Breast Cancer Res Treat (2015) 154:591–608
DOI 10.1007/s10549-015-3643-5

EPIDEMIOLOGY

Correlation between progression-free survival and overall survival in metastatic breast cancer patients receiving anthracyclines, taxanes, or targeted therapies: a trial-level meta-analysis

George Adunlin[1] · John W. W. Cyrus[1] · George Dranitsaris[2]

Received: 16 October 2015 / Accepted: 13 November 2015 / Published online: 23 November 2015

Abstract Over the past decade, several new drugs have received regulatory approval for metastatic breast cancer (MBC). However, some of these approvals were based on improvement in progression-free survival (PFS), without a concomitant increase in overall survival (OS). This has led some to question the utility of using PFS as a measure for drug approval. To address the uncertainty of using PFS as a surrogate for OS in MBC, a systematic literature review followed by a trial-level correlative analysis was conducted in patients receiving anthracyclines, taxanes, or targeted therapies. Electronic databases were searched to identify randomized trials published between January 1990 and August 2015. Data extraction included hazard ratios for PFS (HR_{PFS}) and OS (HR_{OS}) between comparative arms as well as trial-level parameters. Weighted multivariate regression analysis was then used to test the strength of the association between HR_{PFS} and HR_{OS}. 72 trials providing 84 comparative arms met the inclusion criteria. HR_{PFS} was a significant predictor of HR_{OS} (model coefficient = 0.18, $p = 0.04$). However, only 31 % (i.e., model R^2) of the variability between the PFS–OS association was accounted for. When trials were limited to ≥2nd-line setting, the strength of the association improved (model coefficient = 0.40, $p < 0.001$) and the model R^2 increased to 55 %. However, the HR_{PFS}–HR_{OS} association was no longer significant when only 1st-line trials were considered ($p = 0.90$). HR_{PFS} is a predictor for HR_{OS} in MBC randomized trials. However, the effect was driven by trials in the ≥2nd-line setting. Therefore, PFS can be a suitable surrogate for OS in trials evaluating new treatments in the 2nd setting and beyond. The use of PFS alone as a primary trial endpoint in the 1st-line setting is not recommended.

Keywords Breast cancer · Surrogate endpoints · Progression-free survival · Overall survival

Introduction

The past decade has been witness to impressive advances in the treatment of metastatic breast cancer (MBC). MBC remains incurable, but from the initial diagnosis of metastatic disease, many patients are now living beyond a median of 3 years [1, 2]. Contributors to the improvement in overall survival (OS) have been better supportive care and the approval of novel anticancer agents and targeted therapies. However, some of the drugs that have been approved by regulators such as the Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have not readily received reimbursement from public and private payers [3]. One contributing factor has been an improvement in progression-free survival (PFS) with a new drug, but without the concomitant increase in OS. The health policy impact of this phenomenon was demonstrated when the U.S. FDA removed the breast cancer indication from bevacizumab after two randomized trials failed to show an improvement in OS [4]. Given the high cost of the newer anticancer agents, it has been suggested that the oncology community needs to consider the value offered by a new drug, with a clinically meaningful OS benefit being a key component to the value

✉ George Dranitsaris
george@augmentium.com

[1] Virginia Commonwealth University School of Medicine, Richmond, VA, USA

[2] Augmentium Pharma Consulting Inc., 283 Danforth Ave, Suite 448, Toronto M4K 1N2, Canada

proposition [5–7]. However, many solid tumor randomized trials including MBC use PFS as the primary endpoint, but PFS has not been fully validated as a surrogate for OS across several tumor types [8].

The FDA definition of PFS is a documented disease progression [i.e., greater than 20 % increase in tumor size, based on the response evaluation *criteria* in solid tumors (*RECIST*)1.1] or death from any cause from the time of randomization [9]. Demonstrating an OS benefit remains a challenge in cancer drug development. An important factor contributing to the difficulty in detecting an OS benefit is survival post progression (SPP) [8]. The longer the SPP, the harder it is to detect an OS benefit. In one simulation study, it was demonstrated that even when an OS benefit existed, it could not be statistically detected if the SPP was 12 months or longer [10]. The major events that mask potential OS benefits during the SPP period include mandated patient cross over into the experimental arm of the trial upon progression, secondary therapies, and heterogeneity in access to effective supportive care [11, 12]. As a result, a longer duration of SPP increases the opportunity for these and other factors to dilute any of the incremental survival benefits that may be associated with the new treatment under investigation.

Under these considerations, some investigators have argued that surrogate endpoints such as improvements in PFS should be accepted by regulatory agencies and payers because it would save drug development time and costs, and ultimately improve patient access to effective new drugs [11, 13]. The advantage of using PFS over OS is that secondary interventions cannot contaminate the former measurement. In addition, a patient cannot be crossed over into the experimental therapy until disease progression has occurred. Arguments against the use of PFS as a primary endpoint for drug approval and reimbursement are the potential for interobserver variability in measuring tumor shrinkage, and PFS is only a measure of drug effect during administration and is poorly correlated with survival and quality of life [8, 14].

The use of PFS as a surrogate for OS has been validated in metastatic colorectal cancer with the both patient- and trial-level analyses [15–18]. As a result, the FDA and other regulatory agencies have accepted PFS as a surrogate endpoint for the drug approval in metastatic colorectal cancer. However, uncertainty remains in MBC. Meta-analyses at both the patient and trial level have yielded conflicting results [11, 19, 20]. Furthermore, these studies did not evaluate all lines of MBC therapy, nor did they consider the impact of targeted therapies. To address this uncertainty and to test the hypothesis that PFS is a valid surrogate endpoint for OS in MBC randomized trials, a systematic review of the literature followed by a trial-level correlational analysis was conducted in MBC patients receiving anthracyclines, taxanes, and targeted therapies.

Methods

Systematic review of randomized trials

We searched PubMed/Medline, EMBASE, and the Cochrane Central Register of Controlled Trials for randomized controlled trials evaluating anthracyclines, taxanes, and targeted therapies in patients with MBC published between January 1, 1990 and August 1, 2015. Electronic searches of the major conference proceedings were also conducted. Validated filters for randomized clinical trials were used for EMBASE and Medline [21, 22].

There was no restriction on the line of therapy being tested in each study. Trials evaluating 1st-, 2nd-, and beyond 2nd-line therapy were considered. There was also no restriction on trials evaluating single-agent or combination therapy. The trial must have utilized a parallel group-randomized design with at least 65 MBC patients enrolled into each arm. At least one of the arms must have included an anthracycline, a taxane, or a targeted therapy. A measure of progression-free and OS outcomes data must also have been reported in each study arm. Trials that reported time to progression (TTP) or time to treatment failure (TTF) were considered. However, the exact definition used in the trial was documented for subsequent statistical adjustment. Trials that only reported hazard ratios (HR) for PFS and OS were also included. Trials evaluating hormonal therapies were not incorporated into the analysis because these agents are a different class of drugs with a unique mechanism of action.

Studies were selected on the basis of the predetermined criteria and agreed upon by two evaluators. Any disagreement on specific studies between the two evaluators was resolved through discussion. Once trials meeting the inclusion criteria were identified, the following data were extracted: sample size, year of publication, regions involved (e.g., North American, European, global), line of therapy being evaluated, chemotherapy regimen, dosage, duration of therapy, definition of primary and secondary endpoints, how tumor response was assessed (WHO vs. RECIST criteria), trial duration, median number of cycles delivered, if patient cross over was allowed, if the progression-free and OS outcomes were censored, definition of PFS and all relevant clinical outcomes such as median PFS, TTP, TTF, OS, and the associated HR. The extracted data were recorded into a database for the subsequent statistical analysis.

Statistical analysis

The two co-primary endpoints for evaluating the association between PFS and OS were the correlation between the HR for PFS (HR_{PFS}) and OS (HR_{PFS}) as well as the correlation between differences in the median PFS (Δ PFS) and OS (Δ OS) between the experimental and control arms of the trials. The association between PFS and OS is related to the prediction of the endpoint of interest (e.g., HR_{OS} or Δ OS) from the surrogate (e.g., HR_{PFS} or Δ PFS). Hence, the stronger the correlation, the more valid the surrogate. The objective of the study was to assess the validity of using PFS as a surrogate endpoint for OS in patients with MBC. For each trial that met the inclusion criteria, the association between PFS (HR_{PFS} or Δ PFS) and OS (HR_{OS} and Δ OS) was initially measured using the Spearman rank correlation coefficient. This was then followed with a weighted multivariable regression analysis.

In two separate analyses, weighted (on the total trial sample size) multivariable regression analysis was used to measure the association between the HR_{PFS} (primary predictor variable) and HR_{OS} (dependent variable). In the second analysis, Δ PFS was the main predictor variable and Δ OS was the dependent variable. These approaches provided a measure of the model R^2 statistic, which is the proportion of variability in the dependent variable accounted for by the model. Whenever HRs for PFS and OS were not reported in a given trial, they were calculated using the following formulas: HR_{OR} = median OS in experimental group/median OS in the control group; HR_{PFS} = median PFS in experimental group/median PFS in the control group.

Other independent variables considered in the regression models included line of therapy, combination versus single-agent therapy, year of trial publication, region where the study was conducted (U.S. vs. European vs. global), what the primary trial endpoint was (i.e., PFS, TTP, TTF or OS), if the PFS measurement in the trial was consistent with the current FDA definition, if the trial incorporated data censoring into the analysis, and if patient cross over was permitted from the control into the experimental arm. Normality in the distribution of the dependent variables was made through a comparison of means and medians as well as the application of the Skew test. The independent variables were retained in the final model through a backwards elimination process ($p < 0.05$ to retain). The models were also adjusted for clustering on the primary study citation in cases where trials had multiple experimental arms.

