Reproducibility of residual cancer burden for prognostic assessment of breast cancer after neoadjuvant chemotherapy

Florentia Peintinger[1,2], Bruno Sinn[3], Christos Hatzis[4], Constance Albarracin[3], Erinn Downs-Kelly[5], Jerzy Morkowski[6], Rebekah Gould[3] and W Fraser Symmans[3]

[1]Institute of Pathology, Medical University of Graz, Graz, Austria; [2]Department of Gynecology, General Hospital Leoben, Leoben, Austria; [3]Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; [4]Yale Comprehensive Cancer Center, Yale School of Medicine, New Haven, CT, USA; [5]Department of Pathology, Huntsman Cancer Hospital, University of Utah, Salt Lake City, UT, USA and [6]MLD Pathology, Houston, TX, USA

The residual cancer burden index was developed as a method to quantify residual disease ranging from pathological complete response to extensive residual disease. The aim of this study was to evaluate the inter-Pathologist reproducibility in the residual cancer burden index score and category, and in their long-term prognostic utility. Pathology slides and pathology reports of 100 cases from patients treated in a randomized neoadjuvant trial were reviewed independently by five pathologists. The size of tumor bed, average percent overall tumor cellularity, average percent of the *in situ* cancer within the tumor bed, size of largest axillary metastasis, and number of involved nodes were assessed separately by each pathologist and residual cancer burden categories were assigned to each case following calculation of the numerical residual cancer burden index score. Inter-Pathologist agreement in the assessment of the continuous residual cancer burden score and its components and agreement in the residual cancer burden category assignments were analyzed. The overall concordance correlation coefficient for the agreement in residual cancer burden score among pathologists was 0.931 (95% confidence interval (CI) 0.908–0.949). Overall accuracy of the residual cancer burden score determination was 0.989. The kappa coefficient for overall agreement in the residual cancer burden category assignments was 0.583 (95% CI 0.539–0.626). The metastatic component of the residual cancer burden index showed stronger concordance between pathologists (overall concordance correlation coefficient = 0.980; 95% CI 0.954–0.992), than the primary component (overall concordance correlation coefficient = 0.795; 95% CI 0.716–0.853). At a median follow-up of 12 years residual cancer burden determined by each of the pathologists had the same prognostic accuracy for distant recurrence-free and survival (overall concordance correlation coefficient = 0.995; 95% CI 0.989–0.998). Residual cancer burden assessment is highly reproducible, with reproducible long-term prognostic significance.

Modern Pathology (2015) **28,** 913–920; doi:10.1038/modpathol.2015.53; published online 1 May 2015

Neoadjuvant chemotherapy is often used in patients with locally advanced breast cancer to downstage the tumor and to evaluate *in vivo* chemosensitivity.[1,2] Pathological complete response is defined as the absence of invasive cancer in the breast and in the nodes after completion of neoadjuvant chemotherapy. A recent meta-analysis of 12 randomized trials by the Collaborative Trials in Neoadjuvant Breast Cancer confirmed pathologic complete response as a surrogate endpoint for event-free and overall survival. In particular, pathologic complete response was associated with a 52% reduction in the probability of an event and a 64% reduction in the probability of death.[3] Thus, pathologic complete response has been used as the primary endpoint in a number of trials evaluating efficacy of different drugs. Breast cancer of certain subtypes may have an excellent chemosensitivity but may also show a spectrum of post-neoadjuvant chemotherapy residual disease ranging from minimal (near pathologic complete response) to extensive residual disease. At present, a variety of

non-standardized procedures are used for the evaluation of pathological response after neoadjuvant treatment and this can impair the quality and reliability of pathology assessment across different institutions. Evaluation of the Neo-tAnGo study showed that only 45% of the pathology reports from patients with residual disease indicated the chemotherapy effect and less than 10% quantified response at all.[4] Residual disease can be subtle and/or scattered, and in these cases pathology reports tend to collect more descriptive rather than quantitative information in the absence of standardized guidelines to measure and report the extent of residual cancer. In addition, the reproducibility and prognostic significance of reported residual disease assessment across different pathologists is difficult to study and has not been formally tested.

At M.D Anderson Cancer Center we developed residual cancer burden (RCB) as a method to quantify residual disease after neoadjuvant chemotherapy for breast cancer.[5] RCB can be calculated through a web-based calculator either as a numerical score (index) or as a category.[6] The RCB index is based on histopathological variables such as number of involved nodes, size of the largest nodal metastasis and, size and percent cellularity of the primary tumor bed.

