



ELSEVIER

Contents lists available at ScienceDirect

EClinicalMedicine

journal homepage: <https://www.journals.elsevier.com/eclinicalmedicine>

## Research Paper

# Evaluating the evidence behind the surrogate measures included in the FDA's table of surrogate endpoints as supporting approval of cancer drugs

Bishal Gyawali<sup>a,b,\*</sup>, Spencer P. Hey<sup>a,c</sup>, Aaron S. Kesselheim<sup>a,c</sup>

<sup>a</sup> Program On Regulation, Therapeutics, And Law (PORTAL), Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, US

<sup>b</sup> Department of Oncology, Department of Public Health Sciences and Division of Cancer Care and Epidemiology, Queen's University, Kingston, Canada

<sup>c</sup> Harvard Center for Bioethics, Harvard Medical School, Boston, MA, US

## ARTICLE INFO

## Article History:

Received 24 January 2020

Revised 17 March 2020

Accepted 18 March 2020

Available online 13 April 2020

## Keywords:

Surrogate endpoints

FDA

Breast cancer

Progression-free survival

Overall survival

Response rates

Event-free survival

Disease-free survival

Pathological complete

## ABSTRACT

**Background:** In July 2018, the FDA first published a table listing all surrogate measures that it has used, and may accept for future use, in regulatory approval. However, the strength of surrogacy for those measures was not formally assessed. Using the case example of breast cancer, we aimed to evaluate the strength of correlation of surrogate measures listed in the FDA's Table with overall survival.

**Methods:** This cross-sectional study of the FDA's Table of Surrogate Endpoints was conducted in May 2019. All surrogate measures listed in the FDA table as appropriate for accelerated or regular approval for breast cancer were extracted. We identified studies evaluating the correlation of treatment benefit in the surrogate with treatment benefit in overall survival and extracted results from the correlation analysis.

**Findings:** Five surrogate endpoints were listed for breast cancer in the FDA website: pathological complete response rates (pCR), event-free survival (EFS), disease-free survival (DFS), objective response rates (ORR), and progression-free survival (PFS), of which pCR was listed as appropriate only for accelerated approval, while the rest were considered appropriate for accelerated or regular approval. No correlation study evaluated the correlation of treatment effects on EFS with that on OS. The results from correlation studies evaluating pCR, DFS, ORR, and PFS suggest that the treatment effects on none of these surrogate measures were strongly correlated with treatment effects on OS ( $r < 0.85$  or  $R^2 < 0.7$ , except for DFS in HER2 positive early breast cancer ( $R^2 = 0.75$ )).

**Interpretation:** Using breast cancer as an example, we evaluated the underlying evidence for the surrogate endpoints for solid tumors listed in the FDA's Table of Surrogate Endpoints and found weak or missing correlations of treatment effects on these surrogates with treatment effects on OS. Surrogate measures should be predictive of clinical benefit to be useful in supporting regular FDA approval.

**Funding:** Work on this project was funded by the Arnold Ventures. Dr. Kesselheim is also supported by the Harvard-MIT Center for Regulatory Science. 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.

© 2020 The Authors. 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/>)

## Introduction

Surrogate measures are intermediate endpoints that serve as substitutes for direct measures of how patients feel, function, or survive in clinical trials [1]. In oncology, improving how a patient feels (quality of life) or how long a patient lives (overall survival) are clinical outcomes, while reduction in tumor size or prolongation of time

until the tumor has grown are surrogate measures. Cancer trials may use clinical outcomes or surrogate measures. The use of surrogate measures in oncology trials can reduce the trial duration by 11–19 months [2].

The use of the surrogate measures for regulatory approval has been a matter of substantial debate in the oncology, health policy, and regulatory communities. The U.S Food and Drug Administration (FDA) accepts validated surrogate endpoints as evidence of benefit, but as it notes on its website: "Clinical trials are needed to show that surrogate endpoints can be relied upon to predict, or correlate with, clinical benefit. Surrogate endpoints that have undergone this testing are called validated surrogate endpoints" [3].

