancer: An ALTTO Trial [BIG 2-06/NCCTG NO63D (Alliance)] ring study. Br Cancer Res Treat 2014; **143**: 485-92.

[11] Provenzano E, Bossuyt V, Viale G, et al. Standardization of pathological evaluation and reporting of postneoadjuvant specimens in clinical trials of breast cancer: recommendations from an international working group. Modern Pathology 2015; **28**: 1185-1201

[12] Provenzano E, Vallier A-L, Champ R, et al. A central review of histopathology reports after breast cancer neoadjuvant chemotherapy in the neo-tango trial. Br J Cancer 2013; **108**: 866-72.

[13] Fisher B, Brown A, Mamounas E et al. Effect of preoperative chemotherapy on local-regional disease in women with operable breast cancer: findings from National Surgical Adjuvant Breast and Bowel Project B-18. J Clin Oncol 1997; **15**: 2483 – 93.

[14] Pozdnyakova O, Rodig S, Bhandarkar S, Wu K, Thiele J , Hasserjian R. The importance of central pathology review in international trials: a comparison of local versus central bone marrow reticulin grading. Leukemia 2015; **29**, 241–244; doi:10.1038/leu.2014.262

[15] Symmans WF, Wei C, Gould R, et al. Long-term prognostic value of residual cancer burden (RCB) classification following neoadjuvant chemotherapy [abstract S6-02]. Paper presented at: San Antonio Breast Cancer Symposium; 10–14 Dec 2013, 2013; SanAntonio, TX.

[16] Sheri A, Smith IE, Johnston SR, et al. Residual proliferative cancer burden to predict longterm outcome following neoadjuvant chemotherapy. Annals of Oncology 2015; **26**: 75–80. doi:10.1093/annonc/mdu508

[17] Bossuyt, V. & Symmans, W.F. Standardizing of Pathology in Patients Receiving Neoadjuvant Chemotherapy. Ann Surg Oncol 2016; **23**: 3153. doi:10.1245/s10434-016-5317-x