The RCB categories have been shown to correlate with long-term survival outcomes across breast cancer subtypes and a number of clinical study groups such as I-SPY (1,2), GEICAM, ACOSOG (Z11103), CALGB (40601, 40603), NSABP (B-40, B-41), and ABCSG (34) have incorporated RCB as the primary or secondary endpoint of chemotherapy response in prospective neoadjuvant trials.[7] Concerns have been raised that parameters used for RCB calculation are not part of a standardized pathology report and may be somewhat subjective for evaluation among different observers, especially when reporting the extent of residual tumor cellularity.[8]

The aim of this study was to evaluate the inter-pathologist reproducibility of RCB index score and category, and of the long-term prognostic utility, when assessed by five different pathologists in a blinded "round–robin" analysis of retrospective reports and slides from patients with residual *in situ*, invasive, and/or nodal disease after six months of neoadjuvant taxane-anthracycline based chemotherapy.

Materials and methods

We selected 100 random cases with residual *in situ*, invasive, or metastatic carcinoma in the axillary nodes from patients who were treated in a published randomized neoadjuvant trial (protocol MDACC DM 98-240) with a regimen including paclitaxel followed by fluorouracil, doxorubicin, and cyclophosphamide.[9] These cases included 60 hormone-receptor positive, 23 HER2-positive and 17 triple-negative tumors. The gross pathologic reviews, tissue

sampling, description of gross findings and tissue sections had been performed in the past using legacy clinical methods (that is, without standardization and before RCB had been conceived). The pathology slides and original pathology reports were reviewed independently by the five pathologists including two fellows, one visitor, and two faculty members at M.D Anderson Cancer Center. The original pathology reports at M.D Anderson Cancer Center routinely included two-dimensional measurements of the macroscopic tumor dimensions, number of involved nodes and the diameter of the largest metastasis. In cases of multicentric disease, the largest tumor bed was measured. Pathologists were free to infer results from reports or their interpretation of the slides from these retrospective materials, as they saw fit.

Pathologist A's RCB results were derived from the original development cohort that was published in 2007.[5] Pathologists B, C, D, and E were blinded to other results or outcomes and were assessing RCB for the first time in their career when they participated in this study. They did not receive individual coaching nor were they trained in RCB evaluation at the microscope. Pathologists B, C, and D assessed the 100 cases soon after publication of the original RCB paper (in 2007), and pathologist E assessed the same cases one year later (five cases were missing from pathologist E). They were provided with the published materials and the corresponding website for appended instructions and protocol for pathologists and RCB calculation.

Microscopic and macroscopic pathological components, namely size of tumor bed (mm), average percent overall tumor cellularity (invasive and *in situ*), average percent of the cancer within the tumor bed that is *in situ*, size of largest axillary metastasis (mm) and number of involved nodes were assessed separately by each pathologist and RCB categories were assigned to each case following calculation of the numerical RCB index score.[6]

Inter-pathologist agreement in the assessment of the continuous RCB score and its components was evaluated based on the overall concordance correlation coefficient, and agreement in the RCB category assignments was assessed based on the kappa coefficient.

The agreement between the continuous scores obtained by the five pathologists was evaluated based on the overall concordance correlation coefficient for multiple observers.[10] When disagreement was present, it was assessed in terms of a systematic shift (inaccuracy) component and a random error (imprecision) component. The 95% confidence interval (CI) for the overall concordance correlation coefficient were obtained based on U-statistics. Each tumor was also assigned into one of four pre-defined RCB categories according to the RCB score. The agreement in the RCB category assignments by the five pathologists was evaluated based on the simple (unweighted) kappa statistic for multiclass observations. CIs were obtained based on the asymptotic

Table 1 Estimated concordance correlation coefficients for continuous RCB scores from 95 breast cancer specimens

Comparison	Pairwise concordance correlation coefficients	95% Confidence interval	Accuracy	Precision
Path A *vs* path B	0.952	0.926–0.969	0.992	0.959
Path A *vs* path C	0.948	0.929–0.963	0.998	0.950
Path A *vs* path D	0.938	0.914–0.956	0.980	0.958
Path A *vs* path E	0.908	0.849–0.945	0.970	0.937
Path B *vs* path C	0.950	0.923–0.968	0.998	0.953
Path B *vs* path D	0.950	0.922–0.968	0.997	0.953
Path B *vs* path E	0.921	0.870–0.952	0.990	0.930
Path C *vs* path D	0.915	0.885–0.937	0.990	0.925
Path C *vs* path E	0.906	0.854–0.940	0.983	0.922
Path D *vs* path E	0.923	0.872–0.954	0.995	0.927
Overall	0.931	0.908–0.949	0.989	0.941

variance of the statistic. Computations were performed using R 3.1.[11]

Distant recurrence-free survival was defined as the interval from diagnosis until distant disease recurrence or death from any cause. Overall survival was defined as the interval from diagnosis until death from any cause. Survival analyses were computed using the R package survival.[12] The Kaplan–Meier estimator and the log rank test was used to assess the effect of RCB classes on survival outcome. Significance of the effect of the continuous RCB score on survival outcome was evaluated by Cox regression analysis adjusted for hormone-receptor status.

