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

Peer-reviewed author version

Saad, Everardo D.; Squifflet, Pierre; BURZYKOWSKI, Tomasz; Quinaux, Emmanuel; Delaloge, Suzette; Mavroudis, Dimitris; Perez, Edith; Piccart-Gebhart, Martine; Schneider, Bryan P.; Slamon, Dennis; Wolmark, Norman & BUYSE, Marc (2019)
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. In: LANCET ONCOLOGY, 20(3), p. 361-370.

DOI: 10.1016/S1470-2045(18)30750-2
Handle: http://hdl.handle.net/1942/30248

Disease-free survival as a surrogate for overall survival in HER-2-positive, early breast cancer trials of adjuvant trastuzumab for up to 1 year

Running head: DFS as a surrogate for OS in HER-2-positive breast cancer

Everardo D. Saad, MD[1], Pierre Squifflet, MSc[1], Tomasz Burzykowski, PhD[1,2], Emmanuel Quinaux, MSc[1], Suzette Delaloge, MD[3], Dimitris Mavroudis, MD[4], Edith Perez, MD[5], Martine Piccart-Gebhart, MD, PhD[6], Bryan P. Schneider, MD[7], Dennis Slamon, MD, PhD[8], Norman Wolmark, MD[9], and Marc Buyse, ScD[2,10]

[1]International Drug Development Institute (IDDI), Louvain-la-Neuve, Belgium; [2]Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), Hasselt University, Diepenbeek, Belgium; [3]Unicancer Breast Group, Paris, France; [4]University of Crete, Heraklion, Greece; [5]Mayo Clinic, Jacskonville, FL; [6]Institut Jules Bordet, Université Libre de Bruxelles (ULB), Brussels, Belgium; [7]Indiana University, Indianapolis, IN; [8]University of California, Los Angeles (UCLA), Los Angeles, CA; [9]NRG Oncology/Pittsburgh, Pittsburgh, PA[10]; International Drug Development Institute (IDDI), San Francisco, CA.

Address for correspondence:

Everardo D. Saad, MD

International Drug Development Institute

Avenue Provinciale 30 – 1340 Louvain-la-Neuve, Belgium

Phone: +32 10 61 44 44 / Fax: +32 10 61 88 88

Email: everardo.saad@iddi.com

ABSTRACT

Background: Although frequently used as primary end point, disease-free survival (DFS) has not been validated as a surrogate for overall survival (OS) in early breast cancer. We investigated surrogacy in the adjuvant setting of anti-HER-2 antibodies.

Methods: In a systematic review, we identified trials with completed accrual and available DFS and OS results as of 09/16. Eligibility required at least one arm to have the antibody planned for 12 months, and at least one control arm with (1) chemotherapy without the antibody, (2) a lower total dose or duration of the antibody, or (3) observation alone. Units of analysis were 'contrasts': two-arm trials gave rise to one contrast, while trials with more than two arms gave rise to more than one contrast. We measured the association between DFS and OS using Spearman's correlation coefficient (ρ), and the association between hazard ratios (HRs) for DFS and OS using R^2. We computed the surrogate threshold effect (STE), the maximum HR for DFS that statistically predicts a HR for OS < 1.00 in a future trial.

Findings: Eight trials (N=21,480 patients) gave rise to a full set (12 contrasts) and to a reduced set (11 contrasts) that excluded one trial (N=481), with trastuzumab used in all cases. In both sets, patient-level associations were strong (ρ=0·90). Trial-level associations gave rise to values of R^2 of 0·75 (95% confidence interval [CI], 0·50 to 1·00) for the full set and 0·84 (95% CI, 0·67 to 1·00) for the reduced set. Subgroups defined by nodal and hormone-receptor statuses yielded qualitatively similar results. Depending on the expected number of deaths in a future trial, the STEs ranged from 0·59 to 0·84.

Interpretation: We suggest that it is appropriate to continue to use DFS as a surrogate for OS in trials in HER-2-positive, early breast cancer.

Funding: Roche Pharma AG.

INTRODUCTION

Breast cancer is currently divided into molecular subtypes with relevant prognostic and predictive implications for clinical practice.[1] In HER-2-positive breast cancer, adjuvant therapy with trastuzumab improves outcomes.[2-4] Although shorter[5-7] and longer duration of trastuzumab therapy,[2,8] as well as other anti-HER-2 agents,[9-11] have been investigated in the adjuvant setting, 1 year of trastuzumab remains the most common option.[1,12] The development of novel adjuvant regimens is a lengthy process, and the analysis of overall survival (OS) requires long follow-up. One possibility to expedite drug development and patient access to improved regimens is the use of neoadjuvant therapy as a platform for testing novel agents, given the hypothesis that superior regimens in this context are more promising in the adjuvant setting than regimens that add no improvements in clinical or pathologic responses. However, doubts remain about the predictive ability of the neoadjuvant platform,[13,14] and another possibility to expedite the development of adjuvant therapy is to use surrogates for OS.[15] Although disease-free survival (DFS) has often been used as primary end point in adjuvant trials of breast cancer, to our knowledge it has not been formally validated in this setting, as it has in others.[16-18] Meta-analyses of individual-patient data from randomized trials provides two measures of association between the potential surrogate and the final end point of interest: the patient-level and the trial-level associations.[19] The former denotes the prognostic role of the surrogate (*e.g.*, whether patients with prolonged DFS are also more likely to experience prolonged OS), whereas the trial-level association provides predictive information (*i.e.*, treatment-induced changes of a certain magnitude in the surrogate are accompanied by proportional changes in the final end point). These two associations are independent, as illustrated by a strong patient-level association, but weak trial-level association, between pathologic complete response and OS in the

neoadjuvant therapy of breast cancer.[14] The current study was conducted to evaluate the role of DFS as a surrogate for OS in the adjuvant treatment of HER-2-positive breast cancer.

METHODS

Study ethics and oversight

This study was designed and conducted by the authors, and the protocol was approved by the Ethics Committee of Hasselt University, Belgium. Since the project consisted of re-analysis of data from Institutional Review Board-approved clinical trials, no informed consent was sought from patients. However, data from one eligible trial that finished accrual in 2010 (PHARE) were not made available because of concerns on the part of the French National Cancer Institute about an alleged need to obtain repeated informed consent from patients.[6] For all trials included in the analysis, data were shared with the authors after approval from the original trial sponsor, whether academic or industry. The costs associated with literature search and with data collection, management and analysis were defrayed by the financial support provided by Roche Pharma AG, Germany. The manuscript was drafted and reviewed by the authors. The financial sponsor was given the opportunity to provide courtesy review of the manuscript, whose final content is the responsibility of the authors alone.

