



Surrogate threshold effect based on a meta-analysis for the predictive value of progression-free survival for overall survival in hormone receptor-positive, HER2-negative metastatic breast cancer

Michael Patrick Lux^{1,2,3,4} · Sarah Böhme⁵ · Stephanie Hücherig⁶ · Ulli Jeratsch⁶ · Niclas Kürschner⁵ · Diana Lüftner⁷

Received: 20 February 2019 / Accepted: 26 April 2019 / Published online: 7 May 2019
© The Author(s) 2019

Abstract

Purpose Clinical trials investigating therapies for metastatic breast cancer (mBC) generally use progression-free survival (PFS) as primary endpoint, which is not accepted as patient-relevant outcome within the German benefit assessment. Hence a validation of PFS as surrogate endpoint for overall survival (OS) is needed, e.g., in the indication of HR+, HER2-negative mBC.

Methods A systematic search was conducted. RCT were included if at least one study arm investigated fulvestrant, letrozole, tamoxifen, exemestane, or anastrozole. Additionally, hazard ratios reported for OS/PFS including confidence interval or standard error were mandatory. Pearson correlation coefficient was calculated to estimate the relation of surrogate endpoint PFS and patient-relevant outcome OS as well as the surrogate threshold effect (STE) which is used to determine thresholds for the estimate of the surrogate endpoint.

Results 16 studies with 5324 patients in total were included in the analyses. The correlation between hazard ratios of PFS and OS was statistically significant ($r=0.72$, 95% CI 0.35–0.90) representing a positive linear relationship. STE analysis was applied. The meta-regression model revealed a STE for HR_{PFS} of 0.60 and sensitivity analyses underlined robustness of the results.

Conclusions Based on the derived STE, it is possible to draw conclusions on a significant effect in OS for a hypothetical trial demonstrating an upper confidence limit of HR_{PFS} < 0.60 in PFS. However, only final OS results are able to confirm if a clinical relevant difference in survival time can be achieved.

Keywords Breast neoplasms · Surrogate threshold effect · Progression-free survival · Overall survival · Surrogate validation

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10549-019-05262-4>) contains supplementary material, which is available to authorized users.

✉ Michael Patrick Lux
m.lux@vincenz.de

Sarah Böhme
sarah.boehme@pfizer.com

Stephanie Hücherig
stephanie.huecherig@ams-europe.com

Ulli Jeratsch
Ulli.Jeratsch@ams-europe.com

Niclas Kürschner
niclas.kuerschner@pfizer.com

Diana Lüftner
diana.lueftner@charite.de

¹ Kooperatives Brustzentrum Paderborn, Paderborn, Germany

² Frauenklinik St. Louise, Paderborn, Germany

³ St. Josefs-Krankenhaus, Salzkotten, Germany

⁴ Frauen- und Kinderklinik St. Louise, Husener Str. 81, 33098 Paderborn, Germany

⁵ Pfizer Deutschland GmbH, Linkstraße 10, 10785 Berlin, Germany

⁶ AMS Advanced Medical Services GmbH, Rosa-Bavarese-Str. 5, 80639 Munich, Germany

⁷ Klinik für Hämatologie, Onkologie und Tumorimmunologie, Charité – Universitätsmedizin Berlin, Campus Benjamin Franklin, Hindenburgdamm 30, 12200 Berlin, Germany

Introduction

Endocrine therapies are the mainstay of treatment in hormone receptor-positive (HR+), human epidermal growth factor receptor 2 (HER2)-negative metastatic breast cancer (mBC) except in life-threatening situations qualifying the patient to receive chemotherapy [1].

Clinical trials investigating therapies for mBC often use progression-free survival (PFS) as primary endpoint [2], since patients with mBC have a relatively long survival time of around 3 years in median. With the desire to rapidly translate promising new agents into clinical practice, there is the need for endpoints which can be measured in a timely manner. Therefore, it is currently discussed whether endpoints based on disease progression, including PFS, time-to-progression (TTP), or time-to-treatment failure (TTF), are appropriate to demonstrate clinical benefit. These endpoints ensure an early availability of study outcomes and can serve as sensitive parameters for the benefit of a study medication as they are not influenced by subsequent lines of therapy or cross-over [2, 3]. Further advantages are the widespread use and comparability of PFS and TTP since they are most frequently used as primary endpoints in phase III trials and are worldwide accepted for the approval of new drugs [4–6].

However, the prolongation of overall survival (OS) is one of the most important therapeutic goals [7]. OS is regarded as unambiguous criterion, but there are certain disadvantages of OS as primary endpoint in the metastatic setting of breast cancer: the need for large numbers of patients, the long duration of follow-up phases until results become available, and the need for multiple subsequent therapies, which can confound OS. These limitations particularly cause difficulties in first-line studies [8, 9].

Health technology assessment (HTA) agencies worldwide generally accept PFS as endpoint in clinical trials [10], whereas the German Institute for Quality and Efficiency in Health Care (IQWiG) and the Federal Joint Committee (Gemeinsamer Bundesausschuss, G-BA) do not accept endpoints based on disease progression as a patient-relevant outcome within the benefit assessment of pharmaceuticals because they are measured by imaging techniques. Patient relevance of such endpoints might be accepted when measured via symptoms experienced by the patient. This would, however, lead to an omission of the re-evaluation of metastases in the course of clinical trials, which is considered unethical by physicians and does not comply with guideline recommendations [11]. Possible solutions for these different requirements have to be developed.

IQWiG suggested methods for the validation of surrogate endpoints in HTA context [12]. The aim of this study was the application of these methods in the indication of HR+, HER2-negative mBC to validate PFS as surrogate endpoint for OS.

Materials and methods

Literature search

A systematic search was conducted on the basis of the databases MEDLINE and EMBASE as well as in five EBM Reviews sources in September 2016 and was performed in accordance with PRISMA guidelines (Appendix A.1). The following keywords and associated subject headings were used: “breast cancer” and “metastatic” or “locally advanced” in combination with “fulvestrant” or “letrozole” or “tamoxifen” or “exemestane” or “anastrozole” (Online Appendices A.2–A.4). Inclusion criteria for trials are listed in Table 1.