Results

Agreement in Continuous RCB Score

Five cases were excluded from the analysis due to missing data from any one of the five observers. Therefore, a total of ninety-five cases were evaluated for consistency or agreement between the five pathologic measurements of residual cancer by assessing agreement between the corresponding RCB score on the continuous scale. In this analysis the five pathologists were treated symmetrically, that is, none of them was considered as providing a reference score. Agreement in this setting implies that the observations by any of the pathologists can be used interchangeably. Table 1 shows the estimated pairwise concordance correlation coefficients for each pair of pathologists. There was generally good agreement between pathologists with the pairwise concordance correlation coefficients ranging from 0.91 to 0.95. To evaluate the source of disagreement, the concordance correlation coefficient is typically expressed as the product of two terms, accuracy and precision, which can be estimated separately. The estimated accuracy coefficients for all pairwise comparisons were very close to 1 indicating that the marginal distributions of RCB score between two pathologists were equal, that is, both means and variances were equal. The source of

disagreement seems to be due to reduced precision, which is measured by the Pearson correlation coefficient between pairs of observations. The overall concordance correlation coefficient for the agreement among all five pathologists was 0.931 (95% CI = 0.908–0.949). Overall accuracy of the RCB score determination was 0.989, suggesting negligible shift (location or scale) in the distribution of RCB score among pathologists. However, the overall precision was 0.941, which indicates appreciable within-sample variation or random error in the evaluations of the pathologists. The overall concordance, accuracy, and precision were 0.923 (95% CI 0.892–0.946), 0.982, and 0.941 in the hormone receptor positive; and 0.942 (95% CI 0.899–0.967), 0.993, and 0.949 in the hormone-receptor-negative subset.

Agreement Among RCB Category Assignments

Instead of using the RCB score of a tumor directly for prognosis, each tumor is typically assigned to one of four RCB categories ranging from no residual cancer (RCB-0 or pathologic complete response), minimal (RCB-I), moderate (RCB-II) or extensive residual cancer (RCB-III) based on published cutoff points on the RCB score.[5] Agreement between the RCB categories was evaluated using the simple kappa statistic for multinomial observations. Table 2 shows the number of calls in each of the four RCB categories made by the five pathologists. The marginal distributions by the first two pathologists look similar, but those by pathologists C, D, and E appear more deviant at the tails. The kappa statistic for inter-rater agreement with respect to classification into a single RCB category is also shown in Table 2. Values of the kappa statistic around 0.6 indicate moderate to substantial agreement. Classification of extensive residual disease (RCB-III) appeared to be very consistent among the five pathologists, with a kappa of 0.666, whereas classification of minimal residual disease (RCB-I) appears to be the least concordant with a kappa of 0.533. The overall kappa was 0.583 (95% CI = 0.539–0.626), indicating good overall agreement.

Agreement in Pathologic Components of the RCB Score

The overall concordance analysis showed excellent accuracy in the determination of the RCB score by the five pathologists, but indicated reduced precision due to a sizable within-sample variability or random error component. In order to understand the source of this variability, we evaluated the concordance separately for the primary and metastatic components of the RCB score. The primary RCB component was the main source of within-sample variability of the RCB score (overall concordance correlation coefficient = 0.795; 95% CI = 0.716–0.853), whereas the metastatic component shows perfect concordance between pathologists (overall concordance correlation coefficient = 0.980; 95% CI = 0.954–0.992). Further

Table 2 Pathologists' observed marginal distributions for RCB categories obtained from 95 breast cancer specimens

Observer	Pathologic complete response	RCB-I	RCB-II	RCB-III
A	4	18	61	12
B	3	18	55	19
C	1	25	52	17
D	6	11	53	25
E	4	11	55	25
Kappa	0.682	0.533	0.542	0.666
95% Confidence interval	0.618–0.745	0.469–0.596	0.478–0.606	0.602–0.730
Overall Kappa	0.583			
95% Confidence interval	0.539–0.626			

evaluation of the pathologic measurements contributing to the primary RCB component revealed that estimation of the primary tumor bed size (concordance correlation coefficient = 0.704; 95% CI = 0.550–0.812) and of the fraction of invasive cancer (concordance correlation coefficient = 0.699; 95% CI = 0.621–0.763) affects precision and accuracy. The precision of the two estimates was similar (0.781 for invasive carcinoma *vs* 0.742 for tumor bed size), but estimation of the fraction of invasive cancer was less accurate (0.894) or more biased compared with estimation of the tumor bed size (accuracy = 0.949).