\* Corresponding author.

E-mail address: [bg.bishalgyawali@gmail.com](mailto:bg.bishalgyawali@gmail.com) (B. Gyawali).

## Research in context

### *Evidence before this study*

The FDA has recently published a table to highlight all surrogate endpoints that it has used, and may accept for future use, in regulatory approval as a fulfillment of the requirement of the 21st Century Cures Act to publish a list of “surrogate endpoints which were the basis of approval or licensure (as applicable) of a drug or a biological product.” However, the FDA table doesn’t provide any empirical evidence as to the validity or strength of correlation for these studies. Previous studies, published before the FDA Table, have studied the strength of correlation of various surrogates, but none in relation to the FDA Table.

### *Added value of this study*

Using breast cancer as an example, we evaluated the underlying evidence for the surrogate endpoints for solid tumors listed in the FDA Table of Surrogate Endpoints. Although EFS is listed on the table, no study evaluating the correlation of EFS was found while the strength of correlation for treatment effects on DFS, ORR and PFS with treatment effects on OS were weak despite being listed as appropriate for even traditional approval. The only exception was DFS in HER2 positive early breast cancer which had strong correlation with OS. pCR was listed as appropriate only for accelerated approval.

### *Implications of all the available evidence*

The strength of correlation for each surrogate in each tumor type varies widely. All surrogate endpoints must be formally evaluated for validity of surrogacy before listing them as appropriate for traditional approval in the FDA Table.

## Methods

### *Data source*

In May 2019, we extracted the rows related to solid tumors from the FDA’s Table of Surrogate Endpoints, which is publicly accessible on the FDA website (Table 1) [6]. The FDA table contains 5 columns: Column 1 describes the disease (solid tumors). Column 2 describes the patient populations (listing various tumor types). Column 3 specifies the surrogate measure of interest (e.g., progression-free survival). Column 4 states whether the surrogate measure is considered to be appropriate for accelerated approval, regular approval, or both. Accelerated approval is a pathway in which approval can be gained by demonstrating an effect on a surrogate measure or intermediate clinical endpoint that is only “reasonably likely” to predict a real clinical endpoint [4]. It is supposed to be reserved for serious conditions and post-approved studies are required to verify the anticipated clinical benefits [8]. Column 5 describes the mechanisms of action of interventions for which the surrogate measure is believed to be an appropriate clinical trial endpoint. For oncology, column 5 described all surrogates as “mechanism agnostic” at the time of data extraction.

There were six surrogate endpoints listed for solid tumors in the table: durable overall response rate (ORR), progression-free survival (PFS), Disease-free survival (DFS), event-free survival (EFS), pathological complete response (pCR) and metastasis-free survival. We selected breast cancer as a case example for this study because five of the six surrogate measures (all but metastasis-free survival) in the Table were described as appropriate for clinical trials of new breast cancer drugs. Later, a new surrogate endpoint of “plasma testosterone levels” was also added to the FDA Table as a surrogate for traditional approval in patients with advanced prostate cancer, the mechanism of action limited to gonadotropin-releasing hormone antagonist.

### *Data search and extraction*

For each surrogate measure described as appropriate for a new breast cancer drug trial, we searched the literature for surrogacy correlation studies. Two recent studies have systematically searched for all such correlation studies of surrogate measures in oncology [9,10]. The more recent had systematically searched for correlation studies until 25 January 2018. We first used the tables from these two systematic reviews to identify studies related to breast cancer. To incorporate findings from any new studies, we repeated our search for correlation studies between 2018 January 1 and 2020 March 1 for any new correlation studies in breast cancer in Pubmed, Cochrane, and Google scholar databases. We focused our search on systematic reviews, meta-analyses, and correlation studies, as these are the study types that can prove the validity of surrogate measures. The search strategy is provided in the supplementary appendix.

