



Research Paper

A correlation analysis to assess event-free survival as a trial-level surrogate for overall survival in early breast cancer

Bishal Gyawali^{a,b,*}, Elvira D'Andrea^a, Jessica M. Franklin^a, Aaron S. Kesselheim^{a,c}

^a Program On Regulation, Therapeutics, And Law (PORTAL), Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States

^b Department of Oncology, Department of Public Health Sciences and Division of Cancer Care and Epidemiology, Queen's University, Kingston, Canada

^c Harvard Center for Bioethics, Harvard Medical School, Boston, MA, United States

ARTICLE INFO

Article History:

Received 6 November 2020

Revised 8 January 2021

Accepted 14 January 2021

Available online 29 January 2021

Keywords:

event-free survival

overall survival

neoadjuvant

breast cancer

FDA

surrogates

ABSTRACT

Background: Event-free survival (EFS) has been listed on the FDA Table of Surrogate Endpoints as a surrogate measure that can be considered for accelerated or traditional approval in breast cancer. However, no studies have evaluated the correlation between the treatment effects on EFS and treatment effects on overall survival (OS).

Methods: We performed a systematic search of the literature until May 2020 according to the PRISMA guideline for all published randomized controlled trials (RCTs) in early breast cancer in the neoadjuvant setting. Data on EFS and OS, including the hazard ratio (HR) and 95% confidence intervals (CI), were extracted from each study and the association between the trial-level EFS HR and the trial-level OS HR was estimated using a linear mixed-effects model on the log scale.

Findings: Of the 7 RCTs ($N = 2211$) included in the analysis, 5 included patients with HER2 positive tumor type. The estimated linear association between log HR EFS and log HR OS indicated a positive slope ($\beta = 0.58$ [95% CI: $-0.32-1.48$]) and the coefficient of determination confirmed a moderate trial-level association between log HRs for OS and EFS ($R^2 0.76$ [95% CI $0.34-1.00$], but with wide confidence intervals).

Interpretation: Treatment effects in EFS are moderately correlated with treatment effects in OS in early breast cancer in the neoadjuvant setting, but the association was not significant. Thus, there is currently insufficient evidence to support EFS for use as a surrogate endpoint for traditional approval, although it may be considered for accelerated approval.

Funding: Arnold Ventures.

© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

The US Food and Drug Administration (FDA) approves many new cancer drugs on the basis of pivotal trials showing changes to surrogate measures, which are laboratory values or other tumor-specific measurements such as response rate [1]. While some surrogate measures have proven to be useful in predicting positive changes to real clinical endpoints among patients with cancer (such as overall survival (OS) or symptom control), many in cancer trials are not well linked to clinical endpoints or have not yet been validated before they are used by the FDA [2]. When cancer drugs are approved based on these kind of surrogate measures and then prescribed to patients, they may not be worth their substantial financial cost, or their risks might outweigh any real clinical benefits they provide.

To help guide patients, physicians, and clinical trial investigators about the range of possible surrogate measures used in drug approvals, the FDA in 2018 first published a table of “surrogate endpoints which were the basis of approval or licensure (as applicable) of a drug or a biological product [3].” This list provides information about endpoints that may be considered and discussed with FDA for drug development programs. For example, for solid tumors, surrogate measures listed on the FDA table include objective response rates, pathological complete response rates, disease-free survival (DFS), event-free survival (EFS), metastases-free survival, progression-free survival and plasma testosterone levels. However, the table does not provide any evidence about the validity of using these surrogates in pivotal trials intended to support a drug's FDA approval.

EFS is a particularly notable case. According to the FDA guidance, EFS is used in the neo-adjuvant setting and is defined as time from randomization to any of the following events: progression of disease that precludes surgery, local or distant recurrence, or death due to any cause. EFS bears resemblance to DFS except that randomization takes place

* Corresponding author at: Queen's Cancer Research Institute, Cancer Care and Epidemiology, 10 Stuart Street, Level 2, Kingston K7L3N6, ON, Canada.

