



Evaluation of pathological complete response as surrogate endpoint in neoadjuvant randomised clinical trials of early stage breast cancer: systematic review and meta-analysis

Fabio Conforti,¹ Laura Pala,¹ Isabella Sala,² Chiara Oriecuia,³ Tommaso De Pas,¹ Claudia Specchia,⁴ Rossella Graffeo,⁵ Eleonora Pagan,² Paola Queirolo,¹ Elisabetta Pennacchioli,¹ Marco Colleoni,⁶ Giuseppe Viale,^{7,8} Vincenzo Bagnardi,² Richard D Gelber⁹

For numbered affiliations see end of the article

Correspondence to: F Conforti fabio.conforti@ieo.it (or @ieoufficiale and @fabioconforti1 on Twitter; ORCID 0000-0002-6233-0502)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2021;375:e066381 <http://dx.doi.org/10.1136/bmj-2021-066381>

Accepted: 16 November 2021

ABSTRACT OBJECTIVE

To evaluate pathological complete response as a surrogate endpoint for disease-free survival and overall survival in regulatory neoadjuvant trials of early stage breast cancer.

DESIGN

Systematic review and meta-analysis.

DATA SOURCES

Medline, Embase, and Scopus to 1 December 2020.

ELIGIBILITY CRITERIA FOR STUDY SELECTION

Randomised clinical trials that tested neoadjuvant chemotherapy given alone or combined with other treatments, including anti-human epidermal growth factor 2 (anti-HER2) drugs, targeted treatments, antivasular agents, bisphosphonates, and immune checkpoint inhibitors.

DATA EXTRACTION AND SYNTHESIS

Trial level associations between the surrogate endpoint pathological complete response and disease-free survival and overall survival.

METHODS

A weighted regression analysis was performed on log transformed treatment effect estimates (hazard ratio for disease-free survival and overall survival and relative risk for pathological complete response), and the coefficient of determination (R^2) was used to quantify the association. The secondary objective was to explore heterogeneity of results in preplanned subgroups analysis, stratifying trials according treatment type in the experimental arm, definition

used for pathological complete response (breast and lymph nodes v breast only), and biological features of the disease (HER2 positive or triple negative breast cancer). The surrogate threshold effect was also evaluated, indicating the minimum value of the relative risk for pathological complete response necessary to confidently predict a non-null effect on hazard ratio for disease-free survival or overall survival.

RESULTS

54 randomised clinical trials comprising a total of 32 611 patients were included in the analysis. A weak association was observed between the log(relative risk) for pathological complete response and log(hazard ratio) for both disease-free survival ($R^2=0.14$, 95% confidence interval 0.00 to 0.29) and overall survival ($R^2=0.08$, 0.00 to 0.22). Similar results were found across all subgroups evaluated, independently of the definition used for pathological complete response, treatment type in the experimental arm, and biological features of the disease. The surrogate threshold effect was 5.19 for disease-free survival but was not estimable for overall survival. Consistent results were confirmed in three sensitivity analyses: excluding small trials (<200 patients enrolled), excluding trials with short median follow-up (<24 months), and replacing the relative risk for pathological complete response with the absolute difference of pathological complete response rates between treatment arms.

CONCLUSION

A lack of surrogacy of pathological complete response was identified at trial level for both disease-free survival and overall survival. The findings suggest that pathological complete response should not be used as primary endpoint in regulatory neoadjuvant trials of early stage breast cancer.

Introduction

The US Food and Drug Administration and European Medicines Agency support the use of pathological complete response in neoadjuvant randomised clinical trials of early stage breast cancer as a surrogate endpoint for long term patients' clinical outcome (event-free or disease-free survival and overall survival), in the accelerated approval process of new drugs; and the current FDA table of surrogate endpoints includes pathological complete response for breast cancer.^{1 2} This decision addressed the need to expedite drug approvals, allowing patients to have access to effective treatments faster, more efficiently,

WHAT IS ALREADY KNOWN ON THIS TOPIC

Pathological complete response is a US Food and Drug Administration approved surrogate endpoint for disease-free survival and overall survival in randomised clinical trials testing neoadjuvant treatments in early stage breast cancer

Previous meta-analyses including a limited number of trials showed a strong correlation between pathological complete response and disease-free survival and overall survival at patient level but not trial level

The surrogacy value of pathological complete response is controversial

WHAT THIS STUDY ADDS

This meta-analysis showed a weak association between pathological complete response and disease-free survival and overall survival at trial level

The findings suggest that pathological complete response should not be used as a surrogate endpoint in regulatory neoadjuvant randomised clinical trials of early stage breast cancer

Better surrogate endpoints are needed

and more economically than waiting for the final results of adjuvant or neoadjuvant randomised clinical trials.^{1,2}

Evidence supporting the decision of regulatory agencies was mainly derived from an FDA sponsored meta-analysis of individual patient data from 12 randomised controlled trials.³ This analysis robustly showed a strong correlation between pathological complete response and both disease-free survival and overall survival at patient level, but it failed to show a statistically significant association at trial level.³

Buyse et al proposed that “a good surrogate endpoint must be shown to be causally linked to the true endpoint” and “to capture the whole effect of treatment upon the true endpoint.”^{4,5} Operationally this means that a good surrogate endpoint should fulfil the condition of a meaningful association with the true endpoint at both patient and trial level.^{4,5} A strong association at patient level indicates that the surrogate and true endpoint are likely causally linked, whereas a strong association at trial level indicates that the surrogate captures a large proportion of the treatment effect on the true endpoint.^{4,5} The trial level association between endpoints, however, does not simply follow from the patient level association.^{4,5}

Several surrogate endpoints in oncology show a statistically significant association with patients' overall survival at both individual and trial level, such as disease-free survival for human epidermal growth factor 2 (HER2) positive early stage breast cancer or disease-free survival and progression-free survival for early and advanced stage colorectal cancer, respectively.^{6,7} Berry and Hudis suggested that the absence of pathological complete response surrogacy at trial level observed in the FDA meta-analysis for breast cancer could be potentially explained by the limited number of trials analysed, especially given the little spread of treatment effects across the trials (ie, narrow range of the pathological complete response odd ratios and disease-free survival and overall survival hazard ratios reported in the 12 trials included in the analysis).⁸ Furthermore, the power of the surrogacy analysis could be affected by the use of hazard ratio as a measure of the effect of treatments on disease-free survival and overall survival. Whereas the hazard ratio is the gold standard measure for treatment effect in adjuvant and neoadjuvant trials of breast cancer, it could be affected by the loss of the proportional hazards assumption. To account for the potential loss of power owing to non-proportionality, surrogacy analysis should be performed with many trials.

Berruti et al's subsequent meta-analysis of aggregate data from 29 randomised controlled trials also failed to show a statistically significant surrogacy for pathological complete response at trial level.⁹ This analysis, however, had several limitations—one of the most important being that the potential heterogeneity of results according to the biological features of disease has not been evaluated.⁹ Another important limitation shared by the FDA sponsored analysis and that of Berruti et al is that the analyses only included

randomised clinical trials that tested chemotherapy, and, in most cases old regimens, with the exception of only two trials that tested anti-HER2 targeted treatment.^{3,9}

The FDA guidance for the use of complete pathological response as an endpoint to support accelerated approval has highlighted all such limitations affecting the results of the previous analyses and recognised the important value of further analyses to overcome these limitations.¹ The controversy about the surrogacy value of pathological complete response was also discussed at the St Gallen International Breast Cancer Consensus Conference (Vienna, 2021), where only 40% of panellists supported its use as an appropriate endpoint for defining standard adjuvant or neoadjuvant systemic regimens to treat early stage breast cancer.

Because a larger number of trials is now available, we performed a meta-analysis of all the randomised clinical trials that tested neoadjuvant treatments for early stage breast cancer, to assess the utility of pathological complete response as a surrogate for long term patients' outcome at trial level.

Methods

In this study we followed the Preferred Reporting Items for Systematic Reviews and Meta analyses (PRISMA) and Reporting of Surrogate Endpoint Evaluation using Meta analyses (ReSEEM) guidelines.¹⁰ We systematically searched PubMed, Medline, Embase, and Scopus to 1 December 2020 for all randomised clinical trials that tested neoadjuvant chemotherapy given alone or combined with other treatments. The search terms were “breast cancer”, “neoadjuvant therapy”, “preoperative therapy”, and “pathologic complete response”.

Trials were considered eligible for inclusion if they were randomised clinical trials that tested chemotherapy administered alone or in combination with other treatments in a neoadjuvant setting; they contained data on pathological complete response rates and survival outcomes (ie, a combination of disease-free survival, event-free survival, relapse-free survival, and overall survival) in the different treatment arms; and an explicit definition of pathological complete response was reported and based on excision histology.

Any trial in which additional post-surgical adjuvant treatments were delivered were considered to be eligible if all participants received the same treatment. We excluded neoadjuvant trials that tested endocrine treatment because of the low associated rate of pathological complete response. Two investigators (FC and LP) independently reviewed the list of retrieved articles for relevancy, and two investigators (IS and CO) independently extracted data from the studies, with discrepancies resolved by consensus with all investigators. Data were extracted on study design, number of patients enrolled, type of treatment, pathological complete response rate, definition of pathological complete response, number of disease-

free survival and overall survival events, and duration of follow-up. The Cochrane Collaboration's risk of bias tool was used to determine study methodological quality.¹¹

The primary objective was to assess the trial level association between pathological complete response as the surrogate endpoint and long term outcome in patients (ie, disease-free survival, overall survival, or both). The secondary objective was to explore heterogeneity of results according to the type of treatment in the experimental arm, the definition of pathological complete response (breast and lymph nodes *v* breast only), and biological features of the disease (HER2 positive and triple negative breast cancer).

We used the classification reported in the original paper to define treatment arms as experimental or control for each trial. In all analyses we used the endpoints for long term outcome in patients, as provided by the trial investigators. Because the endpoints definition was not standardised in most of the neoadjuvant trials, we considered several time-to-event endpoints to be equivalent to disease-free survival: relapse-free survival, event-free survival, and progression-free survival. In the case of studies reporting results for more than one time-to-event endpoint, we selected only one for the regression analysis using the following hierarchical order: disease-free survival, event-free survival, relapse-free survival, and progression-free survival. The hazard ratio of disease-free survival (or equivalent endpoint) and overall survival between the experimental arm and the control arm was used as the treatment effect estimate for the long term patients' clinical outcome (the true endpoint).

