



Published in final edited form as:

Int J Technol Assess Health Care. 2008 ; 24(4): 371–383. doi:10.1017/S0266462308080495.

Progression-free survival as a surrogate endpoint in advanced breast cancer

Rebecca A. Miksad, MD, MPH [Instructor in Medicine]

(rmiksad@bidmc.harvard.edu), Department of Medicine, Division of Hematology and Oncology, Beth Israel Deaconess Medical Center, Harvard Medical School, Shapiro 9, 330 Brookline Avenue, Boston, MA 02215; Senior Scientist, Institute for Technology Assessment and Department of Radiology, Massachusetts General Hospital, Harvard Medical School, 101 Merrimac Street, Boston, MA 02114

Vera Zietemann, MPH, PhD [Senior Scientist], Raffaella Gothe [Senior Scientist], Ruth Schwarzer, MA, MPH [Research Scientist], Annette Conrads-Frank, PhD [Senior Scientist], Petra Schnell-Inderst, MPH, PhD [Senior Scientist], and Björn Stollenwerk, PhD [Senior Scientist]

(publichealth@umit.at) Department of Public Health, Medical Decision Making and Health Technology Assessment, UMIT—University for Health Sciences, Medical Informatics and Technology, Eduard Wallnoefer Center I, Hall i.T., Austria, A-6060

Uwe Siebert, MD, MPH, MSc, ScD [Chair, Professor of Public Health (UMIT)]

(public-health@umit.at), Department of Public Health, Medical Decision Making and Health Technology Assessment, UMIT—University for Health Sciences, Medical Informatics and Technology, Eduard Wallnoefer Center I, Hall i.T., Austria, A-6060; Associate Professor of Radiology (Harvard University), Institute for Technology Assessment and Department of Radiology, Massachusetts General Hospital, Harvard Medical School, 101 Merrimac Street, Boston, MA 02114

Abstract

Objectives—Progression-free survival (PFS) has not been validated as a surrogate endpoint for overall survival (OS) for anthracycline (A) and taxane-based (T) chemotherapy in advanced breast cancer (ABC). Using trial-level, meta-analytic approaches, we evaluated PFS as a surrogate endpoint.

Methods—A literature review identified randomized, controlled A and T trials for ABC. Progression-based endpoints were classified by prospective definitions. Treatment effects were derived as hazard ratios for PFS (HR_{PFS}) and OS (HR_{OS}). Kappa statistic assessed overall agreement. A fixed-effects regression model was used to predict HR_{OS} from observed HR_{PFS} . Cross-validation was performed. Sensitivity and subgroup analyses were performed for PFS definition, year of last patient recruitment, line of treatment, and constant rate assumption.

Portions of this work were orally presented at the 29th Annual Meeting of the Society for Medical Decision Making on October 23, 2007 (Lee Lusted Prize for outstanding presentation) and at the Eastern Oncology Cooperative Group Young Investigator Symposium on November 11, 2007 (Clinical Research Award for outstanding presentation) and as a poster at the Annual European Congress of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) on October 21, 2007.

Results—Sixteen A and fifteen T trials met inclusion criteria, producing seventeen A ($n = 4,323$) and seventeen T ($n = 5,893$) trial-arm pairs. Agreement (kappa statistic) between the direction of HR_{OS} and HR_{PFS} was 0.71 for A ($p = .0029$) and 0.75 for T ($p = .0028$). While HR_{PFS} was a statistically significant predictor of HR_{OS} for both A ($p = .0019$) and T ($p = .012$), the explained variances were 0.49 (A) and 0.35 (T). In cross-validation, 97 percent of the 95 percent prediction intervals crossed the equivalence line, and the direction of predicted HR_{OS} agreed with observed HR_{OS} in 82 percent (A) and 76 percent (T). Results were robust in sensitivity and subgroup analyses.

Conclusions—This meta-analysis suggests that the trial-level treatment effect on PFS is significantly associated with the trial-level treatment effect on OS. However, prediction of OS based on PFS is surrounded with uncertainty.

Keywords

Progression-free survival; Surrogate endpoint; Meta-analytic validation; Advanced breast cancer; Overall survival

Surrogate endpoints are attractive in clinical trials when the primary endpoint is difficult to measure because of time, costs, the need to test multiple regimens, ethical considerations, or pressure from patient advocacy groups (27;41). In advanced breast cancer, prolongation of survival and symptom improvement are commonly accepted as evidence of clinical benefit and as appropriate primary endpoints (27;65). However, a validated surrogate endpoint allows for inferences about a treatment's benefit when primary endpoint data are not available (6;7;26). One potential surrogate endpoint for overall survival (OS) in advanced breast cancer is progression-free survival (PFS), a composite endpoint defined by the United States (U.S.) Food and Drug Administration (FDA) as the time from randomization to documented progression or death from any cause (25;65). However, the association between PFS and OS has not been systematically evaluated in advanced breast cancer for anthracycline and taxane-based chemotherapy.

PFS includes death as part of its composite endpoint— and, therefore, differs from other progression-based endpoints. For example, time to progression is defined by the U.S. FDA as the time from randomization until tumor progression (death is censored), and time to treatment failure is a composite endpoint defined as the time from randomization until the patient stops the trial treatment for any reason including progression, toxicity, or preference (death is censored) (25;65). The inclusion of death in PFS may be important in advanced breast cancer because tumor growth, whether or not it is documented, usually precedes death. In other words, background mortality contributes relatively little to the 2-year median survival for advanced breast cancer and death likely reflects the treatment's ability, or lack thereof, to control disease progression (30;64;65). Therefore, censored deaths may constitute informative censoring and may impact the ability of other progression-based endpoints to predict OS. Because PFS uniquely overcomes this limitation by including death, its ability to predict OS should be evaluated separately from other progression-based endpoints.

PFS has additional benefits as an endpoint because it is measured before postprogression treatments are initiated. Therefore, PFS data is not affected by postprotocol agents.

Furthermore, PFS is measured earlier and has a higher event frequency compared to OS. Consequently, PFS results may be available sooner and, potentially, with smaller and less costly trials (2;22;33).

However, PFS has several potential limitations that it shares with all progression-based endpoints. The need for frequent radiologic studies raises the potential for assessment bias and increases the complexity of data capture and validation (25;65). PFS may be sensitive to differences in the duration between assessments (25;47), unprotocolled assessments (25;47), variations in censoring (25;46), and the use of cytostatic agents (67). Furthermore, variability in PFS measurement may be magnified when it is used to predict OS (20;26). With multiple effective therapies available for advanced breast cancer, the link between PFS and OS may be distorted by nonprotocolled treatments given after the trial regimen (19;22;33).

Although PFS has not been systematically evaluated in advanced breast cancer for anthracycline and taxane-based chemotherapy, studies suggest that time to progression and response rate have a statistically significant association with OS (1;33). However, only 34 to 37 percent of OS variability was explained by variability in time to progression and response rate, respectively (1;33). An abstract suggested a similar result ($R^2 = .44$) for PFS for trials comparing taxane and anthracycline chemotherapy (13). However, cross-validation was not performed for this narrowly defined set of trials. Although there is controversy about the optimal approach for validating surrogate endpoints (12;15;42), one option is to evaluate performance at the trial level (33).