E-mail address: bgyawali@bwh.harvard.edu (B. Gyawali).

Research in Context

Evidence before this study

Various surrogate endpoints in different cancer types that may be used for regulatory approval have been listed by the FDA in a new Table. A previous study has examined all the surrogate endpoints included in the table for breast cancer and discovered that event-free survival (EFS), although included in the FDA Table, had never been studied as a surrogate for overall survival (OS) in breast cancer.

Added value of this study

This correlation analysis of randomized controlled trials in early breast cancer fills that gap by systematically studying the correlation between treatment effects in EFS and treatment effects in OS. We find moderate level correlation between log hazard ratios of OS and EFS but with wide confidence intervals.

Implications of all the available evidence

Given the moderate, but not significant, correlation between treatment effects in EFS and treatment effects in OS, EFS may not be suitable as a regulatory endpoint for traditional approval but may be considered for accelerated approval of breast cancer drugs.

that used EFS in the neoadjuvant setting, excluding the trials who used the term EFS in the adjuvant setting to imply DFS.

2.2. Data extraction

Data were independently extracted from published reports by two authors (BG and EDA) and verified by the third author (JMF). Trial characteristics extracted from each randomized controlled trial included study name, year of publication, intervention tested, hazard ratios and confidence intervals for EFS and OS. Since failure to undergo surgery is an important event that distinguishes EFS from DFS, the proportion of patients who failed to undergo surgery was also recorded for each trial.

2.3. Statistical analysis

We sought to evaluate the association between the trial-level EFS hazard ratio (HR) and the trial-level OS HR as a measure of trial-level surrogacy. This association indicates whether treatment effects on EFS are predictive of treatment effects on OS, and it is a stronger measure of surrogacy than individual-level correlations between EFS and OS [7]. Trial-level surrogacy was estimated using a linear mixed-effects model that described a linear relationship between the true log HR for OS and the true log HR for EFS and accounts for the uncertainty in the HR estimates, including different sample sizes across trials. Heterogeneity was calculated by the model as the variability in the log HR for OS across trials. If EFS is a reliable trial-level surrogate for OS, β (the slope of the linear relationship) should be positive and large in absolute value. Additionally, a coefficient of determination (R^2) was provided to quantify the proportion of variance in the effects of treatment on the trial-level OS HR that is explained by the surrogate (trial-level EFS HR). The statistical analysis was performed using R statistical software (R source package by Korn et al. [7]). A detailed description of the model is reported in the Supplement.

2.4. Funding

Work on this project was funded by the Arnold Ventures. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

3. Results

Of the 1533 studies in the original sample, 124 were of potential relevance. After screening full texts, 7 randomized controlled trials in neoadjuvant setting met our eligibility criteria and were included in our analysis (Fig. 1). One large trial with 602 patients was excluded because it did not report OS data [8].

3.1. Description of studies

Of the 7 randomized controlled trials ($N = 2211$, intervention arm $n = 1094$, control arm $n = 1117$), 5 were conducted in HER2 positive patient population [9–13], one in HER2 negative population [14], and two in HER2 positive or triple negative population [15] (Table 1). All studies were conducted between 2013 and 2020. Percentage of patients who had an EFS endpoint due to failure to undergo surgery ranged from 1% to 20% across the trials.

3.2. Directions of EFS and OS

The directions of HRs for EFS and OS were concordant in all the randomized controlled trials (Table 2). Two RCTs reported a

before definitive surgery or radiotherapy, while in DFS, randomization happens after surgery or radiation [4]. Hence, a key difference between EFS and DFS is that failure to undergo surgery constitutes an event in EFS but not in DFS. In a previous study of relationship between surrogate measures in the FDA's table and overall survival in patients with breast cancer, EFS was found to have no validated correlation studies to confirm that the treatment effects on EFS predicted treatment effects on OS [5].

Since the FDA considers EFS as a potential surrogate measure for pivotal trials of investigational drugs developed to treat breast cancer, and we could find no formal validation of this surrogate measure, we conducted a systematic review and correlation analysis to assess the relationship between EFS and OS to determine the appropriateness of using EFS as a trial-level surrogate measure for FDA approvals.