Agreement in Prognostic Risk Assessment

RCB scores and RCB classes determined by each of the pathologists were prognostic for DRFS at a median follow-up of 12.16 years. Figure 1a and Table 3 summarize the results of the Cox regression analysis for the continuous RCB data with adjustment for hormone-receptor status. The results demonstrate reproducible estimation of risk by the different pathologists reporting continuous RCB score. Figure 2 shows excellent concordance of the predicted survival proportions derived from these models with an overall concordance correlation coefficient of 0.995 (95% CI = 0.989–0.998). Concerning the categorical RCB classification, Figure 1b and Table 3 show the results from a Cox regression using the categorical RCB data adjusted for hormone-receptor status. Figure 3 shows Kaplan–Meier plots for RCB classes defined by each of the pathologists, suggesting generally good agreement for the survival estimates. Each of the categorizations was prognostic for distant recurrence-free survival, and there was some variation in the estimated 5- and 10-year survival, but the differences were not significant (Table 4). The one outlier appears to be pathologist D, whose RCB-I cases had worse prognosis than expected. However, it should be noted that the result is likely to have resulted from classification of two cases as RCB-I (of 11 total cases) that had an early relapse event and were not classified as RCB-I by the
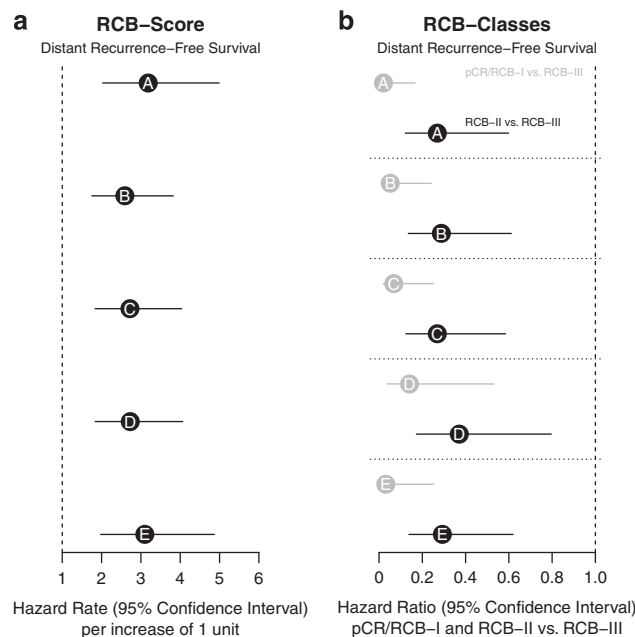

Figure 1 (**a**) Hazard rates with 95% confidence intervals for prediction of distant recurrence-free survival for pathologists A–E. The hazard rates correspond to 1 unit increase of the continuous RCB score. (**b**) Hazard ratios for the categorical RCB classes. The risk for pathologic complete response (pCR)/RCB-I (gray) and RCB-II (black) is reported relative to that for RCB-III. The analysis was adjusted for hormone-receptor status.

other pathologists. This was also reflected in the estimates of 5- and 10-year distant recurrence-free survival for pathologist D (Table 4). The other observed difference in prognosis relates to the RCB-III category. In general, the estimates for 5- and 10-year distant recurrence-free survival were higher for RCB-III (Table 4) for the pathologists who more frequently classified tumors as RCB-III (Table 2). For example, pathologist A identified 13% of cases as RCB-III corresponding to 5-year distant recurrence-free survival of 42%, whereas pathologists D and E identified 26% RCB-III with 5-year DRFS of 60%. The analysis of overall survival showed similar results and is presented in the Supplementary Material.

Discussion

Overall, there was strong concordance across the five pathologists for assessment of the continuous RCB score (concordance correlation coefficient = 93.1%), that is the combination of excellent accuracy (98.9%) and precision (94.1%). However, inter-pathologist agreement for assignment of RCB score to a category (pathologic complete response, RCB-I, RCB-II, or RCB-III) was only good, with overall kappa value of 0.583 (95% CI 0.539–0.626). More variability was seen between pathologists in the assessment of percent tumor cellularity (concordance correlation coefficient = 69.9%) and tumor bed size (concordance correlation coefficient = 70.4%). Importantly, the prognostic assessment of future risk using RCB

Table 3 Hazard rates for continuous RCB score and hazard ratios with 95% CI for prediction of distant recurrence-free survival for observers A–E

Observer	Hazard rate	Lower 95% confidence limit	Upper 95% confidence limit	P-value (LR test)
Continuous RCB-index: distance recurrence-free survival				
A	3.183	2.030	4.990	< 0.001
B	2.591	1.755	3.824	< 0.001
C	2.721	1.834	4.035	< 0.001
D	2.731	1.836	4.062	< 0.001
E	3.102	1.977	4.868	< 0.001