From each trial we extracted the proportion of patients with a pathological complete response per treatment arm as the surrogate endpoint for the analysis. The relative risk of pathological complete response between the experimental and control arm was used as the treatment effect estimate on the surrogate outcome. We recognised specific definitions of what constitutes a pathological complete response: ypT0-ypN0 indicates absence of invasive and intraductal disease in breast and nodes; ypT0/is-ypN0 indicates absence of invasive disease in breast and nodes; ypT0-ypN0/+ indicates absence of invasive and intraductal disease in breast, irrespective of nodes; and ypT0/is-ypN0/+ indicates absence of invasive disease in breast, irrespective of nodes. When studies used more than one definition to report the rates of pathological complete response, we recorded all information and selected the appropriate definition of the primary analysis using the hierarchical order: ypT0-ypN0, ypT0/is-ypN0 (in both cases the endpoint applies to breast and lymph nodes), ypT0-ypN0/+, and ypT0/is-ypN0/+ (in both cases the endpoint applies to breast only).

Finally, we performed three sensitivity analyses: in the first we excluded small trials enrolling fewer than 200 patients, in the second we excluded trials

with a median follow-up shorter than 24 months, and in the third we used the absolute difference of pathological complete response rates between control and experimental arm (rate in experimental arm–rate in control arm) instead of relative risk for pathological complete response as an estimate of treatment effects on the surrogate endpoint.

Statistical analysis

We used a correlation approach to assess surrogacy as previously described.^{4 5 12} To quantify the association between the effect of treatment on the reference endpoints of disease-free survival and overall survival and the effect of treatment on the surrogate endpoint of pathological complete response, we used a weighted linear regression model. From each trial we extracted treatment effects, expressed as hazard ratios for disease-free survival and overall survival and relative risks for pathological complete response, from each trial and considered these on a log scale in the model. Weights were defined as the number of disease-free survival and overall survival events reported or derived from each trial. In addition, as sensitivity analysis we also evaluated two different weighting systems based on the inverse of the variance of the log of the pathological complete response relative risk and on the trial sample size.

The coefficient of determination (R^2) was used to measure the variation of the weighted treatment effects explained by the model and to quantify the surrogacy level of pathological complete response. We used the TrialLevelMA function of the R package Surrogate to calculate R^2 and associated 95% confidence intervals.¹³ According to ReSEEM (Systematic Review and Recommendation for Reporting of Surrogate Endpoint Evaluation using Meta-analyses) guidelines, R^2 values ≥ 0.7 represent strong correlations (and thus suggest surrogacy), values between 0.69 and 0.5 represent moderate correlations, and values < 0.5 represent weak correlations.¹⁰ The slope of the regression line was also reported as an alternative measure of surrogacy.

Leave-one-out cross validation was performed to validate results obtained in the main analysis. Each trial was left out once, and the surrogate model was built with the other trials; this model was then reapplied to the left out trial to predict the effect of treatment on the reference endpoints (disease-free survival or overall survival). The leave-one-out cross validated R^2 was calculated as the correlation between the individual predictions made by the model over all trials and the actual treatment effects.

To assess homogeneity of slopes according to the levels of a defined factor, we included the interaction term between log(relative risk) for pathological complete response and the defined factor in a multivariable meta-regression model and calculated the associated F statistic. Moreover, to adjust the R^2 for trial level covariates, we also fitted a multivariable weighted linear model, including the trial level covariates as adjustable variables. We report the adjusted R^2 —that is, the square of the partial

correlation coefficient obtained from the multivariable model.

Finally, we calculated the surrogate threshold effect, defined as the minimum relative risk of the pathological complete response necessary to predict a statistically significant disease-free survival or overall survival benefit in a future trial. The surrogate threshold effect was located as the intersection of the upper limit of the 95% prediction band and the horizontal line representing the predicted hazard ratio for disease-free survival or overall survival equal to 1 (null effect).¹⁴ The 95% prediction band was calculated from the weighted regression model used to derive the coefficient of determination R^2 and was based on the predicted weight assigned to the hazard ratio for a future trial. Because the regression model in the main analyses was weighted by the number of events, in the calculation of the prediction band and consequently in the identification of the surrogate threshold effect, we considered a future trial with expected number of events to be equal to the average number of events observed in the set of trials included in the model itself. As a sensitivity analysis the surrogate threshold effect was computed for different scenarios, varying the expected number of events for a future trial.

Patient and public involvement

Members of the study group have regular meetings with patient representatives about ongoing scientific projects and activities. During these meetings the project and its objectives are discussed, and we accepted the patients' suggestions, which were mainly focused on the need to make the final version of the paper as clear and less technical as possible, to widely disseminate the results given the relevant implications for research and clinical practice.

Results

Characteristics and quality assessment

Overall, 54 randomised clinical trials comprising a total of 32 611 patients were included in the analysis (supplementary fig S1 and table S1; references in table S1 are cited in the full paper only). Seven trials had three arms and three trials had four arms, for a total of 67 comparisons analysed.¹⁵⁻⁸²

The trials tested different regimens or schedule of neoadjuvant chemotherapy. Ten trials evaluated an anthracycline based regimen versus an anthracycline or taxane based chemotherapy or the two combined.^{15-25 82} Ten compared a dose dense or intensified chemotherapy regimen or both with a standard dose regimen.²⁶⁻³⁶ Six trials tested the addition of capecitabine,^{37-43 78} three of carboplatin,^{57-60 75} two of nab-paclitaxel, two of gemcitabine, and one of vinorelbine to a standard anthracycline or taxane based regimen or the two combined.^{61 62 71 77 78} Twelve trials tested the combination of chemotherapy with anti-HER2 targeted treatment,^{44-53 68 69 74 80 83} five with bevacizumab,^{57 70 72 78 81} two with anti-programmed death 1 (PD1) or anti-PDL1 drugs,^{63 64} two with everolimus,^{73 76} and one with zoledronic acid.⁶⁵

Forty studies applied a pathological complete response definition to breast and lymph nodes and 13 to breast only. Study specific pathological complete response rates ranged between 2% and 68%, relative risks for pathological complete response ranged between 0.52 and 3.0 (pooled relative risk 1.21, 95% confidence interval 1.15 to 1.27) and the hazard ratios for disease-free survival ranged between 0.26 and 2.61 (pooled hazard ratio 0.91, 95% confidence interval 0.85 to 0.96) and for overall survival ranged between 0.19 and 2.27 (pooled hazard ratio 0.89, 0.84 to 0.94; supplementary table S1).

The endpoint for time to recurrence was disease-free survival in 35 trials, event-free survival in nine trials, relapse-free survival in seven trials, and progression-free survival in three trials (supplementary table S2). The median follow-up across trials was 56 months (range 15.5-120 months).

Randomised treatment allocation sequences were generated in all trials. Six trials were double blinded. Supplementary table 3 lists the quality scores according to the risk of bias tool for each trial. No trial was scored as low quality.

Main analysis

The effects of breast cancer treatment on pathological complete response compared with on disease-free survival or overall survival was assessed and a regression equation was estimated based on data from all the trials included in the analysis. A weak association was found between the log(relative risk) for pathological complete response and the log(hazard ratio) for disease-free survival ($R^2=0.14$, 95% confidence interval 0.00 to 0.29), and the slope of the regression line was -0.27 (fig 1 and table 1). The corresponding association for overall survival was similarly weak ($R^2=0.08$, 0.00 to 0.22), and the slope of the regression line was -0.20 (fig 1 and table 1). After adjustment for trial level covariates, such as definition of pathological complete response, type of treatment, and size of trial, the R^2 values did not materially change (disease-free survival $R^2=0.11$, 0.00 to 0.25, overall survival $R^2=0.08$, 0.00 to 0.22).

The leave-one-out cross validation analysis confirmed that the surrogacy of pathological complete response was weak for both disease-free survival and overall survival: the leave one-out cross validated R^2 was 0.07 for disease-free survival and 0.02 for overall survival. The R^2 values obtained in the leave-one-out models ranged from 0.11 to 0.20 for disease-free survival and from 0.06 to 0.12 for overall survival (supplementary fig S2A and B).

Subgroup and sensitivity analyses

The surrogacy of pathological complete response was explored in preplanned analyses stratifying trials according to the type of treatment in the experimental arm, definition of pathological complete response, and biological features of the disease. The different systemic treatments administered in the experimental arm were classified according to five

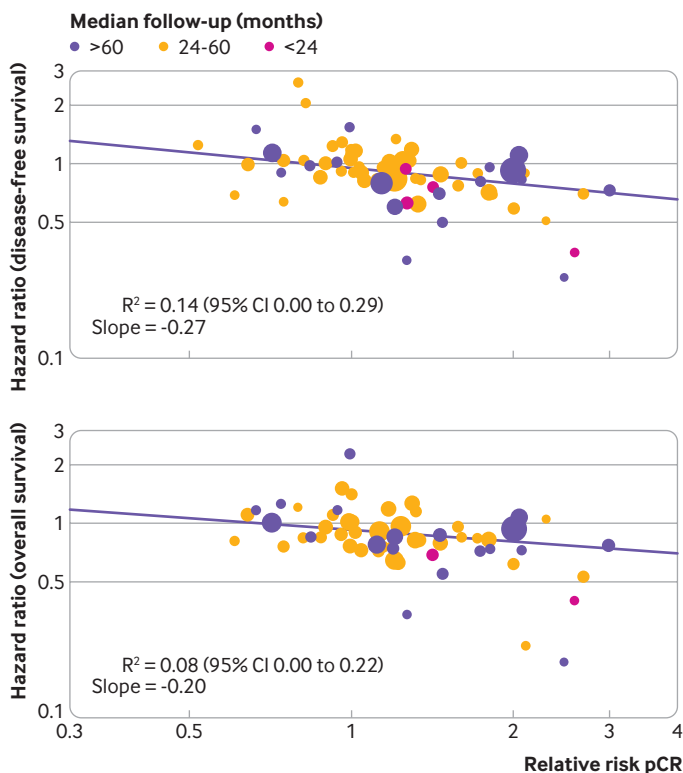


Fig 1 | Correlations between effects of breast cancer treatment on pathological complete response (pCR) and disease-free survival and overall survival. Each circle represents a trial, and the surface area of the circle is proportional to the number of events observed in the corresponding trial. Straight lines represent weighted regression lines

groups (supplementary table S1): regimens using anthracycline or taxane based chemotherapy, or both (10 trials; fig.2A and fig2D), dose dense or intensified chemotherapy regimens (10 trials; fig 2B and fig2E), regimens containing capecitabine (six trials; fig 2G and fig2J), chemotherapy in combination with anti-HER2 targeted treatments (12 trials; fig 2C and fig2F), chemotherapy in combination with bevacizumab (five trials; fig 2H and fig2K), and other treatments (17 trials; fig 2I and fig2L). The association between the log(relative risk) for pathological complete response and log(hazard ratio) for both disease-free survival and overall survival was weak in all the treatment subgroups explored (F test for homogeneity of slopes: $P=0.03$ for disease-free survival and $P=0.17$ for overall survival; table 1), with the only exception represented by the two subgroups of trials testing, respectively, the dose dense or intensified regimens, in which the association was moderate for disease-free survival ($R^2=0.62$, 95% confidence interval 0.18 to 1.00; fig 2B) but weak for overall survival ($R^2=0.45$, 0.00 to 0.99; fig 2E and table 1) and regimens containing capecitabine, in which the association was weak for disease-free survival ($R^2=0.21$, 0.00 to 1.00; fig 2G) and moderate for overall survival ($R^2=0.64$, 0.00 to 1.00; fig.2J and table 1).