2. Methods

2.1. Study identification

We conducted a systematic search from the initiation of each database to May 2020 of PubMed, the Cochrane Library, and Google Scholar, in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) reporting guideline [6], for all randomized controlled trials of breast cancer that measured EFS as an endpoint. We updated the search in December 2020 for additional studies published in the interim. We used the search terms “event-free survival” or “event free survival” or “event free” or “event-free” or “neoadjuvant” and “breast neoplasms” or “breast cancer”. We limited our search to randomized controlled trials and findings published in English. After title and abstract screening by two authors (BG and EDA) acting independently, the full texts of potentially relevant studies were downloaded and reviewed for the inclusion and exclusion criteria.

We included studies testing for drug interventions in a randomized design and excluded trials of surgery, radiation or bone marrow transplants. To assess the clinical value of EFS as a trial-level surrogate measure, studies had to report both EFS and OS. We excluded studies that did not report either EFS or OS results and duplicate publications from the same randomized trial. Finally, to be consistent with the FDA guidance, we limited our analysis to only those studies

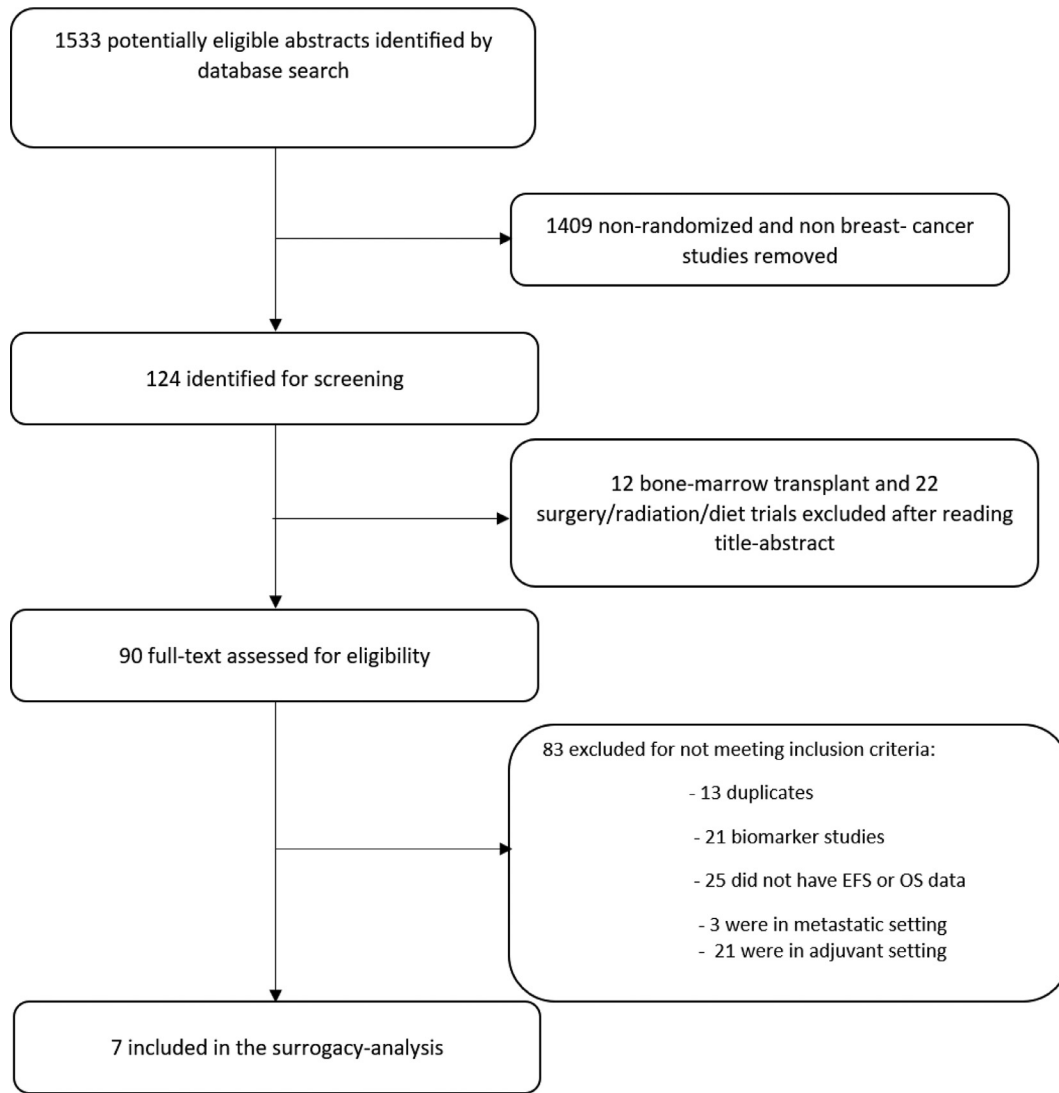


Fig. 1. PRISMA flow diagram of included studies.

significant worsening of EFS, both associated with a non-significant worsening of OS [12,15]. One RCT demonstrated a significant benefit in EFS that did not translate to significant benefit in OS [11]. Four other RCTs demonstrated a hazard ratio below 1 for both the EFS and OS but none were significant [9,10,13,14].

3.3. Correlation between EFS and OS