Table 1 Inclusion criteria for trials in the systematic literature search

Population	Women with hormone receptor-positive and/or estrogen receptor-positive and/or progesterone receptor-positive, HER2-negative, locally advanced [not amenable to resection or radiotherapy with curative intent] or metastatic breast cancer regardless of line of treatment for locally advanced or metastatic disease
Intervention	At least one study arm investigated: fulvestrant, letrozole, tamoxifen, exemestane, or anastrozole
Comparator	Any drug intervention as single agent or in combination therapy
Endpoints	Overall survival and progression-free survival (according to RECIST) ^a reported as hazard ratio of interventional study drug vs. control from a cox proportional hazard model and confidence interval or standard error
Type of study	Randomized controlled trials (all phases)
Publication type	Randomized controlled trials reported in accordance with CONSORT guidelines
Language	English, German

CONSORT Consolidated Standards of Reporting Trials, HER2 human epidermal growth factor receptor 2, RECIST Response Evaluation Criteria In Solid Tumors, TTP time-to-progression

^aTTP or comparable endpoints were considered if the definition was identical to PFS (time from randomization to objective disease progression or death from any cause)

Randomized controlled trials (RCT) were included if at least 80% of the study population met the inclusion criteria. In case of missing information regarding HER2 status or HR status, the proportion of patients meeting the inclusion criteria was extrapolated based on epidemiological data. In case HER2 status was unknown, a proportion of 81.9% of HR+ patients was assumed to be HER2-negative [13]; for patients with both unknown HER2 status and hormone receptor status, a proportion of 64.5% was assumed to be HR+ and HER2-negative [13]. Trials with TTP or comparable endpoints were considered if the definition was identical to PFS (time from randomization to objective disease progression or death from any cause). Only studies reporting PFS according to Response Evaluation Criteria In Solid Tumors (RECIST) [14] were included to ensure standardized and comparable endpoint evaluation. Overall survival had to be reported in the studies and should be defined as the time from the date of randomization to the date of death from any cause.

Two reviewers independently assessed citations to determine relevance to the research question. Included studies were cross-checked for relevance by physicians. If several publications for one study were available, data from the latest publication or publications reporting final data cuts were used. Data from included studies were extracted by one reviewer; another reviewer checked for consistency against the original source. Risk of bias on study level was assessed and summarized for all included individual studies (Online Appendix A.5).

Statistical methods

As part of a rapid report, the German IQWiG presented methods for surrogate endpoints validation and recommendations for correlation-based procedures [12]. Health technology assessments are based on these methods in Germany. The methods include the evaluation of the certainty of conclusion of study results and the correlation between effect estimates of surrogate endpoint (e.g., PFS) and true outcome (e.g., OS) on trial level, whereas correlation is estimated by sample Pearson correlation coefficient r . Requirements for a successful surrogate validation are a high correlation (lower confidence limit (LCL) of $r > 0.85$) and a high certainty of conclusion of results of included studies. If the correlation is low (upper confidence limit < 0.7), no statement of surrogate validation is possible. In all other cases—where correlation is in the medium range and validity of surrogate endpoint is therefore unclear according to IQWiG methodology—they propose to apply the concept of STE [15], allowing conclusions on true endpoints by means of surrogate endpoints. STE is defined as minimal treatment effect on the surrogate endpoint explaining a non-zero (i.e., significant) effect on the true endpoint. In this context, STE represents the maximum value of the hazard ratio for PFS (HR_{PFS}) that needs

to be observed in a trial to ensure the possibility to draw conclusion of a significant effect on OS.

First, we tested the correlation between both outcomes ($H_0: r = 0$ vs. $H_1: r \neq 0$). Second, if correlation was medium, we fitted a random effects mixed-model to the data with moderator HR_{PFS} and outcome variable hazard ratio of OS (HR_{OS}) weighted by standard error (SE) of OS using the restricted maximum likelihood (REML) estimator for the amount of heterogeneity. Since SE is usually not reported, we recalculated it by means of 95% confidence interval (CI) of hazard ratio with $(\log(HR) - \log(HR_{LCL})) / z_{(0.975)}$, whereas $z_{(0.975)}$ is the 97.5 percentile of the standard normal distribution. Based on the regression fit, we calculated a prediction band to a significance level of $\alpha = 0.05$ for HR_{OS} . Meta-regression model and prediction values were implemented with R [16] using functions `rma.uni` and `predict.rma` from `metafor` package [17]. The STE resulted from the intersection of the upper prediction limit curve and the horizontal where $HR_{OS} = 1$ (zero effect).

In sensitivity analyses, we investigated if factors HER2 status (reported vs. not reported), line of treatment (only first-line vs. others), and therapy option (studies comparing combination therapy with monotherapy vs. studies comparing two monotherapies) accounted for heterogeneity.