Both definitions of pathological complete response showed a weak association with disease-free survival and with overall survival: R^2 for disease-free survival

and overall survival was, respectively, 0.02 (0.00 to 0.19) and 0.01 (0.00 to 0.12) for pathological complete response applied to breast only (fig S3A and C), and 0.15 (0.00 to 0.34) and 0.10 (0.00 to 0.28) for pathological complete response applied to breast and lymph nodes (fig S3B and D; F test for homogeneity of slopes: $P=0.23$ for disease-free survival and $P=0.35$ for overall survival; table 1).

The association between the log(relative risk) for pathological complete response and log(hazard ratio) for both disease-free survival and overall survival was weak in both triple negative and HER2 positive breast cancer: R^2 for disease-free survival and overall survival was, respectively, 0.42 (0.05 to 0.79; fig 3A) and 0.17 (0.00 to 0.55; fig 3C) for triple negative breast cancer, and 0.37 (0.05 to 0.69; fig 3B) and <0.01 (0.00 to 0.05; fig 3D) for HER2 positive disease (F test for homogeneity of slopes: $P=0.56$ for disease-free survival and $P=0.33$ for overall survival; table 1).

A post hoc analysis was also performed with trials stratified according to the type of time-to-event endpoint used (disease-free survival, event-free survival, or other endpoints, including relapse-free survival and progression-free survival; table 1): the R^2 for the association between the log(relative risk) for pathological complete response and log(hazard ratio) was 0.03 (0.00 to 0.15) for disease-free survival (supplementary fig S4A), 0.40 (0.00 to 0.85) for event-free survival (fig S4B), and 0.39 (0.00 to 0.82) for the other endpoints (fig S4C).

Sensitivity analyses were performed excluding small trials (11 trials with a sample size <200 patients; supplementary fig S5A and B and table 1); excluding trials with short median follow-up (five trials with median follow-up <24 months; table 1); using absolute difference of pathological complete response between treatment arms instead of the relative risk for pathological complete response (supplementary fig S6A and B and table S4, and table 1); using the sample size as weighting systems in the regression model, instead of number of disease-free survival and overall survival events (table 1); and using the inverse of the variance of the log of the pathological complete response relative risk as weighting systems in the regression model (table 1). The results of the sensitivity analyses were comparable to those of the main analysis.

Assessment of surrogate threshold effect of pathological complete response

The surrogate threshold effect was calculated, indicating the minimum relative risk of the pathological complete response necessary to confidently predict a non-null effect on hazard ratios for disease-free survival or overall survival in a future randomised trial. In this calculation, a future trial was considered to have an expected number of events equal to the average number of events observed in the main analysis including all the trials (131 disease-free survival events and 91 overall survival events). Because of the weak association observed between pathological complete

Table 1 | Results of main analyses, subgroup analyses, and sensitivity analyses

Subgroups: long term outcomes	No of comparisons analysed	R ² (95% CI)	Slope of regression line	STE	F test for homogeneity of slopes	
					Disease-free survival	Overall survival
Main analyses						
Disease-free survival	67	0.14 (0.00 to 0.29)	-0.27	5.19		
Overall survival	59	0.08 (0.00 to 0.22)	-0.20	-		
Subgroups analyses						
pCR definition:						
Breast only						
Disease-free survival	14	0.02 (0.00 to 0.19)	-0.08	-	0.23	0.35
Overall survival	14	0.01 (0.00 to 0.12)	-0.06	-		
Breast and lymph nodes:						
Disease-free survival	52	0.15 (0.00 to 0.34)	-0.32	4.44		
Overall survival	44	0.10 (0.00 to 0.28)	-0.26	-		
Treatment arms						
Antracycline and taxane based v antracycline based regimens:						
Disease-free survival	10	0.002 (0.00 to 0.06)	0.02	-		
Overall survival	9	0.01 (0.00 to 0.14)	-0.03	-		
Intensified/dose dense v standard dose regimens:						
Disease-free survival	10	0.62 (0.18 to 1.00)	-0.49	2.43		
Overall survival	10	0.45 (0.00 to 0.99)	-0.38	4.88		
Chemotherapy plus anti-HER2 targeted treatment:						
Disease-free survival	17	0.46 (0.08 to 0.85)	-0.80	2.74		
Overall survival	13	0.31 (0.00 to 0.78)	-0.69	-		
Capecitabine-containing v standard regimens						
Disease-free survival	6	0.21 (0.00 to 1.00)	-0.64	-	0.03	0.17
Overall survival	6	0.64 (0.00 to 1.00)	-1.05	-		
Chemotherapy plus bevacizumab:						
Disease-free survival	5	0.06 (0.00 to 0.72)	0.52	-		
Overall survival	5	0.10 (0.00 to 0.87)	1.00	-		
Other comparisons:						
Disease-free survival	19	0.02 (0.00 to 0.14)	-0.08	-		
Overall survival	16	0.02 (0.00 to 0.15)	-0.08	-		
Biological features of disease						
Triple negative:						
Disease-free survival	19	0.42 (0.05 to 0.79)	-0.63	2.24	0.56	0.33
Overall survival	16	0.17 (0.00 to 0.55)	-0.37	-		
HER2 positive:						
Disease-free survival	25	0.37 (0.05 to 0.69)	-0.80	2.43		
Overall survival	18	0.002 (0.00 to 0.05)	0.08	-		
Time-to-recurrence endpoint definition:						
Disease free survival						
Disease-free survival	38	0.03 (0.00 to 0.15)	-0.11	-		
Overall survival	37	0.02 (0.00 to 0.10)	-0.09	-		
Event-free survival						
Event-free survival	14	0.40 (0.00 to 0.85)	-0.86	2.05	0.03	0.34
Overall survival	10	0.25 (0.00 to 0.81)	-0.34	-		
Others						
Others	15	0.39 (0.00 to 0.82)	-0.44	4.39		
Overall survival	12	0.29 (0.00 to 0.79)	-0.38	-		
Sensitivity analyses						
Delta-pCR as surrogate endpoint:						
Disease-free survival	67	0.24 (0.06 to 0.42)	-1.83	0.22		
Overall survival	59	0.13 (0.00 to 0.30)	-1.38	0.31		
Randomised controlled trials with sample size <200 patients excluded:						
Disease-free survival	56	0.13 (0.00 to 0.30)	-0.26	-		
Overall survival	49	0.06 (0.00 to 0.20)	-0.18	-		
Trial sample size as weight in the regression model:						
Disease-free survival	67	0.14 (0.00 to 0.30)	-0.34	5.00		
Overall survival	59	0.09 (0.00 to 0.23)	-0.26	-		
Inverse of variance of log of pCR relative risk as weight in regression model:						
Disease-free survival	67	0.25 (0.07 to 0.44)	-0.69	2.32		
Overall survival	59	0.14 (0.00 to 0.31)	-0.43	3.29		
Randomised controlled trials with <24 months of follow-up excluded:						
Disease-free survival	63	0.13 (0.00 to 0.29)	-0.25	6.05		
Overall survival	57	0.07 (0.00 to 0.20)	-0.19	-		

STE=surrogate threshold effect; pCR=pathological complete response; anti-HER2=anti-human epidermal growth factor 2.