The slope of the regression line of the final model provided an estimate of how much of a risk reduction (i.e., via the HR) in PFS contributes to a decrease in the risk of death for patients who were randomized into the experimental arm of the trial. In the case of the model that used Δ PFS and Δ OS as the predictor and depended variables, the final model coefficient estimated the incremental OS benefit per incremental month of PFS reported for the experimental arm of the trial. The stability of the base case results for each modeling analysis was then evaluated in a series of one-way sensitivity analyses. All statistical analyses were performed using Stata, release 14.0 (Stata Corp., College Station, Texas, USA).

Results

The systematic literature search identified 3167 relevant references consisting of 3119 records from the database search and 48 additional records from other sources. From this initial pool of references, 880 duplicates were discarded. Following the title and abstract review, 1528 studies were rejected for being out of scope. Of the remaining references subject to the full-text review, 759 were removed using the exclusion criteria. The final set of bibliographic records that fulfilled the eligibility criteria comprised 72 randomized trials (Appendix "List of studies included in the meta analysis"), which provided 84 trial comparative arms, with median sample sizes in the control and experimental arms being 149 and 144 patients, respectively. Figure 1 shows the flowchart of study selection process.

Trial characteristics are summarized in Table 1. The publication years spanned from 1991 to 2015, with a maximum of 11 publications in 2011. The majority of trials ($n = 41$) were conducted globally and 55 of 72 (76.4 %) evaluated new treatments in the first-line setting. The most common progression endpoint was TTP ($n = 44$), 33 and 7 of the trials used PFS and TTF, respectively. Overall, 44 studies used the FDA definition of PFS, which is based on the RECIST 1.1 criteria [9]. OS was reported in 78 of the 84 comparative arms, with 52 studies utilizing data censoring and 21 of 84 study arms allowing crossover to the experimental regimen upon disease progression.

The univariate Spearman Rank correlation coefficient suggested a modest association between HR_{OS} and HR_{PFS} (Spearman's rho = 0.46; $p < 0.001$) as well as Δ OS and Δ PFS (Spearman's rho = 0.52; $p < 0.001$). As illustrated by Fig. 2, there was a positive trend in the association where a lower HR_{PFS} between the experimental and control groups indicated a reduction in the HR_{OS}. A HR_{OS} below 1.0 between the experimental and control groups would suggest a reduction in the risk of death in the former group of patients. Similarly, a larger Δ PFS was positively correlated with a greater Δ OS, indicating an improvement in overall survival between the experimental and control groups (Fig. 3).

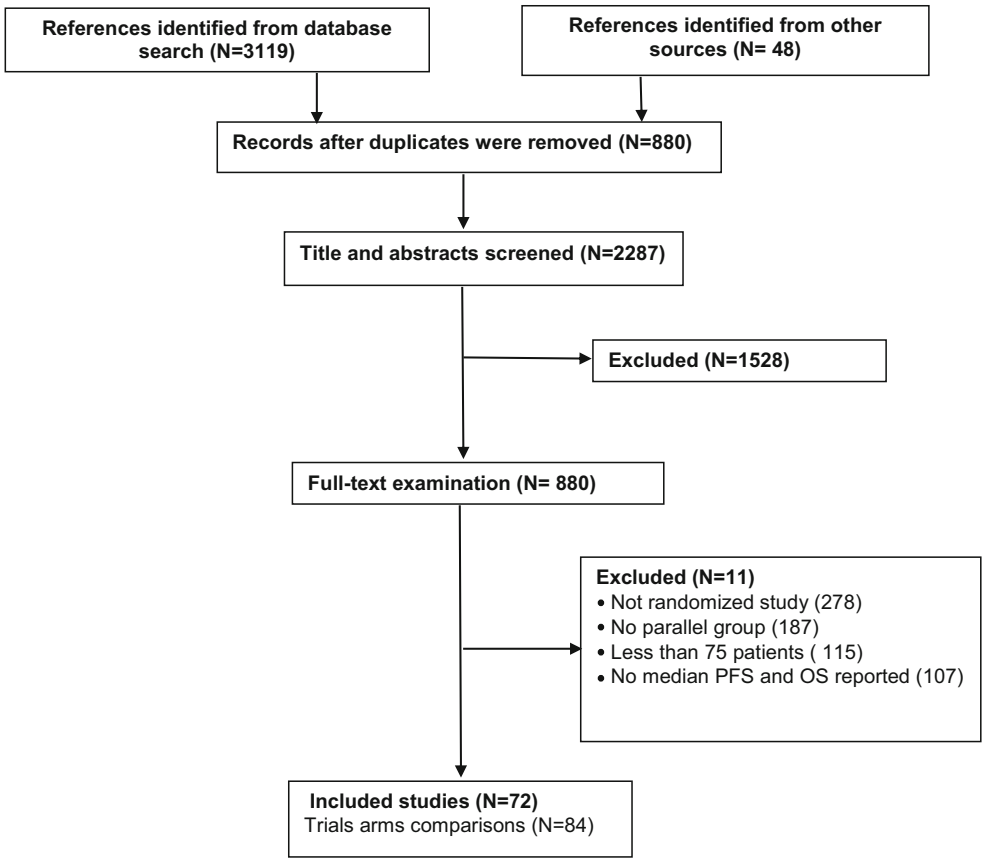

Fig. 1 Flow diagram of study selection process

The weighted multivariable regression modeling confirmed the findings of the univariate correlational analysis. Through the backwards elimination process, the final variables that were retained in the model correlating HR_{OS} with HR_{PFS} were region where the trial was conducted and patient cross over into the experimental arm. Other potentially important variables such as line of therapy, type of therapy (i.e., chemotherapy and targeted therapy alone or in combination) and type of progression endpoint used in the trial were not retained in the final model. Overall, the model R^2 was 0.31, indicating that only 31 % of the variability in the HR_{OS} was accounted for by the three independent variables retained in the model. Therefore, there are other important variables that contributed to the observed variability in the HR_{OS} that was reported in randomized trials evaluating new drugs in MBC.

The model coefficient between HR_{PFS} and HR_{OS} was statistically significant indicating a positive association between these two variables where a reduction in HR_{PFS} from an effective experimental therapy reduced the risk of death in MBC patients across all lines of therapy (Table 2). The findings also revealed that relative to trials conducted exclusively in Europe, global trials yielded a lower HR_{OS} by approximately 16 %. Stated differently, globally conducted trials were more likely to report an OS benefit compared to trials conducted in Europe. It is tempting to speculate that this finding may be related to a lower propensity to offer multiple lines of chemotherapy to patients from regions such as Latin America, Asia, and Southern Africa. The difference in reported HR_{OS} between European and North American trials was not statistically significant.

The allowance of patient cross over also had a statistically significant effect on the HR_{OS}. Trials that allowed cross over reported a 7.4 % reduction in the risk of death between the experimental and control groups compared to trials that did not allow cross over (Table 2). This finding is consistent with the expectation that cross over would only be offered in cases where the experimental agent under investigation appears to be highly effective.

The findings of the weighted multivariable regression analysis investigating the association between Δ PFS and Δ OS were consistent with the former evaluation. The model indicated that for every additional month of PFS, there would be a gain of 0.79 months in OS in the experimental group relative to the control group (Table 3). Other independent variables that were retained in the model by statistical means consisted of region, the allowance of patient