Trial identification and eligibility

The initial search for eligible trials was conducted by a third party, HealthEcon (Basel, Switzerland), in 10/2015 and updated 10/2016. As shown in Figure 1S and accompanying notes (Supplementary Materials), the systematic search was performed in several databases. Two eligible trials were identified for which data would not be

available in a timely fashion, but whose results have been presented when the current project was ongoing.[20,21] The results from these two trials and those from the PHARE trial[6] were used in an exploratory analysis to verify some of the predictions resulting from this study. Moreover, a third eligible trial was ongoing and had its results presented in a scientific meeting when the current analyses had been finalized.[22] No additional eligible trials have been identified, but one trial published only in abstract form was found incidentally, and in this case contact could not be established with investigators; this trial, conducted in India and not reported in any of the registries accessed, enrolled 134 patients and only reported 3-year DFS results.[23]

Eligible trials were randomized studies of adjuvant therapy for stage I to III breast cancer, with randomization performed after surgery and accrual completed as of 09/2016; patients had to have HER-2-positive disease, either exclusively or with stratification for HER-2 positivity ascertained by accepted methods; anti-HER-2 antibody use had been planned for a total duration of 1 year in at least one of the trial arms, and at least one arm had (1) observation alone, (2) chemotherapy alone, or (3) a lower total dose (per cycle or in terms of treatment duration) of the anti-HER-2 antibody; and the trial research question must have involved the antibody. Excluded were trials enrolling patients with recurrent, metastatic or non-invasive disease, and those testing neoadjuvant therapy exclusively (if both neoadjuvant and adjuvant therapies were allowed in a trial, randomization must have been after surgery). One of the eight eligible trials, with a much shorter follow-up and smaller number of events than all other trials, was excluded from some of the analyses.[7] The DFS and OS curves presented in the publication of this trial had different follow-up times and distributions of censored observations between the two treatment arms, as a result of differential

exclusion of randomized patients during the first year, precluding proper intention-to-treat (ITT) analysis.[7]

Study objectives

The primary objective was to assess surrogacy of DFS for OS in trials using an adjuvant anti-HER-2 antibody and considering the ITT population of each trial. Key secondary objectives were to estimate these associations in trials that had trastuzumab for 1 year in at least one experimental arm, and at least one non-anti-HER-2-containing arm, with the same chemotherapy in both of these arms; to estimate these associations in patient subgroups defined by hormone-receptor expression and nodal status; to conduct sensitivity analyses according to previous use of neoadjuvant therapy and different definitions of DFS[24]; and to conduct exploratory analyses of the association between DFS and breast-cancer-specific survival (BCSS). Since no information was generally available on causes of death, this exploratory analysis was conducted with deaths preceded by recurrence considered to be due to breast cancer.

Statistical methods

A two-level modeling approach was used to estimate the association between DFS and OS and between the treatment effects on these end points.[19] At the patient level, the joint distribution of the surrogate and the true end point was estimated by using a copula-based model. Three different distributions, corresponding to three different copulas (Clayton's, Hougaard's, or Plackett's) were considered, and the one providing the maximum likelihood value was selected for inference. Subsequently, the strength of the association between the surrogate and the true endpoint was quantified by the value of Spearman's rank correlation coefficient (ρ) corresponding to the selected

copula. For the trial-level assessment, the proportional hazards model was used to jointly estimate the hazard ratios (HR) for DFS and OS. A linear regression was then fitted through the points representing the logarithms of the hazard ratio (logHR) for DFS and for OS from each unit of analysis, or contrast. For simplicity of interpretation, the graphs show HRs (rather than logHRs) in their axes. Given that all identified trials had trastuzumab as the anti-HER-2 antibody, and in order to have more homogeneous comparisons, trastuzumab for 1 year was used in the numerator for DFS and OS for all contrasts analyzed; this was done regardless of whether trastuzumab was used alone,[2] with chemotherapy,[3,4,7,9,25-27] or with lapatinib.[9] For the trial-level assessment, each two-arm trial gave rise to one contrast. Two trials with three arms[4,27] had their chemotherapy-alone arm randomly split to generate two or more contrasts. One trial with four arms gave rise to two contrasts with no need for random splits.[9] Randomly splitting the control arm does not lead to multiplicity issues, because no extra significance tests are generated.

In all analyses, an attempt was made to fit the regression models while taking into account the estimation error present in the estimated HRs for OS and DFS by using a measurement-error model. [19] In case of numerical problems with fitting the models, weighted regression models were considered by using as weights the number of deaths in each contrast. The linear regression fitted through the estimated treatment effects provides a coefficient of determination (R^2), which quantifies the proportion of variance in the effects of treatment on the true end point that is explained by the surrogate. Additionally, the fitted regression line (whether obtained by using the measurement-error modelling or weighted regression) allows construction of a 95% prediction interval for the HR for OS corresponding to a particular value of HR for DFS.

Note that the 95% prediction interval based on the weighted regression model depends on the weight assigned to the contrast for which HR is being predicted (usually taken proportional to the number of patients contributing to the contrast).

To assess model accuracy, a leave-one-out cross-validation strategy was used, with each contrast left out once and the linear model re-fitted to the remaining contrasts. This model was re-applied to the left-out contrast in order to compare the predicted and observed treatment effect on OS. Finally, the surrogate threshold effect (STE) was investigated. The STE is the minimum treatment effect on the surrogate required to predict a non-zero treatment effect on the final end point in a future randomized trial.[28] Unrealistically large/small values of STE, compared with treatment effects on the surrogate observed in previous clinical trials, indicate poor validity of the surrogate.[28]

RESULTS

Characteristics of trials

Data were available from a total of 21,480 patients from the eight trials analyzed. Table 1 displays selected characteristics of these trials, which gave rise to 12 contrasts. Individual trials had slightly varying definitions for DFS; since in many cases no separate information was available on non-invasive recurrences, DFS henceforth refers to any type of recurrence (invasive or non-invasive) or death from any cause. Given the methodological issues with one of the trials, some of the analyses were conducted in a reduced set with seven trials and 11 contrasts. Table 2 shows selected results from each contrast; these results may differ from those in original publications due to the use of different contrasts than used in the original trial or longer follow-up at the time of the current analysis. Eight of the 12 contrasts consisted in comparisons of chemotherapy or

observation versus the same plus 12 months of trastuzumab; in three contrasts, the comparison was between 12 months of trastuzumab and shorter durations of the antibody, combined with chemotherapy alone or chemotherapy plus lapatinib; and one contrast was the comparison of lapatinib for 12 months versus lapatinib plus trastuzumab for 12 months.