The primary aim of this study was to evaluate the association between the direction and magnitude of the treatment effect on PFS compared with the treatment effect on OS for anthracycline- and taxane-based chemotherapy regimens in advanced breast cancer using a trial-based, meta-analytic approach. We limited our analysis (and the generalization of our results) to anthracycline- and taxane-based chemotherapy for advanced breast cancer because the relationship between PFS and OS may differ for different anti-cancer agents and in different cancers due to underlying biologic factors, the interaction of each anti-cancer agent with each cancer, and differences in the availability of active subsequent line treatments (26).

METHODS

Literature Search, Inclusion Criteria, and Data Abstraction

A systematic literature search in Medline (January 2007) based on published search strategies (32;33) was performed to identify English-language publications of randomized controlled advanced breast cancer trials of anthracycline-based (5-fluorouracil, epirubicin, and cyclophosphamide [FEC]; and 5-fluorouracil, adriamycin, and cyclophosphamide [FAC]) or taxane-based (any combination) chemotherapy. Abstract databases for the American Society of Clinical Oncology and the San Antonio Breast Cancer Symposium annual meetings were also searched.

The inclusion criteria were as follows: (i) PFS data and OS endpoint data reported; (ii) evidence of adequate randomization and blinding; (iii) 80 percent of the sample with advanced breast cancer; (iv) at least one trial regimen included anthracycline- (FEC/FAC) or taxane-based chemotherapy (33). To maximize inclusion, all publications reporting any progression-based data were reviewed and endpoints were classified according to U.S. FDA definitions. Any publication reporting a progression-based endpoint that met the “strict” U.S. FDA definition of PFS (time from randomization to progression or death from any cause) (25;65) was included in the analysis, even if terminology in the publication differed.

Publications using PFS terminology but not providing a definition of the progression-based endpoint were included if there was no evidence of deviation from the U.S. FDA definition after careful review of the entire the publication. These trials were excluded from subgroup analysis of trials meeting the “strict” definition of PFS. Those trials whose endpoint consisted of both progression and death but had a minor deviation from the U.S. FDA definition were included in the primary analysis but excluded from the subgroup of trials meeting the “strict” PFS definition.

Classification of progression-based endpoints and data extraction was independently performed by at least two authors. All discrepancies were refereed by two authors (R.M., U.S.). The extracted data were: (i) definition of progressionbased endpoint; (ii) deviation from prospectively defined PFS definition; (iii) treatment regimen in each arm; (iv) median OS and PFS for each arm; (v) number of patients used to calculate PFS; (vi) year the last patient was recruited; (vii) line of trial treatment; (viii) published hazard ratios; and (ix) description of censoring technique(s) and therapies received subsequent to trial protocol. For trials with multiple publications, data from the longest follow-up period was used.

Statistical Methods

Calculation of Hazard Ratios—The hazard ratio (HR) for the treatment effect on OS (HR_{OS}) and on PFS (HR_{PFS}) was estimated by calculating the median OS and PFS ratios for each pair of trials arms. An exponential distribution was assumed for the survival function. The trial arm containing the chemotherapy of interest (anthracycline or taxane) was designated as Group 2 for that analysis: $HR_{OS} = \text{Median OS of Group 1} / \text{Median OS of Group 2}$ (Table 1). If a trial compared two different regimens for the chemotherapy of interest, the treatment arm most similar to the current U.S. FDA approved regimen was designated as Group 2. For trials with multiple arms, each pair of arms was evaluated and included in the analysis. The anthracycline and taxane trials were evaluated separately.

Agreement in Direction of PFS and OS Treatment Effects—For both anthracycline- and taxane-based chemotherapy trials, the level of agreement (kappa coefficient) between the direction of the HR_{PFS} and HR_{OS} was calculated along with a p value. Kappa tests the null hypothesis that there is no more agreement between HR_s than might occur by chance alone (4;37).

Prediction of the Magnitude of the Treatment Effect on OS—The explanatory power of the trial-level treatment effect on PFS for the trial-level treatment effect on OS was evaluated using a meta-analytic, fixedeffects weighted linear regression model: $\log_{10}(HR_{OS})$

$= \alpha + \beta * \log_{10}(\text{HR}_{\text{PFS}})$. The intercept term α was included to avoid spurious associations from forcing the regression through the origin and to facilitate comparison with prior studies (33). Each trial arm pair was weighted by the total number of patients assessed for PFS (usually the intent to treat population). For trials with multiple arms, each arm was downweighted (downweighting factor = number of independent trial arm pairs/number of total trial arm pairs) to adjust for multiple comparisons. The coefficient of determination (R^2) was calculated for each model to measure the proportion of variability in HR_{OS} explained by variability in HR_{PFS} .

Leave-one-out cross-validation—The validity of the regression model was tested using leave-one-out crossvalidation. The regression model was re-fitted with the N^{th} trial arm pair excluded. For each excluded trial arm pair, the observed HR_{PFS} was used in the refitted regression equation to calculate the predicted HR_{OS} and 95 percent prediction interval (5;23). This process was repeated for every trial and each predicted HR_{OS} was compared with the respective observed HR_{OS} .

Sensitivity Analyses and Subgroup Analyses—The robustness of the regression model was assessed by evaluating key parameters in repeat analyses: (i) when available, the published HR was substituted for the calculated HR; and (ii) only those trials explicitly meeting the strictest definition of PFS were evaluated.

The potential impact of therapies given after the trial regimen was assessed using two proxy measures: year of last patient entry (before versus after 1990) and line of trial therapy (first versus subsequent-line). These proxies were chosen because the numbers of active treatment options available after progression tend to be greater after first-line protocols and after 1990 (33). The regression model was refitted with an interaction term between HR_{PFS} and an indicator variable for the respective proxy.

All statistical analyses were performed using SAS 9.1 (SAS Institute Inc. Cary, NC, USA). A two-sided p value of $< .05$ was considered statistically significant.

RESULTS

Literature Search and Description of Included Studies

We identified 420 individual publications of anthracyclinebased (FEC/FAC) and 996 of taxane-based trials for advanced breast cancer (Figure 1). Seventy-three anthracycline and thirty-seven taxane publications reported a progressionbased endpoint and met other inclusion criteria. These publications were reviewed to identify trials with PFS data as defined by the U.S. FDA. Six anthracycline and three taxane publications were excluded because the trial was updated with a later publication included in the review. Forty-two percent (31/73) anthracycline and 30 percent (11/37) taxane publications did not provide a detailed description of how the progression-based endpoint was determined: of these publications 6 anthracycline and no taxane trial arm pairs reported PFS data for which scrutiny of the methods and results did not reveal obvious deviations from the U.S. FDA definition of PFS. (11;17;18;24;45;61) The majority of these publications (4/6) were

published in 1990 or earlier. These six trials were excluded in a sensitivity analysis of trials meeting the “strict” definition of PFS.

Of those publications providing a definition of the progression-based endpoint, 66 percent (28/42) of the anthracycline and 54 percent (14/26) of the taxane publications did not have a description consistent with U.S. FDA definitions and terminology. These publications were reclassified according to U.S. FDA endpoint definitions. The majority of re-classifications occurred from time to progression to PFS and from time to treatment failure to time to progression. The most common reason for exclusion from this analysis was that the endpoint did not include all causes of death and, therefore, did not meet the definition of PFS.

To maximize the number of studies in the analysis, four trials were included that reported a composite progressionbased endpoint of death and progression but differed from the strict PFS definition in the following ways: (i) all deaths were included except those occurring after a predefined time interval from date of last study treatment (36;58;68); (ii) all deaths were included except those not attributed to study drug or breast cancer (58;68); and (iii) the last follow-up visit was defined as a PFS event (57). These four trials were excluded in a sensitivity analysis of trials meeting the “strict” definition of PFS.