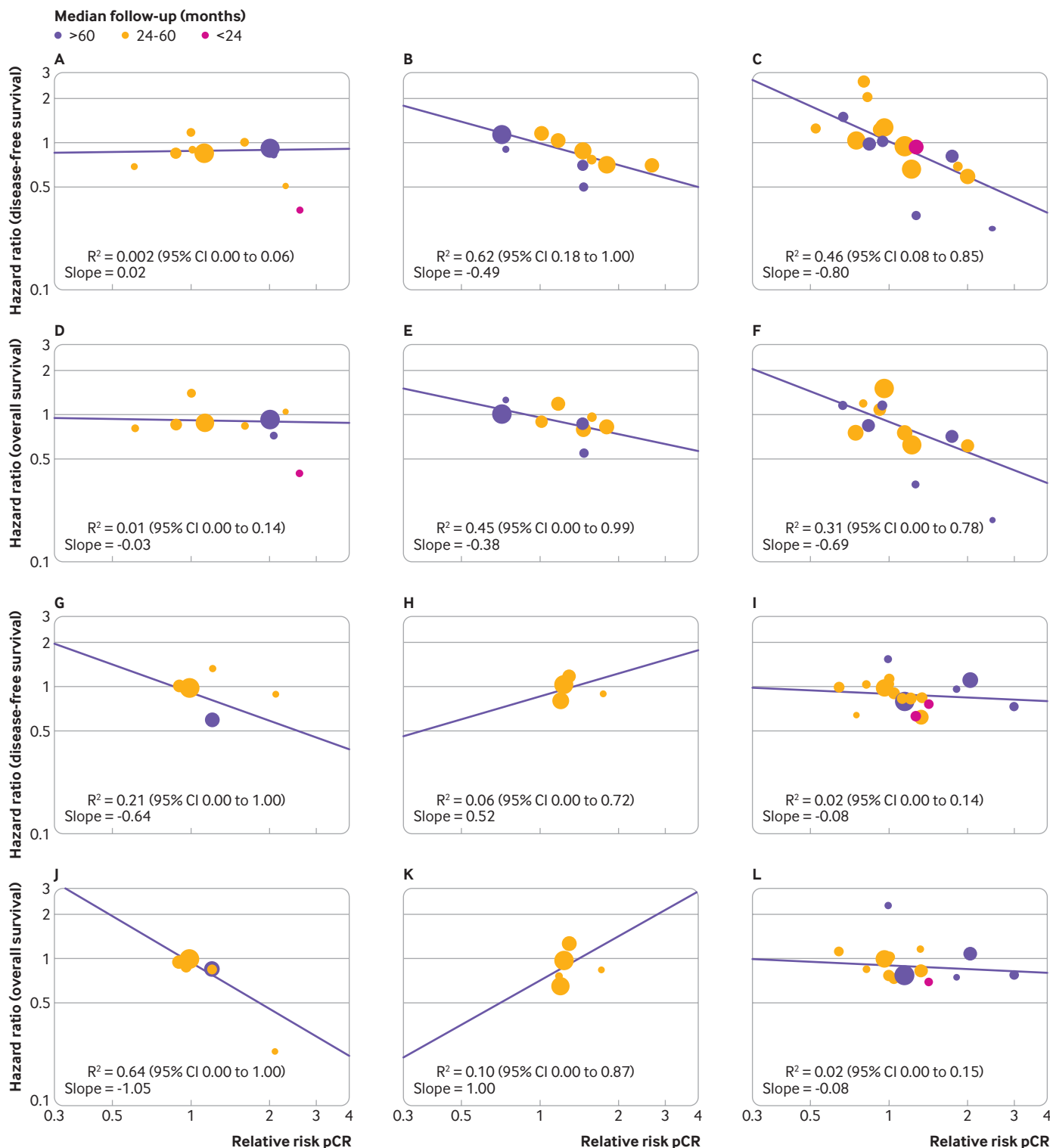


Fig 2 | Correlation between effects of breast cancer treatment on pathological complete response (pCR) and disease-free survival (panels A-C and G-I) and overall survival (panels D-F and J-L). A and D=Antracycline and taxane based v antracycline based regimens; B and E=dose dense or intensified v standard dose regimens; C and F=chemotherapy plus anti-human epidermal growth factor 2 targeted treatment; G and J=regimens containing capecitabine v standard regimens; H and K=chemotherapy plus bevacizumab; I and L=other treatments. Each circle represents a trial, and the surface area of the circle is proportional to the number of events observed in the corresponding trial. Straight lines represent weighted regression lines

response and disease-free survival and overall survival, the surrogate threshold effect was not estimable or was high for the main analysis including all the trials (fig 4A and fig 4B), as well as for all the subgroups explored

(supplementary figs S7-S10 and table 1). The surrogate threshold effect calculated for absolute difference of pathological complete response instead of the relative risk was, respectively, 0.22 for disease-free survival

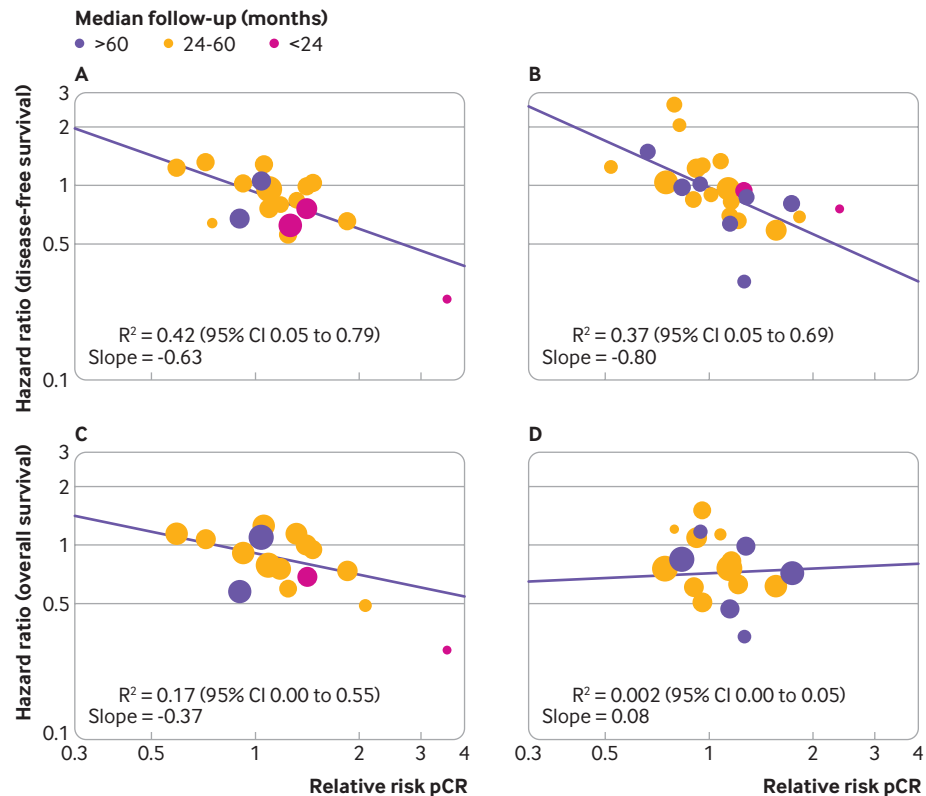


Fig 3 | Correlation between effects of breast cancer treatment on pathological complete response and disease-free survival (panels A and B) and overall survival (panels C and D). A and C=triple negative breast cancer; B and D=human epidermal growth factor 2 positive breast cancer. Each circle represents a trial, and the surface area of the circle is proportional to the number of events observed in the corresponding trial. Straight lines represent weighted regression lines

and 0.31 for overall survival (table 1). Supplementary table S5 and figures S11-S14 show the surrogate threshold effect obtained by varying the number of expected events in a future randomised controlled trial—for example, in a randomised controlled trial with, respectively, 800 disease-free survival or overall survival events, a relative risk greater than 1.67 for pathological complete response predicts significant gains in disease-free survival, whereas a relative risk greater than 1.51 predicts significant gains in overall survival.

Discussion

The findings from this meta-analysis do not support the use of pathological complete response as a surrogate endpoint for disease-free survival and overall survival in neoadjuvant trials of early stage breast cancer. We found that the coefficient of determination of the association between pathological complete response and overall survival was 0.08 (95% confidence interval 0.00 to 0.22), indicating that only 8% of the variability among treatment effects on overall survival is explained by the effects observed with pathological complete response. This coefficient was even lower when estimated using leave-one-out cross validation. Much has been discussed about when a surrogate endpoint could be theoretically considered validated, but the consensus is that a candidate surrogate endpoint would

be valid only if the coefficient of determination (R^2) is at least equal to or higher than 0.7.^{4,5,10} Furthermore, our subgroup analysis confirmed that the weak association between pathological complete response and long term clinical outcomes was evident for all the subgroups explored, independently of the type of treatment, the definition of pathological complete response, and the biological features of the disease. Finally, results of the surrogate threshold effect analysis suggested that a statistically significant effect on overall survival could be confidently predicted only if a very high relative risk for pathological complete response was observed.

Several explanations might account for the lack of pathological complete response surrogacy at trial level. One hypothesis is that pathological complete response measures the effect of a treatment only on the primary tumour and not on micrometastatic systemic disease, which is the main target of adjuvant and neoadjuvant treatments. The surrogacy assumption is that responses of primary tumours and micrometastases are comparable, but the validity of such an assumption could be affected by the disease itself and type of neoadjuvant treatments.^{84,85} In our opinion, the strong association observed between pathological complete response and long term outcomes at patient level in early stage breast cancer, and the excellent prognosis of patients achieving a pathological complete response, do not support such an hypothesis.³

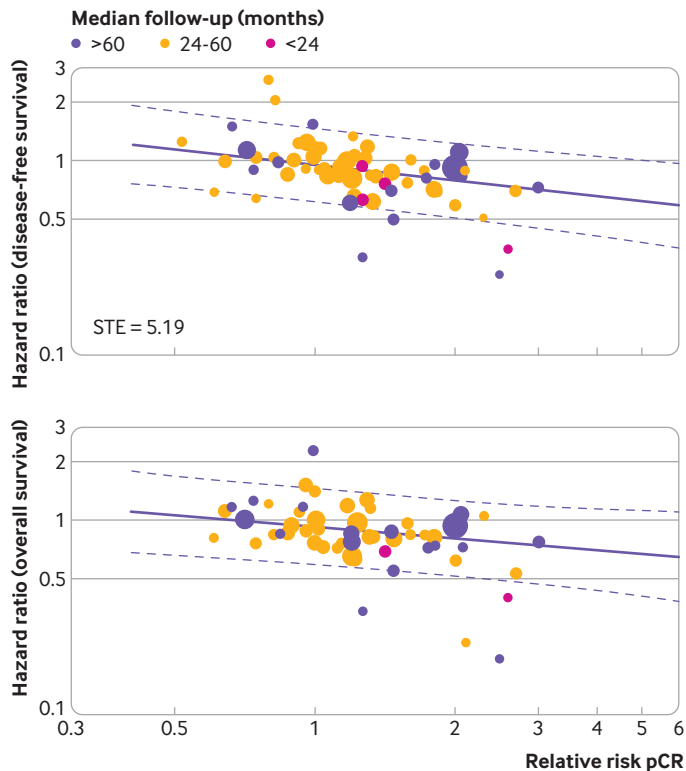


Fig 4 | Surrogate threshold effect (STE) for disease-free survival and overall survival. Each circle represents a trial, and the surface area of the circle is proportional to the number of events observed in the corresponding trial. Straight lines represent weighted regression lines and dashed lines represent 95% prediction bands (prediction bands were based on the values predicted by the weighted regression model). In the predictions, the median weight was considered. STE, when definable, is represented by the intersection point between the horizontal line $y=1$ and upper 95% prediction band