Patient-level surrogacy of DFS for OS

Patient-level associations between DFS and OS were strong: in both the full and in the reduced sets, the ρ value was 0·90 (95% confidence interval [CI], 0·89 to 0·90 in both cases). Figure 2S (Supplementary Materials) shows the Kaplan-Meier curves for DFS and OS in the experimental and control arms defined for the current analyses.

Trial-level surrogacy of DFS for OS

In the full set, analyses weighted by the number of deaths gave rise to an R^2 value of 0·75 (95% CI, 0·50 to 1·00), whereas in the reduced set the value of R^2 was 0·84 (95% CI, 0·67 to 1·00). Figure 1 displays the linear regression model weighted by the number of deaths in the reduced set, which yielded the following regression equation: $\ln (HR_{OS}) = -0{\cdot}005 + 0{\cdot}910 \cdot \ln (HR_{DFS})$, with standard errors of the intercept and slope estimated as 0·042 and 0·124, respectively. Figure 3S (Supplementary Materials) displays the regression model in the full set. In neither case could the regression model be fitted with adjustment for the magnitude of the estimation errors of the treatment effects on DFS and OS by using a measurement-error model. Note that the 95% CIs for R^2 were relatively wide. Cross-validation performed in both the full and reduced sets showed that only the observed HR for OS for the excluded trial fell outside the prediction interval (Table 3). Moreover, except for the NCCTG N9831 A vs B (C closed) contrast, the

conclusion regarding the significance of the treatment effect on OS resulting from the 95% CI for the predicted HR always agreed with the conclusion based on the 95% CI for the observed HR.

Secondary analyses conducted in subgroups defined by lymph-node and hormone-receptor statuses led to qualitatively similar results to those found for the analyses in the full and reduced sets (Table S1, Supplementary Materials). In the subset of trials that had trastuzumab for 1 year in at least one experimental arm, and at least one non-anti-HER-2-containing arm, with the same chemotherapy in both arms, seven contrasts could be formed with a total of 11,309 patients and 2,248 deaths. In this analysis, the trial-level association was weaker than in the analyses in the full and reduced sets, with an R^2 of 0·46 (95% CI, 0·00 to 1·00), possibly because of the exclusion of contrasts with more extreme HRs. In an analysis excluding 1,082 patients with previous use of neoadjuvant therapy, which was allowed in only three trials, the trial-level associations were very similar to those in the analyses in the full and reduced sets (data not shown).

Surrogate threshold effect

Since the full and reduced sets were analyzed by regression models weighted by the number of deaths, the STE was computed for different scenarios based on the expected number of deaths in a future trial. As shown in Table 4, HRs for DFS below 0·82 would predict significant gains in OS in a randomized trial with approximately 800 deaths, whereas HRs for DFS below around 0·70 would predictably be followed by significant gains in OS in trials with approximately 200 deaths.

Exploratory analyses

The association between DFS and BCSS was evaluated using the 2,744 deaths occurring after a recurrence (85% of the total 3,233 deaths). In this case, the results for the reduced set could be obtained from a model adjusted for the magnitude of the estimation errors of the treatment effects by using a measurement-error model. As a result, the 95% prediction limits could be estimated in a uniform way, i.e., irrespectively of the number of deaths (Figure 2). In this analysis, the patient-level association was marked by the ρ value of 0·98 (95% CI, 0·98 to 0·98), whereas the trial-level association was characterized by an R^2 of 0·95 (95% CI, 0·65 to 1·00) and by the following regression equation: $\ln(HR_{BCSS}) = -0{\cdot}037 + 0{\cdot}929 \cdot \ln(HR_{DFS})$, with standard errors of the intercept and slope estimated as 0·073 and 0·237, respectively. Other results from this exploratory analysis are shown in Table S2 (Supplementary Materials). Finally, DFS results from three trials with unavailable data as of this writing were used to verify some OS predictions based on the reduced-set model described above (Table 5). In all cases, the observed HR for OS is included in the prediction interval and the conclusion regarding the significance of the treatment effect on OS resulting from the prediction interval agrees with the conclusion based on the 95% CI for the observed HR.

DISCUSSION

The current study is the first to formally assess DFS as a surrogate for OS in the adjuvant treatment of HER-2-positive breast cancer. In a study presented recently in abstract form, data from nearly 12 thousand patients enrolled in five phase III trials were used to assess various potential surrogates for OS in the adjuvant treatment of breast cancer.[29] Two of the trials analyzed in that study are also included here,[2,26] but the others did not assess anti-HER-2 therapy. The authors of that study found that invasive DFS had the strongest association with OS at the trial level, but concluded that further evaluation on

a larger set of trials was required to improve the precision of their estimations. Moreover, it is conceivable that the association between end points differs according to breast-cancer phenotype, as suggested in a meta-analysis in the neoadjuvant setting.[14] For this reason, we believe our results pertain to HER-2-positive disease, and separate studies should be conducted for luminal and triple-negative phenotypes. Of note, most trials enrolling patients with these phenotypes have been designed on the basis of treatment type and not the phenotype, and with few exceptions patients with luminal and triple-negative disease represent subgroups among the totality of enrolled patients in trials of hormone therapy and chemotherapy. Obtaining specific data from those patients is a foreseeable difficulty in future surrogacy work related to the HER-2-negative phenotypes.