Results

Systematic literature search

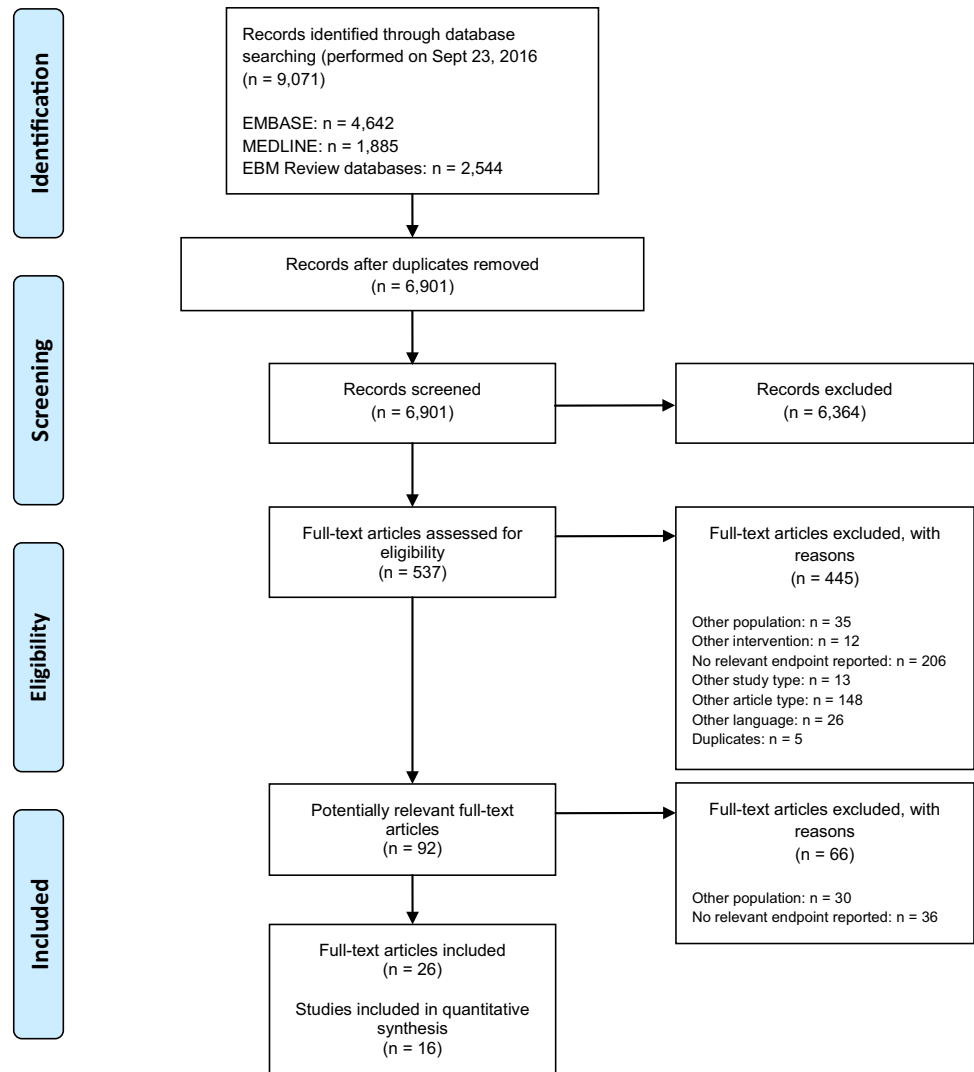
The search identified 9071 citations from MEDLINE®, EMBASE, and EBM Review databases. We included 16 studies (26 full texts) for analysis (Fig. 1).

Characteristics for included trials are summarized in Table 2. The 16 trials included 5324 patients in total. In ten trials, HER2 status was reported for the entire study population. Six trials were included in the analysis since 80% of the study population met the inclusion criteria due to calculations according to epidemiological data (see methods). Six trials (2875 patients) evaluated treatments exclusively in the first-line setting for locally advanced or metastatic disease, and ten trials (2449 patients) included pretreated patients or patients in various lines of treatment. Almost all trials included postmenopausal women except for two trials which included a small (2.9%) [18] or unknown [19] number of premenopausal women treated with GnRH agonists.

Twelve trials compared combination therapy with monotherapy, while four trials compared monotherapy versus monotherapy. Combination treatments were add-on to hormone therapy and comprised different compound classes in comparison to endocrine therapy.

Endpoints were reported for intention-to-treat population (seven trials), full analysis set (three trials), modified

Fig. 1 Flow diagram of study selection process. *N* Number of patients



intention-to-treat (two trials), or for all randomized patients (three trials). For one trial, no information was given on the analysis population.

Statistical analysis

In the main analysis (pool of 16 identified trials), the correlation between hazard ratios of PFS and OS was statistically significant ($r = 0.72$, 95% CI 0.35–0.90, $p = 0.0016$) representing a positive linear relationship of surrogate endpoint and by this patient-relevant endpoint. According to the definition in IQWiG's rapid report, correlation was merely medium-sized and therefore the validity of the surrogate endpoint is unclear and a STE analysis is applied. The meta-regression showed low residual heterogeneity ($\tau^2 = 0.009$, $I^2 = 25\%$) and provided a significant result of the moderator coefficient β_{PFS} ($p = 0.0206$). STE for HR_{PFS} was 0.60 (Fig. 2), and thus for trials meeting the above-mentioned inclusion criteria in this specific indication and

upper confidence limit of HR_{PFS} below STE, it is possible to draw the conclusion of a significant effect on OS by means of surrogate endpoint PFS.

Sensitivity analyses to check the robustness in the main analysis were performed to account for available information about HER2 status (sensitivity analysis 1), line of treatment (sensitivity analysis 2), or therapy option (sensitivity analysis 3) (Table 3). Due to the smaller sample sizes in the subpools, STE values deviate from the value in the main analysis, but correlation in all subpools is positive and at least of a medium magnitude, confirming the positive relationship between OS and PFS. In all subpools STE is below 1 except for sensitivity analysis 2b (Table 3). In this case, STE cannot be calculated (upper confidence limit of $\text{HR}_{\text{OS}} > 1$ for any value of HR_{PFS}). Hence, meta-regression analyses in all specified subpools did not show heterogeneity regarding the observed factors and confirm the results of the main analysis.