Observer	*Pathologic complete response/RCB-I vs RCB-III*		*RCB-II vs RCB-III*		P-value (LR test)
	Hazard ratio	95% CI	Hazard ratio	95% CI	
Categorial RCB: distance recurrence-free survival-reference class RCB-III					
A	0.021	0.003–0.166	0.268	0.121–0.599	< 0.001
B	0.052	0.011–0.242	0.288	0.135–0.611	< 0.001
C	0.068	0.018–0.251	0.269	0.124–0.585	< 0.001
D	0.141	0.037–0.532	0.371	0.173–0.796	< 0.001
E	0.031	0.004–0.253	0.292	0.138–0.619	< 0.001

Abbreviations: CI, confidence interval; LR, likelihood ratio.
The estimates are adjusted for hormone-receptor status.

was highly reproducible, whether by continuous score or category (Figures 1 and 2).

One can be encouraged that the inter-pathologist concordance of RCB index scores represents the "buffering" effect of a multivariate index that is not dependent on any single measurement of a single parameter. Similarly, the RCB index scores from the five different pathologists had significant prognostic value for distant recurrence-free survival (hazard rates 2.6–3.2) and OS (hazard rates 2.1–2.5, see Supplementary Information) adjusted for hormone-receptor status. We observed that the two trainees performed at least as well as their more experienced colleagues, suggesting that learning from educational materials attention to detail are as important as experience.

It is also not surprising that imperfect precision (94.1%) would correspond to reduced inter-pathologist agreement on the category of RCB (58.3%) because the four RCB categories do not really represent independently different pathological outcomes, but are defined by two different arbitrary thresholds applied to a distribution of RCB scores. Minor variation in RCB score therefore leads to higher rate of disagreement among RCB classes, even if the prognostic relevance of the disagreements are actually trivial. For example, the prognostic risk for a high RCB-I is similar to that of a low RCB-II, and even a high rate of imprecision between pathologists would still be likely to assign such cases to either category. Consequently, the Kaplan–Meier plots for the four RCB groups were generally quite similar across the five pathologists. However, we did observe that the pathologists who classified pathologic complete response or minimal residual disease less frequently tended to classify RCB-III more frequently

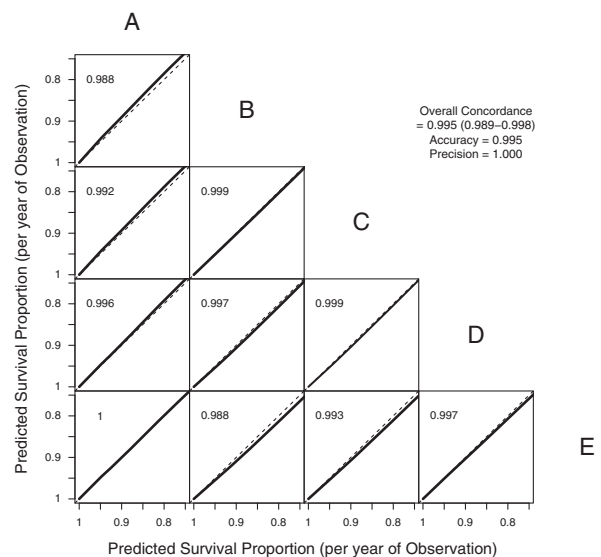RCB–Score: Concordance of Predicted Distant Recurrence–Free Survival

Overall Concordance
= 0.995 (0.989–0.998)
Accuracy = 0.995
Precision = 1.000

Figure 2 Concordance of predicted distant recurrence-free survival from the continuous RCB score. The predicted survival proportions per year of observation are derived from Cox models with adjustment for hormone-receptor status and plotted pairwise for the different observers.