Ultimately, sixteen anthracycline and fifteen taxane trials reported PFS data and met the other inclusion criteria (Table 1). All included anthracycline trials evaluated firstline therapy and 53 percent (9/17) recruited the last patient after 1990. A total of seventeen trial-arm pairs ($n = 4,323$) were available to evaluate HR_{OS} and HR_{PFS} for anthracycline regimens. All included taxane trials recruited the last patient after 1990 and 47 percent (8/17) were second/subsequentline regimens. A total of 17 trial-arm pairs ($n = 5,893$) were available to evaluate HR_{OS} and HR_{PFS} for taxane regimens. The majority of trial-arm pairs exactly met the U.S. FDA (strict) definition of PFS: 59 percent (10/17) of anthracycline (FEC/FAC) and 71 percent (12/17) of taxane trial arm pairs.

Information about censoring techniques was presented for less than half of the trial arm pairs: 41 percent (7/17) of anthracycline (FEC/FAC) and 35 percent (6/17) of taxane trial arm pairs. Details about subsequent treatments were provided in 41 percent (7/17) of anthracycline and 53 percent (9/17) of taxane trial arm pairs. An HR estimate for OS and PFS was published for 12 percent (2/17) of anthracycline (2;35) and 35 percent (6/17) of taxane trial arm pairs (9;16;35;36;38;48).

Agreement in Direction of PFS and OS Treatment Effects

For both regimens, there was good agreement between the direction of PFS and OS (Table 2). The kappa statistic for the anthracycline trial arm pairs was 0.71 (95 percent confidence interval (CI), 0.36–1.00; $p = .0029$) and was 0.75 (95 percent CI, 0.42–1.00; $p = .0028$) for taxane trial arm pairs.

Prediction of the Magnitude of the Treatment Effect on OS

In the primary meta-analytic, fixed-effects regression analysis of anthracycline-based (FEC/FAC) chemotherapy, HR_{PFS} was a significant predictor for HR_{OS} ($p = .0019$), with an explained variance (R^2) of .49 (Table 3 and Figure 2). Similarly, HR_{PFS} was a significant

predictor for HR_{OS} ($p = .012$) for taxane-based chemotherapy with an R^2 of .35. Elimination of a potential outlier (11) did not change results.

Leave-One-Out Cross-Validation

All 95 percent prediction intervals for the predicted HR_{OS} based on the observed HR_{PFS} were wide and all but one in the taxane analysis crossed the equivalence (HR_{OS} = 1) line (Figure 3). All observed HR_{OS} fell within the 95 percent prediction intervals for the anthracyclines (FEC/FAC) trial arm pairs, and 88 percent (15/17) did so for the taxane trial arm pairs. However, the observed and predicted HR_{OS} fell on the opposite side of the equivalence (HR_{OS} = 1) line for 18 percent (3/17) of anthracycline and 24 percent (4/17) of taxane trial-arm pairs. The majority of the predicted HR_{OS} were closer to the equivalence line than the observed HR_{OS}: 82 percent (14/17) of taxane and 71 percent (12/17) of anthracycline trial-arm pairs.

Sensitivity Analyses and Subgroup Analyses

When two available published HR values for the anthracyclines were substituted for calculated HRs, the primary model R^2 increased to .56 and HR_{PFS} remained significant ($p = .0005$) (Table 3). In contrast, when six available published HR values for taxane trial-arm pairs were substituted, the R^2 decreased to .25 and the HR_{PFS} remained significant ($p = .040$).

To assess the impact of adhering to a strict definition of PFS based on information in the publication, the analysis was repeated after excluding those trial arm pairs that did not provide a definition of PFS or deviated from the strict PFS definition. PFS remained significant in both models, although the R^2 decreased to .46 ($p = .032$) for anthracyclines and increased to .59 for taxanes ($p = .0036$) (Table 3).

Analysis of the potential impact of subsequent-line treatments for anthracyclines showed a nonsignificant interaction ($p = .44$) between HR_{PFS} and year of last patient recruitment (before versus after 1990) (Table 3). The anthracycline subgroup of trials recruiting patients after 1990 had a lower R^2 (.34) than the primary model and HR_{PFS} was not statistically significant ($p = .13$). However, the anthracycline subgroup of trials recruiting the last patient before 1990 was similar to the primary model ($R^2 = .51$), and HR_{PFS} remained statistically significant ($p = .03$).

Similarly, there was a statistically nonsignificant interaction between line of trial treatment and HR_{PFS} for taxanes ($p = .60$). However, PFS remained a significant predictor in each taxane subgroup ($p = .022$ and $.038$ for eight first-line and nine subsequent-line trials, respectively). Each taxane subgroup model had a higher R^2 than the primary model: .61 and .48 for first-line and subsequent-line, respectively.

DISCUSSION

Based on our meta-analytic evaluation of anthracycline(FEC/FAC) and taxane-based chemotherapy trials in advanced breast cancer, there is a statistically significant association between both the direction and the magnitude of the trial-level treatment effect on PFS and

the trial-level treatment effect on OS. However, prediction of OS based on PFS is surrounded with uncertainty. Although the direction of the observed HR_{PFS} and the observed HR_{OS} showed good agreement by the kappa statistic (4), 18 percent (anthracyclines) to 24 percent (taxanes) of the predicted HR_{OS} fell on the opposite side of the equivalence line compared with the observed HR_{OS} (i.e., the converse conclusion with respect to which regimen is superior). All of the 95 percent prediction intervals were wide and the majority crossed the HR equivalence line. Approximately half (anthracycline) to one-third (taxanes) of the variance in the treatment effect on OS is explained by the variance in the treatment effect on PFS. In light of our results, the finding of a significant PFS but an insignificant OS in a recent large advanced breast cancer trial (39) published after the conclusion of our study is not unexpected.

The results of this study were robust in cross-validation as well as in sensitivity analyses. Using limited data, heterogeneity of results could not be explained by the constant rate assumption, differences in PFS definition, year of last patient recruitment, or line of therapy. However, exploratory subgroup analyses indicated that PFS may perform better for anthracycline trials that recruited the last patient before 1990 and for taxanes trials stratified by line of treatment.

This study adds to and expands prior work on surrogate endpoint validation (1;13;33). We (i) systematically classified all published trial-level progression-based endpoints of anthracycline- and taxane-based (FEC/FAC) advanced breast cancer trials according to U.S. FDA and EMEA definitions; (ii) assessed trial-level PFS as a surrogate for taxane-based chemotherapy in advanced breast cancer using meta-analytic, trial-level regression analysis techniques; and (iii) evaluated trial-level PFS for taxane and anthracycline trials with leave-one-out crossvalidation and sensitivity analyses. We extended prior work on trial-level PFS for anthracycline-based (FEC/FAC) chemotherapy (33) by including additional trials, verifying endpoint definitions and performing cross-validation. We also enhanced the generalizability of prior work using individual level data (14) by not limiting our evaluation to those trials with this type of data available. The R^2 findings of this study for PFS are consistent with prior studies of other progression-based surrogate endpoints in advanced breast cancer, including the recent individual level analysis with fewer trials (1;13;14;33). We extended the prior work by performing a cross-validation to assess the robustness of the results. Our findings for PFS as a surrogate endpoint differ from those found for colon cancer, suggesting that different diseases, and, perhaps, different agent types, need separate evaluation (54;55;62). Future work on surrogate endpoints for different types of chemotherapy agents and different types of cancers may further elucidate the factors influencing the relationship of PFS and OS.