Another potential explanation is that patients who do not achieve a pathological complete response might not be disadvantaged, as shown by those with endocrine responsive breast cancers who derive important survival benefit from endocrine treatments but rarely obtain a pathological complete response.⁸⁶ In fact, several more granular definitions of pathological response that could capture treatment effects better than pathological complete response have been proposed as surrogate endpoints, such as residual cancer burden in breast cancer or major pathological response (<10% vital tumour cells) in lung cancer and melanoma. Although a strong association between such surrogate endpoints and long term outcomes has been found at patient level, no evidence has been provided yet on their surrogacy value at trial level.^{87 88}

Another explanation could be that a surrogate endpoint that exclusively relies on comparing pathological complete response rates between treatment arms overlooks relevant information from most of the other patients who do not achieve a pathological complete response and who might experience a large spectrum of responses, including primary resistance and disease progression during neoadjuvant treatment. Such a broad spectrum of responses might not be equally distributed between

treatment arms, affecting the overall prognosis of the population more than pathological complete response rates. This is the scenario described in the study by Fleming et al, in which false negative and false positive conclusions about clinical efficacy of a new intervention compared with standard treatment could arise if a surrogate endpoint only captures the effects of interventions on one causal pathway of the disease process (ie, a substantial reduction of relapses in patients achieving a pathological complete response), while the interventions also have an impact on other principal causal pathways (ie, the ability of treatments to modify the clinical course of disease and thus the risk of relapse independently of achieving a pathological complete response).⁸⁹ This could explain results such as those observed in the large GeparTrio trial, which found no difference in pathological complete response rates between treatment arms but reported a survival advantage for the experimental treatments.⁴⁰ A composite surrogate endpoint that takes into account differences between arms not only in pathological complete response rates but also in the rate of the other types of response, including progression of disease, might have greater surrogacy value at trial level.^{4 5} In a recent retrospective analysis of 938 women treated in the neoadjuvant I-SPY2 trial, patients' event-free survival was found to worsen significantly for each unit of residual cancer burden, regardless of tumour subtype and type of neoadjuvant treatment. Comparing distributions of residual cancer burden as a continuous measure of response obtained by treatment arms in randomised controlled trials, would probably provide additional information beyond pathological complete response rate and would better capture the effect of treatments on long term patients' clinical outcomes.⁸³ Furthermore, recently, meta-analytical methods allowing for use of multiple surrogate endpoints jointly have been proposed with the potential benefit of reducing the uncertainty around predictions.⁹⁰

All such hypotheses to explain lack of trial level surrogacy in the presence of strong patient level surrogacy are speculative and remain to be shown. Moreover, this discrepancy can simply occur because of causal inference mechanisms, such as the confounding effect by known and unknown prognostic factors that have a similar influence on both the surrogate and the final endpoints creating a correlation between them at individual level, even when the association is weak at trial level.^{4 5}

Policy implications

To date, the FDA has approved two drugs in the neoadjuvant setting for breast cancer under the accelerated approval pathway, based on results on surrogate endpoints: pertuzumab for HER2 positive disease and pembrolizumab for triple negative breast cancer. Although the follow-up for overall survival of the Keynote-522 trial, leading to accelerated approval of pembrolizumab, is too short to draw any conclusions, the discrepancy observed in the Adjuvant

Pertuzumab and Herceptin IN Initial TherapY in Breast Cancer (APHINITY) trial between the statistically significant and large improvement of pathological complete response rate and the lack of evidence of survival benefit for patients treated with pertuzumab, pointed to the risk of using pathological complete response as a surrogate endpoint. In oncology, many drugs were originally approved on the basis of substantial improvement of a supposed—but actually not fully shown—surrogate endpoint, which in later studies failed to show evidence of survival benefit, such as bevacizumab for breast cancer, olaratumab for sarcoma, and atezolizumab for urothelial carcinoma.^{91 92} These and numerous other examples suggest a fundamental flaw in the use of surrogate endpoints for drug approvals and the need for rigorous evidence of the surrogacy value of drugs before use.^{91 92} Despite the caveats, the reliance of regulatory agencies on surrogate endpoints for drug approval has increased considerably in recent years.^{91 92}

The lack of surrogacy at trial level showed here, substantially limits the possibility of using pathological complete response to confidently anticipate the results of randomised controlled trials and to predict long term outcome of the populations enrolled and thus to support accelerated drugs approval. However, all this does not undermine the value of pathological complete response when used for other reasons, as well as the importance of neoadjuvant trials.⁹³ Indeed, given the strong association between pathological complete response and overall survival shown at patient level, pathological complete response represents the best biomarker available to predict patients' residual risk of relapse after neoadjuvant therapy and has utility in identifying those at substantial risk who require escalation of adjuvant therapy, as shown for HER2 positive disease in the KATHERYNE trial and for triple negative breast cancer in the Capecitabine for Residual Cancer as Adjuvant Therapy (CREATE-X) trial.^{94 95}

Limitations of this study

Our study has several limitations. Our analysis is based on aggregate data from trials, and not on individual patient data (IPD). IPD analyses allow for checking the plausibility of randomisation sequences, verifying data integrity and consistency, fitting bivariate and copula based models that are among the preferred methods of assessment of trial level associations, adjusting the analyses for baseline prognostic covariates, and taking into account the fact that each within trial surrogate outcome is estimated with error. Nevertheless, the specific aim of our analysis was to assess surrogacy at trial level, and we used only data from randomised clinical trials of high quality, making it unlikely that an IPD analysis would substantially change our conclusions.⁹⁶⁻⁹⁸ We also did not explore potential differences of the pathological complete response surrogacy value within the subgroups of HER2 positive disease defined by hormone receptor status. Finally, the terminology of time to recurrence endpoints used across trials is heterogenous. However, in many

cases—particularly in the earliest trials—the definition provided by authors in the original papers for both disease-free survival and progression-free survival endpoints substantially resembled the FDA definition of event-free survival.^{1 50} An IPD meta-analysis of a large number of randomised controlled trials to assess the surrogacy value of pathological complete response as well as of other intermediate endpoints in adjuvant and neoadjuvant trials would complement our analyses, and we hope that the Early Breast Cancer Trialists' Collaborative Group might support such an analysis in the future.

Conclusions

Our meta-analysis found lack of surrogacy of pathological complete response for long term patients' outcome at trial level. Although this finding does not affect the role of pathological complete response to estimate patients' residual risk of relapse after neoadjuvant treatment and to identify those patients who are candidates for further adjuvant treatments, use of pathological complete response to predict long term outcomes of patient populations enrolled in neoadjuvant randomised clinical trials is questionable. For this reason, we suggest that pathological complete response should not be used as a primary endpoint in regulatory neoadjuvant trials in early stage breast cancer.

AUTHOR AFFILIATIONS

¹Division of Melanoma, Sarcomas and Rare Tumors, IEO, European Institute of Oncology, IRCCS, Milan, Italy

²Department of Statistics and Quantitative Methods, University of Milan-Bicocca, Milan, Italy

³Department of Clinical and Experimental Sciences, University of Brescia, Brescia, Italy

⁴Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy

⁵Breast Unit of Southern Switzerland, Oncology Institute of Southern Switzerland, Bellinzona, Switzerland

⁶Division of Medical Senology, IEO, European Institute of Oncology IRCCS, Milan, Italy

⁷Department of Pathology, IEO, European Institute of Oncology, IRCCS, Milan, Italy

⁸University of Milan, Milan, Italy

⁹Medical School, Harvard T H Chan School of Public Health, and Frontier Science and Technology Research Foundation, Boston, MA, USA

FC and LP thank Aron Goldhirsch for his mentorship.

Contributors: FC and LP are joint first authors. VB and RDG are joint last authors. FC, LP, VB, and RDG conceived, designed, planned, and managed the study, acquired data, interpreted the results, drafted the manuscript, and critically reviewed or revised the manuscript for important intellectual content. VB, FC, LP, IS, CO, and CS managed the study, acquired data, and performed the statistical analyses. All other authors supervised the data analysis, provided the interpretation of results, and contributed to the drafting and critical review of the manuscript. All authors approved the final draft. FC is the guarantor. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: None received.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the

submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Data sharing: Detailed data on the included studies are available on reasonable request to the corresponding author.

The lead author (FC) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Dissemination to participants and related patient and public communities: We plan to disseminate the findings and conclusions from this study through a lay language summary of our findings (see supplementary file), which will be widely promoted by our respective institutions (the European Institute of Oncology, Bicocca University of Milan, and Harvard University) through press releases, social media (such as Twitter), and the websites of our institutions. We also plan to present the results of our study at international scientific conferences.