Figure 1: Level of agreement across two pathologists' rating of RCB class

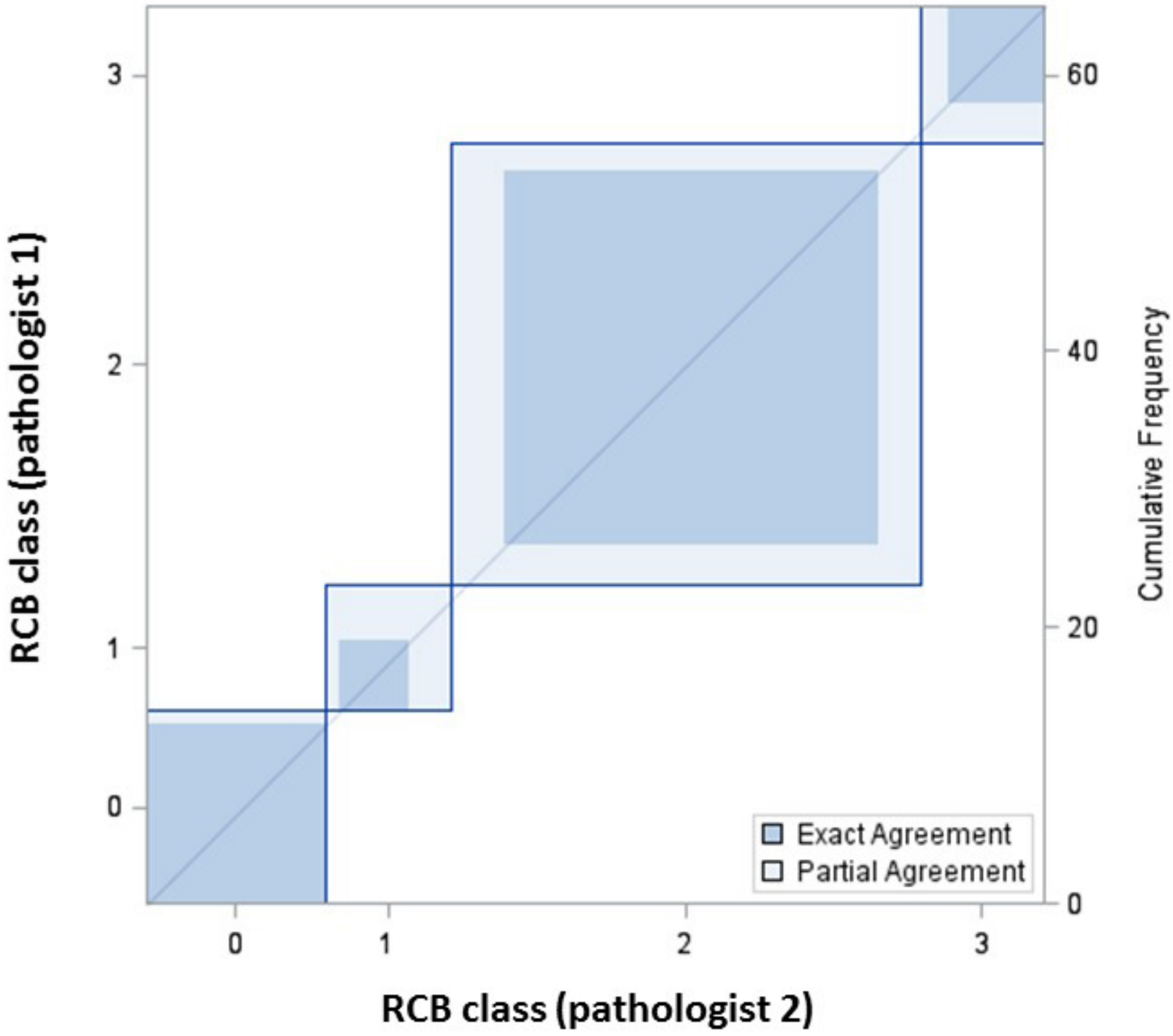