Despite the biological rationale for considering DFS as a surrogate for OS in early breast cancer, patients with breast cancer are often elderly and die from other causes. Therefore, there is a need to confirm whether DFS and OS are associated both at the patient and at the trial levels. The results of this meta-analysis suggest a strong association between DFS and OS, and between the treatment effects on these two end points. The measure of treatment-level association (R^2) is equal to or above 0.75 in both the full and the reduced sets analyzed herein; this is a commonly used threshold for accepting the validity of a surrogate for OS.[19,30] On the other hand, the 95% CIs for these estimates are relatively wide, precluding any definitive conclusions. Unfortunately, models taking into account the magnitude of the estimation error in the estimated treatment effects on DFS and OS could not be fitted for the analyses having OS as the final end point, which may have led to biased estimates of the strength of the association between the treatment effects. On the other hand, the use of the model that

adjusted for the magnitude of the estimation error in the treatment estimates was possible for BCSS as the final end point. This model yielded an R^2 value of 0.95, lending additional support to the notion that recurrences are on the causal pathway to death in early breast cancer.

The chief limitation of this study is that regression analyses are very sensitive to outliers. In the current work, exclusion of one trial with short follow-up and few events, and with different censoring patterns during the first year of follow-up from some of the analyses indeed led to quantitatively different results than in the full set. Nevertheless, the values of R^2 equal to or above 0·75 in the full and reduced sets are reassuring in this regard. Additional limitations exist, one of which relating to our definition of trial eligibility. At the time the study was planned, and to this date in several countries, 1 year of trastuzumab remains the standard of care for HER-2-positive disease. As a result, our findings cannot be expanded to different settings, such as longer treatment with trastuzumab or extension of adjuvant therapy through the use of neratinib.[10] Another limitation relates to our analysis of BCSS, which has emerged as an endpoint more recently. Since BCSS had not been assessed systematically in most of the trials analyzed here, we used as a proxy the cases of death preceded by a recurrence. Thus, the analysis of BCSS remains exploratory, and it will be important to compare the predictive ability of DFS and BCSS when both have been collected systematically in a sufficient number of trials. Finally, the number of contrasts in the current work precludes meaningful analyses in subsets defined by different patient subgroups or trial types. Thus, given our inability to differentiate DFS from invasive DFS due to the heterogeneity of definitions across trials, our surrogacy results apply to DFS broadly defined, but future studies should try to compare the predictive ability of these two end

points. Likewise, future studies should investigate separately the trials that compared 1 year of trastuzumab versus other durations of the antibody, most of which are unavailable to us at present.[6,7,9,20-22]

The current results are in line with those obtained in colorectal, gastric and non-small-cell lung cancer, in which DFS was found to be an acceptable surrogate for OS in the adjuvant setting.[16-18] Similar conclusions were drawn for relapse-free survival in the adjuvant therapy of melanoma.[31] In those studies, the estimated values of R^2 ranged from 0·91 and 0·96 in their respective main analyses. Such values suggest stronger correlations between treatment effects in those settings than found here. Whether this is due to specific features of the trials analyzed, biological differences between these clinical settings, or the play of chance, remains speculative. Arguably, the efficacy of anti-HER-2 therapy in the metastatic setting, and the fact that patients with early breast cancer often die from other causes, may attenuate the association between treatment effects in breast cancer, in comparison with other settings. The analyses using a proxy for BCSS, which showed higher values of R^2 than the analyses using OS as the final end point, provides indirect support to this argument.

In summary, our results suggest that DFS can be used as a surrogate for OS in the adjuvant treatment of HER-2-positive, early breast cancer. These results, which apply mainly to the adjuvant use of trastuzumab for 12 months, indicate levels of association, both at the patient and at the trial level, that are desirable from the point of view of replacing a final endpoint such as OS.

Research in context

Evidence before this study

Although disease-free survival (DFS) has often been used as primary end point in adjuvant trials of breast cancer, it has not been formally validated as a surrogate for overall survival (OS). This meta-analysis was conducted to evaluate the role of DFS as a surrogate for OS in the adjuvant treatment of HER-2-positive breast cancer. In 10/2015 and 10/2016, several databases were used to search eligible randomized trials. Data were provided by investigators for all but one of the nine eligible trials. A two-level modeling approach was used to estimate the association between DFS and OS (patient-level association) and between the treatment effects on these end points (trial-level association). Patient-level associations were strong. Trial-level associations were moderate or strong, depending on the set analyzed.

Added value of this study

These results suggest that DFS can be used as a surrogate for OS in the adjuvant treatment of HER-2-positive, early breast cancer. These results apply mainly to the adjuvant use of trastuzumab for 12 months,

Implications of all the available evidence

The levels of association found here seem sufficient for the purpose of replacing a final endpoint such as OS. Further studies should assess DFS as a surrogate for OS in other breast-cancer phenotypes.

Acknowledgements

We thank Dr. Pascale Holzerny and Johannes Ziemen, from HealthEcon (Basel, Switzerland), for conducting the systematic search.

REFERENCES

1. Curigliano G, Burstein HJ, Winer EP, et al. De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. *Ann Oncol* 2017; **28**(8): 1700-12.

2. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* 2005; **353**(16): 1659-72.

3. Romond EH, Perez EA, Bryant J, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* 2005; **353**(16): 1673-84.

4. Slamon D, Eiermann W, Robert N, et al. Adjuvant trastuzumab in HER2-positive breast cancer. *N Engl J Med* 2011; **365**(14): 1273-83.

5. Joensuu H, Kellokumpu-Lehtinen PL, Bono P, et al. Adjuvant docetaxel or vinorelbine with or without trastuzumab for breast cancer. *N Engl J Med* 2006; **354**(8): 809-20.

6. Pivot X, Romieu G, Debled M, et al. 6 months versus 12 months of adjuvant trastuzumab for patients with HER2-positive early breast cancer (PHARE): a randomised phase 3 trial. *Lancet Oncol* 2013; **14**(8): 741-8.

7. Mavroudis D, Saloustros E, Malamos N, et al. Six versus 12 months of adjuvant trastuzumab in combination with dose-dense chemotherapy for women with HER2-positive breast cancer: a multicenter randomized study by the Hellenic Oncology Research Group (HORG). *Ann Oncol* 2015; **26**(7): 1333-40.

8. Cameron D, Piccart-Gebhart MJ, Gelber RD, et al. 11 years' follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive early breast cancer: final analysis of the HERceptin Adjuvant (HERA) trial. *Lancet* 2017; **389**(10075): 1195-205.