Table 1 Description of studies included in the analysis

Author, year of publication	Study drug versus control	Control (*N*)	Drug (*N*)	Line of therapy	Median PFS control	Median PFS drug	Median OS control	Median OS drug	HR reported	HR PFS drug	HR OS drug
Cowan (1991)	*D* = bisantrene 320 mg/m^2 *q3w* *C* = doxorubicin 60 mg/m^2 *q3w*	117	128	1 & 2	4.4	2.2	10.5	9.7	N	2	0.92
Cowan (1991)	*D* = mitoxantrone 14 mg/m^2 *q3w* *C* = doxorubicin 60 mg/m^2 *q3w*	117	120	1 & 2	4.4	2.3	10.5	5.9	N	1.91	1.5
Blomqvist (1993)	*D* = fluorouracil 500 mg/m^2 *q4w* + epirubicin 60 mg/m^2 *q4w* + cyclophosphamide 500 mg/m^2 *q4w* *C* = fluorouracil 500 mg/m^2 *q1w* + epirubicin 60 mg/m^2 *q1w* + cyclophosphamide 500 mg/m^2 *q1w*	86	84	1	9.2	5.4	11.8	21.2	N	1.7	
Hausmaninger (1995)	*D* = epirubicin 40 mg/m^2 *q3w* + vindesine 3 mg/m^2 *q3w* *C* = mitoxantrone 10 mg/m^2 *q3w* + vindesine 3 mg/m^2 *q3w*	149	146	All	5.4	5.1	–	–	N	1.06	
Bastholt (1996)	*D* = epirubicin IV 60 mg/m^2 *q3w* *C* = epirubicin IV 40 mg/m^2 *q3w*	75	73	1	4.4	4.7	13.6	14	N	0.94	0.97
Bastholt (1996)	*D* = epirubicin IV 90 mg/m^2 *q3w* *C* = epirubicin IV 40 mg/m^2 *q3w*	75	71	1	4.4	8.4	13.6	14.6	N	0.52	0.93
Bastholt (1996)	*D* = epirubicin IV 135 mg/m^2 *q3w* *C* = epirubicin IV 40 mg/m^2 *q3w*	75	67	1	4.4	8.4	13.6	11.3	N	0.52	1.2
Brufman (1997)	*D* = epirubicin 100 mg/m^2 *q3w* + 5-FU 500 mg/m^2 *q3w* + cyclophosphamide 500 mg/m^2 *q3w* *C* = epirubicin 50 mg/m^2 *q3w* + 5-FU 500 mg/m^2 *q3w* + cyclophosphamide 500 mg/m^2 *q3w*	242	212	1	5.3	5.8	–	–	N	0.91	
Joensuu (1998)	*D* = cyclophosphamide 500 mg/m^2 + epirubicin 60 mg/m 2 *q4w* + fluorouracil 500 mg/m^2 *q3w* + mitomycin 8 mg/m^2 *q4w* + vinblastine 6 mg/m^2 *C* = epirubicin 20 mg/m^2 *q4w* + mitomycin 8 mg/m^2 *q4w*	150	153	1 & 2	10	8	18	16	N	1.25	1.13
Bontenbal (1998)	*D* = epirubicin 90 mg/m^2 *q3w* *C* = doxorubicin 75 mg/m^2 *q3w*	128	131	2	5.3	4.4	10.9	10.2	N	1.2	1.07
Bishop (1999)	*D* = paclitaxel 200 mg/m^2 IV *q3w* *C* = cyclophosphamide 100 mg/m^2 + methotrexate 40 mg/m^2 IV + fluorouracil 600 mg/m^2 IV + prednisone 40 mg/m^2/d orally q 28 days	102	107	1	6.4	5.3	13.9	17.3	N	1.21	0.8
Blajman (1999)	*D* = 5-fluorouracil 500 mg/m^2 IV + doxorubicin 50 mg/m^2 IV + cyclophosphamide 500 mg/m^2 IV *C* = vinorelbine 25 mg/m^2 IV + doxorubicin 50 mg/m^2 IV	85	85	1	9	7.5	17.3	17.8	N	1.2	0.97
Chan (1999)	*D* = docetaxel 100 mg/m^2 *q3w* *C* = doxorubicin 75 mg/m^2 *q3w*	165	161	1 & 2	4.9	6	14	15	N	0.82	0.93

Table 1 continued

Author, year of publication	Study drug versus control	Control (*N*)	Drug (*N*)	Line of therapy	Median PFS control	Median PFS drug	Median OS control	Median OS drug	HR reported	HR PFS drug	HR OS drug
Estaban (1999)	*D* = 5-fluorouracil 500 mg/m^2 IV + cyclophosphamide 100 mg/m^2 + epirubicin 30 mg/m^2	76	75		7.7	11.9	12	17.3	N	0.65	0.69
	C = 5-fluorouracil 500 mg/m^2 IV + cyclophosphamide 100 mg/m^2 + Mitoxantrone 6 mg/m^2										
Nabholtz (1999)	*D* = docetaxel 100 mg/m^2 IV *q3w*	189	203	1 & 2	2.6	4.4	8.7	11.4	N	0.59	0.76
	C = mitomycin 12 mg/m^2 IV q6w + vinblastine 6 mg/m^2 IV *q3w*										
Sjostrom (1999)	*D* = docetaxel 100 mg/m^2 *q3w*	140	143	1 & 2	3	6.3	11	10.4	N	0.48	1.06
	C = methotrexate 200 mg/m^2 *q3w* + 5-fluorouracil 600 mg/m^2 *q3w*										
Smith (1999)	*D* = paclitaxel 250 mg/m^2 *q3w*	279	284	1	6.3	7.2	21.1	21.9	N	0.88	0.96
	C = paclitaxel 250 mg/m^2 *q3w*										
French Epi Gp (2000)	*D* = fluorouracil 500 mg/m^2 *q3w* + epirubicin 100 mg/m^2 *q3w* + cyclophosphamide 500 mg/m^2 *q3w*	139	145	1	10.3	8.3	17.9	18.9	N	1.24	0.95
	C = fluorouracil 500 mg/m^2 *q3w* + epirubicin 75 mg/m^2 *q3w* + cyclophosphamide 500 mg/m^2 *q3w*										
French Epi Gp (2000)	*D* = fluorouracil 500 mg/m^2 *q3w* + epirubicin 100 mg/m^2 *q3w* + cyclophosphamide 500 mg/m^2 *q3w*	139	133	1	10.3	6.2	17.9	16.3	N	1.66	1.1
	C = fluorouracil 500 mg/m^2 *q3w* + epirubicin 75 mg/m^2 *q3w* + cyclophosphamide 500 mg/m^2 *q3w*										
Norris (2000)	*D* = doxorubicin 50 mg/m^2 IV + vinorelbine 25 mg/m^2 IV	149	151	1&2	6.1	6.2	14.4	13.8	N	0.98	1.04
	C = doxorubicin 70 mg/m^2 IV										
Paridaens (2000)	*D* = Paclitaxel 200 mg/m^2 IV *q3w*	165	166	1	3.9	7.5	15.6	18.3	N	0.52	0.85
	C = doxorubicin 75 mg/m^2 IV *q3w*										
Ackland (2001)	*D* = cyclophosphamide 400 mg/m^2 IV + epirubicin 50 mg/m^2 IV + fluorouracil 500 mg/m^2 IV	237	223	1	6.3	8.7	18.2	20.1	Y	0.73	0.87
	C = cyclophosphamide 500 mg/m^2 IV + methotrexate 40 mg/m^2 IV + fluorouracil 600 mg/m^2 IV										
Batist (2001)	*D* = myocet 60 mg/m^2 *q3w* + cyclophosphamide of 600 mg/m^2 *q3w*	155	142	1	5.5	5.1	16	19	Y	1.03	1.04
	C = doxorubicin *q3w* + cyclophosphamide of 600 mg/m^2 *q3w*										
Del Mastro (2001)	*D* = cyclophosphamide 1000 mg/m^2 *q2w* + epirubicin 80 mg/m^2 *q2w*+ fluorouracil 600 mg/m^2 q2ws	74	77	1	14.3	12.8	32.7	27.2	N	1.12	1.2
	C = cyclophosphamide 600 mg/m^2 *q3w* + epirubicin 60 mg/m^2 *q3w* + fluorouracil 600 mg/m^2 *q3w*										
Fountzilas (2001)	*D* = epirubicin 110 mg/m^2 *q2w*+ paclitaxel 225 mg/m^2 q2w	90	93	1	8.5	10	20	21.5	N	0.85	0.93
	C = epirubicin 80 mg/m^2 *q3w* + paclitaxel 175 mg/m^2										

Table 1 continued

Author, year of publication	Study drug versus control	Control (*N*)	Drug (*N*)	Line of therapy	Median PFS control	Median PFS drug	Median OS control	Median OS drug	HR reported	HR PFS drug	HR OS drug
Jassem (2001)	*D* = doxorubicin 50 mg/m^2 *q3w* + paclitaxel 220 mg/m^2 *q3w* *C* = 5-fluorouracil 500 mg/m^2 *q3w* + doxorubicin 50 mg/m^2 *q3w* + cyclophosphamide 500 mg/m^2 *q3w*	133	134	1	6.2	8.1	18.3	23	Y	0.74	0.68
Namer (2001)	*D* = mitoxantrone 12 mg/m^2 IV + vinorelbine 25 mg/m^2 IV *C* = 5-fluorouracil 500 mg/m^2 IV + cyclophosphamide 500 mg/m^2 IV + doxorubicin or epirubicin 50 mg/m^2 IV	139	142	1	7	7	20	17	N	1	1.18
Slamon (2001)	*D* = paclitaxel 175 mg/m^2 *q3w* + trastuzumab 4 mg/kg IV q1w *C* = paclitaxel 175 mg/m^2 *q3w*	96	92	1	3	6.9	18.4	22.1	N	0.43	0.83
Slamon (2001)	*D* = doxorubicin 60 mg/m^2 or epirubicin 75 mg/m^2 + cyclophosphamide 600 mg/m^2 *C* = paclitaxel 175 mg/m^2 *q3w*	96	138	1	3	6.09	18.4	21.4	N	0.49	0.86
Slamon (2001)	*D* = doxorubicin 60 mg/m^2 or epirubicin 75 mg/m^2 + cyclophosphamide 600 mg/m^2 + trastuzumab 4 mg/kg IV q1w *C* = paclitaxel 175 mg/m^2 *q3w*	96	143	1	3	7.8	18.4	26.8	N	0.38	0.69
Biganzoli (2002)	*D* = doxorubicin 60 mg/m^2 IV *q3w* + paclitaxel 175 mg/m^2 *q3w* *C* = doxorubicin 60 mg/m^2 IV *q3w* + cyclophosphamide 600 mg/m^2 *q3w*	137	138	1	6	6	20.5	20.6	N	1	1
Bonneterre (2002)	*D* = docetaxel 100 mg/m^2 *q3w* *C* = 5-fluorouracil 750 mg/m^2 *q3w* + vinorelbine 25 mg/m^2 *q3w*	90	88	2	5.1	6.5	15	16	N	0.78	0.94
Harris (2002)	*D* = liposome-encapsulated doxorubicin (TLC D-99) 75 mg/m^2 *q3w* *C* = doxorubicin 75 mg/m^2 *q3w*	116	108	1	4.3	3.8	20	16	Y	0.92	0.76
Heidemann (2002)	*D* = mitoxantrone 12 mg/m^2 IV *q3w* *C* = fluorouracil 500 mg/m^2 IV *q3w* + epirubicin 50 mg/m^2 IV *q3w* + cyclophosphamide 500 mg/m^2 IV *q3w*	133	133	≥ 2	6.15	4.4	15.8	14.1	N	1.4	1.12
O'Shaughnessy (2002)	*D* = capecitabine 1250 mg/m^2 + docetaxel 75 mg/m^2 *C* = capecitabine 1250 mg/m^2 + docetaxel 100 mg/m^2	256	255	All	4.2	6.1	11.5	14.5	Y	0.65	0.77
Nabholtz (2003)	*D* = doxorubicin 50 mg/m^2 *q3w* + docetaxel 75 mg/m^2 *q3w* *C* = doxorubicin 60 mg/m^2 + cyclophosphamide 600 mg/m^2	215	214	1	7.44	8.7	21.7	22.5	Y	0.76	0.96
Parnes (2003)	*D* = cyclophosphamide 500 mg/m^2 + doxorubicin 40 mg/m^2 + fluorouracil 200 mg/m^2 IV + leucovorin 200 mg/m^2 in 100 mL IV *C* = cyclophosphamide 500 mg/m^2 + doxorubicin 40 mg/m^2 + fluorouracil 200 mg/m^2 IV	121	121	1	9	9.48	20.64	20.52	N	0.95	1.01