Provenance and peer review: Not commissioned; externally peer reviewed.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- US Department of Health and Human Services. US Food and Drug Administration, Center for Drug Evaluation and Research (CDER): Guidance for Industry: Pathological Complete Response in Neoadjuvant Treatment of High-Risk Early-Stage Breast Cancer—Use as an Endpoint to Support Accelerated Approval. www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm305501.pdf
- European Medicines Agency. EMA/CHMP/ 151853/2014: Draft guideline on the role of the pathological complete response as an endpoint in neoadjuvant breast cancer studies. www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/04/WC500165781.pdf
- Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet* 2014;384:164-72. doi: 10.1016/S0140-6736(13)62422-8
- Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* 1998;54:1014-29. doi: 10.2307/2533853
- Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 2000;1:49-67. doi: 10.1093/biostatistics/1.1.49
- Sargent DJ, Wieand HS, Haller DG, et al. Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *J Clin Oncol* 2005;23:8664-70. doi: 10.1200/JCO.2005.01.6071
- Saad ED, Squifflet P, Burzykowski T, et al. Disease-free survival as a surrogate for overall survival in patients with HER2-positive, early breast cancer in trials of adjuvant trastuzumab for up to 1 year: a systematic review and meta-analysis. *Lancet Oncol* 2019;20:361-70. doi: 10.1016/S1470-2045(18)30750-2
- Berry DA, Hudis CA. Neoadjuvant therapy in breast cancer as a basis for drug approval. *JAMA Oncol* 2015;1:875-6. doi: 10.1001/jamaoncol.2015.1293
- Berruti A, Amoroso V, Gallo F, et al. Pathologic complete response as a potential surrogate for the clinical outcome in patients with breast cancer after neoadjuvant therapy: a meta-regression of 29 randomized prospective studies. *J Clin Oncol* 2014;32:3883-91. doi: 10.1200/JCO.2014.55.2836
- Xie W, Halabi S, Tierney JF et al, A systematic review and Recommendation for reporting of surrogate endpoint evaluation using meta-analyses JNCI CS 2019; 3:pkz002.
- Higgins JPT, Altman DG. Assessing risk of bias in included studies. In: Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions*. Wiley, 2008: 187-241. doi: 10.1002/9780470712184.ch8
- Burzykowski T, Molenberghs G, Buyse M, Geys H, Renard D. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. *J R Stat Soc Ser C Appl Stat* 2001;50:405-22. doi: 10.1111/1467-9876.00244
- Der Elst WW, Meyvisch P, Alonso A, Ensor HM, Molenberghs CJWG. Surrogate: Evaluation of Surrogate Endpoints in Clinical Trials. R package version 1.8 2020 <https://CRAN.R-project.org/package=Surrogate>.
- Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat* 2006;5:173-86. doi: 10.1002/pst.207
- Buzdar AU, Singletary SE, Theriault RL, et al. Prospective evaluation of paclitaxel versus combination chemotherapy with fluorouracil, doxorubicin, and cyclophosphamide as neoadjuvant therapy in patients with operable breast cancer. *J Clin Oncol* 1999;17:3412-7. doi: 10.1200/JCO.1999.17.11.3412
- Diéras V, Fumoleau P, Romieu G, et al. Randomized parallel study of doxorubicin plus paclitaxel and doxorubicin plus cyclophosphamide as neoadjuvant treatment of patients with breast cancer. *J Clin Oncol* 2004;22:4958-65. doi: 10.1200/JCO.2004.02.122
- Smith IC, Heys SD, Hutcheon AW, et al. Neoadjuvant chemotherapy in breast cancer: significantly enhanced response with docetaxel. *J Clin Oncol* 2002;20:1456-66. doi: 10.1200/JCO.2002.20.6.1456
- Heys SD, Sarkar T, Hutcheon AW. Primary docetaxel chemotherapy in patients with breast cancer: impact on response and survival. *Breast Cancer Res Treat* 2005;90:169-85. doi: 10.1007/s10549-004-1001-0
- Evans TR, Yellowlees A, Foster E, et al. Phase III randomized trial of doxorubicin and docetaxel versus doxorubicin and cyclophosphamide as primary medical therapy in women with breast cancer: an anglo-celtic cooperative oncology group study. *J Clin Oncol* 2005;23:2988-95. doi: 10.1200/JCO.2005.06.156
- Mansi JL, Yellowlees A, Lipscombe J, et al. Five-year outcome for women randomised in a phase III trial comparing doxorubicin and cyclophosphamide with doxorubicin and docetaxel as primary medical therapy in early breast cancer: an Anglo-Celtic Cooperative Oncology Group study. *Breast Cancer Res Treat* 2010;122:787-94. doi: 10.1007/s10549-010-0989-6
- Rastogi P, Anderson SJ, Bear HD, et al. Preoperative chemotherapy: updates of National Surgical Adjuvant Breast and Bowel Project Protocols B-18 and B-27. *J Clin Oncol* 2008;26:778-85. doi: 10.1200/JCO.2007.15.0235
- Cortés-Flores AO, Morgan-Villela G, Castro-Cervantes JM, Vázquez-Camacho G, Fuentes-Orozco C, González-Ojeda A. [Neoadjuvant treatment for locally advanced breast cancer. Comparison of two schemes based on docetaxel-epirubicin vs. 5-fluorouracil-epirubicin-cyclophosphamide]. *Cir Cir* 2008;76:23-8.
- Bonnefoi H, Piccart M, Bogaerts J, et al. EORTC 10994/BIG 1-00 Study Investigators. TP53 status for prediction of sensitivity to taxane versus non-taxane neoadjuvant chemotherapy in breast cancer (EORTC 10994/BIG 1-00): a randomised phase 3 trial. *Lancet Oncol* 2011;12:527-39. doi: 10.1016/S1470-2045(11)70094-8
- Chua S, Smith IE, A'Hern RP, et al. TOPIC Trial Group. Neoadjuvant vinorelbine/epirubicin (VE) versus standard adriamycin/cyclophosphamide (AC) in operable breast cancer: analysis of response and tolerability in a randomised phase III trial (TOPIC 2). *Ann Oncol* 2005;16:1435-41. doi: 10.1093/annonc/mdi276
- Chen X, Ye G, Zhang C, et al. Significantly superior outcome after neoadjuvant chemotherapy with docetaxel, anthracycline and cyclophosphamide versus docetaxel plus cyclophosphamide: Results from the NATT trial in triple negative or HER2 positive breast cancer. *Ann Oncol* 2012;23. doi: 10.1016/S0923-7534(20)32909-4
- Baldini E, Gardin G, Giannessi PG, et al. Accelerated versus standard cyclophosphamide, epirubicin and 5-fluorouracil or cyclophosphamide, methotrexate and 5-fluorouracil: a randomized phase III trial in locally advanced breast cancer. *Ann Oncol* 2003;14:227-32. doi: 10.1093/annonc/mdg069
- Therasse P, Mauriac L, Welnicka-Jaskiewicz M, et al. EORTC. Final results of a randomized phase III trial comparing cyclophosphamide, epirubicin, and fluorouracil with a dose-intensified epirubicin and cyclophosphamide + filgrastim as neoadjuvant treatment in locally advanced breast cancer: an EORTC-NCIC-SAKK multicenter study. *J Clin Oncol* 2003;21:843-50. doi: 10.1200/JCO.2003.05.135
- Fraci G, D'Aiuto G, Comella P, et al. Southern Italy Cooperative Oncology Group (SICOG) Italy. Preoperative weekly cisplatin, epirubicin, and paclitaxel (PET) improves prognosis in locally advanced breast cancer patients: an update of the Southern Italy Cooperative Oncology Group (SICOG) randomised trial 9908. *Ann Oncol* 2010;21:707-16. doi: 10.1093/annonc/mdp356
- Untch M, Möbus V, Kuhn W, et al. Intensive dose-dense compared with conventionally scheduled preoperative chemotherapy for high-risk primary breast cancer. *J Clin Oncol* 2009;27:2938-45. doi: 10.1200/JCO.2008.20.3133
- Arun BK, Dhingra K, Valero V, et al. Phase III randomized trial of dose intensive neoadjuvant chemotherapy with or without G-CSF in locally advanced breast cancer: long-term results. *Oncologist* 2011;16:1527-34. doi: 10.1634/theoncologist.2011-0134
- Untch M, Fasching PA, Konecny GE, et al. Arbeitsgemeinschaft Gynäkologische Onkologie PREPARE investigators. PREPARE trial: a randomized phase III trial comparing preoperative, dose-dense, dose-intensified chemotherapy with epirubicin, paclitaxel and CMF versus a standard-dosed epirubicin/cyclophosphamide followed by paclitaxel ± darbepoetin alfa in primary breast cancer—results at the time of surgery. *Ann Oncol* 2011;22:1988-98. doi: 10.1093/annonc/mdq709