Fig. 2 shows the association between trial-level HR for EFS effects and trial-level HR for OS effects observed in the 7 trials with data on both outcomes. The solid line represents equality between OS and

Table 1
Studies included in the analysis.

Study Name	First Author	Year	Setting	Number of patients	EFS HR (95% CI)	OS HR (95% CI)	Percentage of patients who did not undergo surgery for various reasons	Reference (PMID)
SWOG S0800	Nahleh	2016	HER2 -ve	211	0.89 (0.48–1.65)	0.84 (0.41–1.73)	17/215 (8%)	PMID: 27393622
HannaH	Jackisch	2019	HER2 +ve	591	0.98 (0.74–1.29)	0.94 (0.61–1.45)	45/596(8%)	PMID: 30998824
NeoALTO	de Azambuja	2014	HER2 +ve	301	0.78 (0.47–1.28)	0.62 (0.30–1.25)	28/455(6%)	PMID: 25130998
NOAH	Gianni	2014	HER2 +ve	235	0.64 (0.44–0.93)	0.66 (0.43–1.01)	47/235(20%)	PMID: 24657003
NATT	Chen	2013	HER2 +ve or triple negative	96	2.42 (1.11–5.30)	2.52 (0.41–15.38)	1/96(1%)	PMID: 24292815
KRISTINE	Hurvitz	2019	HER2 +ve	444	2.61 (1.36–4.98)	1.21(0.37–3.96)	26/444(6%)	PMID: 31157583
KEYNOTE-522*	Schmid	2020	Triple negative	602	0.63 (0.43–0.93)	Not reported	36/1174(3%)	PMID: 32101663
IMpassion031	Mittendorf	2020	Triple negative	333	0.76 (0.40–1.44)	0.69 (0.25–1.87)	26/333(8%)	PMID: 32966830

EFS: Event-free survival, OS: Overall survival; HR: Hazard Ratio, CI: confidence interval, PMID: Pubmed Unique Identifier, LN +ve: lymph node positive, HR: hormone receptor.

* Keynote-522 is not included in the meta-analysis due to lack of OS data. It is included in the table for the sake of completeness as it has the largest sample size.

Table 2

Parameters and standard error for linear mixed-effects model for the association of log HR for OS and log HR for EFS across 7 included trials in neoadjuvant setting.

Parameter	Estimate	Standard error
α	-0.156	0.146
β	0.584	0.460
μ	0.045	0.188
g	0	0.001
σ^2	0.192	0.165

α = intercept of the linear relationship between the log HR for OS and log HR for EFS (when the HR for EFS is 1.0, the estimated HR for OS is $e^{-0.156} = 0.86$).