(Table 3), but their corresponding survival estimates for this most resistant category were more favorable. This illustrates how any bias toward over-estimation of cellularity and/or tumor bed size would diminish the prognostic meaning of the RCB-III category. It is important to estimate the average cellularity across the tumor bed area (described in the protocol) rather than the maximum cellularity or the average of the more cellular areas in the tumor bed.
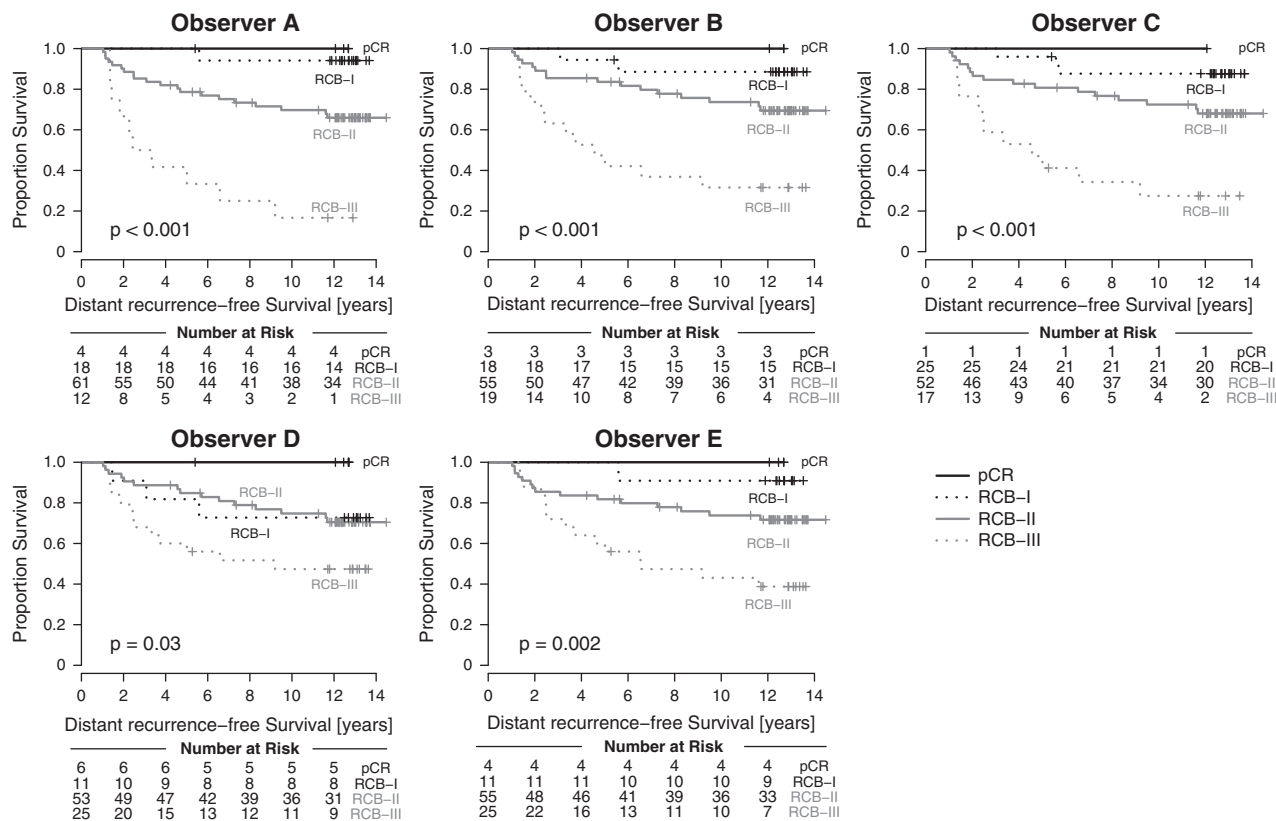
Figure 3 Kaplan–Meier plots for distant recurrence-free survival for RCB classes determined by observers A–E.

We strictly designed this study to not allow coaching or training of pathologists in the performance of RCB assessment, but only provide the published materials from the original manuscript and the accompanying website (http://www3.mdanderson.org/app/medcalc/index.cfm?pagename = jsconvert3). The purpose was to simulate adoption of this method by others from publicly available education materials and to learn about both the analytical and the prognostic reproducibility of RCB assessment by different pathologists.[5,10] One can sometimes recognize a case as minimal, moderate or extensive residual disease based on first impressions after reviewing the slides. However, the very high concordance between pathologists when measuring RCB index, and even higher concordance of the prognostic information derived from those measurements, provides reliable and more specific information to the treating physician and surgeon. That justifies the utilization of this method.

Our study design has several important limitations. First, the perfect comparative study can never be achieved because it is impossible to have different pathologists receive, examine, sample, and interpret the extent of residual disease from identical surgical resection specimens. That would test the entire standard operating procedure for this method of pathologic assessment. This study tested the interpretation of slides and reports from archival samples

that could never be seen or fully understood by the study pathologists, and were devoid of any relevant clinical or radiologic information. Moreover, the cases were selected to have long follow-up (in order to study prognostic relevance) and so they pre-dated current procedures for the gross assessment of a post-neoadjuvant resection specimen. Thus, there were no radiographs, photographs or diagrams of the specimens, no precise maps of how the slides related to the specimen or to each other, no standardized procedures for sampling at that time (other than the intent to determine pathologic complete response from residual disease), the primary tumor beds were not routinely marked with radiologic clips at that time, and sentinel node biopsies were not performed in patients with clinically node-positive disease at that time. Indeed, we might anticipate even stronger prognostic utility when using current standard operating procedures.