This study also highlights the variability in terminology and definition for progression-based endpoints such as PFS. Although most of the trials without a definition of PFS were published in 1990 or earlier, half of the trials with minor deviations from the U.S. FDA definition were published after 2000. While unique logistical and biologic considerations (and pretrial discussions with regulatory agencies) may have led to the definition choice, we believe it is important that the health-services research and clinical community be aware that these differences exist.

Our study must be considered within the context of its limitations. To focus on PFS as a unique endpoint differing from other progression-based endpoints, only a modest number of trials met eligibility criteria and power was limited. A systematic review of the published literature and abstract databases was performed to maximize the number of eligible trials and to minimize potential publication bias. In addition, publications of all trials with progression-based endpoints were reviewed to determine whether the endpoint met the U.S. FDA definition of PFS, even if this terminology was not used in the publication. Although individual level data were not available and we used median survival and PFS values, we followed the work of Parmer et al. to develop hazard ratios based on summary statistics (49). A sensitivity analysis was performed to assess the potential limitation of using calculated HRs.

Variability in the definition of and the use of progressionbased endpoints in the original trials may have introduced heterogeneity into our analysis. To limit potential heterogeneity, each trial was classified according to the U.S. FDA definitions, which were prospectively established. A sensitivity analysis was performed to determine whether exclusion of trials with differences from the strict definition of PFS changed results. Because some trials did not publish PFS data, the number of trials eligible for inclusion in this triallevel analysis was limited. Although great care was taken to include only trials with PFS data, it was impossible to discern if evaluation for disease progression occurred at similar time points for each arm of the trial. Therefore, variability in evaluation schedules between arms may have altered the observed PFS treatment differences. Insufficient data limited analysis of the impact of censoring, the radiologic standard used to determine progression, and follow-up periods: If the true PFS and OS length was not reached due to short follow-up, the predictive value of PFS may be underestimated. Finally, PFS is not the only potential composite endpoint for breast cancer. In contrast to PFS, other composite endpoints such as QTWiST give differential weights to fatal and nonfatal endpoints and may have a different relationship to OS (28;29).

Some authors argue that the association between PFS and OS may be underestimated because of numerous subsequent-line therapies and cross-over from the control arm to active treatment (56). Nonrandomized use of secondline agents may alter the relationship between PFS and OS by distorting the causal pathway between the intervention and the primary endpoint. If the use of subsequent-line treatments is random with respect to PFS (nondifferential bias), the performance of PFS as a surrogate may be diminished. However, if the use of subsequent-line treatments is not random with respect to PFS or is protocolled conditional on PFS (differential bias), then the relationship between PFS and OS reflects the entire treatment strategy and the potential bias may be in either direction. Therefore, it is possible that differences in the predictive ability of HR_{PFS} are due to differences in the impact of subsequent therapies. Because most trial publications did not directly report the use of subsequent line or cross-over therapies, we used proxy variables for the availability of subsequent-line therapies to explore this issue.

Although prediction of OS from PFS was surrounded by uncertainty, PFS is a statistically significant predictor in our trial-based meta-analytic regression model and progression is also a biologically plausible precursor to death. To reduce uncertainty about PFS as a

surrogate, the hypothesis that subsequent-line agents or changes in supportive care over time impact the ability of PFS to reliably predict OS could be directly tested. As suggested by a recent editorial, PFS is a biologically plausible measure of efficacy and a goal of breast cancer treatment is to improve both quality and quantity of life (56). Therefore, PFS may be an appropriate primary endpoint even if is not a formal surrogate endpoint with an ability to predict OS (56). Analyzing the association between PFS and quality-of-life and/or freedom from symptom progression data would further test the value of PFS as an independent, clinically meaningful primary endpoint.

The high percentage of progression-based endpoints that were re-classified to meet U.S. FDA and EMEA definitions and the sensitivity analysis results suggest that the definition of PFS should be universal and carefully detailed in trial protocols and publications. The U.S. FDA and EMEA draft proposals regarding surrogate endpoints have the same PFS definition. However, recommendations for censoring, determination of the date of progression, and missed visits differ (25;65). The Consolidated Standards of Reporting Trials (CONSORT) statement (40) and the Common Terminology Criteria for Adverse Events (63) may provide a model for standardizing PFS definition, measurement, and reporting.

POLICY IMPLICATIONS

This meta-analytic, trial-based analysis of anthracycline and taxane-based chemotherapy for advanced breast cancer suggests that while the trial-level treatment effect on PFS is significantly associated with the trial-level treatment effect on OS, predictions based on trial-level PFS are surrounded with uncertainty for these treatments in advanced breast cancer. However, use of standardized endpoint definitions may increase the reliability and validity of surrogate endpoint data in advanced breast cancer.

Acknowledgments

This study was in part funded by an unrestricted educational grant from Hoffmann La-Roche Ltd., Basel, Switzerland. The authors had complete and independent control over the following: design and conduct of the study; collection, management, analysis and interpretation of the data; and preparation, review, and publication of the manuscript. Dr. Miksad also received funding as a research fellow with the United States National Institutes of Health's Program in Cancer Outcomes Research Training [R25-CA92203] at the Institute for Technology Assessment at the Massachusetts General Hospital. We acknowledge the assistance of Allan Hackshaw and Davina Ghersi for providing detailed information about their work.

REFERENCES

1. A'Hern RP, Ebbs SR, Baum MB. Does chemotherapy improve survival in advanced breast cancer? A statistical overview. *Br J Cancer*. 1988; 57:615–618. [PubMed: 3044434]
2. Ackland SP, Anton A, Breitbach GP, et al. Dose-intensive epirubicin-based chemotherapy is superior to an intensive intravenous cyclophosphamide, methotrexate, and fluorouracil regimen in metastatic breast cancer: A randomized multinational study. *J Clin Oncol*. 2001; 19:943–953. [PubMed: 11181656]
3. Aisner J, Cirrincione C, Perloff M, et al. Combination chemotherapy for metastatic or recurrent carcinoma of the breast—A randomized phase III trial comparing CAF versus VATH versus VATH alternating with CMFVP: Cancer and Leukemia Group B Study 8281. *J Clin Oncol*. 1995; 13:1443–52. [PubMed: 7751891]
4. Altman, DG. *Practical statistics for medical research*. 1st. Chapman and Hall; New York: 1991.