Table 1 continued

Author, year of publication	Study drug versus control	Control (*N*)	Drug (*N*)	Line of therapy	Median PFS control	Median PFS drug	Median OS control	Median OS drug	HR reported	HR PFS drug	HR OS drug
Sledge (2003)	*D* = paclitaxel 150 mg/m^2 *q3w* *C* = doxorubicin 60 mg/m^2 IV *q3w*	224	229	1 & 2	6	6.3	19.1	22.5	N	0.95	0.85
Sledge (2003)	*D* = doxorubicin 60 mg/m^2 IV *q3w* + paclitaxel 175 mg/m^2 *q3w* *C* = doxorubicin 60 mg/m^2 IV *q3w*	224	230	1 & 2	6	8.2	19.1	22.4	N	0.73	0.85
Chan (2004)	*D* = myocet 75 mg/m^2 + cyclophosphamide 600 mg/m^2 *C* = epirubicin 75 mg/m^2 + cyclophosphamide 600 mg/m^2	80	80	1	5.6	7.7	16	18.3	Y	0.66	0.87
Conte (2004)	*D* = epirubicin 120 mg/m^2 + paclitaxel 250 mg/m^2 *q3w* *C* = epirubicin 90 mg/m^2 + paclitaxel 200 mg/m^2	108	94	1	11	10.8	20	26	N	1.02	0.77
Ejlertsen (2004)	*D* = epirubicin 90 mg/m^2 IV *q3w* + vinorelbine 25 mg/m^2 *q3w* *C* = epirubicin 90 mg/m^2 IV *q3w*	194	193	1	8.2	10.1	18	19.1	N	0.81	0.94
Fountzilas (2004)	*D* = paclitaxel 175 mg/m^2 *q3w* + carboplatin AUC of 6 mg min/ml *q3w* *C* = paclitaxel 175 mg/m^2 *q3w* + epirubicin 80 mg/m^2 *q3w*	163	164	1	8.1	10.8	22.7	27.8	N	1.49	0.82
Keller (2004)	*D* = liposomal doxorubicin 50 mg/m^2 IV *q4w* *C* = vinorelbine 30 mg/m^2 or mitomycin C 10 mg/m^2 *q4w* + vinblastine 5 mg/m^2 q6w to q8w	151	150	2 & 3	2.9	2.5	10.4	9	Y	1.27	1.07
O'Brien (2004)	*D* = pegylated liposomal doxorubicin HCl *q4w* *C* = doxorubicin 60 mg/m^2 *q4w*	255	254	All	7.8	6.9	22	21	Y	1	1.11
Winer (2004)	*D* = paclitaxel 210 mg/m^2 *C* = paclitaxel 175 mg/m^2	158	156	1 & 2	3.9	4.1	11	12	N	0.95	0.92
Winer (2004)	*D* = paclitaxel 250 mg/m^2 *C* = paclitaxel 175 mg/m^2	158	155	1 & 2	3.9	4.9	11	14	N	0.8	0.79
Bontenbal (2005)	*D* = doxorubicin 50 mg/m^2 *q3w* + docetaxel 75 mg/m^2 *C* = fluorouracil 500 mg/m^2 *q3w* + doxorubicin 50 mg/m^2 *q3w* + cyclophosphamide 500 mg/m^2 *q3w*	107	109	1	6.6	8	16.2	22.6	Y	1.5	1.43
Feher (2005)	*D* = gemcitabine 1200 mg/m^2 IV *C* = epirubicin 35 mg/m^2	199	198	1	6.1	3.4	19.1	11.8	N	1.67	0.62
Gradishar (2005)	*D* = paclitaxel 050 mg/m^2 *C* = doxorubicin 3.90 mg/m^2 IV	225	229	All	3.9	5.3	12.9	15.1	N	0.75	0.82
Icli (2005)	*D* = cisplatin 70 mg/m IV *q3w* + oral etoposide 50 mg *q3w* *C* = paclitaxel 175 mg/m^2 IV *q3w*	101	100	All	3.9	5.5	9.5	14	N	0.71	0.68
Jones (2005)	*D* = paclitaxel 050 mg/m^2 *C* = doxorubicin 3.60 mg/m^2 IV	213	211	1	3.6	5.7	12.7	15.4	Y	1.64	1.41

Table 1 continued

Author, year of publication	Study drug versus control	Control (*N*)	Drug (*N*)	Line of therapy	Median PFS control	Median PFS drug	Median OS control	Median OS drug	HR reported	HR PFS drug	HR OS drug
Langley (2005)	*D* = epirubicin 75 mg/m^2 *q3w* + paclitaxel 200 mg/m^2 *q3w* *C* = epirubicin 75 mg/m^2 *q3w* + cyclophosphamide 600 mg/m^2 *q3w*	352	353	1	7.1	7	14	13	Y	1.07	1.02
Marty (2005)	*D* = docetaxel 100 mg/m^2 IV *q3w* + trastuzumab 4 mg/kg *C* = docetaxel 100 mg/m^2 IV *q3w*	94	94	1	6.1	11.7	22.7	31.2	N	0.52	0.73
Miller (2005)	*D* = capecitabine 2500 mg/m^2 *q3w* + bevacizumab 15 mg/kg *q3w* *C* = capecitabine 2500 mg/m^2 *q3w*	230	232	All	4.17	4.86	14.5	15.1	N	0.98	0.96
Zielinski (2005)	*D* = gemcitabine 1000 mg/m^2 + epirubicin 90 mg/m^2 + paclitaxel 175 mg/m^2 *C* = fluorouracil 500 mg/m^2 + epirubicin 90 mg/m^2 + cyclophosphamide 500 mg/m^2	135	124	1	9	9.1	24.9	29.5	N	0.9	0.84
Harvey (2006)	*D* = docetaxel 75 mg/m^2 IV *q3w* *C* = docetaxel 100 mg/m^2 IV *q3w*	188	188	2	2.95	3.49	10.6	10.3	N	0.85	1.03
Harvey (2006)	*D* = docetaxel 60 mg/m^2 IV *q3w* *C* = docetaxel 100 mg/m^2 IV *q3w*	188	151	2	2.92	3.9	10.6	12.3	N	0.75	0.86
Robert (2006)	*D* = trastuzumab 4 mg/kg; 2 mg/kg *q3w* + paclitaxel 175 mg/m^2 + carboplatin 6 mg/mL *q3w* *C* = trastuzumab 4 mg/kg; 2 mg/kg *q3w* + paclitaxel 175 mg/m2 *q3w*	98	98	1	7.1	10.7	32.2	35.7	Y	0.66	0.9
Gray (2009)	*D* = paclitaxel 90 mg/m^2 IV *q3w* + bevacizumab 10 mg/kg *q3w* *C* = paclitaxel 90 mg/m^2 IV *q3w*	354	368	1	5.8	11.8	25.2	26.7	Y	0.48	0.88
Albain (2008)	*D* = gemcitabine 1250 IV mg/m^2 *q3w* + paclitaxel 175 mg/m^2 *q3w* *C* = paclitaxel 175 mg/m^2 IV *q3w*	263	266		3.9	5.9	15.8	18.6	Y	0.73	0.82
Cameron (2008)	*D* = anthracycline or taxane, or trastuzumab + lapatinib 1250 mg/day + capecitabine 2000 mg/m^2 *C* = anthracycline or taxane, or trastuzumab + capecitabine 2500 mg/m^2	201	198	2 & 3	4.3	6.2	15.3	15.6	Y	0.57	0.78
Cassier (2008)	*D* = doxorubicin 50 mg/m^2 *q3w* + docetaxel 75 mg/m^2 *q3w* *C* = doxorubicin 50 mg/m^2 *q3w* + paclitaxel 175 mg/m^2 *q3w*	103	107		8	8.7	27.3	21.4	N	0.92	1.28
Di Leo (2008)	*D* = paclitaxel 175 mg/m^2 IV *q3w* + lapatinib 1500 mg/d *C* = paclitaxel 175 mg/m^2 IV *q3w* + placebo	288	291	1	5.3	5.8	20.2	23	Y	0.9	0.86
Chan (2009)	*D* = gemcitabine 1000 mg/m^2 IV *C* = capecitabine 1250 mg/m^2 *q3w* + docetaxel 75 mg/m^2 *q3w*	152	153	All	7.98	8.05	21.45	19.29	Y	1.2	1.11
Fountzilas (2009)	*D* = paclitaxel 175 mg/m^2 *q3w* + carboplatin AUC of 6, in 500 mL *q3w* *C* = paclitaxel 80 mg/m^2 *q3w*	136	136	1	11.4	11.5	41	29.9	N	0.99	1.9