The results from this retrospective study of archival materials are informative for the interpretation of clinical trials where RCB is proposed as a primary or secondary endpoint. One can appreciate several scenarios with different implications for the expected quality of results. Some trials have prospectively included a standard operating procedure to standardize pathology assessment for RCB, pathologic complete response, and AJCC/UICC staging and proactively trained at least one dedicated pathologist

Table 4 Kaplan–Meier estimates of distant recurrence-free survival rates at 5 and 10 years (95% CI) for RCB classes obtained from observers A through E

	Observer	Pathologic complete response	RCB-I	RCB-II	RCB-III
5-year distant recurrence-free survival	A	100	100	79 (69–90)	42 (21–81)
5-year distant recurrence-free survival	B	100	94 (84–100)	84 (74–94)	47 (30–76)
5-year distant recurrence-free survival	C	100	96 (89–100)	81 (71–92)	47 (28–78)
5-year distant recurrence-free survival	D	100	82 (62–100)	85 (75–95)	60 (44–83)
5-year distant recurrence-free survival	E	100	100	82 (72–93)	60 (44–83)
10-year distant recurrence-free survival	A	100	94 (84–100)	70 (59–82)	17 (5–60)
10-year distant recurrence-free survival	B	100	89 (75–100)	74 (63–87)	32 (16–61)
10-year distant recurrence-free survival	C	100	88 (75–100)	72 (61–86)	27 (12–61)
10-year distant recurrence-free survival	D	100	73 (51–100)	75 (64–88)	47 (31–72)
10-year distant recurrence-free survival	E	100	91 (75–100)	74 (63–87)	43 (27–68)

at each site to incorporate the protocol and prospectively interpret and report RCB (for example, I-SPY2, ABCSG, Kristine). This approach even allows for internal auditing by central review of part or all of the subjects. Others include a relevant standard operating procedure for pathology, but intend to retrospectively collect slides and reports for a central review, or request prospective assessment of RCB without training of pathologists at each site or real-time confirmation of performance. Yet others do not include any standardized procedures specific to the assessment of RCB but intend to obtain RCB assessments by retrospective review of available materials. Our study represents the results that might be expected from the last scenario. The rate of inter-pathologist disagreement in their interpretations of the tumor bed size and cellularity is likely to be overstated compared with a higher quality approach (first scenario). These would reasonably be expected to improve with standardized procedures for macroscopic assessment and mapping of tissue sections to the gross specimen that are recommended in the published standard operating procedure for RCB assessment in a prospective setting.[5,13] Consequently, one might expect that inter-pathologist agreement in RCB assessments would also be higher with standardization of more consistent methods to map the residual tumor bed for extent of disease and definition of the area in which to estimate percent cancer cellularity (for example, protocol for pathologists, www.mdanderson.org/breastcancer_RCB). What happens prospectively in the grossing room (and is recorded by images, maps and description) profoundly affects the quality of interpretation and reporting of the extent and burden of residual disease.

Assessment of residual nodal disease was highly reproducible (overall concordance correlation coefficient = 0.980), despite fibrotic chemotherapy effects that might contribute to variable interpretation. Indeed, it is well established in the pathology community that thorough examination of the axillary specimen is essential to avoid underestimation of nodal residual disease, and so nodal assessment is already well standardized.[14] Nevertheless, the excellent concordance would be bolstered by a majority of cases with zero nodal disease and the limitation that no study could ever compare the sampling of the same axillary contents by multiple pathologists.

Despite the limitations imposed on this study of archival materials by the outdated methods for evaluation of resection specimens, the concordance of RCB measurements and agreement of RCB categories still compare very favorably to those described for common diagnostic procedures such as receptor testing, Ki 67 immunohistochemistry for assessment of proliferation index, or inter-observer measurements of tumor diameter using imaging methods. For example, concordance studies for estrogen-receptor protein testing by immunohistochemistry have shown an overall concordance rate of 87, 90, and 97% between primary institution and central testing.[15–17] Similarly, in an international reproducibility study assessment of proliferation using Ki 67 revealed an intra-laboratory reproducibility of 94% and an inter-laboratory reproducibility of 59–71% (central and local staining, respectively).[18] In another study assessment of the inter-observer variability among pathologists showed that agreement was 89% for evaluation of Ki 67 in predicting response to neoadjuvant chemotherapy.[19] Also, inter-observer agreement among radiologists evaluating conventionally the largest tumor diameter after neoadjuvant chemotherapy using MRI showed a concordance correlation coefficient of 93%.[20]

In conclusion, there was strong reproducibility of measurements and similar long-term prognostic meaning of RCB when evaluated by five pathologists. Minor imprecision in scoring had a larger effect on the assignment to RCB categories, but those differences had little prognostic relevance. This study demonstrates that it would be feasible and reasonable to retrospectively evaluate RCB from the slides and reports from subjects in a clinical trial. However, we predict that incorporation of a standardized protocol, with formal training of site pathologists to prospectively evaluate RCB, would produce even stronger results (because assessment of tumor bed

area and cellularity would be improved), such that inter-pathologist reproducibility would be excellent and prognostic meaning of the RCB results would be even better than we report here.