Table 2 Overview of trial characteristics

No.	Primary Author, year of publication	Enrollment period	Patients with HER2-negative status (%)	Hormone receptor-positive status (%)	Double-blinded	Line of therapy ^a	Therapy option (c = combi vs. mono; m = mono vs. mono)	Intervention (study drug vs. control)	Median follow-up	HR OS (95% CI), <i>n</i> _A = number of events of study drug, <i>n</i> _B = number of events control	HR PFS (95% CI), <i>n</i> _A = number of events of study drug, <i>n</i> _B = number of events control
(1)	Bachelot, 2012 [20] ^b	March 2008–May 2009	95.5	100 (ER and/or PR)	No	1+2	c	Tamoxifen + everolimus (<i>n</i> = 54) vs. tamoxifen (<i>n</i> = 57)	Tamoxifen + everolimus: 23.7 months (range 2.6–32.7 months), tamoxifen: 24.2 months (range 0.9–36.2 months)	0.45 (0.24, 0.81), <i>n</i> _A = 16, <i>n</i> _B = 31	TTP, 0.54 (0.36, 0.81), <i>n</i> _A = NA, <i>n</i> _B = NA
(2)	Bergh, 2012 (FACT) [18] ^b	January 2004–March 2008	NA ^c	100 (ER and/or PR)	No	1	c	Anastrozole + fulvestrant (<i>n</i> = 258) vs. anastrozole (<i>n</i> = 256)	8.9 months (range 0–54 months)	1.00 (0.76, 1.32), <i>n</i> _A = 102, <i>n</i> _B = 102	TTP, 0.99 (0.81, 1.20), <i>n</i> _A = 200, <i>n</i> _B = 200
(3)	Burstein, 2014 (CALGB 40302) [21] ^b	September 2006–July 2010	80.8	100 (ER and/or PR)	Yes	1+2+3	c	Fulvestrant + lapatinib (<i>n</i> = 146) vs. fulvestrant + placebo (<i>n</i> = 145)	33.6 months	1.10 (0.83, 1.47) ^d , <i>n</i> _A = NA, <i>n</i> _B = NA	0.96 (0.75, 1.22) ^d , <i>n</i> _A = NA, <i>n</i> _B = NA
(4)	Clemons, 2014 (ZAM-BONEY) [22] ^b	October 2009–October 2011	95.3	100 (ER and/or PR of primary tumor)	Yes	1+2+3	c	Fulvestrant + vandetanib (<i>n</i> = 61) vs. fulvestrant + Placebo (<i>n</i> = 68)	NA	0.69 (0.37, 1.31), <i>n</i> _A = 18, <i>n</i> _B = 23	0.94 (0.64, 1.36), <i>n</i> _A = 55, <i>n</i> _B = 63
(5)	Di Leo, 2014 (CONFIRM) ^e [23] ^b [24] ^b	February 2005–August 2007	NA ^c	100 ER	Yes	1+2	m	Fulvestrant 500 mg (<i>n</i> = 362) vs. fulvestrant 250 mg (<i>n</i> = 374)	NA	0.79 (0.67, 0.94), <i>n</i> _A = 261, <i>n</i> _B = 293	0.79 (0.68, 0.93) ^f , <i>n</i> _A = 297, <i>n</i> _B = 321
(6)	Dickler, 2016 (CALGB 40503) [19] ^b	May 2008–November 2011	90.7	100 (ER and/or PR)	Yes	1+2	c	Letrozole + bevacizumab (<i>n</i> = 173) vs. letrozole + Placebo (<i>n</i> = 170)	39 months for PFS, 42 months for OS	0.87 (0.65, 1.18), <i>n</i> _A = 81, <i>n</i> _B = 90	0.75 (0.59, 0.96), <i>n</i> _A = 126, <i>n</i> _B = 138

Table 2 (continued)

No.	Primary Author, year of publication	Enrollment period	Patients with HER2-negative status (%)	Hormone receptor-positive status (%)	Double-blinded	Line of therapy ^a	Therapy option (c = combi vs. mono; m = mono vs. mono)	Intervention (study drug vs. control)	Median follow-up	HR OS (95% CI), <i>n</i> _A = number of events study drug, <i>n</i> _B = number of events control	HR PFS (95% CI), <i>n</i> _A = number of events study drug, <i>n</i> _B = number of events control
(7)	Finn, 2015 (PALOMA-1) [25] ^b	December 2009–May 2012	100	100 (ER)	No	I	c	Letrozole + palbociclib (<i>n</i> = 84) vs. letrozole (<i>n</i> = 81)	Letrozole + palbociclib: 29.6 months (95% CI 27.9–36.0), letrozole: 27.9 months (95% CI 25.5–31.1)	0.81 (0.49, 1.34), <i>n</i> _A = 30, <i>n</i> _B = 31	0.49 (0.32, 0.75), <i>n</i> _A = 41, <i>n</i> _B = 59
(8)	Iwata, 2013 [26] ^b	April 2005–December 2010	~86.6 ^c	88.66 (ER and/or PR)	Yes	I	m	Exemestane (<i>n</i> = 147) vs. anastrozole (<i>n</i> = 145)	NA	1.06 (0.73, 1.54), <i>n</i> _A = 57, <i>n</i> _B = 55	TTP, 1.01 (0.77, 1.32), <i>n</i> _A = 103, <i>n</i> _B = 114
(9)	Johnston, 2013 (SoFEA) ^g [27] ^b	March 2004–August 2010	~86.6 ^h	99.6% (ER and/or PR)	Yes (anastrozole and placebo)	1+2	c	Fulvestrant + anastrozole (<i>n</i> = 243) vs. fulvestrant + Placebo (<i>n</i> = 231)	37.9 months (interquartile range 23.1–50.8 months)	0.95 (0.76, 1.17), <i>n</i> _A = 168, <i>n</i> _B = 167	1.00 (0.83, 1.21), <i>n</i> _A = 235, <i>n</i> _B = 221
(10)	Llombart-Cusac, 2012 [28] ^b	September 2001–May 2003	NA ^c	100 (ER and/or PR)	No	I	m	Exemestane (<i>n</i> = 49) vs. Anastrozole (<i>n</i> = 51)	9.1 months (range, 0.07–79.96 months)	1.33 (0.78, 2.25), <i>n</i> _A = NA, <i>n</i> _B = NA	TTP, 1.13 (0.75, 1.72), <i>n</i> _A = 44, <i>n</i> _B = 44
(11)	Martin, 2015 (LEA) [29] ^b	November 2007–August 2011	100	100	No	I	c	Letrozole/fulvestrant + bevacizumab (<i>n</i> = 190) vs. letrozole/fulvestrant (<i>n</i> = 184)	23.7 months (range, 0–58.2 months)	0.87 (0.58, 1.32), <i>n</i> _A = 47, <i>n</i> _B = 45	0.83 (0.65, 1.06), <i>n</i> _A = 128, <i>n</i> _B = 135
(12)	Mehta, 2012 [30] ^b	June 2004–July 2009	89.3 ^h	100 (ER and/or PR)	No	I	c	Anastrozole + fulvestrant (<i>n</i> = 349) vs. anastrozole (<i>n</i> = 345)	35 months (range, 3–78) for PFS	0.81 (0.65, 1.00), <i>n</i> _A = 154, <i>n</i> _B = 176	0.80 (0.68, 0.94), <i>n</i> _A = 268, <i>n</i> _B = 297