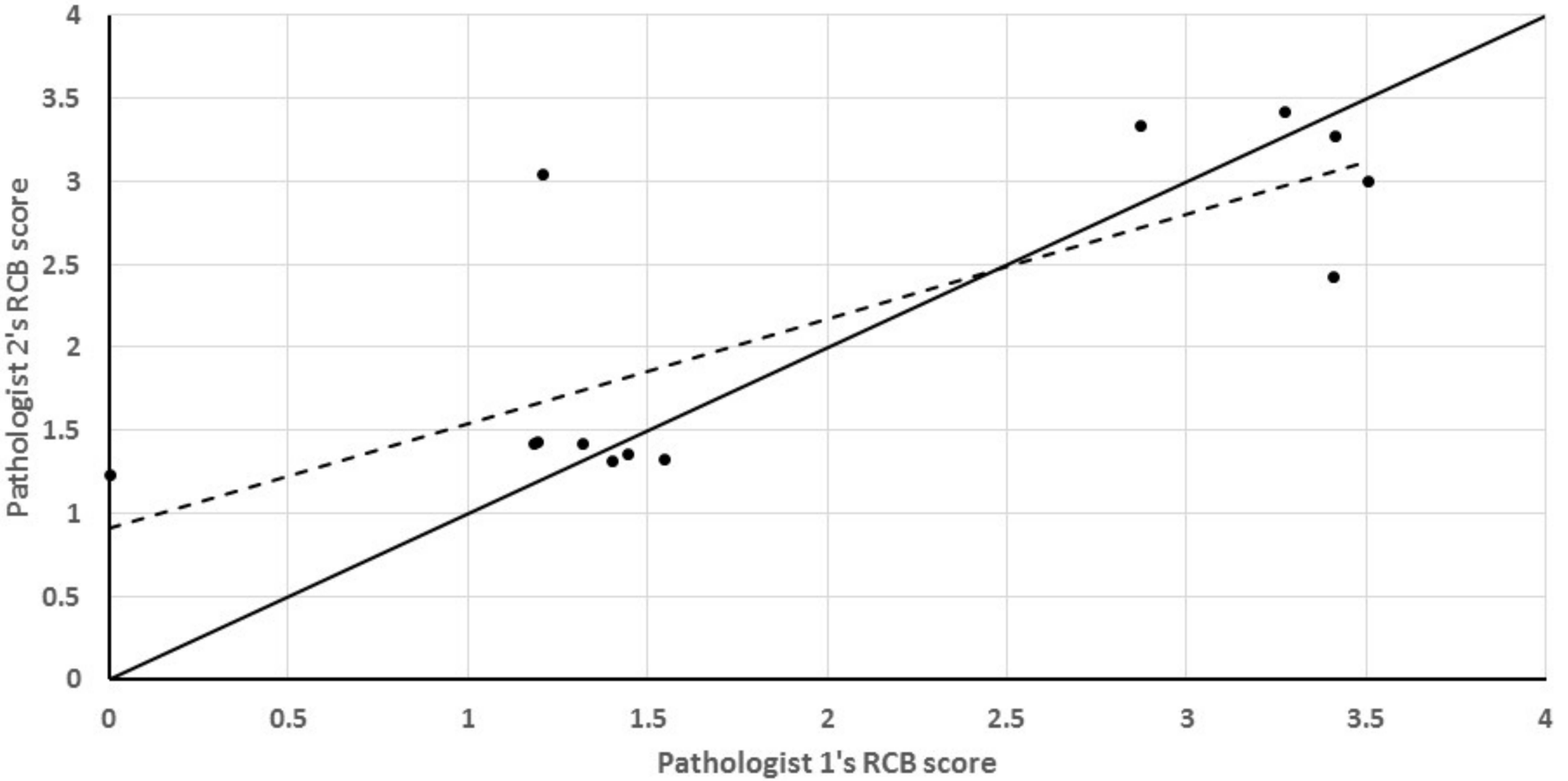
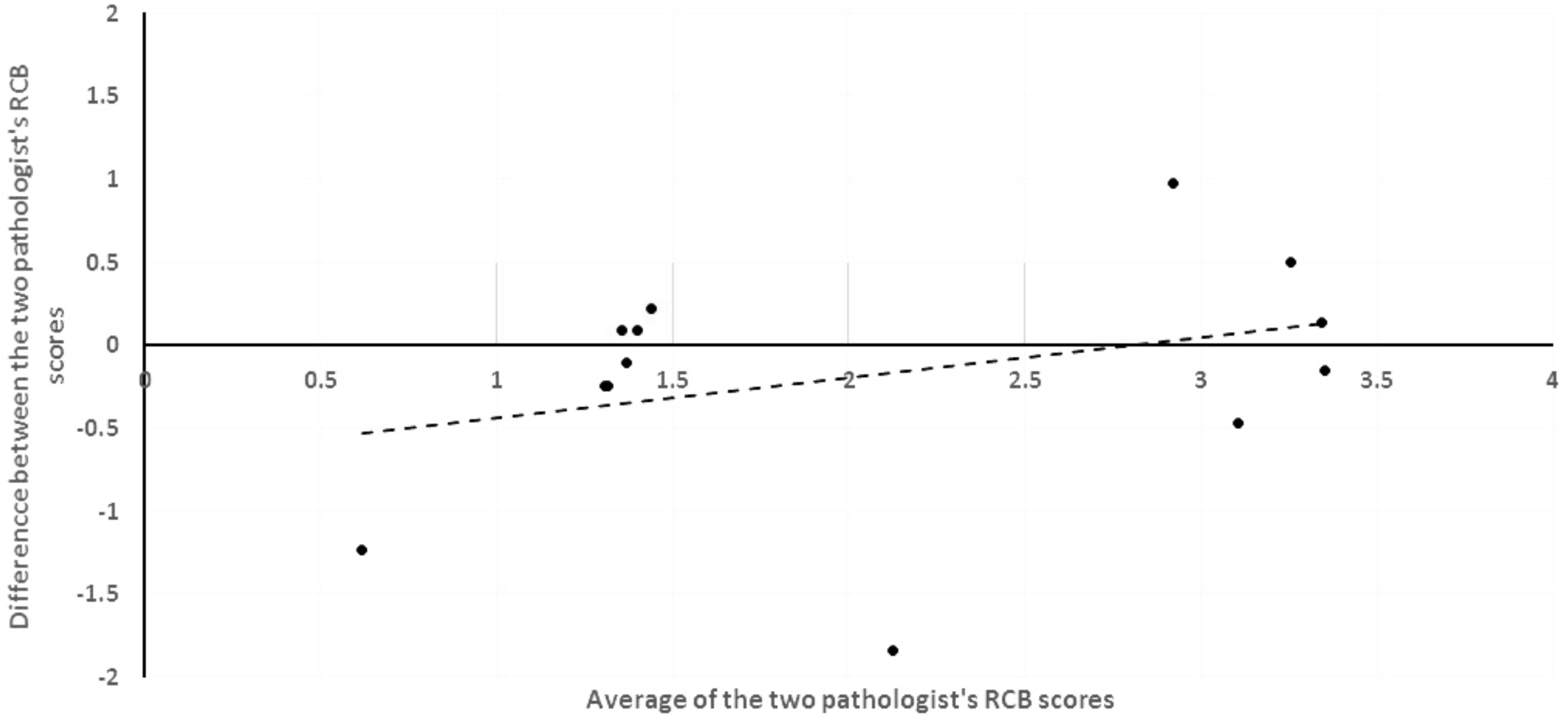