β = coefficient of the linear relationship between the log HR for OS and log HR for EFS.

μ = average log HR for EFS across trials (corresponding to a HR of $e^{0.101} = 1.05$).

g = variance of log HR for OS across trials that is not explained by EFS.

σ^2 = variance of log HR for EFS across trials.

EFS effects, while the dashed line represents the estimated linear association from the random effects model. The pattern of circles clustering around the dashed line indicates a positive slope ($\beta = 0.58$ [95% CI: -0.32–1.48]), meaning that a one unit increase in log HR for EFS was associated with a 0.58 unit increase in log HR for OS. Although the absolute value of the estimated slope indicates that trial-level EFS and trial-level OS are moderately correlated, the confidence interval indicates the data are also consistent with a negative slope. The coefficient of determination confirmed a moderate trial-level association between log HRs for OS and EFS (R^2 0.76 [95% CI 0.34–1.00]), although the confidence intervals provided a large range

of plausible values. This wide uncertainty is due to the small sample size (seven trials).

The model also shows that nearly all the heterogeneity in OS treatment effects between studies can be explained by EFS or random variability, as there was little residual heterogeneity ($g = 0$).

4. Discussion

EFS has been listed by the FDA as a trial-level surrogate for OS without any formal validation or evidence of its link to real clinical outcomes [16]. This systematic review of randomized controlled trials of cancer drugs for early breast cancer found that the treatment effects in EFS may or may not predict treatment effects in OS since the analysis was limited by a small sample size and a wide confidence interval. Since additional studies are needed to confirm these results, it may be premature to consider EFS a valid surrogate measure for use by the FDA in approving drugs for breast cancer via the traditional approval pathway. Similar analyses should be performed to confirm or reject each surrogate measures currently listed in the FDA's table, and the results linked directly to the table to help patients, physicians, and clinical trial investigators relying on that information.

Using surrogate measures as endpoints has become far more common in recent years among oncology trials [1]. However, when non-clinical endpoints are used to make decisions about the efficacy of an intervention, it can be fraught with several risks, especially for cancer drugs that have serious side effects, high costs, and long treatment duration. Thus, it is important to ascertain that the measure is a valid