Table 2 (continued)

No.	Primary Author, year of publication	Enrollment period	Patients with HER2-negative status (%)	Hormone receptor-positive status (%)	Double-blinded	Line of therapy ^a	Therapy option (c = combi vs. mono; m = mono vs. mono)	Intervention (study drug vs. control)	Median follow-up	HR OS (95% CI), n_A = number of events study drug, n_B = number of events control	HR PFS (95% CI), n_A = number of events study drug, n_B = number of events control
(13)	Piccart, 2014 (BOLERO-2) [31] ^b	June 2009 – January 2011	100	100 (ER and/or PR)	Yes	1 + 2	c	Exemestane + everolimus ($n = 485$) vs. exemestane + placebo ($n = 239$)	NA	0.89 (0.73, 1.10), $n_A = 267$, $n_B = 143$	0.45 (0.38, 0.54), $n_A = NA$, $n_B = NA$
(14)	Robertson, 2013 [32] ^b	March 2008 – July 2009	94.9	100 (ER and/or PR)	Yes (ganitumab/placebo)	1 + 2	c	Exemestane/fulvestrant + ganitumab ($n = 106$) vs. exemestane/fulvestrant + placebo ($n = 50$)	NA	1.78 (1.06, 2.98) ^j , $n_A = 63$, $n_B = 19$	1.17 (0.80, 1.72) ^j , $n_A = NA$, $n_B = NA$
(15)	Yamamoto, 2013 [33] ^b	October 2008 – November 2011	91.2	100 (ER and/or PR)	No	≥ 2	m	Toremifene ($n = 46$) vs. exemestane ($n = 45$)	NA	0.60 (0.26, 1.39), $n_A = NA$, $n_B = NA$	0.61 (0.38, 0.99), $n_A = NA$, $n_B = NA$
(16)	Yardley, 2013 [34] ^b	June 2008 – July 2010	90.8	98.5% (ER) 78.5% (PR)	Yes	≥ 1 NA	c	Exemestane + entinostat ($n = 64$) vs. exemestane + placebo ($n = 66$)	Exemestane + entinostat: 24.0 months, exemestane + placebo: 26.4 months, for OS, respectively	0.59 (0.36, 0.97), $n_A = NA$, $n_B = NA$	0.73 (0.50, 1.07), $n_A = NA$, $n_B = NA$

Table 2 (continued)

CI confidence interval, HR hazard ratio of interventional study drug vs. control, NA not available

^aLine of therapy for locally advanced or metastatic disease, previous therapy included endocrine and/or chemotherapy

^bPublication reporting hazard ratio for relevant endpoints

^cAccording to registry data, a proportion of 81.9% of the hormone receptor-positive patients was assumed to be HER2-negative in case HER2 status was unknown [13]. For patients with both unknown HER2 status and hormone receptor status, a proportion of 64.5% was assumed to be hormone receptor-positive and HER2-negative [13]

^dRecalculated hazard ratio for lapatinib vs. placebo. Burstein, 2014 originally reported inverse effect measures of placebo to lapatinib: OS HR: 0.91, 95% CI: 0.68–1.21, PFS HR: 1.04, 95% CI: 0.82–1.33

^eCONFIRM results for endpoint OS are published by Di Leo, 2010 and results for endpoint PFS are published by Di Leo, 2014

^fCorrected HR and CI used for analyses [23]

^gSoFEEA also examined third treatment arm exemestane and compared fulvestrant + placebo vs. exemestane which is not included in STE analysis

^hStudy included both patients with known HER2 status as well as patients with unknown HER2 status (hormone receptor status of patients positive for both groups). For the latter group of patients, 81.9% was assumed to be HER2-negative; see footnote (d)

ⁱMehta, 2012 reported comparison of anastrozole vs. anastrozole + fulvestrant. To be consistent with other trials in the study pool comparing a combination therapy with a mono therapy, we calculated inverse hazard ratios to represent the comparison of anastrozole + fulvestrant vs. anastrozole

^jRecalculated 95% CI to be consistent with other results of this table. Robertson, 2013 originally reported 80% CI for OS (HR 1.78, 80% CI 1.27–2.50; $p = 0.025$) and PFS (HR 1.17, 80% CI 0.91–1.50; $p = 0.44$)

Discussion

PFS is an accepted endpoint with a definition based on standardized criteria according to RECIST [14]. The outcome of PFS is not influenced by subsequent therapies, and results are timely available and a lower number of patients are needed than for OS. In addition, results are widely accepted for the approval [4, 5] as well as the HTA evaluation of new drugs [10] except from German HTA bodies due to an assumption of missing proof of patient relevance due to evaluation of PFS by imaging and not by symptoms.

From a physician's point of view, PFS has a high relevance for patients. In case of a progression, the patient's therapy needs to be changed, which entails different adverse effects and requires new procedures and adjustments of schedules. A proven progression also has a significant impact on the psychological well-being and quality of life [35].

Additionally, a prolongation of OS and maintaining quality of life continues to be the focus of treatment in the metastatic situation of breast cancer [7]. To quickly transfer results on PFS from trials with innovative therapies to clinical practice, it would be advantageous if a validation of progression-based endpoints as surrogate endpoint for OS is available, which was the aim of this study.

Methods used in this work have some limitations. It is possible that the pool of included studies does not include all publicly available data because the search was limited to three literature databases and included no further sources. In addition, several aspects often lead to exclusion of studies. One reason was poor reporting, for example if data for only one of the required endpoint were published. Lack of information regarding HER2 status leading to non-conformity with the defined patient population and no PFS/TTP assessment according to RECIST criteria were other reasons. Especially older studies were often not in accordance with the inclusion criteria.

The sensitivity analyses show that the STE values vary strongly when only very small study subsets are considered. Nevertheless, the values are not so far apart that they would point completely in the other direction, i.e., $STE > 1$. Furthermore, the STE is sensitive to outlier observations when number of studies in the model is low. The generation of randomization and whether allocation concealment was adequately carried out was rarely reported in the individual studies. To what extent this has an impact on the endpoints OS and PFS and finally on the STE remains unclear.