- 32 Untch M, von Minckwitz G, Konecny GE, et al. Arbeitsgemeinschaft Gynäkologische Onkologie PREPARE investigators. PREPARE trial: a randomized phase III trial comparing preoperative, dose-dense, dose-intensified chemotherapy with epirubicin, paclitaxel, and CMF versus a standard-dosed epirubicin-cyclophosphamide followed by paclitaxel with or without darbepoetin alfa in primary breast cancer--outcome on prognosis. *Ann Oncol* 2011;22:1999-2006. doi: 10.1093/annonc/mdq713
- 33 Ellis GK, Barlow WE, Gralow JR, et al. Phase III comparison of standard doxorubicin and cyclophosphamide versus weekly doxorubicin and daily oral cyclophosphamide plus granulocyte colony-stimulating factor as neoadjuvant therapy for inflammatory and locally advanced breast cancer: SWOG 0012. *J Clin Oncol* 2011;29:1014-21. doi: 10.1200/JCO.2009.27.6543
- 34 Walker LG, Eremin JM, Aloysius MM, et al. Effects on quality of life, anti-cancer responses, breast conserving surgery and survival with neoadjuvant docetaxel: a randomised study of sequential weekly versus three-weekly docetaxel following neoadjuvant doxorubicin and cyclophosphamide in women with primary breast cancer. *BMC Cancer* 2011;11:179. doi: 10.1186/1471-2407-11-179
- 35 Schneeweiss A, Möbus V, Tesch T, et al. Survival analysis of the randomized phase III GeparOcto trial comparing neoadjuvant chemotherapy (NACT) of iddEPC versus weekly paclitaxel, liposomal doxorubicin (plus carboplatin in triple-negative breast cancer, TNBC) (PM(Cb)) for patients (pts) with high-risk early breast cancer (BC). *Ann Oncol* 2020;supplement 4:S303-4. doi: 10.1016/j.annonc.2020.08.282
- 36 Vriens BEP, Vriens JH, Aarts MJB, et al. Breast Cancer Trialists' Group of the Netherlands (BOOG). Improved survival for sequentially as opposed to concurrently delivered neoadjuvant chemotherapy in non-metastatic breast cancer. *Breast Cancer Res Treat* 2017;165:593-600. doi: 10.1007/s10549-017-4364-8
- 37 Lee KS, Ro J, Nam BH, et al. A randomized phase-III trial of docetaxel/capecitabine versus doxorubicin/cyclophosphamide as primary chemotherapy for patients with stage II/III breast cancer. *Breast Cancer Res Treat* 2008;109:481-9. doi: 10.1007/s10549-007-9672-y
- 38 Kelly CM, Green MC, Broglio K, et al. Phase III trial evaluating weekly paclitaxel versus docetaxel in combination with capecitabine in operable breast cancer. *J Clin Oncol* 2012;30:930-5. doi: 10.1200/JCO.2011.36.2079
- 39 von Minckwitz G, Kümmel S, Vogel P, et al. German Breast Group. Neoadjuvant vinorelbine-capecitabine versus docetaxel-doxorubicin-cyclophosphamide in early nonresponsive breast cancer: phase III randomized GeparTrio trial. *J Natl Cancer Inst* 2008;100:542-51. doi: 10.1093/jnci/djn085
- 40 von Minckwitz G, Blohmer JU, Costa SD, et al. Response-guided neoadjuvant chemotherapy for breast cancer. *J Clin Oncol* 2013;31:3623-30. doi: 10.1200/JCO.2012.45.0940
- 41 Toi M, Ohno S, Sato N, et al. Preoperative docetaxel (T) with or without capecitabine (X) following epirubicin, 5-fluorouracil and cyclophosphamide (FEC) in patients with operable breast cancer (OOCR N003): Results of comparative study and predictive marker analysis. *Cancer Res* 2012;72. doi: 10.1158/0008-5472.SABCS12-P1-14-02
- 42 von Minckwitz G, Rezaei M, Loibl S, et al. Capecitabine in addition to anthracycline- and taxane-based neoadjuvant treatment in patients with primary breast cancer: phase III GeparQuattro study. *J Clin Oncol* 2010;28:2015-23. doi: 10.1200/JCO.2009.23.8303
- 43 von Minckwitz G, Rezaei M, Fasching PA, et al. Survival after adding capecitabine and trastuzumab to neoadjuvant anthracycline-taxane-based chemotherapy for primary breast cancer (GBG 40-GeparQuattro). *Ann Oncol* 2014;25:81-9. doi: 10.1093/annonc/mdt410
- 44 Buzdar AU, Ibrahim NK, Francis D, et al. Significantly higher pathologic complete remission rate after neoadjuvant therapy with trastuzumab, paclitaxel, and epirubicin chemotherapy: results of a randomized trial in human epidermal growth factor receptor 2-positive operable breast cancer. *J Clin Oncol* 2005;23:3676-85. doi: 10.1200/JCO.2005.07.032
- 45 Buzdar AU, Valero V, Ibrahim NK, et al. Neoadjuvant therapy with paclitaxel followed by 5-fluorouracil, epirubicin, and cyclophosphamide chemotherapy and concurrent trastuzumab in human epidermal growth factor receptor 2-positive operable breast cancer: an update of the initial randomized study population and data of additional patients treated with the same regimen. *Clin Cancer Res* 2007;13:228-33. doi: 10.1158/1078-0432.CCR-06-1345
- 46 Gianni L, Eiermann W, Semiglazov V, et al. Neoadjuvant chemotherapy with trastuzumab followed by adjuvant trastuzumab versus neoadjuvant chemotherapy alone, in patients with HER2-positive locally advanced breast cancer (the NOAH trial): a randomised controlled superiority trial with a parallel HER2-negative cohort. *Lancet* 2010;375:377-84. doi: 10.1016/S0140-6736(09)61964-4
- 47 Baselga J, Bradbury I, Eidtmann H, et al. NeoALTTO Study Team. Lapatinib with trastuzumab for HER2-positive early breast cancer (NeoALTTO): a randomised, open-label, multicentre, phase 3 trial. *Lancet* 2012;379:633-40. doi: 10.1016/S0140-6736(11)61847-3
- 48 Huober J, Holmes E, Baselga J, et al. Survival outcomes of the NeoALTTO study (BIG 1e06): updated results of a randomised multicenter phase III neoadjuvant clinical trial in patients with HER2-positive primary breast cancer. *Euro J Cancer* 2019;118:169e177. doi: 10.1016/j.ejca.2019.04.038
- 49 Gianni L, Pienkowski T, Im YH, et al. Efficacy and safety of neoadjuvant pertuzumab and trastuzumab in women with locally advanced, inflammatory, or early HER2-positive breast cancer (NeoSphere): a randomised multicentre, open-label, phase 2 trial. *Lancet Oncol* 2012;13:25-32. doi: 10.1016/S1470-2045(11)70336-9
- 50 Gianni L, Pienkowski T, Im YH, et al. 5-year analysis of neoadjuvant pertuzumab and trastuzumab in patients with locally advanced, inflammatory, or early-stage HER2-positive breast cancer (NeoSphere): a multicentre, open-label, phase 2 randomised trial. *Lancet Oncol* 2016;17:791-800. doi: 10.1016/S1470-2045(16)00163-7
- 51 Hurvitz SA, Martin M, Jung KH, et al. Neoadjuvant Trastuzumab Emtansine and Pertuzumab in Human Epidermal Growth Factor Receptor 2-Positive Breast Cancer: Three-Year Outcomes From the Phase III KRISTINE Study. *J Clin Oncol* 2019;37:2206-16. doi: 10.1200/JCO.19.00882
- 52 Jackisch C, Hegg R, Stroyakovskiy D, et al. HannaH phase III randomised study: Association of total pathological complete response with event-free survival in HER2-positive early breast cancer treated with neoadjuvant-adjuvant trastuzumab after 2 years of treatment-free follow-up. *Eur J Cancer* 2016;62:62-75. doi: 10.1016/j.ejca.2016.03.087
- 53 Fernandez-Martinez A, Krop IE, Hillman DW, et al. Survival, Pathologic Response, and Genomics in CALGB 40601 (Alliance), a Neoadjuvant Phase III Trial of Paclitaxel-Trastuzumab With or Without Lapatinib in HER2-Positive Breast Cancer. *J Clin Oncol* 2020;38:4184-93. doi: 10.1200/JCO.20.01276
- 54 Cocconi G, Di Blasio B, Boni C, et al. Italian Oncology Group for Clinical Research (GOIRC), Parma, Italy. Primary chemotherapy in operable breast carcinoma comparing CMF (cyclophosphamide, methotrexate, 5-fluorouracil) with an anthracycline-containing regimen: short-term responses translated into long-term outcomes. *Ann Oncol* 2005;16:1469-76. doi: 10.1093/annonc/mdi278
- 55 Smith IE, A'Hern RP, Coombes GA, et al. TOPIC Trial Group. A novel continuous infusional 5-fluorouracil-based chemotherapy regimen compared with conventional chemotherapy in the neo-adjuvant treatment of early breast cancer: 5 year results of the TOPIC trial. *Ann Oncol* 2004;15:751-8. doi: 10.1093/annonc/mdh175
- 56 Kaufmann M, Eiermann W, Schuette M, et al. Long-term results from the neoadjuvant GeparDuo trial: A randomized, multicenter, open phase III study comparing a dose-intensified 8-week schedule of doxorubicin hydrochloride and docetaxel (ADoc) with a sequential 24-week schedule of doxorubicin hydrochloride/cyclophosphamide followed by docetaxel (AC-Doc) regimen as preoperative therapy (NACT) in patients (pts) with operable breast cancer (BC). *J Clin Oncol* 2010;28:76s. doi: 10.1200/jco.2010.28.15_suppl.537
- 57 Sikov WM, Berry DA, Perou CM, et al. Impact of the addition of carboplatin and/or bevacizumab to neoadjuvant once-per-week paclitaxel followed by dose-dense doxorubicin and cyclophosphamide on pathologic complete response rates in stage II to III triple-negative breast cancer: CALGB 40603 (Alliance). *J Clin Oncol* 2015;33:13-21. doi: 10.1200/JCO.2014.4.57.0572
- 58 Sikov WM, Berry DA, Perou CM, et al. Abstract S2-05: Event-free and overall survival following neoadjuvant weekly paclitaxel and dose-dense AC +/- carboplatin and/or bevacizumab in triple-negative breast cancer: Outcomes from CALGB 40603 (Alliance). *Cancer Res* 2016;76(4 Supplement). doi: 10.1158/1538-7445.SABCS15-S2-05
- 59 von Minckwitz G, Schneeweiss A, Loibl S, et al. Neoadjuvant carboplatin in patients with triple-negative and HER2-positive early breast cancer (GeparSixto; GBG 66): a randomised phase 2 trial. *Lancet Oncol* 2014;15:747-56. doi: 10.1016/S1470-2045(14)70160-3
- 60 Loibl S, Weber KE, Timms KM, et al. Survival analysis of carboplatin added to an anthracycline/taxane-based neoadjuvant chemotherapy and HRD score as predictor of response-final results from GeparSixto. *Ann Oncol* 2018;29:2341-7. doi: 10.1093/annonc/mdy460
- 61 Untch M, Jackisch C, Schneeweiss A, et al. German Breast Group (GBG), Arbeitsgemeinschaft Gynäkologische Onkologie-Breast (AGO-B) Investigators. Nab-paclitaxel versus solvent-based paclitaxel in neoadjuvant chemotherapy for early breast cancer (GeparSepto-GBG 69): a randomised, phase 3 trial. *Lancet Oncol* 2016;17:345-56. doi: 10.1016/S1470-2045(15)00542-2
- 62 Untch M, Jackisch C, Schneeweiss A, et al. NAB-Paclitaxel Improves Disease-Free Survival in Early Breast Cancer: GBG 69-GeparSepto. *J Clin Oncol* 2019;37:2226-34. doi: 10.1200/JCO.18.01842
- 63 Schmid P, Cortes J, Pusztai L, et al. KEYNOTE-522 Investigators. Pembrolizumab for Early Triple-Negative Breast Cancer. *N Engl J Med* 2020;382:810-21. doi: 10.1056/NEJMoa1910549