Acknowledgments

Disclosure/conflict of interest

Dr Symmans has filed Residual Cancer Burden (RCB) as intellectual property (Nuvera Biosciences), patenting the RCB equation. (The RCB calculator is freely available on the worldwide web.) Dr Symmans reports current stock in Nuvera Biosciences and past stock in Amgen. The remaining authors declare no conflict of interest.

References

1 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: 2672–2685.

2 Kuerer HM, Neuman LA, Smith TL *et al.* Clinical course of breast cancer patients with complete pathologic primary tumor and axillary lymph node response to doxorubicin-based neoadjuvant chemotherapy. J Clin Oncol 1999;17:460–469.

3 Cortazar P, Zhang L, Untch M *et al.* Pathological complete response and long-term clinical benefit in breast cancer: the CTneoBC pooled analysis. The Lancet 2014;384:164–172.

4 Provenzano E, Vallier AL, Champ R *et al.* A central review of histopathology reports after breast cancer neoadjuvant chemotherapy in the neo-tango trial. Br J Cancer 2013;108:866–872.

5 Symmans WF, Peintinger F, Hatzis C *et al.* Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. J Clin Oncol 2007;25:4414–4422.

6 The University of Texas MD Anderson Cancer Center, Residual Cancer Burden Calculator. http://www3. mdanderson.org/app/medcalc/index.cfm?pagename = jsconvert3 2014.

7 Symmans WF, Wei C, Gould R *et al.* Long-term prognostic value of residual cancer burden (RCB) classification following neoadjuvant chemotherapy. Cancer Res 2013;73, S6-02.

8 Abrial C, Thivat E, Tacca O *et al.* Measurement of residual disease after neoadjuvant chemotherapy. J Clin Oncol 2008;26:3094.

9 Green MC, Buzdar AU, Smith T *et al.* Weekly paclitaxel improves pathologic complete remission in operable breast cancer when compared with paclitaxel once every 3 weeks. J Clin Oncol 2005;23:5983–5992.

10 Barnhart HX, Haber M, Song J. Overall concordance correlation coefficient for evaluating agreement among multiple observers. Biometrics 2002;58:1020–1027.

11 R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, AustriaVienna, AustriaURL http://www.R-project.org/ 2014.

12 Therneau T, A Package for Survival Analysis in S. R package version 2.37-7, URL http://CRAN.R-project. org/package = survival/ 2014.

13 Detailed Pathology Methods for Using Residual Cancer Burden http://www.mdanderson.org/education-and-research/resources-for-professionals/clinical-tools-and-resources/clinical-calculators/calculators-rcb-pathology-protocol2.pdf MD Anderson Cancer Center, Houston, TX, USA.

14 Boughey JC, Donohue JH, Jakub JW *et al.* Number of lymph nodes identified at axillary dissection: effect of neoadjuvant chemotherapy and other factors. Cancer 2010;116:3322–3329.

15 Gelber RD, Gelber S. International Breast Cancer Study Group (IBCSG) and the Breast International Group (BIG): facilitating consensus by examining patterns of treatment effects. Breast 2009;18:S2–S8.

16 Badve SS, Baehner FL, Gray RP *et al.* Estrogen- and progesterone-receptor status in ECOG 2197: comparison of immunohistochemistry by local and central laboratories and quantitative reverse transcription polymerase chain reaction by central laboratory. J Clin Oncol 2008;26:2473–2481.

17 Viale G, Regan MM, Maiorano E *et al.* Prognostic and predictive value of centrally reviewed expression of estrogen and progesterone receptors in a randomized trial comparing letrozole and tamoxifen adjuvant therapy for postmenopausal early breast cancer: BIG 1-98. J Clin Oncol 2007;25:3846–3852.

18 Polley MY, Leung SC, McShane LM *et al.* An international Ki67 reproducibility study. J Natl Cancer Inst 2013;105:1897–1906.

19 Manucha V, Zhang X, Thomas RM. The satisfactory reproducibility of the Ki-67 index in breast carcinoma, and it's correlation with the recurrence score. Clin Cancer Investig J 2014;3:310–314.

20 Takeda K1, Kanao S, Okada T *et al.* Assessment of CAD-generated tumor volumes measured using MRI in breast cancers before and after neoadjuvant chemotherapy. Eur J Radiol 2012;81:2627–2631.