Figure 2a



Solid line = line of equality. Dashed line = linear line of best fit.

Figure 2b



Dashed line = linear line of best fit.

Figure 3: Level of agreement across the two methods

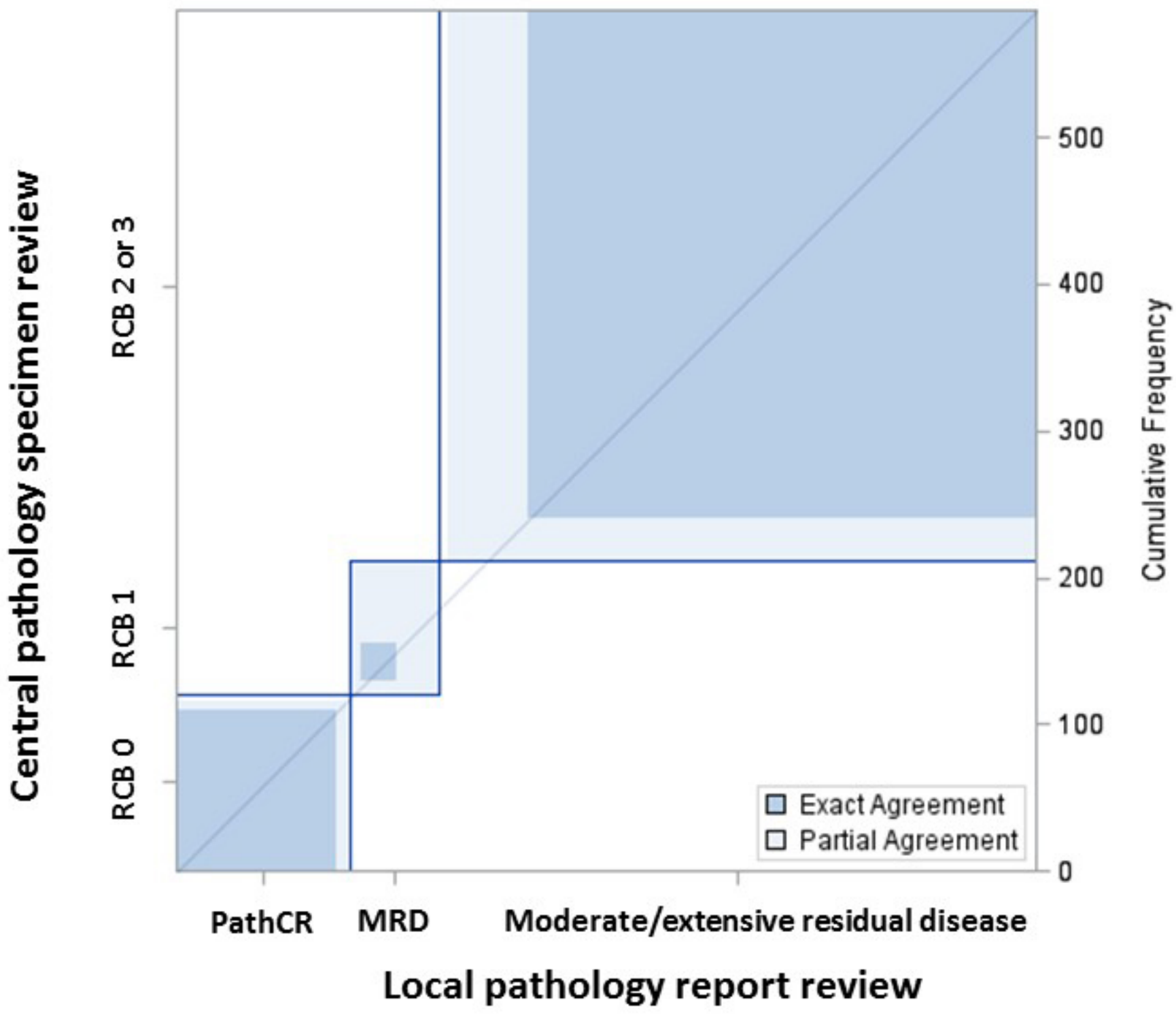


Table 1: Patient characteristics

		Full trial population (n=800) N (%)	Central pathology review sample (n=587) N (%)	Inter-rater reliability sample (n=65) N (%)
Randomised treatment	Bev+D-FEC	399 (50%)	290 (49%)	29 (45%)
	D-FEC	401 (50%)	297 (51%)	36 (55%)
Age	≤50 years old	543 (68%)	393 (67%)	44 (68%)
	>50 years old	257 (32%)	194 (33%)	21 (32%)
ER status	Negative (Allred score 0-2)	248 (31%)	194 (33%)	18 (28%)
	Weakly positive (Allred score 3-5)	75 (9%)	60 (10%)	10 (15%)
	Strongly positive (Allred score 6-8)	477 (60%)	333 (57%)	37 (57%)
Tumour size	≤50mm	635 (79%)	472 (80%)	54 (83%)
	>50mm	165 (21%)	115 (20%)	11 (17%)
Clinical involvement of axillary nodes	No	383 (48%)	288 (49%)	38 (58%)
	Yes	417 (52%)	299 (51%)	27 (42%)
Inflammatory or locally advanced disease or both	No	651 (81%)	484 (82%)	51 (78%)
	Yes	149 (19%)	103 (18%)	14 (22%)
RCB category *	0	-	121 (21%)	14 (21%)
	1	-	90 (15%)	9 (14%)
	2	-	290 (49%)	33 (51%)
	3	-	86 (15%)	9 (14%)

* RCB category is only known for those patients included in the central pathology review sample

Table 2: RCB categories for the 65 patients, by the two pathologists

	Pathologist 1				
Pathologist 2	**RCB Cat 0**	**RCB Cat 1**	**RCB Cat 2**	**RCB Cat 3**	**Total**
RCB Cat 0	13	-	-	-	13
RCB Cat 1	1	5	3	-	9
RCB Cat 2	-	4	27	3	34
RCB Cat 3	-	-	2	7	9
Total	14	9	32	10	65

Table 3: Levels of residual cancer at surgery, from the two assessment methods for the 781 patients

Local pathology report review	Central specimen review					Total