5. Armitage, P.; Berry, G.; Matthews, JNS. *Statistical methods in medical research*. 4th. Blackwell Science; Malden, MA: 2001.
6. Baker SG, Kramer BS. A perfect correlate does not a surrogate make. *BMC Med Res Methodol*. 2003; 3:16. [PubMed: 12962545]
7. Baker SG. Surrogate endpoints: Wishful thinking or reality? *J Natl Cancer Inst*. 2006; 98:502–503. [PubMed: 16622112]
8. Bennett JM, Muss HB, Doroshow JH, et al. A randomized multicenter trial comparing mitoxantrone, cyclophosphamide, and fluorouracil with doxorubicin, cyclophosphamide, and fluorouracil in the therapy of metastatic breast carcinoma. *J Clin Oncol*. 1988; 6:1611–1620. [PubMed: 3049953]
9. Biganzoli L, Cufer T, Bruning P, et al. Doxorubicin and paclitaxel versus doxorubicin and cyclophosphamide as first-line chemotherapy in metastatic breast cancer: The European Organization for Research and Treatment of Cancer 10961 Multicenter Phase III Trial. *J Clin Oncol*. 2002; 20:3114–3121. [PubMed: 12118025]
10. Bishop JF, Dewar J, Toner GC, et al. Initial paclitaxel improves outcome compared with CMFP combination chemotherapy as front-line therapy in untreated metastatic breast cancer. *J Clin Oncol*. 1999; 17:2355–2364. [PubMed: 10561297]
11. Boccardo F, Rubagotti A, Rosso R, Santi L. Chemotherapy with or without tamoxifen in postmenopausal patients with late breast cancer. A randomized study. *J Steroid Biochem*. 1985; 23:1123–1127. [PubMed: 3912616]
12. Burzykowski T, Molenberghs G, Buyse M. The validation of surrogate end points by using data from randomized clinical trials: A case-study in advanced colorectal cancer. *J R Stat Soc Ser A Stat Soc*. 2004; 167:103–124.
13. Burzykowski, T.; Piccart, MJ.; Sledge, G., et al. A quantitative study of tumor response and progression-free survival as surrogate endpoints for overall survival in first-line treatment of metastatic breast cancer; 28th Annual San Antonio Breast Cancer Symposium; San Antonio, Texas, United States. 2005. Abstract 6084
14. Burzykowski T, Buyse M, Piccart-Gebhart MJ, et al. Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. *J Clin Oncol*. 2008; 26:1987–1992. [PubMed: 18421050]
15. Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics*. 1998; 54:1014–1029. [PubMed: 9840970]
16. Chan S, Friedrichs K, Noel D, et al. Prospective randomized trial of docetaxel versus doxorubicin in patients with metastatic breast cancer. *J Clin Oncol*. 1999; 17:2341–2354. [PubMed: 10561296]
17. Conte PF, Pronzato P, Rubagotti A, et al. Conventional versus cytotoxic polychemotherapy with estrogenic recruitment in metastatic breast cancer: Results of a randomized cooperative trial. *J Clin Oncol*. 1987; 5:339–347. [PubMed: 3546611]
18. Conte PF, Baldini E, Gardin G, et al. Chemotherapy with or without estrogenic recruitment in metastatic breast cancer. A randomized trial of the Gruppo Oncologico Nord Ovest (GONO). *Ann Oncol*. 1996; 7:487–490. [PubMed: 8839903]
19. De Gruttola V, Fleming T, Lin DY, Coombs R. Perspective: Validating surrogate markers—are we being naive? *J Infect Dis*. 1997; 175:237–246. [PubMed: 9203643]
20. De Gruttola VG, Clax P, DeMets DL, et al. Considerations in the evaluation of surrogate endpoints in clinical trials. Summary of a National Institutes of Health workshop. *Control Clin Trials*. 2001; 22:485–502. [PubMed: 11578783]
21. Del Mastro L, Venturini M, Lionetto R, et al. Accelerated-intensified cyclophosphamide, epirubicin, and fluorouracil (CEF) compared with standard CEF in metastatic breast cancer patients: Results of a multicenter, randomized phase III study of the Italian Gruppo Oncologico Nord-Ovest-Mammella Inter Gruppo Group. *J Clin Oncol*. 2001; 19:2213–2221. [PubMed: 11304774]
22. Di Leo A, Bleiberg H, Buyse M. Overall survival is not a realistic end point for clinical trials of new drugs in advanced solid tumors: A critical assessment based on recently reported phase III trials in colorectal and breast cancer. *J Clin Oncol*. 2003; 21:2045–2047. [PubMed: 12743164]

23. Efron B. Estimating the error rate of a prediction rule: Improvement on cross-validation. *J Am Stat Assoc.* 1983; 78:316–331.
24. Ejlersten B, Pfeiffer P, Pedersen D, et al. Decreased efficacy of cyclophosphamide, epirubicin and 5-fluorouracil in metastatic breast cancer when reducing treatment duration from 18 to 6 months. *Eur J Cancer.* 1993; 29A:527–531. [PubMed: 8435205]
25. European Medicines Agency. Appendix 1 to the guidelines on the evaluation of anticancer medicinal products in man: Methodological considerations for using Progression Free Survival (PFS) as primary endpoint in confirmatory trials for registration [Draft]. Dec. 2006 www.emea.europa.eu/pdfs/human/ewp/26757506en.pdf.
26. Fleming TR, DeMets DL. Surrogate end points in clinical trials: Are we being misled? *Ann Intern Med.* 1996; 125:605–613. [PubMed: 8815760]
27. Fleming TR. Surrogate endpoints and FDA's accelerated approval process. *Health Aff (Millwood).* 2005; 24:67–78. [PubMed: 15647217]
28. Gelber RD, Goldhirsch A. A new endpoint for the assessment of adjuvant therapy in postmenopausal women with operable breast cancer. *J Clin Oncol.* 1986; 4:1772–1779. [PubMed: 3783203]
29. Gelber RD, Gelber S. Quality-of-life assessment in clinical trials. *Cancer Treat Res.* 1995; 75:225–246. [PubMed: 7640165]
30. Gennari A, Conte P, Rosso R, et al. Survival of metastatic breast carcinoma patients over a 20-year period: A retrospective analysis based on individual patient data from six consecutive studies. *Cancer.* 2005; 104:1742–1750. [PubMed: 16149088]
31. Gennari A, Amadori D, De Lena M, et al. Lack of benefit of maintenance paclitaxel in first-line chemotherapy in metastatic breast cancer. *J Clin Oncol.* 2006; 24:3912–3918. [PubMed: 16921042]
32. Ghersi D, Wilcken N, Simes RJ. A systematic review of taxane-containing regimens for metastatic breast cancer. *Br J Cancer.* 2005; 93:293–301. [PubMed: 16052223]
33. Hackshaw A, Knight A, Barrett-Lee P, Leonard R. Surrogate markers and survival in women receiving first-line combination anthracycline chemotherapy for advanced breast cancer. *Br J Cancer.* 2005; 93:1215–1221. [PubMed: 16278665]
34. Jassem J, Pienkowski T, Pluzanska A, et al. Doxorubicin and paclitaxel versus fluorouracil, doxorubicin, and cyclophosphamide as first-line therapy for women with metastatic breast cancer: Final results of a randomized phase III multicenter trial. *J Clin Oncol.* 2001; 19:1707–1715. [PubMed: 11251000]
35. Jassem, J.; Pluzanska, A.; Pienkowski, T., et al. Randomized phase III multicenter trial of doxorubicin and paclitaxel (AT) versus fluorouracil-doxorubicin-cyclophosphamide (FAC) as first-line therapy metastatic breast cancer (MBC): Long-term efficacy and external review of results. *San Antonio Breast Cancer Symposium; San Antonio, Texas, USA: 2004. Abstract 5043*
36. Jones SE, Erban J, Overmoyer B, et al. Randomized phase III study of docetaxel compared with paclitaxel in metastatic breast cancer [see comment]. *J Clin Oncol.* 2005; 23:5542–5551. [PubMed: 16110015]
37. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33:159–174. [PubMed: 843571]
38. Langley RE, Carmichael J, Jones AL, et al. Phase III trial of epirubicin plus paclitaxel compared with epirubicin plus cyclophosphamide as first-line chemotherapy for metastatic breast cancer: United Kingdom National Cancer Research Institute trial AB01. *J Clin Oncol.* 2005; 23:8322–8330. [PubMed: 16293863]
39. Miller K, Wang M, Gralow J, et al. Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. *N Engl J Med.* 2007; 357:2666–2676. [PubMed: 18160686]
40. Moher D, Schulz KF, Altman D. The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA.* 2001; 285:1987–1991. [PubMed: 11308435]
41. Molenberghs G, Burzykowski T, Alonso A, Buyse M. A perspective on surrogate endpoints in controlled clinical trials. *Stat Methods Med Res.* 2004; 13:177–206. [PubMed: 15198486]