9. Piccart-Gebhart M, Holmes E, Baselga J, et al. Adjuvant Lapatinib and Trastuzumab for Early Human Epidermal Growth Factor Receptor 2-Positive Breast

Cancer: Results From the Randomized Phase III Adjuvant Lapatinib and/or Trastuzumab Treatment Optimization Trial. *J Clin Oncol* 2016; **34**(10): 1034-42.

10. Chan A, Buyse M, Yao B. Neratinib after trastuzumab in patients with HER2-positive breast cancer - Author's reply. *Lancet Oncol* 2016; **17**(5): e176-7.

11. von Minckwitz G, Procter M, de Azambuja E, et al. Adjuvant Pertuzumab and Trastuzumab in Early HER2-Positive Breast Cancer. *N Engl J Med* 2017; **377**(2): 122-31.

12. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology. Breast Cancer, version 4.2017. Available at http://www.nccn.org (Accessed 16 March 2018).

13. Burzykowski T, Saad ED, Buyse M. Adoption of Pathologic Complete Response as a Surrogate End Point in Neoadjuvant Trials in HER2-Positive Breast Cancer Still an Open Question. *JAMA Oncol* 2017; **3**(3): 416.

14. 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**(9938): 164-72.

15. Gill S, Sargent D. End points for adjuvant therapy trials: has the time come to accept disease-free survival as a surrogate end point for overall survival? *Oncologist* 2006; **11**(6): 624-9.

16. 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**(34): 8664-70.

17. Oba K, Paoletti X, Alberts S, et al. Disease-free survival as a surrogate for overall survival in adjuvant trials of gastric cancer: a meta-analysis. *J Natl Cancer Inst* 2013; **105**(21): 1600-7.

18. Mauguen A, Pignon JP, Burdett S, et al. Surrogate endpoints for overall survival in chemotherapy and radiotherapy trials in operable and locally advanced lung cancer: a re-analysis of meta-analyses of individual patients' data. *Lancet Oncol* 2013; **14**(7): 619-26.

19. Burzykowski T, Molenberghs G, Buyse M. The Evaluation of Surrogate Endpoints. New York, Springer, 2005.

20. Conte PF, Bisagni G, Frassoldati A, Brandes AA, Anselmi E, et al. 9 weeks vs 1 year adjuvant trastuzumab in combination with chemotherapy: Results of the phase III multicentric Italian study Short-HER. Journal of Clinical Oncology 35, no. 15_suppl (May 2017) 501-501.

21. Joensuu H, Fraser J, Wildiers H, Huovinen R, Auvinen P, Utriainen M, et al. A randomized phase III study of adjuvant trastuzumab for a duration of 9 weeks versus 1 year, combined with adjuvant taxane-anthracycline chemotherapy, for early HER2-positive breast cancer (the SOLD study). Cancer Res 2018;78(4 Supplement):GS3-04.

22. Earl HM, Hiller L, Vallier A-L, Loi S, Howe D, Higgins HB et al. PERSEPHONE: 6 versus 12 months (m) of adjuvant trastuzumab in patients (pts) with HER2 positive (+) early breast cancer (EBC): Randomised phase 3 non-inferiority trial with definitive 4-year (yr) disease-free survival (DFS) results. 506.

23. Singhal MK, Kapoor A, Narayan S, Maharia S, Nirban RK, Beniwal SK, Kumar HS. A phase III randomized study of paclitaxel and trastuzumab versus paclitaxel alone for early stage, ER and PR receptor negative and HER2-positive breast cancer as adjuvant treatment. Ann Oncol 26 (Supplement 3): iii1–iii3, 2015.

24. Gourgou-Bourgade S, Cameron D, Poortmans P, et al. Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative

(Definition for the Assessment of Time-to-event Endpoints in CANcer trials)dagger. *Ann Oncol* 2015; **26**(5): 873-9.

25. Schneider BP, O'Neill A, Shen F, et al. Pilot trial of paclitaxel-trastuzumab adjuvant therapy for early stage breast cancer: a trial of the ECOG-ACRIN cancer research group (E2198). *Br J Cancer* 2015; **113**(12): 1651-7.

26. Spielmann M, Roche H, Delozier T, et al. Trastuzumab for patients with axillary-node-positive breast cancer: results of the FNCLCC-PACS 04 trial. *J Clin Oncol* 2009; **27**(36): 6129-34.

27. Perez EA, Suman VJ, Davidson NE, et al. Sequential versus concurrent trastuzumab in adjuvant chemotherapy for breast cancer. *J Clin Oncol* 2011; **29**(34): 4491-7.

28. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharmaceutical statistics* 2006; **5**(3): 173-86.

29. Savina M, Jacot W, Mathoulin-Pélissier S, Laghzali Y, Bellera C, Gourgou S. Surrogate endpoints for Overall Survival in Randomized Controlled Trials evaluating adjuvant treatment for breast cancer: a meta-analysis. Annals of Oncology 2017; 28 (suppl_5): v43-v67. 10.1093/annonc/mdx362.

30. Lassere MN. The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. *Stat Methods Med Res* 2008; **17**(3): 303-40.

31. Suciu S, Eggermont AMM, Lorigan P, et al. Relapse-Free Survival as a Surrogate for Overall Survival in the Evaluation of Stage II-III Melanoma Adjuvant Therapy. *J Natl Cancer Inst* 2018; **110**(1).

Table 1. Selected characteristics of trials analyzed and designation of control and experimental arms for the current work.