Table 1 continued

Author, year of publication	Study drug versus control	Control (*N*)	Drug (*N*)	Line of therapy	Median PFS control	Median PFS drug	Median OS control	Median OS drug	HR reported	HR PFS drug	HR OS drug
Fountzilas (2009)	*D* = gemcitabine 1000 mg/m^2 *q3w* + docetaxel 75 mg/m^2 *q3w* *C* = paclitaxel 80 mg/m^2 *q3w*	136	144	1	11.4	10.4	41	26.9	N	1.1	1.5
Gradishar (2009)	*D* = nab-paclitaxel 300 mg/m^2 *q3w* *C* = docetaxel 100 mg/m^2 *q3w*	76	76	1	7.5	11	26.6	26.6	N	0.63	1
Gradishar (2009)	*D* = nab-paclitaxel 150 mg/m^2 *q3w* *C* = docetaxel 100 mg/m^2 *q3w*	76	76	1	7.5	12.8	26.6	22.2	N	0.59	1.2
Gradishar (2009)	*D* = nab-paclitaxel 100 mg/m^2 *q3w* *C* = docetaxel 100 mg/m^2 *q3w*	76	74	1	7.5	12.9	26.6	28.7	N	0.46	0.93
Kaufman (2009)	*D* = trastuzumab 4 mg/kg IV; 2 mg/kg + anastrozole 1 mg/d *C* = anastrozole 1 mg/d	104	103	1	2.4	4.8	23.9	28.5	Y	0.63	0.85
Sparano (2009)	*D* = pegylated liposomal doxorubicin 30 mg/m^2 *q3w* + docetaxel 60 mg/m^2 *q3w* *C* = docetaxel 75 mg/m^2 *q3w*	373	378	All	7	9.8	20.6	20.5	Y	0.66	1.02
Yardley (2009)	*D* = liposomal doxorubicin 40 mg/m^2 IV *q4w* *C* = docetaxel 36 mg/m^2 IV *q4w*	52	50	1	5.5	6.5	15.4	16.2	N	0.84	0.95
Blackwell (2010)	*D* = lapatinib 1000 mg daily + trastuzumab IV 4 mg/kg; 2 mg/kg *C* = lapatinib 1500 mg daily	148	148	≥2	1.89	2.8	9.1	12.03	Y	0.71	0.75
Mavroudis (2010)	*D* = docetaxel 75 mg/m^2 IV *q3w* + epirubicin 75 mg/m^2 *q3w* *C* = docetaxel 75 mg/m^2 IV *q3w* + capecitabine 950 mg/m^2 *q3w*	141	145	1	10.6	11	37.6	35.7	N	0.96	1.05
Wardley (2010)	*D* = trastuzumab 8 mg/kg; 6 mg/kg *q3w* + Docetaxel 75 mg/m^2 *q3w* + capecitabine 950 mg/m^2 *q3w* *C* = trastuzumab 8 mg/kg; 6 mg/kg *q3w* + docetaxel 100 mg/m^2 *q3w*	112	113	1	12.8	17.9			N	0.72	
Verma (2012)	*D* = trastuzumab emtansine 3.6 mg/kg IV *C* = lapatinib 1250 mg/d + capecitabine 1000 mg/m^2	496	495	All	6.4	9.6	25.1	30.9	Y	0.65	0.682
Baselga (2012); Swain (2013)	*D* = pertuzumab 840 mg/d + trastuzumab 8 mg/kg + docetaxel 75 mg/m^2 *C* = placebo + trastuzumab 8 mg/kg + docetaxel 75 mg/m^2	406	402	1	12.4	18.5	40.8	56.5	Y	0.62	0.68
Baselga (2012); Piccart (2014)	*D* = everolimus 10 mg/d + exemestane 5 mg/d *C* = placebo + exemestane 5 mg/d	239	485	1	4.1	10.6	26.6	31	Y	0.36	0.89
André (2014)	*D* = everolimus 5 mg/d *q3w* + trastuzumab 2 mg/kg *q3w* + vinorelbine 25 mg/m^2 *q3w* *C* = placebo + trastuzumab 2 mg/kg *q3w* + vinorelbine 25 mg/m^2 *q3w*	285	284	1	5.78	7			Y	0.78	

Table 1 continued

Author, year of publication	Study drug versus control	Control (*N*)	Drug (*N*)	Line of therapy	Median PFS control	Median PFS drug	Median OS control	Median OS drug	HR reported	HR PFS drug	HR OS drug
Hurvitz (2014)	*D* = everolimus 10 mg/d + trastuzumab 4 mg/kg; 2 mg/kg *q4w* + paclitaxel 80 mg/m^2 *q4w* *C* = placebo + trastuzumab 4 mg/kg; 2 mg/kg *q4w* + paclitaxel 80 mg/m^2 *q4w*	480	239	1	14.49	14.95			Y	0.89	
Krop (2014)	*D* = trastuzumab emtansine 3.6 mg/kg IV *q3w* *C* = treatment of physician's choice	198	404	≥2	3.3	6.2			Y	0.528	
Finn (2015)	*D* = palbociclib 125 mg/d *q3w* + letrozole 2.5 mg/d *q3w* *C* = letrozole 2.5 mg/d *q3w*	81	84	1	27.9	29.6			Y	0.488	
Turner (2015)	*D* = palbociclib 125 mg/d + fulvestrant 500 mg *C* = placebo + fulvestrant 500 mg	174	347		3.8	9.2			Y	0.42	

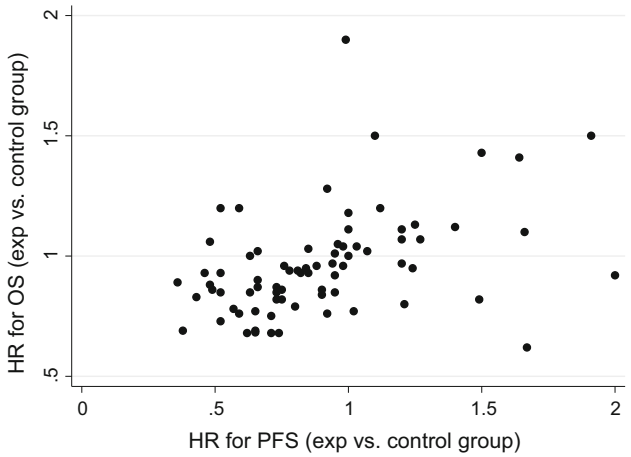

Fig. 2 Association between the HR_{PFS} and the HR_{OS} the experimental and control groups. The Spearman rho coefficient was 0.46; $p < 0.001$

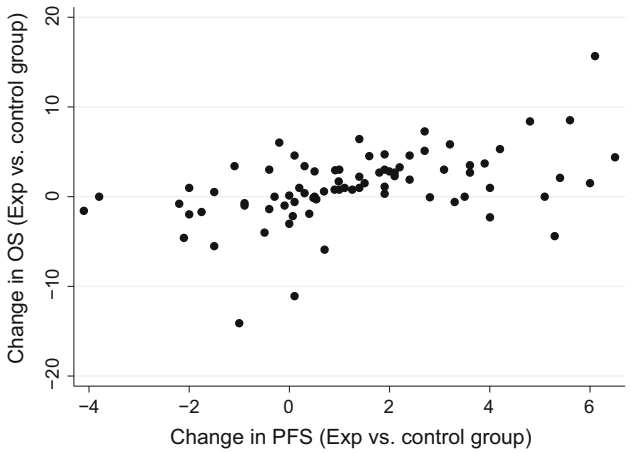

Fig. 3 Association between Δ PFS and Δ OS the experimental and control groups. The Spearman rho coefficient was 0.52; $p < 0.001$

Table 2 Multivariable regression analysis on the association between the HR for OS and PFS

Private variable	Coefficient	SEM	*p* value
Intercept	−0.90	0.10	
HR for PFS	0.18	0.086	0.04
Region (vs. European)			
Global	−0.16	0.07	0.025
North American	−0.077	0.069	0.26
Cross over permitted	−0.074	0.035	0.039
Model R^2	0.31		

Dependent variable HR for OS between the experimental and control groups of the study

HR hazard ratio, *PFS* progression-free survival, *OS* overall survival, *Model* R^2 proportion of variability in the dependent variable that is accounted for by the model

Table 3 Multivariable regression analysis on the association between change in OS and change in PFS between the experimental and control group

Private variable	Coefficient	SEM	*p* value
Intercept	−1.49	1.1	
Change in median PFS	0.79	0.24	0.001
Region (vs. European)			
Global	2.54	1.07	0.021
North American	0.39	1.22	0.75
Cross over permitted	2.73	1.22	0.029
≥3rd-line versus 1st/2nd-line trials	−3.1	1.33	0.023
Model R^2	0.44		

Dependent variable: change in median OS between the experimental and control groups of the study

PFS progression-free survival, *OS* overall survival, *Adjusted* R^2 proportion of variability in the dependent variable that is accounted for by the model

cross over and line of therapy (1st- or 2nd- vs. ≥3nd-line trials). Globally conducted trials reported an OS gain of 2.5 months compared to trials conducted in Europe. Trials allowing cross over were associated with a 2.73-month increment in OS. Furthermore, trials evaluating new treatments in the 3rd-line setting and beyond reported a reduced OS benefit by approximately 3.1 months ($p = 0.023$). Overall, the model R^2 was 0.44, indicating that only 44 % of the variability in Δ OS was accounted for by the four independent variables retained in the final model (Table 3).