42. Molenberghs, G.; Buyse, ME.; Burzykowski, T. The history of surrogate endpoint validation. In: Molenberghs, G.; Buyse, ME.; Burzykowski, T., editors. *Statistics for biology and health: The evaluation of surrogate endpoints*. New York: Springer: 2005. p. 67-82.
43. Nabholz JM, Gelmon K, Bontenbal M, et al. Multicenter, randomized comparative study of two doses of paclitaxel in patients with metastatic breast cancer [see comment]. *J Clin Oncol*. 1996; 14:1858–1867. [PubMed: 8656254]
44. Nabholz JM, Senn HJ, Bezwoda WR, et al. Prospective randomized trial of docetaxel versus mitomycin plus vinblastine in patients with metastatic breast cancer progressing despite previous anthracycline-containing chemotherapy. 304 Study Group. *J Clin Oncol*. 1999; 17:1413–1424. [PubMed: 10334526]
45. Namer M, Soler-Michel P, Turpin F, et al. Results of a phase III prospective, randomised trial, comparing mitoxantrone and vinorelbine (MV) in combination with standard FAC/FEC in front-line therapy of metastatic breast cancer. *Eur J Cancer*. 2001; 37:1132–1140. [PubMed: 11378344]
46. Niimi M, Yamamoto S, Fukuda H, et al. The Influence of handling censored data on estimating progression-free survival in cancer clinical trials (JCOG9913-A). *Jpn J Clin Oncol*. 2002; 32:19–26. [PubMed: 11932358]
47. Panageas KS, Ben-Porat L, Dickler MN, et al. When you look matters: The effect of assessment schedule on progression-free survival. *J Natl Cancer Inst*. 2007; 99:428–432. [PubMed: 17374832]
48. Paridaens R, Biganzoli L, Bruning P, et al. Paclitaxel versus doxorubicin as first-line single-agent chemotherapy for metastatic breast cancer: A European Organization for Research and Treatment of Cancer Randomized Study with cross-over. *J Clin Oncol*. 2000; 18:724–733. [PubMed: 10673513]
49. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med*. 1998; 17:2815–2834. [PubMed: 9921604]
50. Parnes HL, Cirrincione C, Aisner J, et al. Phase III study of cyclophosphamide, doxorubicin, and fluorouracil (CAF) plus leucovorin versus CAF for metastatic breast cancer: Cancer and Leukemia Group B 9140. *J Clin Oncol*. 2003; 21:1819–1824. [PubMed: 12721259]
51. Perry MC, Kardinal CG, Korzun AH, et al. Chemohormonal therapy in advanced carcinoma of the breast: Cancer and Leukemia Group B protocol 8081. *J Clin Oncol*. 1987; 5:1534–1545. [PubMed: 3655856]
52. Pierga JY, Jouve M, Asselain B, et al. Randomized trial comparing two different modalities of administration of the same cytotoxic drugs in metastatic breast cancer. *J Infus Chemother*. 1995; 5:197–200. [PubMed: 8934727]
53. Pierga JY, Jouve M, Asselain B, et al. Randomized trial comparing protracted infusion of 5-fluorouracil with weekly doxorubicin and cyclophosphamide with a monthly bolus FAC regimen in metastatic breast carcinoma (SPM90). *Br J Cancer*. 1998; 77:1474–1479. [PubMed: 9652764]
54. 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; 23:8664–8670. [PubMed: 16260700]
55. Sargent DJ, Patiyil S, Yothers G, et al. End points for colon cancer adjuvant trials: Observations and recommendations based on individual patient data from 20,898 patients enrolled onto 18 randomized trials from the ACCENT Group. *J Clin Oncol*. 2007; 25:4569–4574. [PubMed: 17876008]
56. Sargent DJ, Hayes DF. Assessing the measure of a new drug: Is survival the only thing that matters? *J Clin Oncol*. 2008; 26:1922–1923. [PubMed: 18421044]
57. Sjostrom J, Blomqvist C, Mouridsen H, et al. Docetaxel compared with sequential methotrexate and 5-fluorouracil in patients with advanced breast cancer after anthracycline failure: A randomised phase III study with crossover on progression by the Scandinavian Breast Group. *Eur J Cancer*. 1999; 35:1194–1201. [PubMed: 10615229]
58. Sledge GW, Neuberg D, Bernardo P, et al. Phase III trial of doxorubicin, paclitaxel, and the combination of doxorubicin and paclitaxel as front-line chemotherapy for metastatic breast cancer: An intergroup trial (E1193). *J Clin Oncol*. 2003; 21:588–592. [PubMed: 12586793]

59. Sledge GW Jr, Hu P, Falkson G, et al. Comparison of chemotherapy with chemohormonal therapy as first-line therapy for metastatic, hormone-sensitive breast cancer: An Eastern Cooperative Oncology Group study. *J Clin Oncol.* 2000; 18:262–266. [PubMed: 10637238]
60. Smith RE, Brown AM, Mamounas EP, et al. Randomized trial of 3-hour versus 24-hour infusion of high-dose paclitaxel in patients with metastatic or locally advanced breast cancer: National Surgical Adjuvant Breast and Bowel Project Protocol B-26. *J Clin Oncol.* 1999; 17:3403–3411. [PubMed: 10550134]
61. Speyer JL, Green MD, Zeleniuch-Jacquotte A, et al. ICRF-187 permits longer treatment with doxorubicin in women with breast cancer. *J Clin Oncol.* 1992; 10:117–127. [PubMed: 1727913]
62. Tang PA, Bentzen SM, Chen EX, Siu LL. Surrogate end points for median overall survival in metastatic colorectal cancer: literature-based analysis from 39 randomized controlled trials of first-line chemotherapy. *J Clin Oncol.* 2007; 25:4562–4568. [PubMed: 17876010]
63. Trotti A, Colevas AD, Setser A, et al. CTCAE v3.0: Development of a comprehensive grading system for the adverse effects of cancer treatment. *Semin Radiat Oncol.* 2003; 13:176–181. [PubMed: 12903007]
64. V. S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) & Center for Biologics Evaluation and Research (CBER). Oncologic Drugs Advisory Committee Meeting; Dec 16. 2003
65. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) & Center for Biologics Evaluation and Research (CBER). Guidance for Industry. Clinical trial endpoints for the approval of cancer drugs and biologics. 2005:1–26. [Draft Guidance].
66. Winer EP, Berry DA, Woolf S, et al. Failure of higher-dose paclitaxel to improve outcome in patients with metastatic breast cancer: Cancer and leukemia group B trial 9342. *J Clin Oncol.* 2004; 22:2061–2068. [PubMed: 15169793]
67. Yu RX, Holmgren E. Endpoints for agents that slow tumor growth. *Contemp Clin Trials.* 2007; 28:18–24. [PubMed: 16829208]
68. Zielinski C, Beslija S, Mrsic-Krmpotic Z, et al. Gemcitabine, epirubicin, and paclitaxel versus fluorouracil, epirubicin, and cyclophosphamide as first-line chemotherapy in metastatic breast cancer: A Central European Cooperative Oncology Group International, multicenter, prospective, randomized phase III trial. *J Clin Oncol.* 2005; 23:1401–1408. [PubMed: 15735116]