Trial alias and registration nr.	N eligible patients	N eligible arms/contrasts	Events defining disease-free survival	Control arm(s) for current work	Experimental arm(s) for current work
ALTTO[9] NCT00490139	8,381	4/2	Recurrence of invasive breast cancer at local, regional, or distant sites; contralateral invasive breast cancer; second nonbreast malignancy ; or death from any cause.	(1) Chemotherapy[a] plus lapatinib for 12 months (2) Chemotherapy[a] plus trastuzumab for 12 months	(1) Chemotherapy[a] plus lapatininb and trastuzumab for 12 months (2) Chemotherapy[a] plus trastuzumab for 12 weeks and lapatinib for 34 weeks
BCIRG 006[4] NCT00021255	3,222	3/2	Breast-cancer recurrence, a second primary cancer (excluding contralateral ductal carcinoma *in situ*), or death from any cause.	Doxorubicin, cyclophosphamide and docetaxel	(1) Doxorubicin, cyclophosphamide, docetaxel and trastuzumab for 12 months (2) docetaxel, carboplatin and trastuzumab for 12 months
E2198[25] NCT 00003992	234	2/1	Disease recurrence, development of invasive second primary, or death.	Paclitaxel and trastuzumab for 12 weeks followed by doxorubicin and cyclophosphamide for four cycles	Same as control, plus trastuzumab for 12 months

HERA[2] NCT00045032	3,401	2/1	Recurrence of breast cancer at any site, development of ipsilateral or contralateral breast cancer (including ductal carcinoma *in situ* but not lobular carcinoma *in situ*), second nonbreast malignancy (other than basal-cell or squamous-cell carcinoma of the skin or carcinoma *in situ* of the cervix), or death from any cause.	Observation	Trastuzumab for 12 months
HORG[7] NCT00615602	481	2/1	Breast cancer recurrence (either locoregional or distant), contralateral breast cancer, second nonbreast malignancy, or death from any cause.	Dose-dense fluorouracil, epirubicin and cyclophosphamide followed by dose-dense docetaxel plus trastuzumab for 6 months	Dose-dense fluorouracil, epirubicin and cyclophosphamide followed by dose-dense docetaxel plus trastuzumab for 12 months

NCCTG N9831[3] NCT00005970	1,885	3/3	Local, regional, or distant recurrence; contralateral breast cancer, including ductal carcinoma *in situ*; other second primary cancers; or death before recurrence or a second primary cancer.	Doxorubicin and cyclophosphamide followed by weekly paclitaxel	(1) Same as control plus sequential trastuzumab for 12 months while arm C[b] was open (2) Same as control plus concurrent trastuzumab for 12 months (3) Same as control plus sequential trastuzumab for 12 months while arm C[b] was closed
NSABP B-31[3] NCT00004067	2,102	2/1	Local, regional, or distant recurrence; contralateral breast cancer, including ductal carcinoma *in situ*; other second primary cancers; or death before recurrence or a second primary cancer.	Doxorubicin and cyclophosphamide followed by paclitaxel every 3 weeks	Same as control, plus trastuzumab for 12 months
PACS 04[26] NCT00054587	527	2/1	Local or regional recurrence, distant metastases,	Fluorouracil, epirubicin and cyclophosphamide, or epirubicin and docetaxel	Same as control, plus trastuzumab for 12 months

			contralateral breast cancer, or death from any cause.		

[a]Chemotherapy of choice prior to randomization (design 1), anthracycline-based regimen followed by taxane and concurrent trial therapy (design 2A), or docetaxel, carboplatin and concurrent trial therapy (design 2B).

[b]This trial had a period during which accrual was closed to one of the arms (C, with concurrent chemotherapy and trastuzumab), thus generating the need to create contrasts that only had concurrently randomized patients.

ALTTO, Adjuvant Lapatinib And/Or Trastuzumab Treatment Optimisation trial; BCIRG, Breast Cancer International Research Group; E2198, ECOG-ACRIN Cancer Research Group trial 2198 (supported by the National Cancer Institute grant numbers CA180820 and CA180795); HERA, HERceptin Adjuvant trial; HORG, Hellenic Oncology Research Group; NCCTG, North Central Cancer Treatment Group; NSABP, National Surgical Adjuvant Breast and Bowel Project (funded by the National Cancer Institute under grants U10CA180868 and 180822); PACS, *Programmes d'Actions Concertées Sein*.

Table 2. Selected results for each contrast analyzed.

Trial alias	Contrast	N patients	Disease-free survival			Overall survival		
			Median follow-up	N events	HR	Median follow-up	N events	HR
ALTTO[9]	Chemotherapy plus lapatinib for 12 months vs chemotherapy plus lapatinib plus trastuzumab for 12 months	4,193	82	776	0·68	83	396	0·69
	Chemotherapy plus sequential trastuzumab/lapatinib vs chemotherapy plus trastuzumab for 12 months	4,188	83	730	1·08	83	362	1·16
BCIRG 006[4]	Doxorubicin, cyclophosphamide and docetaxel vs same chemotherapy plus trastuzumab for 12 months	1,611	126	419	0·68	126	247	0·62
	Doxorubicin, cyclophosphamide and docetaxel vs docetaxel, carboplatin and trastuzumab for 12 months	1,611	126	424	0·81	126	264	0·81
E2198[25]	Paclitaxel and trastuzumab for 12 weeks followed by doxorubicin and cyclophosphamide vs same as plus trastuzumab for 12 months	234	76	62	1·25	77	42	1·25
HERA[2]	Observation vs trastuzumab for 12 months	3,401	132	1,113	0·76	132	725	0·74

HORG[7]	Fluorouracil, epirubicin, cyclophosphamide and docetaxel plus trastuzumab for 6 months vs fluorouracil, epirubicin, cyclophosphamide and docetaxel plus trastuzumab for 12 months	481	49	45	0·63	50	18	1·45
NCCTG N9831[3]	Doxorubicin, cyclophosphamide and paclitaxel vs doxorubicin, cyclophosphamide and paclitaxel plus sequential trastuzumab for 12 months, arm C open	1,423	146	412	0·82	153	295	0·79
	Doxorubicin, cyclophosphamide and paclitaxel vs doxorubicin, cyclophosphamide and paclitaxel plus concurrent trastuzumab for 12 months	1,418	146	399	0·70	153	259	0·75
	Doxorubicin, cyclophosphamide and paclitaxel vs doxorubicin, cyclophosphamide and paclitaxel plus sequential trastuzumab for 12 months, arm C closed	291	165	105	0·56	175	75	0·73
NSABP B-31[3]	Doxorubicin, cyclophosphamide and paclitaxel vs same plus trastuzumab for 12 months	2,102	119	696	0·59	119	445	0·66
PACS 04[26]	Fluorouracil, epirubicin and cyclophosphamide, or epirubicin and docetaxel vs same plus trastuzumab for 12 months	527	112	190	0·76	113	106	0·81

HR, hazard ratio.