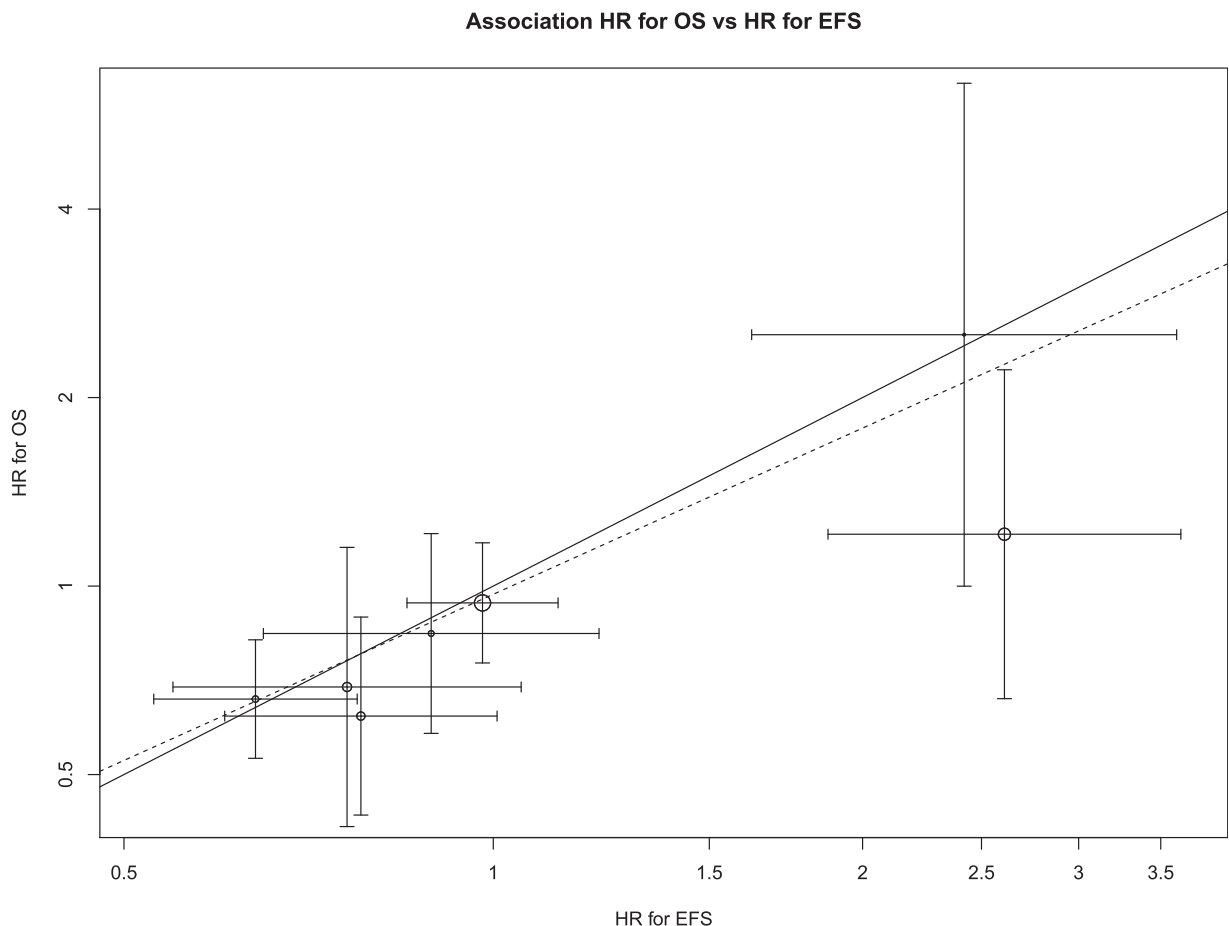


Fig. 2. Association of trial-level EFS hazard ratio and OS hazard ratio across 6 included trials.

The graph shows the association of trial-level EFS and OS effects expressed as hazard ratios. The solid line represents equality between OS and EFS effects. The dashed line represents the estimated slope of the linear association from the random effects model. Areas of circles are proportional to trial sample sizes and horizontal and vertical line segments represent 95% confidence intervals for the trial-level hazard ratios.

surrogate for clinical outcomes before using it as an endpoint in randomized controlled trials to make regulatory approval decisions. Clinical outcomes for patients with cancer can be measured in terms of OS or quality of life. However, quality of life is not consistently measured or reported in oncology trials [17], often leaving OS as the main clinical outcome for which rigorous correlation studies are needed to validate potential surrogates.

A surrogate endpoint may still be used for early drug approval via the accelerated approval pathway since that pathway requires the drug to confirm the clinical benefit in a confirmatory trial. However, the traditional pathway does not have such post approval requirements and thus, the bar for traditional approval should require confirmation of clinical benefit either by showing improved OS or validation of the surrogate endpoint. Therefore, the level of surrogate correlation expected of an endpoint for accelerated approval and traditional approval maybe different. We found a moderate correlation between HRs of EFS and OS, and therefore acknowledge that EFS may be appropriately used for accelerated approval. Its use for traditional approval maybe premature however because whether the treatment effects on EFS could be considered a trial level surrogate for treatment effects on OS remains an unanswered question, since the confidence intervals were wide and included negative slope, probably due to small sample size. One of the largest neoadjuvant trials using an EFS endpoint has yet to report OS information [8]. It will be necessary to revisit the surrogacy validation by updating these results once the OS data from that trial or other newer trials are reported.