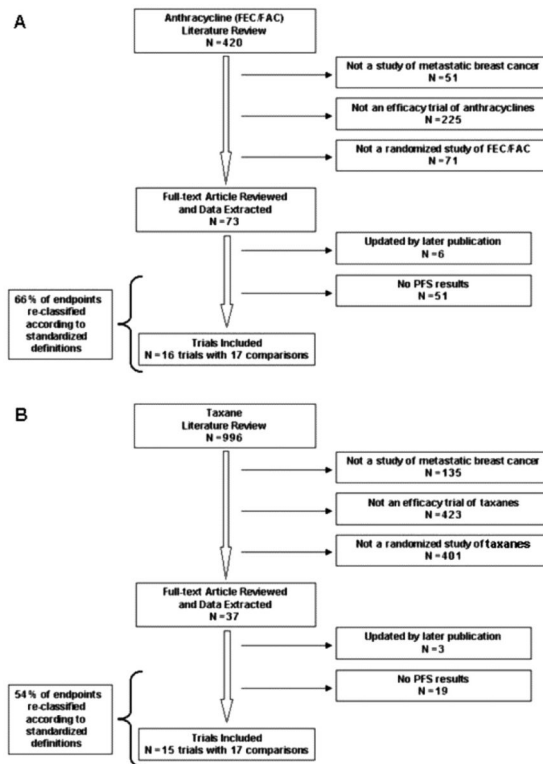


Figure 1. Systematic literature review for anthracycline (A) and taxane-based chemotherapy (B) for advanced breast cancer. Results of systematic review of the literature. FEC, 5-fluorouracil, epirubicin and cyclophosphamide; FAC, 5-fluorouracil, adriamycin and cyclophosphamide; N = number.

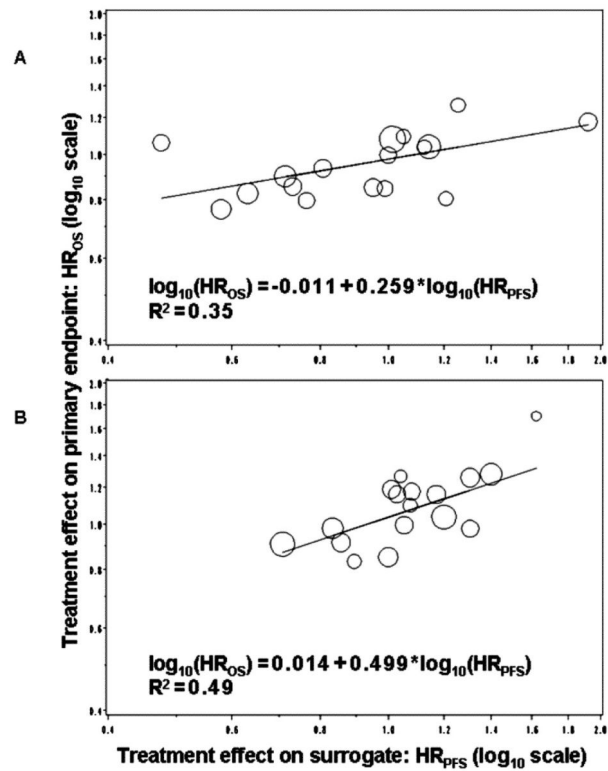


Figure 2.

Primary meta-analytic, fixed-effects regression model for anthracycline (A) and taxane-based chemotherapy (B). Primary, fixed-effects meta-analytic regression analysis assessing the association of HR_{PFS} and HR_{OS} . Size of circle indicates relative sample size of each trial-arm pair. Regression equation noted in figure, along with R^2 value. HR, hazard ratio; OS, overall survival; PFS, progression free survival; log, logarithm.

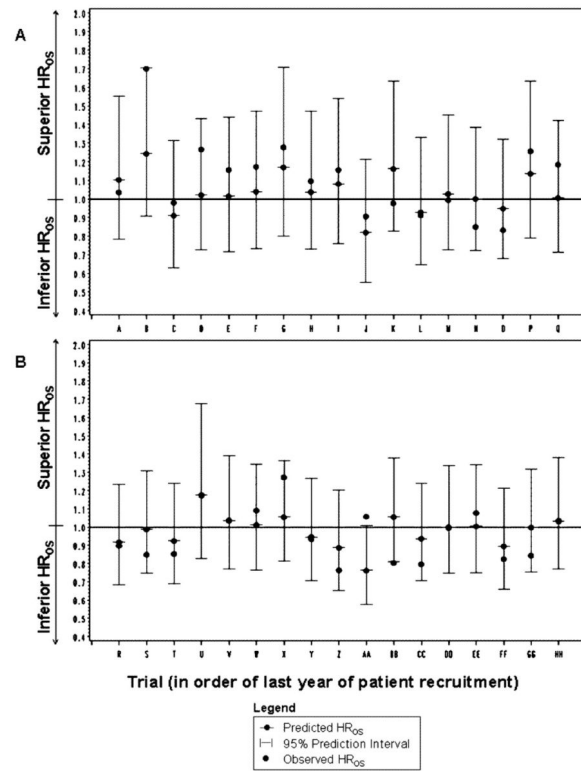


Figure 3. Leave-one-out cross-validation: anthracycline- (A) and taxane-based chemotherapy (B). Observed HR_{OS} for each trial-arm pair is plotted against the predicted HR_{OS} and 95% prediction intervals calculated from the HR_{PFS} in the primary meta-analytic, fixed-effects regression model. HR, hazard ratio; OS, overall survival; PFS, progression free survival.

Table 1
 Anthracycline (FEC/FAC) and Taxane-Based Chemotherapy Trials Included in Analysis