	RCB Cat 0	RCB Cat 1	RCB Cat 2	RCB Cat 3	Not available	
pathCR *	109	9	1	0	34	153
MRD	7	25	29	0	16	77
Moderate/extensive residual disease	5	56	260	86	144	551
Total	121	90	290	86	194	781

* pCR in all breast tumours and absence of disease in all removed axillary lymph nodes (ypT0/Tis ypN0)
Shaded cells indicate agreement

Table 4A: Treatment arm comparison using local pathology report review data (n=781 patients) [1]

Local pathology report review	**D→FEC** **% (95%CI)**	**Bev+D→FEC** **% (95%CI)**	p *****
	(n=66/393)	(n=87/388)	
pCR in all breast tumours AND absence of disease in all removed ax LNs (ypT0/Tis ypN0)	**17%** (13-21%)	**22%** (18-27%)	**0·03**
ER neg (Allred 0-2) (n=241)	**31%** (23-40)	**45%** (36-55)	
ER weak pos (Allred 3-5) (n=74)	**30%** (16-47)	**51%** (34-68)	
ER strong pos (Allred 6-8) (n=466)	**7%** (4-11)	**6%** (3-10)	
‡Grade 1/2 (n=293)	**8%** (5-14)	**6%** (3-11)	
Grade 3 (n=403)	**23%** (17-29)	**36%** (29-43)	

***** Adjusted for the five stratification variables

‡Tumour grade of each patient's largest breast tumour at baseline.

Table 4B: Treatment arm comparison using central pathology specimen review data (n=587 patients)

Central pathology review	**D→FEC** **% (95%CI)**	**Bev+D→FEC** **% (95%CI)**	p *****
	(n=49/297)	(n=72/290)	
RCB 0	**16%** (12-21%)	**25%** (20-30%)	**0·02**
ER neg (Allred 0-2) (n=194)	**33%** (24-44%)	**45%** (35-55%)	
ER weak pos (Allred 3-5) (n=60)	**22%** (9-42%)	**45%** (28-64%)	
ER strong pos (Allred 6-8) (n=333)	**6%** (3-11%)	**8%** (4-14%)	
‡Grade 1/2 (n=216)	**9%** (4-16%)	**6%** (2-12%)	
Grade 3 (n=313)	**23%** (17-31%)	**39%** (32-47%)	

***** Adjusted for the five stratification variables

‡Tumour grade of each patient's largest breast tumour at baseline.