This study highlights the difficulty to establish surrogacy even after multiple RCTs have been conducted. In this analysis, even after seven RCTs, we are still unsure about the validity of EFS as a surrogate for OS. Our previous study assessing the surrogacy of PFS for OS with bevacizumab in breast cancer had also led to similar conclusion, i.e., uncertainty in the validity of a surrogate measure despite multiple RCTs available [18]. Together, these data suggest that the confidence in the surrogacy of an intermediate endpoint maybe possible only after many RCTs and it may be far more efficient to measure OS in a given RCT than to predict OS based on surrogacy from other RCTs.

Surrogate measures used in breast cancer generally have poor correlation with OS [5]. The FDA table should include evidence from correlation studies similar to this one to support the inclusion of a surrogate for a given tumor type. This would increase the relevance of the FDA's table to the research and clinical communities. Furthermore, this evidence based on correlation studies should be updated when information from new trials is available. Whether a certain surrogate should be included as appropriate for accelerated approval or regular approval (or should be excluded from the table) should depend on the strength of correlation. It is also important to clearly define, using evidence from such correlation studies, whether a certain surrogate can be used as an endpoint in the confirmatory trials of drugs that gained accelerated approval on the basis of a different surrogate measure [19,18]. Based on our study, EFS may be considered as an endpoint for accelerated approval but would need confirmation of surrogacy before being acceptable as an endpoint sufficient for traditional approval.

There has been some debate about what constitutes surrogacy in cancer drug trials. As Shyr and Shyr recently highlighted, proof of surrogacy requires demonstration of association (correlation) at the individual level and at the trial level [20]. An association at the individual level (longer EFS correlates with longer OS) reflects the prognostic importance of a surrogacy, while an association at the trial level (HR for EFS correlates with HR for OS) reflects that the treatment effects on the surrogate can predict treatment effects on the clinical outcomes. Since the demonstration of association at an individual level requires individual patient data, we were able to test only for trial-level association in this analysis. This lack of ability to demonstrate individual level correlation is an important limitation of our work. Furthermore, although the trials in our analysis provided

information on EFS as surrogate measures, experimental arms and controls differed across the randomized controlled trials and this heterogeneity may make our estimates for EFS or OS less reliable.

Treatment effects in EFS are moderately but not significantly correlated with treatment effects in OS in breast cancer. Thus, although EFS maybe considered appropriate for accelerated approval, its validity for traditional approval remains to be demonstrated. The FDA should ensure that all surrogate measures are analyzed using similar methodologies before they can be relied on as the basis for regulatory approval.

Declaration of Interests

Dr. Kesselheim reports receiving grant support from the Harvard-MIT Center for Regulatory Science and Arnold Ventures. All other authors have nothing to declare.

Author contribution statement

BG and ASK conceived the study. BG did the literature search, extracted the data and wrote the first draft of the manuscript. EDA and JMF determined the methodology for the study and conducted statistical analysis. All authors contributed to the interpretation of the results. All authors contributed substantially to the editing and revising of the manuscript. All authors agreed to final submission.

Funding

Work on this project was funded by the Arnold Ventures. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data sharing statement

All data necessary to complete this analysis are listed in [Table 1](#).

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.eclim.2021.100730](https://doi.org/10.1016/j.eclim.2021.100730).

References

- [1] Gyawali B, Sharma S, Booth CM. Is the number of cancer drug approvals a surrogate for regulatory success? *J Cancer Policy* 2019;22:100202.
- [2] Haslam A, Hey SP, Gill J, Prasad V. A systematic review of trial-level meta-analyses measuring the strength of association between surrogate end-points and overall survival in oncology. *Eur J Cancer* 2019;106:196–211.
- [3] U.S. Food and Drug Administration. Table of surrogate endpoints that were the basis of drug approval or licensure. <https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure> latest on.
- [4] Food US, Administration Drug. Clinical trial endpoints for the approval of cancer drugs and biologics: guidance for industry. <https://www.fda.gov/media/71195/download> on.
- [5] Gyawali B., Hey S.P., Kesselheim A.S. Evaluating the evidence behind the surrogate measures included in the FDA's table of surrogate endpoints as supporting approval of cancer drugs. *EClinicalMedicine*.
- [6] Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700.
- [7] Korn EL, Sachs MC, McShane LM. Statistical controversies in clinical research: assessing pathologic complete response as a trial-level surrogate end point for early-stage breast cancer. *Ann Oncol: Off J Eur Soc Med Oncol* 2016;27(1):10–5.
- [8] Schmid P, Cortes J, Pusztai L, et al. Pembrolizumab for early triple-negative breast cancer. *N Engl J Med* 2020;382(9):810–21.
- [9] Jackisch C, Stroyakovskiy D, Pivov X, et al. Subcutaneous vs Intravenous Trastuzumab for patients with ERBB2-positive early breast cancer: final analysis of the HannaH Phase 3 randomized clinical trial. *JAMA Oncol* 2019;5(5):e190339.