Sensitivity analysis on the primary findings

A one-way sensitivity analysis was conducted to evaluate the stability of the primary results generated from both multivariate analyses. This was characterized by focusing on trials that were published within the last 12 years (i.e., from 2004 onward), limiting the analysis to 1st- or ≥2nd-line trials only, studies that used the FDA definition of PFS, had utilized data censoring, allowed cross over and were conducted globally. Of the seven sensitivity analyses performed, the statistically significant association between HR_{PFS} and HR_{OS} was retained in only three cases; ≥2nd-line trials, those that used the FDA definition of PFS and those allowing cross over (Table 4). Of these, trials ≥2nd-line setting had the highest model R^2 at 0.55 and the model coefficient between HR_{PFS} and HR_{OS} increased from 0.18 in the base case to 0.40 ($p < 0.001$). In contrast, the model coefficient for trials in the 1st-line setting dropped to 0.01 ($p = 0.90$), indicating that such trials are unlikely to ever yield a statistically significant HR_{OS}. Trials allowing cross over to truly efficacious new drugs were also more likely to yield a HR_{OS} in favor of the experimental treatment (Table 4).

The same series of sensitivity analyses were also performed for the multivariate models evaluating Δ PFS and Δ OS. In contrast to the former series of sensitivity analyses, the current series revealed that the significant association between Δ PFS and Δ OS was maintained in 6 of the 7 performed. The only case where the association was lost was when the analysis was limited to trials that used the FDA definition of PFS (Table 4). In their entirety, these findings imply that Δ PFS may be a better surrogate to OS than HR_{PFS}. However, as a measure of effect size, HR_{PFS} is preferred because it considers the entire time horizon of the Kaplan–Meier survival curve. It was also interesting to note that limiting the analysis to trials that allowed cross over increased the model coefficient for Δ PFS from 0.79 in the base case analysis to 1.57, with the model R^2 increasing to 75 %. The finding that cross over trials yielded stronger and more consistent associations between improvements in PFS and OS suggests that trials allowing cross over are somehow different than those that do not.

Discussion

In order to increase the likelihood of a new drug receiving regulatory approval and eventual reimbursement, a statistically and clinically meaningful increment in OS relative to an accepted standard of care should be demonstrated [3, 7]. However, demonstrating an OS benefit is challenging in solid tumors, particularly in earlier line trials where multiple effective therapies and modern supportive care are available upon progression [10]. To avoid the contaminating effects of these subsequent therapies, drug developers have used surrogate endpoints of patient benefit such as PFS, TTP, and TTF. For a surrogate endpoint to be used as a measure for drug approval, there should be at least some evidence that supports its correlation to OS.

In disease sites such as metastatic colorectal cancer, improvements in PFS have been shown to be statistically correlated to improvements in OS in both trial-level and patient-level analyses [15, 16, 18]. However, studies evaluating PFS as a surrogate to OS in MBC have generated conflicting results and uncertainty remains [11]. In one report, Miksad and colleagues conducted a trial-level analysis to measure the association between the HR_{PFS} and HR_{OS} in advanced-stage breast cancer patients who received anthracycline- or taxane-based chemotherapy [20]. The investigators found that HR_{PFS} was a statistically significant predictor of HR_{OS} with up to 48 % of the variance accounted for [20]. In contrast to these findings, Burzykowski et al. conducted a patient-level analysis on 3953 patients from 11 randomized trials evaluating

Table 4 Summary of sensitivity analysis on the base case results

Sensitivity analysis	Sample size	Model coefficient, 95 % CI and p value	Model R^2
HR for OS and PFS			
Base case	75	0.18 (0.009 to 0.35); $p = 0.04$	0.31
Year of publication ≥2004	40	0.12 (−0.22 to 0.48); $p = 0.48$	0.31
First-line trials only	48	0.01 (−0.19 to 0.22); $p = 0.90$	0.30
Trials in 2nd line and beyond	27	0.40 (0.21 to 0.58); $p < 0.001$	0.55
Trials meeting the modern FDA definition of PFS	38	0.26 (0.05 to 0.42); $p = 0.015$	0.38
Trials with censored data	47	0.18 (−0.05 to 0.42); $p = 0.12$	0.35
Trials allowing cross over	20	0.24 (0.14 to 0.33); $p < 0.001$	0.49
Only globally conducted trials	34	0.031 (−0.27 to 0.33); $p = 0.83$	0.10
Δ in OS and Δ in PFS			
Base case	79	0.79 (0.32 to 1.26); $p = 0.001$	0.44
Year of publication ≥2004	41	0.68 (0.08 to 1.28); $p = 0.027$	0.51
First-line trials only	51	0.77 (0.26 to 1.28); $p = 0.004$	0.47
Trials in 2nd line and beyond	28	1.1 (0.38 to 1.74); $p = 0.004$	0.49
Trials meeting the modern FDA definition of PFS	39	0.74 (0.10 to 1.37); $p = 0.25$	0.50
Trials with censored data	49	0.81 (0.25 to 1.36); $p = 0.006$	0.54
Trials allowing cross over	20	1.57 (0.66 to 2.48); $p = 0.002$	0.75
Globally conducted trials only	36	0.82 (0.14 to 1.51); $p = 0.02$	0.51

HR hazard ratio, *OS* overall survival, *PFS* progression-free survival, *Δ* change, *FDA* Food and Drug Administration

anthracyclines or taxanes in the first-line setting of MBC [19]. The investigators failed to find a statistically significant correlation between HR_{PFS} and HR_{OS}. Burzykowski and colleagues concluded that PFS was not an acceptable surrogate endpoint in this treatment setting [19].

In the current study, a trial-level meta-analysis was conducted to measure the association between PFS and improvements in OS through two different endpoints; HR_{PFS} and Δ PFS. The analysis used a weighted multivariate modelling approach, which allowed additional predictor variables to be evaluated. The findings indicated that both HR_{PFS} and Δ PFS were modestly correlated with improvements in OS, with 31 and 44 % of the variability explained by the respective models. However, the sensitivity analysis indicated that when the analysis was limited to trials evaluating new treatments in the 2nd setting and beyond, the model coefficient for the PFS surrogate measures increased significantly, as did the model R^2. When the analysis was limited to trials in the first-line setting, the statistically significant correlations between the surrogate PFS measures and improvements in OS were lost, consistent with findings of the patient-level analysis conducted by Burzykowski et al. [19]. These observations indicate that PFS can be a suitable surrogate for OS in MBC randomized trials evaluating new treatments in the 2nd setting and beyond. In the 1st setting, PFS as a primary trial endpoint is of limited clinical value and should be supplemented with meaningful patient reported outcome measures such as improvements in performance status, symptom control, and weight gain [8].

There are a number of limitations in the current study that need to be acknowledged. All meta-analyses are affected by the quality of the studies analyzed. For that reason, we limited our review to published prospective randomized trials with sufficient sample size. However, publication bias remains an issue and it must also be remembered that meta-analyses are only associations between trial-level parameters and study outcomes. True causation can only be established with an analysis of patient-level data. The R^2 of the various multivariate models ranged from 31 to 75 %. Therefore, there are additional factors contributing to the variability between the PFS-OS surrogacy that were not accounted. In 55 of the 84 eligible comparative trial arms, the HR for either PFS or OS was not reported. Hence, it had to be manually calculated using the reported medians. Such an approach may not reflect the true HR from a properly conducted survival analysis. In 3 of the accepted trials, we were also unsure if the current FDA definition of PFS was used. Lastly, variability in the evaluation of PFS between trials may also have impacted the observed differences in median PFS.

Despite these limitations, the findings of this correlative meta-analysis of prospective randomized trials were consistent with other trial-level analyses and indicate that

improvements in PFS are correlated with increased OS. However, the effect appears to be driven by trials evaluating new drugs in ≥2nd-line setting. Therefore, PFS can be a suitable surrogate for OS in MBC randomized trials evaluating new treatments in the 2nd setting and beyond. The use of PFS alone as a primary trial endpoint in the 1st-line setting is not recommended.

Funding This study was not supported by external funding.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix: List of studies included in the meta-analysis