ALTTO, Adjuvant Lapatinib And/Or Trastuzumab Treatment Optimisation trial; BCIRG, Breast Cancer International Research Group; E2198, ECOG-ACRIN Cancer Research Group trial 2198 (supported by the National Cancer Institute grant numbers CA180820 and CA180795); HERA, HERceptin

Adjuvant trial; HORG, Hellenic Oncology Research Group; NCCTG, North Central Cancer Treatment Group; NSABP, National Surgical Adjuvant Breast and Bowel Project (funded by the National Cancer Institute under grants U10CA180868 and 180822); PACS, *Programmes d'Actions Concertées Sein*.

Table 3. Cross-validation using "leave-one-out" analysis (see text for explanations).

Analysis / contrast	Observed HR for OS	95% CI for the observed HR	Predicted HR for OS	95% CI for the predicted HR	Observed HR within prediction interval?
Full set (12 contrasts)					
ALTTO / L vs TL	0·69	[0V56;0.84]	0·71	[0·57;0·88]	Yes
ALTTO / T vs T→L	1·16	[0·94;1.43]	0·94	[0·72;1·24]	Yes
BCIRG 006 / ACT vs ACTH	0·62	[0·48;0.79]	0·72	[0·57;0·91]	Yes
BCIRG 006 / ACT vs TCH	0·81	[0·63;1.03]	0·82	[0·64;1·06]	Yes
E 2198	1·25	[0·68;2.29]	1·21	[0·64;2·30]	Yes
HERA	0·74	[0·64;0.86]	0·79	[0·68;0·93]	Yes
HORG	1·45	[0·57;3.67]	0·66	[0·32;1·34]	No
NCCTG N9831 / A vs B (C open)	0·79	[0·62;1.00]	0·84	[0·66;1·06]	Yes
NCCTG N9831 / A vs C	0·75	[0·58;0·96]	0·72	[0·56;0·93]	Yes
NCCTG N9831 / A vs B (C closed)	0·73	[0·46;1·16]	0·58	[0·38;0·91]	Yes
NSABP B-31	0·66	[0·54;0·79]	0·59	[0·47;0·75]	Yes
PACS-04	0·81	[0·55;1·19]	0·78	[0·53;1·15]	Yes
Reduced set (11 contrasts)					
ALTTO / L vs TL	0·69	[0.56;0·84]	0·70	[0·59;0·84]	Yes
ALTTO / T vs T→L	1·16	[0.94;1·43]	0·95	[0·77;1·17]	Yes
BCIRG 006 / ACT vs ACTH	0·62	[0.48;0·79]	0·71	[0·60;0·86]	Yes
BCIRG 006 / ACT vs TCH	0·81	[0.63;1·03]	0·82	[0·66;1·01]	Yes
E 2198	1·25	[0.68;2·29]	1·22	[0·71;2·08]	Yes
HERA	0·74	[0.64;0·86]	0·79	[0·69;0·90]	Yes
NCCTG N9831 / A vs B (C open)	0·79	[0.62;1·00]	0·84	[0·69;1·02]	Yes
NCCTG N9831 / A vs C	0·75	[0.58;0·96]	0·72	[0·58;0·88]	Yes
NCCTG N9831 / A vs B (C closed)	0·73	[0.46;1·16]	0·58	[0·41;0·81]	Yes
NSABP B-31	0·66	[0.54;0·79]	0·58	[0·49;0·70]	Yes
PACS-04	0·81	[0.55;1·19]	0·78	[0·56;1·07]	Yes

ACT, doxorubicin, cyclophosphamide and docetaxel; ACTH, doxorubicin, cyclophosphamide, docetaxel and trastuzumab; ALTTO, Adjuvant Lapatinib And/Or Trastuzumab Treatment Optimisation trial; BCIRG, Breast Cancer International Research Group; CI, confidence interval; DFS, disease-free survival; ECOG, Eastern Cooperative Oncology Group; HERA, HERceptin Adjuvant trial; HORG, Hellenic Oncology Research Group; HR, hazard ratio; L, lapatinib; NCCTG, North Central Cancer Treatment Group; NSABP, National Surgical Adjuvant Breast and Bowel Project; OS, overall survival; PACS, *Programmes d'Actions Concertées Sein*; T, trastuzumab; TCH,

docetaxel, carboplatin and trastuzumab; T→L, trastuzumab followed by lapatinib; TL, trastuzumab plus lapatinib.

Table 4. Surrogate threshold effects in the full and reduced sets (see text), according to the expected number of deaths in a future randomized trial.

Expected number of deaths	**Surrogate threshold effect**	
	Full set (12 trials)	**Reduced set (11 trials)**
100	0·59	0·62
200	0·69	0·71
400	0·77	0·79
800	0·82	0·84

Table 5. Comparison between observed and predicted[a] hazard ratios for overall survival based on published hazard ratio for disease-free survival in three trials with unavailable data.

Trial	Observed HR for DFS	Observed HR for OS	95% CI for the observed HR for OS	Predicted HR for OS	95% CI for the predicted HR for OS
PHARE[b]	1·28	1·46	[1·06;2·01]	1·25	[0·93;1·67]
Short-HER[c]	1·15	1·06	[0·73;1·55]	1·13	[0·78;1·64]
SOLD[c]	1·39	1·36	[0·98;1·89]	1·34	[0·94;1·91]

[a]Predicted hazard ratios are obtained using the regression equation shown in the text for the main analysis.

[b]Hazard ratios with 6 months of trastuzumab as experimental and 12 as control.

[c]Hazard ratios with 9 weeks of trastuzumab as experimental and 12 months as control.