- 64 Mittendorf EA, Zhang H, Barrios CH, et al. Neoadjuvant atezolizumab in combination with sequential nab-paclitaxel and anthracycline-based chemotherapy versus placebo and chemotherapy in patients with early-stage triple-negative breast cancer (IMpassion031): a randomised, double-blind, phase 3 trial. *Lancet* 2020;396:1090-100. doi: 10.1016/S0140-6736(20)31953-X
- 65 de Groot S, Pijl H, Charehbilil A, et al, Dutch Breast Cancer Research Group. Addition of zoledronic acid to neoadjuvant chemotherapy is not beneficial in patients with HER2-negative stage II/III breast cancer: 5-year survival analysis of the NEOZOTAC trial (BOOG 2010-01). *Breast Cancer Res* 2019;21:97. doi: 10.1186/s13058-019-1180-6
- 66 Schneeweiss A, Marmé F, Ruiz A, et al. A randomized phase II trial of doxorubicin plus pemetrexed followed by docetaxel versus doxorubicin plus cyclophosphamide followed by docetaxel as neoadjuvant treatment of early breast cancer. *Ann Oncol* 2011;22:609-17. doi: 10.1093/annonc/mdq400
- 67 Schneeweiss A, Ruiz A, Manikhas A, et al. A randomized phase II trial of doxorubicin plus pemetrexed followed by docetaxel versus doxorubicin plus cyclophosphamide followed by docetaxel as neoadjuvant chemotherapy (NACT) for early breast cancer: Three-year follow-up data. *J Clin Oncol* 2012;30:63s. doi: 10.1200/jco.2012.30.15_suppl.1059
- 68 Pivot X, Bondarenko I, Nowecki Z, et al. A phase III study comparing SB3 (a proposed trastuzumab biosimilar) and trastuzumab reference product in HER2-positive early breast cancer treated with neoadjuvant-adjunct treatment: Final safety, immunogenicity and survival results. *Eur J Cancer* 2018;93:19-27. doi: 10.1016/j.ejca.2018.01.072
- 69 Buzdar AU, Suman VJ, Meric-Bernstam F, et al. Disease-Free and Overall Survival Among Patients With Operable HER2-Positive Breast Cancer Treated With Sequential vs Concurrent Chemotherapy: The ACOSOG Z1041 (Alliance) Randomized Clinical Trial. *JAMA Oncol* 2019;5:45-50. doi: 10.1001/jamaoncol.2018.3691
- 70 Earl HM, Hiller L, Dunn JA, et al, ARTEMIS Investigators Group. Disease-free and overall survival at 3.5 years for neoadjuvant bevacizumab added to docetaxel followed by fluorouracil, epirubicin and cyclophosphamide, for women with HER2 negative early breast cancer: ARTEMIS Trial. *Ann Oncol* 2017;28:1817-24. doi: 10.1093/annonc/mdx173
- 71 Gianni L, Mansutti M, Anton A, et al. Comparing Neoadjuvant Nab-paclitaxel vs Paclitaxel Both Followed by Anthracycline Regimens in Women With ERBB2/HER2-Negative Breast Cancer-The Evaluating Treatment With Neoadjuvant Abraxane (ETNA) Trial: A Randomized Phase 3 Clinical Trial. *JAMA Oncol* 2018;4:302-8. doi: 10.1001/jamaoncol.2017.4612
- 72 von Minckwitz G, Loibl S, Untch M, et al, GBG/AGO-B study groups. Survival after neoadjuvant chemotherapy with or without bevacizumab or everolimus for HER2-negative primary breast cancer (GBG 44-GeparQuinto). *Ann Oncol* 2014;25:2363-72. doi: 10.1093/annonc/mdu455
- 73 Huober J, Fasching PA, Hanusch C, et al. Neoadjuvant chemotherapy with paclitaxel and everolimus in breast cancer patients with non-responsive tumours to epirubicin/cyclophosphamide (EC) ± bevacizumab - results of the randomised GeparQuinto study (GBG 44). *Eur J Cancer* 2013;49:2284-93. doi: 10.1016/j.ejca.2013.02.027
- 74 Untch M, von Minckwitz G, Gerber B, et al, GBG and the AGO-B Study Group. Survival Analysis After Neoadjuvant Chemotherapy With Trastuzumab or Lapatinib in Patients With Human Epidermal Growth Factor Receptor 2-Positive Breast Cancer in the GeparQuinto (G5) Study (GBG 44). *J Clin Oncol* 2018;36:1308-16. doi: 10.1200/JCO.2017.75.9175
- 75 Iwase M, Ando M, Aogi K, et al. Long-term survival analysis of addition of carboplatin to neoadjuvant chemotherapy in HER2-negative breast cancer. *Breast Cancer Res Treat* 2020;180:687-94. doi: 10.1007/s10549-020-05580-y
- 76 Jovanović B, Mayer IA, Mayer EL, et al. A Randomized Phase II Neoadjuvant Study of Cisplatin, Paclitaxel With or Without Everolimus in Patients with Stage II/III Triple-Negative Breast Cancer (TNBC): Responses and Long-term Outcome Correlated with Increased Frequency of DNA Damage Response Gene Mutations, TNBC Subtype, AR Status, and Ki67. *Clin Cancer Res* 2017;23:4035-45. doi: 10.1158/1078-0432.CCR-16-3055
- 77 Earl HM, Vallier AL, Hiller L, et al, Neo-tAnGo Investigators. 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, 2x2 factorial randomised phase 3 trial. *Lancet Oncol* 2014;15:201-12. doi: 10.1016/S1470-2045(13)70554-0
- 78 Bear HD, Tang G, Rastogi P, et al. Neoadjuvant plus adjuvant bevacizumab in early breast cancer (NSABP B-40 [NRG Oncology]): secondary outcomes of a phase 3, randomised controlled trial. *Lancet Oncol* 2015;16:1037-48. doi: 10.1016/S1470-2045(15)00041-8
- 79 Robidoux A, Tang G, Rastogi P, et al. Lapatinib as a component of neoadjuvant therapy for HER2-positive operable breast cancer (NSABP protocol B-41): an open-label, randomised phase 3 trial. *Lancet Oncol* 2013;14:1183-92. doi: 10.1016/S1470-2045(13)70411-X
- 80 Stebbing J, Baranau Y, Baryash V, et al. CT-P6 compared with reference trastuzumab for HER2-positive breast cancer: a randomised, double-blind, active-controlled, phase 3 equivalence trial [Erratum in: *Lancet Oncol* 2017;18:e433]. *Lancet Oncol* 2017;18:917-28. doi: 10.1016/S1470-2045(17)30434-5
- 81 Nahleh Z, Barlow WE, Hayes DF, et al. SWOG S0800 (NCI CD0000636131): addition of bevacizumab to neoadjuvant nab-paclitaxel with dose-dense doxorubicin and cyclophosphamide improves pathologic complete response (pCR) rates in inflammatory or locally advanced breast cancer. *Breast Cancer Res Treat* 2016;158:485-95. doi: 10.1007/s10549-016-3889-6
- 82 van Ramshorst MS, van der Voort A, van Werkhoven ED, et al, Dutch Breast Cancer Research Group (BOOG). Neoadjuvant chemotherapy with or without anthracyclines in the presence of dual HER2 blockade for HER2-positive breast cancer (TRAIN-2): a multicentre, open-label, randomised, phase 3 trial. *Lancet Oncol* 2018;19:1630-40. doi: 10.1016/S1470-2045(18)30570-9
- 83 Symmans WF, Yau C, Chen YY, et al. Assessment of Residual Cancer Burden and Event-Free Survival in Neoadjuvant Treatment for High-risk Breast Cancer: An Analysis of Data From the I-SPY2 Randomized Clinical Trial. *JAMA Oncol* 2021;7:1654-63. doi: 10.1001/jamaoncol.2021.3690
- 84 Rose BS, Winer EP, Mamouni EJ. Perils of the Pathologic Complete Response. *J Clin Oncol* 2016;34:3959-62. doi: 10.1200/JCO.2016.68.1718
- 85 Menzies AM, Amaria RN, Rozeman EA, et al. Pathological response and survival with neoadjuvant therapy in melanoma: a pooled analysis from the International Neoadjuvant Melanoma Consortium (INMC). *Nat Med* 2021;27:301-9. doi: 10.1038/s41591-020-01188-3
- 86 Burstein HJ. Systemic Therapy for Estrogen Receptor-Positive, HER2-Negative Breast Cancer. *N Engl J Med* 2020;383:2557-70. doi: 10.1056/NEJMra1307118
- 87 Yau C, van der Noordaa M, Wei J, et al. Residual cancer burden after neoadjuvant therapy and long-term survival outcomes in breast cancer: A multi-center pooled analysis. 2019 San Antonio Breast Cancer Symposium. *Abstract GS5-01*. Presented December 13, 2019.
- 88 Hellmann MD, Chaffin KE, Williams WN Jr, et al, University of Texas MD Anderson Lung Cancer Collaborative Group. Pathological response after neoadjuvant chemotherapy in resectable non-small-cell lung cancers: proposal for the use of major pathological response as a surrogate endpoint. *Lancet Oncol* 2014;15:e42-50. doi: 10.1016/S1470-2045(13)70334-6
- 89 Fleming TR, Powers JH. Biomarkers and surrogate endpoints in clinical trials. *Stat Med* 2012;31:2973-84. doi: 10.1002/sim.5403
- 90 Bujkiewicz S, Thompson JR, Riley RD, Abrams KR. Bayesian meta-analytical methods to incorporate multiple surrogate endpoints in drug development process. *Stat Med* 2016;35:1063-89. doi: 10.1002/sim.6776
- 91 Lenzer J, Brownlee S. Should regulatory authorities approve drugs based on surrogate endpoints? *BMJ* 2021;374:n2059. doi: 10.1136/bmj.n2059
- 92 Dawoud D, Naci H, Ciani O, Bujkiewicz S. Raising the bar for using surrogate endpoints in drug regulation and health technology assessment. *BMJ* 2021;374:n2191. doi: 10.1136/bmj.n2191
- 93 Hayes DF, Schott AF. Neoadjuvant Chemotherapy: What Are the Benefits for the Patient and for the Investigator? *Natl Cancer Inst Monogr* 2015;2015:36-9. doi: 10.1093/jncimonographs/igv004
- 94 von Minckwitz G, Huang C-S, Mano MS, et al, KATHERINE Investigators. Trastuzumab Emtansine for Residual Invasive HER2-Positive Breast Cancer. *N Engl J Med* 2019;380:617-28. doi: 10.1056/NEJMoa1814017
- 95 Masuda N, Lee S-J, Ohtani S, et al. Adjuvant Capecitabine for Breast Cancer after Preoperative Chemotherapy. *N Engl J Med* 2017;376:2147-59. doi: 10.1056/NEJMoa1612645
- 96 Korn EL, Albert PS, McShane LM. Assessing surrogates as trial endpoints using mixed models. *Stat Med* 2005;24:163-82. doi: 10.1002/sim.1779
- 97 Burzykowski T, Molenberghs G, Buyse M, Geys H, Renard D. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. *Appl Stat* 2001;50:405-2. doi: 10.1111/1467-9876.00244
- 98 Green E, Yothers G, Sargent DJ. Surrogate endpoint validation: statistical elegance versus clinical relevance. *Stat Methods Med Res* 2008;17:477-86. doi: 10.1177/0962280207081863

Supplementary information: Additional tables and figures

Supplementary information: Lay summary of findings