	Author (year published)	Year last subject enrolled	Number of Subjects Randomized	Calculated HR _{OS}	Calculated HR _{PFS}	Progression endpoint terminology in publication	Definition of progression endpoint in publication (Differences in bold)	Predicted HR	Lower bound	Upper bound
Anthracyclines (FEC/FAC)										
A	Perry(51) (1987)	1982	451	1.04	1.20	TTF	Fits strict PFS definition	1.10	0.78	1.55
B	Boccardo(11) (1985)	1982	73	1.70	1.62	PFS	Fits strict PFS definition	1.24	0.91	1.70
C	Bennett(8) (1988)	1985	331	0.98	0.83	TTP	Fits strict PFS definition	0.91	0.63	1.31
D	Conte(17) (1987)	1985	116	1.27	1.04	PFS	No definition provided	1.02	0.73	1.43
E	Aisner (I)(3) (1995)	1987	335	1.16	1.03	TTF	Fits strict PFS definition	1.01	0.72	1.44
F	Aisner (II)(3) (1995)	1987	329	1.17	1.08	TTF	Fits strict PFS definition	1.04	0.73	1.47
G	Ejlertsen(24) (1993)	1989	359	1.28	1.40	PFS	No definition provided	1.17	0.80	1.71
H	Speyer(61) (1992)	1989	150	1.10	1.07	PFS	No definition provided	1.04	0.73	1.47
I	Conte(18) (1996)	1990	258	1.16	1.17	PFS	No definition provided	1.08	0.76	1.54
J	Ackland(2) (2001)	1992	460	0.91	0.71	TTP	Fits strict PFS definition	0.82	0.55	1.21
K	Sledge(59) (2000)	1992	227	0.98	1.31	TTF	Fits strict PFS definition	1.16	0.83	1.63
L	Piarga(52; 53) (1995, 1998)	1993	258	0.91	0.86	TTP	Fits strict PFS definition	0.93	0.65	1.33
M	Parnes(50) (2003)	1995	241	0.99	1.05	TTF	Fits strict PFS definition	1.03	0.73	1.45
N	Namer(45) (2001)	1995	280	0.85	1.00	PFS	No definition provided	1.00	0.73	1.38
O	Del Mastro(21) (2001)	1997	151	0.83	0.90	TTF	No definition provided	0.95	0.68	1.32
P	Jassem(34; 35) (2004, 2001)	1998	267	1.26	1.31	TTP	Fits strict PFS definition	1.14	0.79	1.63
Q	Zielinski(68) (2005)	2002	259	1.18	1.01	Time to Progressive Disease	Difference from strict PFS definition: All deaths included if evidence of progression during treatment period or if clinical diagnosis/post-mortem examination indicated that death related to study drug or breast cancer.	1.01	0.71	1.42
Taxanes										
R	Nabholtz(43) (1996)	1992	471	0.90	0.71	TTP	Fits strict PFS definition	0.92	0.68	1.23
S	Sledge (I)(58) (2003)	1995	487	0.85	0.95	TTF	Difference from strict PFS definition: All death included if due to progression, toxicity or breast cancer within 6 weeks of date last known to be alive with stable	0.99	0.75	1.31

	Author (year published)	Year last subject enrolled	Number of Subjects Randomized	Calculated HR _{OS}	Calculated HR _{PFS}	Progression endpoint terminology in publication	Definition of progression endpoint in publication (Differences in bold)	Predicted HR	Lower bound	Upper bound
disease.										
T	Sledge (II)(58) (2003)	1995	489	0.85	0.73	TTF	As above	0.92	0.69	1.24
U	Paridaens(48) (2000)	1996	331	1.17	1.92	PFS	Fits strict PFS definition	1.18	0.83	1.68
V	Smith(60) (1999)	1996	563	1.04	1.14	Event Free Survival	Fits strict PFS definition	1.03	0.77	1.39
W	Winer (I)(66) (2004)	>1996	314	1.09	1.05	TTP	Fits strict PFS definition	1.01	0.76	1.35
X	Winer (II)(66) (2004)	>1996	313	1.27	1.26	TTP	Fits strict PFS definition	1.05	0.81	1.37
Y	Chan(16) (1999)	1997	326	0.93	0.81	TTP	Fits strict PFS definition	0.95	0.71	1.27
Z	Nabholtz(44) (1999)	1997	392	0.76	0.58	TTP	Fits strict PFS definition	0.89	0.65	1.20
AA	Sjostrom(57) (1999)	1997	282	1.06	0.48	TTP	Difference from strict PFS definition: PFS measured until progression, death or last follow-up visit.	0.76	0.58	1.01
BB	Bishop(10) (1999)	1997	209	0.80	1.21	PFS	Fits strict PFS definition	1.06	0.81	1.38
CC	Jassem(34;35) (2001,2004)	1998	267	0.80	0.77	TTP	Fits strict PFS definition	0.94	0.71	1.24
DD	Biganzoli(9) (2002)	1999	275	1.00	1.00	PFS	Fits strict PFS definition	1.00	0.75	1.34
EE	Langley(38) (2005)	1999	705	1.08	1.01	PFS	Fits strict PFS definition	1.00	0.75	1.34
FF	Jones(36) (2005)	2001	449	0.82	0.63	TTP	Difference from strict PFS definition: All deaths included if occurred less than 6 months after last study drug dose.	0.89	0.66	1.21
GG	Zielinski(68) (2005)	2002	259	0.84	0.99	Time to Progressive Disease	Difference from strict PFS definition. All deaths included if evidence of progression during treatment period or if clinical diagnosis/post-mortem examination indicated that death related to study drug or breast cancer.	1.00	0.75	1.32
HH	Gennari(31) (2006)	2002	215	1.04	1.13	PFS	Fits strict PFS definition	1.03	0.77	1.38

Note. Data extracted during literature review, assessment of progression-based endpoint definition and calculation of hazard ratio (HR) based on published data. Standardized definitions: PFS, progression free survival, time from randomization to progression or death from any cause; TTP, time to disease progression, time from randomization to progression; OS, overall survival, time from randomization until death from any cause; TTF, time to treatment failure, time from randomization to treatment failure for any cause; HR, hazard ratio; I, first trial-arm pair from study; II, second trial-arm pair from study.

Table 2

Agreement in Direction of the Hazard Ratios (HR) for Progression Free Survival (PFS) and Overall Survival (OS)

	Number of trial-arm pairs (%)		
	HR _{OS} <1	HR _{OS} >1	Total
Anthracyclines (FEC/FAC)			
HR _{PFS} <1	4 (25%)	0 (0%)	4 (25%)
HR _{PFS} >1	2 (12.5%)	10 (62.5%)	12 (75%)
Total (%)	6 (37.5%)	10 (62.5%)	16 (100%)
Taxanes			
HR _{PFS} <1	8 (50%)	1 (6.35%)	9 (56.25%)
HR _{PFS} >1	1 (6.25%)	6 (37.5%)	7 (43.75%)
Total (%)	9 (56.25%)	7 (43.75%)	16 (100%)

Note. One trial was excluded from each analysis because HR_{PFS}(45) or HR_{OS}(9) were equal to 1. FEC, 5-fluorouracil, epirubicin and cyclophosphamide; FAC, 5-fluorouracil, adriamycin and cyclophosphamide.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Results for Meta-analytic, Fixed-Effects Regression Analysis: Primary Model, Sensitivity Analysis, and Subgroup Analysis

Model	Trial-arm pairs (N)	Model R ²	<i>p</i> value for HR _{PFS} as a predictor HR _{OS}
Anthracyclines (FEC/FAC)			
Primary model	17	.48	.002
Sensitivity analysis: Strict PFS definition	10	.46	.032
Sensitivity analysis: Published HR substituted (<i>n</i> = 2)	17	.56	.0005
Sub-group analysis: Interaction term for year last patient recruited	17	.51	.009 (interaction <i>p</i> value = .44)
Sub-group analysis: Last patient recruited > 1990	8	.34	.127
Sub-group Analysis: Last patient recruited <1990	9	.51	.030
Taxanes			
Primary model	17	.35	.0120
Sensitivity Analysis: Strict PFS definition	12	.59	.0036
Sensitivity analysis: Published HR substituted (<i>n</i> = 6)	17	.25	.040
Subgroup analysis: Interaction term for first line trial	17	.36	.095 (interaction <i>p</i> value = .60)
Subgroup analysis: First-line trial	8	.61	.022
Subgroup analysis: Second/subsequent-line trial	9	.48	.038

Note. Results of meta-analytic regression analysis, with regression model: $\log_{10} \text{HR (OS)} = \alpha + \beta * \log_{10} \text{HR (PFS)}$. HR, hazard ratio; OS, overall survival; PFS, progression free survival; log, logarithm; FEC, 5-fluorouracil, epirubicin and cyclophosphamide; FAC, 5-fluorouracil, adriamycin and cyclophosphamide.