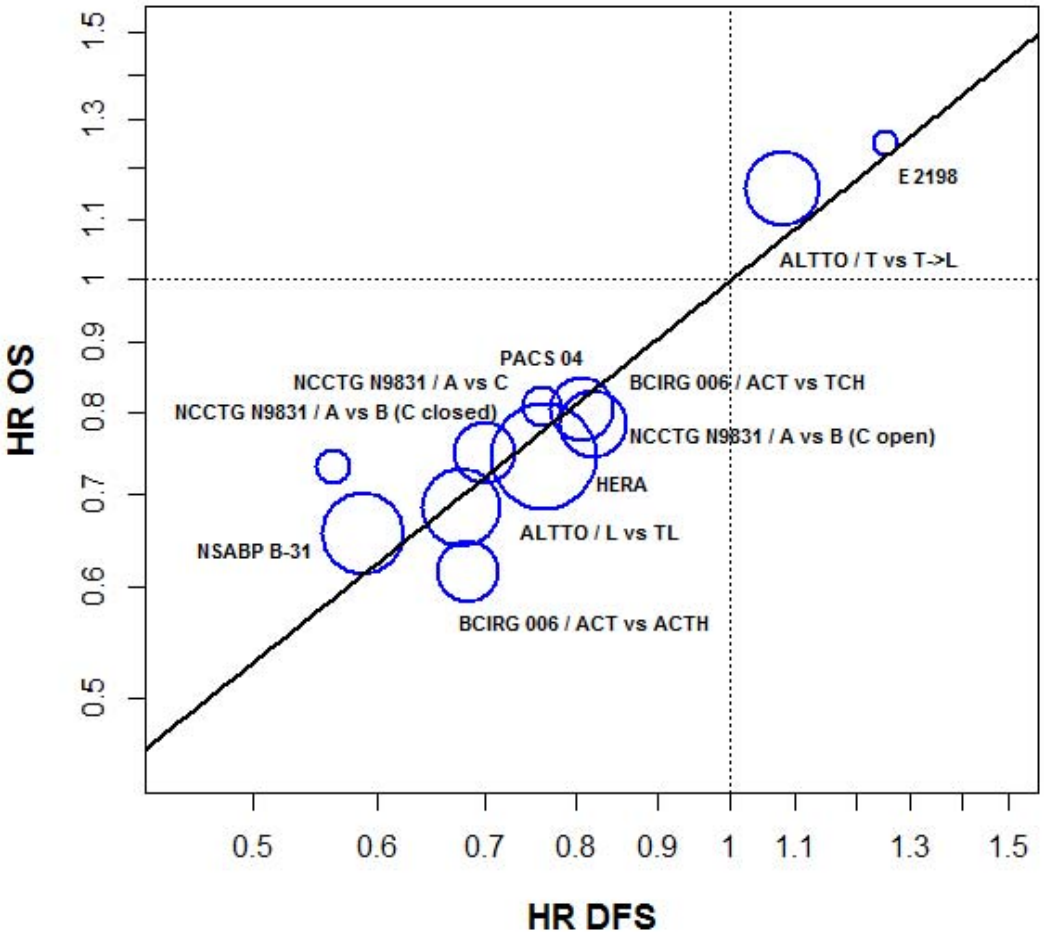
CI, confidence interval; DFS, disease-free survival; HR, hazard ratio; OS, overall survival; PHARE, Protocol of Herceptin Adjuvant with Reduced Exposure; SOLD, Synergism Or Long Duration.

FIGURE CAPTIONS

Figure 1. Trial-level association between the hazard ratio for disease-free survival and the hazard ratio for overall survival in each contrast of the reduced set, analysis weighted by the number of deaths in each contrast.

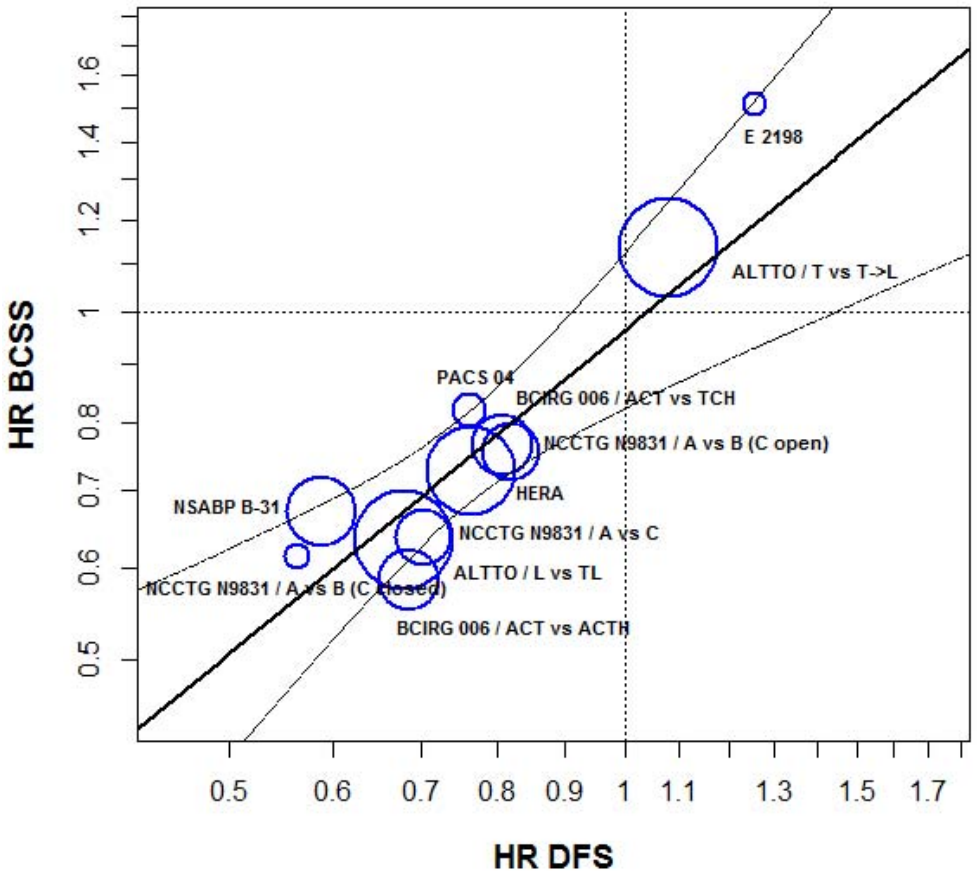
Figure 2. Trial-level association between the hazard ratio for disease-free survival and the hazard ratio for a proxy to breast-cancer-specific survival in each contrast of the reduced set, analysis adjusted for the magnitude of the estimation errors in the treatment effect estimates.

Figure 1



Each circle represents one contrast, with size proportional to the number of deaths.

Figure 2



Each circle represents one contrast, with size proportional to the number of deaths. The curved diagonal lines are the 95% prediction limits for the regression line.