According to IQWiG's method description, the entire 95% CI of PFS has to be below the STE in order to take into account the uncertainty with which both estimators are affected. Gillhaus et al. [36] described that this approach reduces the α error, but also considerably reduces the power

Fig. 2 Meta-regression showing the relationship between hazard ratios of PFS and OS. Expansions of circles were scaled by the inverse of the standard error of HR_{OS} . Numbers in parentheses reflect studies in Table 2. STE is defined as maximum value of HR_{PFS} so that HR_{OS} still is significant, i.e., upper confidence limit of $HR_{OS} < 1$. *CI* Confidence interval, *HR* hazard ratio, *OS* overall survival, *PFS* progression-free survival, *r* Pearson correlation coefficient, *STE* surrogate threshold effect

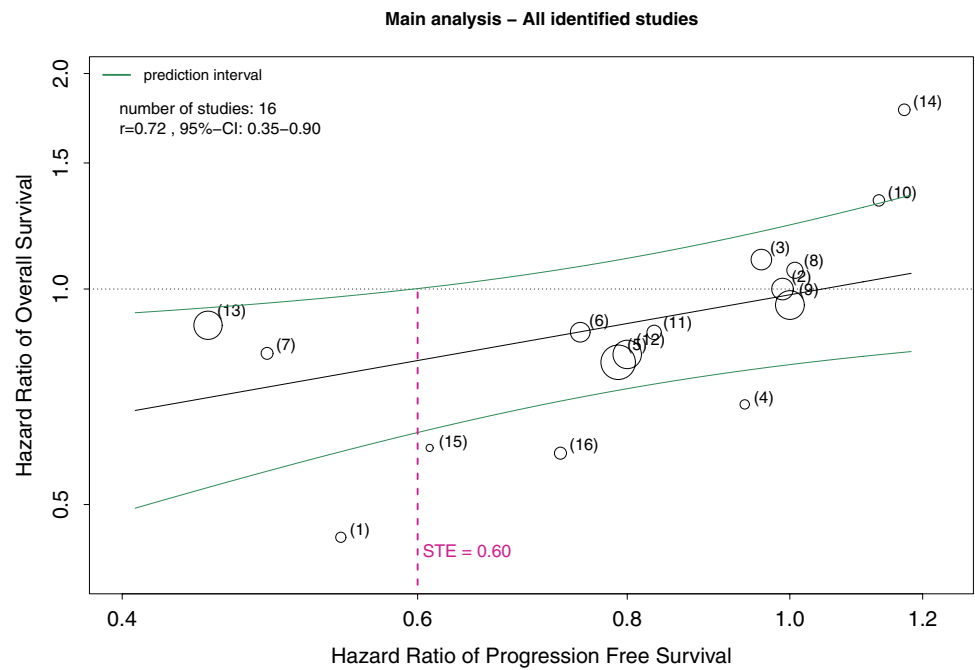


Table 3 Overview of sensitivity analyses

Analysis	Pool	Number of studies	Correlation	STE
Main analysis	All identified studies	16	0.72	0.60
Sensitivity analysis 1	HER2 status			
	(a) Studies reporting HER2 status of patients	13	0.68	0.45
	(b) Studies not reporting HER2 status	3	n. c. ^a	0.86
Sensitivity analysis 2	Line of treatment			
	(a) Studies only including first-line patients	6	0.82	0.75
	(b) Studies including pretreated patients or patients in various lines	10	0.71	n.c. ^b
Sensitivity analysis 3	Therapy option			
	(a) Studies comparing a combination therapy with a mono therapy (combi vs. mono)	12	0.65	0.47
	(b) Studies comparing two monotherapies (mono vs. mono)	4	0.99	0.84

HER2 human epidermal growth factor receptor 2, *n.c.* not calculable

^aIn order to calculate a correlation coefficient, at least 4 studies are needed

^bUpper confidence limit of $HR_{OS} > 1$ for any value of HR_{PFS}

of the STE concept. Statistical power could be increased using a lower α significance level (e.g., 0.1 or 0.2) for the prediction band of HR_{OS} in the meta-regression model. However, this assumption can only be made if the hypothetical trial is conducted in patients with HR+, HER2-negative mBC treated with endocrine therapies alone or in combination with other targeted treatments. The model does not intend to predict the outcome of OS concerning HR or differences in median of OS from the model.

In general, OS results always need a critical appraisal. Especially in mBC, an improvement of OS for a new therapy

option is difficult to measure. Factors like the heterogeneity of the disease, therapy complexity with integration of local therapies (surgery, radiotherapy), and a wide range of systemic therapies as well as a long survival in the metastatic situation with numerous different sequential courses of therapy may have an impact on the results of OS. A model calculation has shown that the probability of demonstrating a significant OS benefit decreases to less than 30% for a post-progression survival (PPS) of more than 12 months [37]. However, survival of several years has been reached especially in mBC. In addition, depending on the required

statistical power, thousands of patients need to be recruited to identify a survival benefit. In the age of individualized therapy with numerous specific subgroups, these studies are hardly feasible. The authors also conclude that the interpretation of OS is only useful, if the PPS is really short [37].

Additional points to take into account are the clinical relevance of OS results. The STE calculated in this publication only allows to draw conclusions on OS in the above-mentioned settings and about the statistical significance of OS. However, it is not possible to predict the differences in median survival times and its clinical relevance. Therefore, it is possible that the final result for OS is statistically significant in a trial but might not be considered clinically relevant. For example, a difference of 3 months in median OS is clinically relevant in an indication with very short survival times like metastatic pancreatic carcinoma [38]. MBC has comparably long survival times of 2–3 years [39] and a difference of 3 months in median OS would normally not be considered clinically relevant. Even if a meaningfully relevant difference in median OS was achieved, a proven prolongation of life with a simultaneous significant deterioration in the quality of life is not always a desirable therapeutic goal [40].

In conclusion, we were able to calculate the STE (0.60) allowing to draw conclusions on OS through the surrogate endpoint PFS besides minor methodological limitations in trials with HR+, HER2-negative mBC treated with endocrine therapies alone or on combination.