1. Ackland SP, Anton A, Breitbach GP et al. (2001) 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 19:943–953.
2. Albain KS, Nag SM, Calderillo-Ruiz G et al. (2008) Gemcitabine plus Paclitaxel versus Paclitaxel monotherapy in patients with metastatic breast cancer and prior anthracycline treatment. J Clin Oncol 26:3950–3957.
3. André F, O'Regan R, Ozguroglu M et al. (2014) Everolimus for women with trastuzumab-resistant, HER2-positive, advanced breast cancer (BOLERO-3): a randomised, double-blind, placebo-controlled phase 3 trial. The lancet oncology 15:580–591.
4. Baselga J, Campone M, Piccart M et al. (2012) Everolimus in postmenopausal hormone-receptor–positive advanced breast cancer. N Engl J Med 366:520–529.
5. Baselga J, Cortés J, Kim S et al. (2012) Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. N Engl J Med 366:109–119.
6. Bastholt L, Dalmark M, Gjedde SB et al. (1996) Dose–response relationship of epirubicin in the treatment of postmenopausal patients with metastatic breast cancer: a randomized study of epirubicin at four different dose levels performed by the Danish Breast Cancer Cooperative Group. J Clin Oncol 14:1146–1155.
7. Batist G, Ramakrishnan G, Rao CS et al. (2001) Reduced cardiotoxicity and preserved antitumor efficacy of liposome-encapsulated doxorubicin and cyclophosphamide compared with conventional doxorubicin and cyclophosphamide in a randomized, multicenter trial of metastatic breast cancer. J Clin Oncol 19:1444–1454.
8. Biganzoli L, Cufer T, Bruning P et al. (2002) 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 20:3114–3121.
9. Bishop JF, Dewar J, Toner GC et al. (1999) Initial paclitaxel improves outcome compared with CMFP combination chemotherapy as front-line therapy in untreated metastatic breast cancer. J Clin Oncol 17:2355–2364.
10. Blackwell KL, Burstein HJ, Storniolo AM et al. (2010) Randomized study of Lapatinib alone or in combination with trastuzumab in women with ErbB2-positive, trastuzumab-refractory metastatic breast cancer. J Clin Oncol 28:1124–1130.
11. Blajman C, Balbiani L, Block J et al. (1999) A prospective, randomized phase III trial comparing combination chemotherapy with cyclophosphamide, doxorubicin, and 5-fluorouracil with vinorelbine plus doxorubicin in the treatment of advanced breast carcinoma. Cancer 85:1091–1097.
12. Blomqvist C, Elomaa I, Rissanen P et al. (1993) Influence of treatment schedule on toxicity and efficacy of cyclophosphamide, epirubicin, and fluorouracil in metastatic breast cancer: a randomized trial comparing weekly and every-4-week administration. J Clin Oncol 11:467–473.
13. Bonneterre J, Roche H, Monnier A et al. (2002) Docetaxel vs 5-fluorouracil plus vinorelbine in metastatic breast cancer after anthracycline therapy failure. Br J Cancer 87:1210–1215.
14. Bontenbal M, Andersson M, Wildiers J et al. (1998) Doxorubicin vs epirubicin, report of a second-line randomized phase II/III study in advanced breast cancer. EORTC Breast Cancer Cooperative Group. Br J Cancer 77:2257–2263.
15. Bontenbal M, Creemers GJ, Braun HJ et al. (2005) Phase II to III study comparing doxorubicin and docetaxel with fluorouracil, doxorubicin, and cyclophosphamide as first-line chemotherapy in patients with metastatic breast cancer: results of a Dutch Community Setting Trial for the Clinical Trial Group of the Comprehensive Cancer Centre. J Clin Oncol 23:7081–7088.
16. Brufman G, Colajori E, Ghilezan N et al. (1997) Doubling epirubicin dose intensity (100 mg/m^2 versus 50 mg/m^2) in the FEC regimen significantly

increases response rates. An international randomised phase III study in metastatic breast cancer. The Epirubicin High Dose (HEPI 010) Study Group. Ann Oncol 8:155–162.
17. Cameron D, Casey M, Press M et al. (2008) A phase III randomized comparison of lapatinib plus capecitabine versus capecitabine alone in women with advanced breast cancer that has progressed on trastuzumab: updated efficacy and biomarker analyses. Breast Cancer Res Treat 112:533–543.
18. Cassier PA, Chabaud S, Trillet-Lenoir V et al. (2008) A phase-III trial of doxorubicin and docetaxel versus doxorubicin and paclitaxel in metastatic breast cancer: results of the ERASME 3 study. Breast Cancer Res Treat 109:343–350.
19. Chan S, Davidson N, Juozaityte E et al. (2004) Phase III trial of liposomal doxorubicin and cyclophosphamide compared with epirubicin and cyclophosphamide as first-line therapy for metastatic breast cancer. Ann Oncol 15:1527–1534.
20. Chan S, Friedrichs K, Noel D et al. (1999) Prospective randomized trial of docetaxel versus doxorubicin in patients with metastatic breast cancer. J Clin Oncol 17:2341–2354.
21. Chan S, Romieu G, Huober J et al. (2009) Phase III study of gemcitabine plus docetaxel compared with capecitabine plus docetaxel for anthracycline-pretreated patients with metastatic breast cancer. J Clin Oncol 27:1753–1760.
22. Conte PF, Guarneri V, Bruzzi P et al. (2004) Concomitant versus sequential administration of epirubicin and paclitaxel as first-line therapy in metastatic breast carcinoma. Cancer 101:704–712.
23. Cowan JD, Neidhart J, McClure S et al. (1991) Randomized trial of doxorubicin, bisantrene, and mitoxantrone in advanced breast cancer: a Southwest Oncology Group study. J Natl Cancer Inst 83:1077–1084.
24. Del Mastro L, Venturini M, Lionetto R et al. (2001) 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-Ouest-Mammella Inter Gruppo Group. J Clin Oncol 19:2213–2221.
25. Di Leo A, Gomez HL, Aziz Z et al. (2008) Phase III, double-blind, randomized study comparing lapatinib plus paclitaxel with placebo plus paclitaxel as first-line treatment for metastatic breast cancer. J Clin Oncol 26:5544–5552.
26. Ejlertsen B, Mouridsen HT, Langkjer ST et al. (2004) Phase III study of intravenous vinorelbine in combination with epirubicin versus epirubicin alone in patients with advanced breast cancer: a Scandinavian Breast Group Trial (SBG9403). J Clin Oncol 22:2313–2320.
27. Esteban E, Lacave A, Fernandez J et al. (1999) Phase III trial of cyclophosphamide, epirubicin, fluorouracil (CEF) versus cyclophosphamide, mitoxantrone, fluorouracil (CNF) in women with metastatic breast cancer. Breast Cancer Res Treat 58:141–150.
28. Feher O, Vodvarka P, Jassem J et al. (2005) First-line gemcitabine versus epirubicin in postmenopausal women aged 60 or older with metastatic breast cancer: a multicenter, randomized, phase III study. Ann Oncol 16:899–908.
29. Finn RS, Crown JP, Lang I et al. (2015) The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): a randomised phase 2 study. The Lancet Oncology 16:25–35.
30. Fountzilas G, Dafni U, Dimopoulos M et al. (2009) A randomized phase III study comparing three anthracycline-free taxane-based regimens, as first line chemotherapy, in metastatic breast cancer. Breast Cancer Res Treat 115:87–99.
31. Fountzilas G, Kalofonos HP, Dafni U et al. (2004) Paclitaxel and epirubicin versus paclitaxel and carboplatin as first-line chemotherapy in patients with advanced breast cancer: a phase III study conducted by the Hellenic Cooperative Oncology Group. Ann Oncol 15:1517–1526.
32. Fountzilas G, Papadimitriou C, Dafni U et al. (2001) Dose-dense sequential chemotherapy with epirubicin and paclitaxel versus the combination, as first-line chemotherapy, in advanced breast cancer: a randomized study conducted by the Hellenic Cooperative Oncology Group. J Clin Oncol 19:2232–2239.
33. French Epirubicin Study Group (2000) Epirubicin-based chemotherapy in metastatic breast cancer patients: role of dose-intensity and duration of treatment. J Clin Oncol 18:3115–3124.
34. Gradishar WJ, Krasnojon D, Cheporov S et al. (2009) Significantly longer progression-free survival with nab-paclitaxel compared with docetaxel as first-line therapy for metastatic breast cancer. J Clin Oncol 27:3611–3619.
35. Gray R, Bhattacharya S, Bowden C et al. (2009) Independent review of E2100: a phase III trial of bevacizumab plus paclitaxel versus paclitaxel in women with metastatic breast cancer. J Clin Oncol 27:4966–4972.
36. Harris L, Batist G, Belt R et al. (2002) Liposome-encapsulated doxorubicin compared with

conventional doxorubicin in a randomized multicenter trial as first-line therapy of metastatic breast carcinoma. Cancer 94:25–36.
37. Harvey V, Mouridsen H, Semiglazov V et al. (2006) Phase III trial comparing three doses of docetaxel for second-line treatment of advanced breast cancer. J Clin Oncol 24:4963–4970.
38. Hausmaninger H, Lehnert M, Steger G et al. (1995) Randomised phase II study of epirubicin-vindesine versus mitoxantrone-vindesine in metastatic breast cancer. Eur J Cancer 31:2169–2173.
39. Heidemann E, Stoeger H, Souchon R et al. (2002) Is first-line single-agent mitoxantrone in the treatment of high-risk metastatic breast cancer patients as effective as combination chemotherapy? No difference in survival but higher quality of life were found in a multicenter randomized trial. Ann Oncol 13:1717–1729.
40. Hurvitz SA, Andre F, Jiang Z et al. (2015) Combination of everolimus with trastuzumab plus paclitaxel as first-line treatment for patients with HER2-positive advanced breast cancer (BOLERO-1): a phase 3, randomised, double-blind, multicentre trial. The Lancet Oncology 16:816–829.
41. Icli F, Akbulut H, Uner A et al. (2005) Cisplatin plus oral etoposide (EoP) combination is more effective than paclitaxel in patients with advanced breast cancer pretreated with anthracyclines: a randomised phase III trial of Turkish Oncology Group. Br J Cancer 92:639–644.
42. Jassem J, Pienkowski T, Pluzanska A et al. (2001) 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 19:1707–1715.
43. Joensuu H, Holli K, Heikkinen M et al. (1998) Combination chemotherapy versus single-agent therapy as first- and second-line treatment in metastatic breast cancer: a prospective randomized trial. J Clin Oncol 16:3720–3730.
44. Kaufman B, Mackey JR, Clemens MR et al. (2009) Trastuzumab plus anastrozole versus anastrozole alone for the treatment of postmenopausal women with human epidermal growth factor receptor 2-positive, hormone receptor-positive metastatic breast cancer: results from the randomized phase III TAnDEM study. J Clin Oncol 27:5529–5537.
45. Keller AM, Mennel RG, Georgoulias VA et al. (2004) Randomized phase III trial of pegylated liposomal doxorubicin versus vinorelbine or mitomycin C plus vinblastine in women with taxane-refractory advanced breast cancer. J Clin Oncol 22:3893–3901.
46. Krop IE, Kim S, González-Martín A et al. (2014) Trastuzumab emtansine versus treatment of physician's choice for pretreated HER2-positive advanced breast cancer (TH3RESA): a randomised, open-label, phase 3 trial. The Lancet Oncology 15:689–699.
47. Langley RE, Carmichael J, Jones AL et al. (2005) 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 23:8322–8330.
48. Marty M, Cognetti F, Maraninchi D et al. (2005) Randomized phase II trial of the efficacy and safety of trastuzumab combined with docetaxel in patients with human epidermal growth factor receptor 2-positive metastatic breast cancer administered as first-line treatment: the M77001 study group. J Clin Oncol 23:4265–4274.
49. Mavroudis D, Papakotoulas P, Ardavanis A et al. (2010) Randomized phase III trial comparing docetaxel plus epirubicin versus docetaxel plus capecitabine as first-line treatment in women with advanced breast cancer. Ann Oncol 21:48–54.
50. Miller KD, Chap LI, Holmes FA et al. (2005) Randomized phase III trial of capecitabine compared with bevacizumab plus capecitabine in patients with previously treated metastatic breast cancer. J Clin Oncol 23:792–799.
51. Nabholtz JM, Falkson C, Campos D et al. (2003) Docetaxel and doxorubicin compared with doxorubicin and cyclophosphamide as first-line chemotherapy for metastatic breast cancer: results of a randomized, multicenter, phase III trial. J Clin Oncol 21:968–975.
52. Nabholtz JM, Senn HJ, Bezwoda WR et al. (1999) 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 17:1413–1424.
53. Namer M, Soler-Michel P, Turpin F et al. (2001) 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 37:1132–1140.
54. Norris B, Pritchard KI, James K et al. (2000) Phase III comparative study of vinorelbine combined with doxorubicin versus doxorubicin alone in disseminated metastatic/recurrent breast cancer: National Cancer Institute of Canada Clinical Trials Group Study MA8. J Clin Oncol 18:2385–2394.