- [10] de Azambuja E, Holmes AP, Piccart-Gebhart M, et al. Lapatinib with trastuzumab for HER2-positive early breast cancer (NeoALTO): survival outcomes of a randomised, open-label, multicentre, phase 3 trial and their association with pathological complete response. *Lancet Oncol* 2014;15(10):1137–46.
- [11] Gianni L, Eiermann W, Semiglazov V, et al. Neoadjuvant and adjuvant trastuzumab in patients with HER2-positive locally advanced breast cancer (NOAH): follow-up of a randomised controlled superiority trial with a parallel HER2-negative cohort. *Lancet Oncol* 2014;15(6):640–7.
- [12] Hurvitz SA, Martin M, Jung KH, et al. Neoadjuvant trastuzumab emtansine and pertuzumab in human epidermal growth factor receptor 2-positive breast cancer: three-year outcomes from the phase III KRISTINE study. *J Clin Oncol* 2019;37(25):2206–16.
- [13] Mittendorf EA, Zhang H, Barrios CH, et al. Neoadjuvant atezolizumab in combination with sequential nab-paclitaxel and anthracycline-based chemotherapy versus placebo and chemotherapy in patients with early-stage triple-negative breast cancer (IMpassion031): a randomised, double-blind, phase 3 trial. *Lancet North Am Ed* 2020;396(10257):1090–100.
- [14] Nahleh ZA, Barlow WE, Hayes DF, et al. SWOG S0800 (NCI CDR0000636131): addition of bevacizumab to neoadjuvant nab-paclitaxel with dose-dense doxorubicin and cyclophosphamide improves pathologic complete response (pCR) rates in inflammatory or locally advanced breast cancer. *Breast Cancer Res Treat* 2016;158(3):485–95.
- [15] Chen X, Ye G, Zhang C, et al. Superior outcome after neoadjuvant chemotherapy with docetaxel, anthracycline, and cyclophosphamide versus docetaxel plus cyclophosphamide: results from the NATT trial in triple negative or HER2 positive breast cancer. *Breast Cancer Res Treat* 2013;142(3):549–58.
- [16] Gyawali B, Hey SP, Kesselheim AS. Evaluating the evidence behind the surrogate measures included in the FDA's table of surrogate endpoints as supporting approval of cancer drugs. *EClinicalMedicine* 2020;21:100332. doi: 10.1016/j.eclinm.2020.100332.
- [17] Hwang TJ, Gyawali B. Association between progression-free survival and patients' quality of life in cancer clinical trials. *Int J Cancer* 2019;144(7):1746–51.
- [18] Hey SP, Gyawali B, D'Andrea E, Kanagaraj M, Franklin JM, Kesselheim AS. A systematic review and meta-analysis of Bevacizumab in first-line metastatic breast cancer: lessons for the research and regulatory enterprises. *J Natl Cancer Inst* 2020;112(4):335–42.
- [19] Gyawali B, Hey SP, Kesselheim AS. Assessment of the clinical benefit of cancer drugs receiving accelerated approval. *JAMA Intern. Med.* 2019;179(7):906–13.
- [20] Shyr Y, Shyr D. What constitutes a valid surrogate end point in cancer clinical trials? *JAMA Oncol* 2020.