This means that for a hypothetical or future trial demonstrating upper confidence limit of $HR_{PFS} < 0.60$ in PFS it is possible to conclude on a significant effect in OS. However, only final OS results can confirm if a clinical relevant difference in survival time is reached. For future prospects, reflecting the current results in regard to ongoing clinical studies examining the addition of CDK 4/6 inhibitors to endocrine therapy will be desirable since they mostly lack of statistical significant, mature OS data for the time being. As long as OS results are not available, conclusions using STE may be drawn from PFS. To gain quick results on a new drug, PFS remains a relevant endpoint with high clinical relevance.

Acknowledgements Sarah Böhme and Ulli Jeratsch thank Ralf Goertz and Johanna Gillhaus for their input and support with statistical methods. Stephanie Hücherig thanks Stefanie Wüstner for her support with the systematic literature review. Sarah Böhme and Niclas Kürschner thank Friedhelm Leverkus and Rainer Niedtner for their valuable input.

Funding This work was supported by Pfizer Deutschland GmbH (no grant numbers applied).

Compliance with ethical standards

Conflict of interest M.P.L. has received honoraria for lectures and advisory boards from Pfizer, Novartis, Lilly, AstraZeneca, Roche, Genomic Health, medac, Grünenthal, Tesaro, and Eisai. D.L. has re-

ceived honoraria for lectures and advisory boards from Hexal, Daiichi Sankyo, Pfizer, Novartis, Lilly, AstraZeneca, Roche, Loreal, and Tesaro. S.B. and N.K. are employees of Pfizer Deutschland. U.J. and S.H. are employees of Advanced Medical Services GmbH and received a research grant from Pfizer Deutschland GmbH.

Research involving human participants and/or animals This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Arbeitsgemeinschaft Gynäkologische Onkologie e.V. (2018) Kommission Mamma. Diagnostik und Therapie von Patientinnen mit primärem und metastasiertem Brustkrebs—Empfehlungen der AGO Kommission Mamma. https://www.ago-online.de/fileadmin/downloads/leitlinien/mamma/2018-03/Gesamt_deutsch/Alle_aktuellen_Empfehlungen_2018.pdf. Accessed 04 June 2018
2. Saad ED, Katz A (2009) Progression-free survival and time to progression as primary end points in advanced breast cancer: often used, sometimes loosely defined. *Ann Oncol* 20(3):460–464
3. Burzykowski T, Buyse M, Piccart-Gebhart MJ, Sledge G, Carmichael J, Luck HJ 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(12):1987–1992
4. U.S. Department of Health and Human Services FDA, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER) (2007) Guidance for Industry Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. <https://www.fda.gov/downloads/drugsGuidanceComplianceRegulatoryInformation/Guidance/UCM071590.pdf>. Accessed 26 July 2018
5. European Medicines Agency (EMA) (2018) Guideline on the evaluation of anticancer medicinal products in man 2017. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2017/11/WC500238764.pdf. Accessed 26 July 2018
6. Seidman AD, Bordeleau L, Fehrenbacher L, Barlow WE, Perlmutter J, Rubinstein L et al (2018) National Cancer Institute Breast Cancer Steering Committee Working Group Report on meaningful and appropriate end points for clinical trials in metastatic breast cancer. *J Clin Oncol*. 36(32):3259–3268
7. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft DK, AWMF) (2017) S3-Leitlinie Früherkennung, Diagnose, Therapie und Nachsorge des Mammakarzinoms, Version 4.0, 2017 AWMF Registernummer: 032-045OL. <https://www.leitlinienprogramm-onkologie.de/leitlinien/mammakarzinom/>. Accessed 26 July 2018
8. Di Leo A, Bleiberg H, Buyse M (2003) 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* 21(10):2045–2047
9. Sargent DJ, Hayes DF (2008) Assessing the measure of a new drug: is survival the only thing that matters? *J Clin Oncol* 26(12):1922–1923