55. O'Brien ME, Wigler N, Inbar M et al. (2004) Reduced cardiotoxicity and comparable efficacy in a phase III trial of pegylated liposomal doxorubicin HCl (CAELYX/Doxil) versus conventional doxorubicin for first-line treatment of metastatic breast cancer. Ann Oncol 15:440–449.
56. O'Shaughnessy J, Miles D, Vukelja S et al. (2002) Superior survival with capecitabine plus docetaxel combination therapy in anthracycline-pretreated patients with advanced breast cancer: phase III trial results. J Clin Oncol 20:2812–2823.
57. Paridaens R, Biganzoli L, Bruning P et al. (2000) 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 18:724–733.
58. Parnes HL, Cirrincione C, Aisner J et al. (2003) 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 21:1819–1824.
59. Piccart M, Hortobagyi GN, Campone M et al. (2014) Everolimus plus exemestane for hormone-receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: overall survival results from BOLERO-2dagger. Ann Oncol 25:2357–2362.
60. Robert N, Leyland-Jones B, Asmar L et al. (2006) Randomized phase III study of trastuzumab, paclitaxel, and carboplatin compared with trastuzumab and paclitaxel in women with HER-2-overexpressing metastatic breast cancer. J Clin Oncol 24:2786–2792.
61. Sjöström J, Blomqvist C, Mouridsen H et al. (1999) 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 35:1194–1201.
62. Slamon DJ, Leyland-Jones B, Shak S et al. (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med 344:783–792.
63. Sledge GW, Neuberg D, Bernardo P et al. (2003) 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). Journal of Clinical Oncology 21:588–592.
64. Smith RE, Brown AM, Mamounas EP et al. (1999) Randomized trial of 3-h versus 24-h 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 17:3403–3411.
65. Sparano JA, Makhson AN, Semiglazov VF et al. (2009) Pegylated liposomal doxorubicin plus docetaxel significantly improves time to progression without additive cardiotoxicity compared with docetaxel monotherapy in patients with advanced breast cancer previously treated with neoadjuvant-adjuvant anthracycline therapy: results from a randomized phase III study. J Clin Oncol 27:4522–4529.
66. Swain SM, Kim S, Cortés J et al. (2013) Pertuzumab, trastuzumab, and docetaxel for HER2-positive metastatic breast cancer (CLEOPATRA study): overall survival results from a randomised, double-blind, placebo-controlled, phase 3 study. The Lancet Oncology 14:461–471.
67. Turner NC, Ro J, André F et al. (2015) Palbociclib in hormone-receptor–positive advanced breast cancer. N Engl J Med 373:209–219.
68. Verma S, Miles D, Gianni L et al. (2012) Trastuzumab emtansine for HER2-positive advanced breast cancer. N Engl J Med 367:1783–1791.
69. Wardley AM, Pivot X, Morales-Vasquez F et al. (2010) Randomized phase II trial of first-line trastuzumab plus docetaxel and capecitabine compared with trastuzumab plus docetaxel in HER2-positive metastatic breast cancer. Journal of Clinical Oncology 28:976–983.
70. Winer EP, Berry DA, Woolf S et al. (2004) Failure of higher-dose paclitaxel to improve outcome in patients with metastatic breast cancer: cancer and leukemia group B trial 9342. Journal of clinical oncology 22:2061–2068.
71. Yardley DA, Burris HA, Spigel DR et al. (2009) A phase II randomized crossover study of liposomal doxorubicin versus weekly docetaxel in the first-line treatment of women with metastatic breast cancer. Clinical breast cancer 9:247–252.
72. Zielinski C, Beslija S, Mrsic-Krmpotic Z et al. (2005) 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 23:1401–1408.

References

1. Baselga J, Cortés J, Kim S et al (2012) Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. N Engl J Med 366:109–119

2. Pierga JY, Delva R, Pivot X et al (2014) Bevacizumab and taxanes in the first-line treatment of metastatic breast cancer: overall survival and subgroup analyses of the ATHENA study in France. Bull Cancer 101:780–788
3. Dranitsaris G, Papadopoulos G (2014) Health Technology Assessment of Cancer Drugs in Canada, the United Kingdom and Australia: Should the United States Take Notice? Applied Health Economics and Health Policy, pp 1–12
4. Lowes R (2011) Avastin no longer approved for breast cancer, FDA says. Medscape Web site. Avastin no longer approved for breast cancer, FDA says. Medscape Web site
5. Fojo T, Grady C (2009) How much is life worth: cetuximab, non-small cell lung cancer, and the $440 billion question. J Natl Cancer Inst 101:1044–1048
6. Hillner BE, Smith TJ (2009) Efficacy does not necessarily translate to cost effectiveness: a case study in the challenges associated with 21st-century cancer drug pricing. J Clin Oncol 27:2111–2113
7. Ellis LM, Bernstein DS, Voest EE et al (2014) American Society of Clinical Oncology perspective: raising the bar for clinical trials by defining clinically meaningful outcomes. J Clin Oncol 32:1277–1280
8. Amir E, Seruga B, Kwong R et al (2012) Poor correlation between progression-free and overall survival in modern clinical trials: are composite endpoints the answer? Eur J Cancer 48:385–388
9. Eisenhauer E, Therasse P, Bogaerts J et al (2009) New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer 45:228–247
10. Broglio KR, Berry DA (2009) Detecting an overall survival benefit that is derived from progression-free survival. J Natl Cancer Inst 101:1642–1649
11. Saad ED, Katz A, Hoff PM et al (2010) Progression-free survival as surrogate and as true end point: insights from the breast and colorectal cancer literature. Ann Oncol 21:7–12
12. Fleming TR, Rothmann MD, Lu HL (2009) Issues in using progression-free survival when evaluating oncology products. J Clin Oncol 27:2874–2880
13. Zhuang SH, Xiu L, Elsayed YA (2009) Overall survival: a gold standard in search of a surrogate: the value of progression-free survival and time to progression as end points of drug efficacy. Cancer J 15:395–400
14. Wilkerson J, Fojo T (2009) Progression-free survival is simply a measure of a drug's effect while administered and is not a surrogate for overall survival. Cancer J 15:379–385
15. Tang PA, Bentzen SM, Chen EX et al (2007) 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 25:4562–4568
16. Buyse M, Burzykowski T, Carroll K et al (2007) Progression-free survival is a surrogate for survival in advanced colorectal cancer. J Clin Oncol 25:5218–5224
17. Shi Q, Sargent DJ (2009) Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. Int J Clin Oncol 14:102–111
18. Shi Q, de Gramont A, Grothey A et al (2015) Individual patient data analysis of progression-free survival versus overall survival as a first-line end point for metastatic colorectal cancer in modern randomized trials: findings from the analysis and research in cancers of the digestive system database. J Clin Oncol 33:22–28
19. Burzykowski T, Buyse M, Piccart-Gebhart MJ et al (2008) 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 26:1987–1992
20. Miksad RA, Zietemann V, Gothe R et al (2008) Progression-free survival as a surrogate endpoint in advanced breast cancer. Int J Technol Assess Health Care 24:371–383
21. Wong SS, Wilczynski NL, Haynes RB (2006) Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE. J Med Libr Assoc 94(1):41–47
22. Haynes RB, McKibbon KA, Wilczynski NL, Walter SD, Werre SR, Hedges Team (2005) Optimal search strategies for retrieving scientifically strong studies of treatment from medline: analytical survey. BMJ 330(7501):1179