10. Chabot I, Rocchi A (2014) Oncology drug health technology assessment recommendations: Canadian versus UK experiences. *Clinicoecon Outcomes Res* 6:357–367
11. Thill M, Liedtke C, Solomayer EF, Muller V, Janni W, Schmidt M et al (2017) AGO recommendations for the diagnosis and treatment of patients with advanced and metastatic breast cancer: update 2017. *Breast Care (Basel)* 12(3):184–191
12. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (2011) IQWiG-Berichte Nr. 80—Aussagekraft von Surrogatendpunkten in der Onkologie. Rapid Report
13. Tumorregister München (2015) Spezielle Auswertungen Mammkarzinom (C50): Metastasierte Patientinnen—Diagnosejahr 2002–2013
14. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R et al (2009) New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 45(2):228–247
15. Burzykowski T, Molenberghs G, Buyse M (2005) *The evaluation of surrogate endpoints*. Springer, New York
16. R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org/>. Accessed 27 Sept 2018
17. Viechtbauer W (2010) Conducting meta-analyses in R with the metafor Package. *J Stat Softw* 36(3):1–48
18. Bergh J, Jonsson PE, Lidbrink EK, Trudeau M, Eiermann W, Brattstrom D et al (2012) FACT: an open-label randomized phase III study of fulvestrant and anastrozole in combination compared with anastrozole alone as first-line therapy for patients with receptor-positive postmenopausal breast cancer. *J Clin Oncol* 30(16):1919–1925
19. Dickler MN, Barry WT, Cirrincione CT, Ellis MJ, Moynahan ME, Innocenti F et al (2016) Phase III trial evaluating letrozole as first-line endocrine therapy with or without bevacizumab for the treatment of postmenopausal women with hormone receptor-positive advanced-stage breast cancer: CALGB 40503 (alliance). *J Clin Oncol* 34(22):2602–2609
20. Bachelot T, Bourcier C, Cropet C, Ray-Coquard I, Ferrero JM, Freyer G et al (2012) Randomized phase II trial of everolimus in combination with tamoxifen in patients with hormone receptor-positive, human epidermal growth factor receptor 2-negative metastatic breast cancer with prior exposure to aromatase inhibitors: a GINECO study. *J Clin Oncol* 30(22):2718–2724
21. Burstein HJ, Cirrincione CT, Barry WT, Chew HK, Tolane SM, Lake DE et al (2014) Endocrine therapy with or without inhibition of epidermal growth factor receptor and human epidermal growth factor receptor 2: a randomized, double-blind, placebo-controlled phase III trial of fulvestrant with or without lapatinib for postmenopausal women with hormone receptor-positive advanced breast cancer-CALGB 40302 (alliance). *J Clin Oncol* 32(35):3959–3966
22. Clemons MJ, Cochrane B, Pond GR, Califaretti N, Chia SK, Dent RA et al (2014) Randomised, phase II, placebo-controlled, trial of fulvestrant plus vandetanib in postmenopausal women with bone only or bone predominant, hormone-receptor-positive metastatic breast cancer (MBC): the OCOG ZAMBONEY study. *Breast Cancer Res Treat* 146(1):153–162
23. Di Leo A, Jerusalem G, Petruzella L, Torres R, Bondarenko IN, Khasanov R et al (2010) Results of the CONFIRM phase III trial comparing fulvestrant 250 mg with fulvestrant 500 mg in postmenopausal women with estrogen receptor-positive advanced breast cancer. *J Clin Oncol* 28(30):4594–4600
24. Di Leo A, Jerusalem G, Petruzella L, Torres R, Bondarenko IN, Khasanov R et al (2014) Final overall survival: fulvestrant 500 mg vs 250 mg in the randomized CONFIRM trial. *J Natl Cancer Inst* 106(1):djt337
25. Finn RS, Crown JP, Lang I, Boer K, Bondarenko IM, Kulyk SO 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. *Lancet Oncol* 16(1):25–35
26. Iwata H, Masuda N, Ohno S, Rai Y, Sato Y, Ohsumi S et al (2013) A randomized, double-blind, controlled study of exemestane versus anastrozole for the first-line treatment of postmenopausal Japanese women with hormone-receptor-positive advanced breast cancer. *Breast Cancer Res Treat* 139(2):441–451
27. Johnston SR, Kilburn LS, Ellis P, Dodwell D, Cameron D, Hayward L et al (2013) Fulvestrant plus anastrozole or placebo versus exemestane alone after progression on non-steroidal aromatase inhibitors in postmenopausal patients with hormone-receptor-positive locally advanced or metastatic breast cancer (SoFEA): a composite, multicentre, phase 3 randomised trial. *Lancet Oncol* 14(10):989–998
28. Llombart-Cussac A, Ruiz A, Anton A, Barnadas A, Antolin S, Ales-Martinez JE et al (2012) Exemestane versus anastrozole as front-line endocrine therapy in postmenopausal patients with hormone receptor-positive, advanced breast cancer: final results from the Spanish Breast Cancer Group 2001-03 phase 2 randomized trial. *Cancer* 118(1):241–247
29. Martin M, Loibl S, von Minckwitz G, Morales S, Martinez N, Guerrero A et al (2015) Phase III trial evaluating the addition of bevacizumab to endocrine therapy as first-line treatment for advanced breast cancer: the letrozole/fulvestrant and avastin (LEA) study. *J Clin Oncol* 33(9):1045–1052
30. Mehta RS, Barlow WE, Albain KS, Vandenberg TA, Dakhil SR, Tirumali NR et al (2012) Combination anastrozole and fulvestrant in metastatic breast cancer. *N Engl J Med* 367(5):435–444
31. Piccart M, Hortobagyi GN, Campone M, Pritchard KI, Lebrun F, Ito Y 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-2 dagger. *Ann Oncol* 25(12):2357–2362
32. Robertson JF, Ferrero JM, Bourgeois H, Kennecke H, de Boer RH, Jacot W et al (2013) Ganitumab with either exemestane or fulvestrant for postmenopausal women with advanced, hormone-receptor-positive breast cancer: a randomised, controlled, double-blind, phase 2 trial. *Lancet Oncol* 14(3):228–235
33. Yamamoto Y, Ishikawa T, Hozumi Y, Ikeda M, Iwata H, Yamashita H et al (2013) Randomized controlled trial of toremifene 120 mg compared with exemestane 25 mg after prior treatment with a non-steroidal aromatase inhibitor in postmenopausal women with hormone receptor-positive metastatic breast cancer. *BMC Cancer* 13:239
34. Yardley DA, Ismail-Khan RR, Melichar B, Lichinitser M, Munster PN, Klein PM et al (2013) Randomized phase II, double-blind, placebo-controlled study of exemestane with or without entinostat in postmenopausal women with locally recurrent or metastatic estrogen receptor-positive breast cancer progressing on treatment with a nonsteroidal aromatase inhibitor. *J Clin Oncol* 31(17):2128–2135
35. Muller V, Nabieva N, Haberle L, Taran FA, Hartkopf AD, Volz B et al (2018) Impact of disease progression on health-related quality of life in patients with metastatic breast cancer in the PRAE-GNANT breast cancer registry. *Breast* 37:154–160
36. Gillhaus J, Goertz R, Jeratsch U, Leverkus F (2017) Validation of surrogates by correlation and surrogate threshold effect—results of simulation studies. *GMS Med Inform Biom Epidemiol*. <https://doi.org/10.3205/mibe000168>
37. Broglio KR, Berry DA (2009) Detecting an overall survival benefit that is derived from progression-free survival. *J Natl Cancer Inst* 101(23):1642–1649
38. Golan T, Sela T, Margalit O, Amit U, Halpern N, Aderka D et al (2017) Short and long-term survival in metastatic

- pancreatic adenocarcinoma, 1993–2013. *J Natl Compr Canc Netw*. 15(8):1022–1027
39. Lobbezoo DJ, van Kampen RJ, Voogd AC, Dercksen MW, van den Berkmortel F, Smilde TJ et al (2013) Prognosis of metastatic breast cancer subtypes: the hormone receptor/HER2-positive subtype is associated with the most favorable outcome. *Breast Cancer Res Treat* 141(3):507–514
40. Smith I (2006) Goals of treatment for patients with metastatic breast cancer. *Semin Oncol* 33(Suppl 2):2–5

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.