### *Data evaluation*

Because we were interested in trial-level surrogacy for regulatory use, we included only those studies that assessed correlation of treatment effects on surrogate measures with treatment effects on overall survival. For example, studies assessing correlation of median progression-free survival (PFS) with median overall survival (OS) would not be included because such studies simply assess the prognostic validity of the surrogate in individual patients rather than trial-level validity of the surrogate for predicting treatment effects on OS. Thus, we included only those studies that studied the correlation between hazard ratio (HR) for a time-to-event surrogate measure and hazard ratio for OS (eg: log HR PFS- log HR OS) or between odds ratios for response rate (RR) and hazard ratio for OS (eg: log OR RR-log HR OS) or those studies that assessed the correlation between the differences in the median PFS (median PFS in treatment arm minus median PFS

If the surrogate measure is not validated as a reliable predictor of clinical benefit in patients with cancer in terms of improvement in survival or quality of life, then using such a measure as a primary endpoint in trials supporting FDA approval of new cancer drugs may result in approved drugs with questionable benefits but often numerous important side effects (and invariably high costs). Therefore, it is essential that surrogate measures used for new cancer drug approvals have compelling evidence of validation for the relevant tumor type and setting [4]. Previous analyses have raised questions about the empirical validation of surrogate endpoints used in FDA approval [5]. To help clarify the FDA’s perspective on the appropriate use of surrogate measures, the Agency recently published a table to highlight surrogate measures that it has used, and may accept for future use, in regulatory approval [6]. The table fulfills the requirement of the 21st Century Cures Act of 2016 to publish a list of “surrogate endpoints which were the basis of approval or licensure (as applicable) of a drug or a biological product.”

We therefore sought to understand the evidence base underlying the information provided in the FDA’s table and make recommendations as to how it can be improved. Although the FDA’s table encompasses all diseases, the information it includes on surrogate measures in cancer deserves special attention because cancer is a disease in which even the most effective therapies still impose serious risks and financial burdens on patients. Cancer drugs also now account for more than one-quarter of FDA drug approvals [7]. In this pilot study, using breast cancer as a case example, we evaluated the strength of evidence for the correlation between the treatment effects on the surrogate measures listed in the FDA table with treatment effects on overall survival (OS).

**Table 1**

Adult surrogate endpoint table (Recreated from the FDA table at <https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure>).

Disease or use	Patient population	Surrogate endpoint	Type of approval appropriate for	Drug mechanism of action
Cancer: solid tumors	Patients with breast cancer; ovarian cancer; renal cell carcinoma; pancreatic neuroendocrine cancer; colorectal cancer; head and neck cancer; non-small cell lung cancer; small cell lung cancer; melanoma; tuberous sclerosis complex; Merkel cell carcinoma; basal cell carcinoma; urothelial carcinoma; cervical cancer; endometrial cancer; hepatocellular carcinoma; fallopian tube cancer; microsatellite instability-high cancer; gastric cancer; thyroid cancer; astrocytoma; AIDS-related Kaposi's sarcoma; urothelial carcinoma; unresectable or metastatic cutaneous squamous cell carcinoma; neurotrophic receptor tyrosine kinase (NTRK) gene fusion without a known acquired resistance mutation	Durable objective response rates (ORR)	Accelerated/ Traditional	Mechanism agnostic
Cancer: solid tumors	Patients with breast cancer; renal cell carcinoma; pancreatic neuroendocrine tumor; soft tissue sarcoma; ovarian, fallopian tube, or primary peritoneal cancer; prostate cancer; thyroid cancer; colorectal cancer; non-small cell lung cancer; head and neck cancer; tuberous sclerosis complex; Merkel cell carcinoma; basal cell carcinoma; urothelial carcinoma; cervical cancer; endometrial cancer; hepatocellular carcinoma; fallopian tube cancer; melanoma; astrocytoma; gastrointestinal stromal tumors	Progression-free survival (PFS)	Accelerated/ Traditional	Mechanism agnostic
Cancer: solid tumors	Patients receiving adjuvant therapy following complete surgical resection of colon cancer; colorectal cancer; melanoma; renal cell cancer; gastrointestinal stromal tumor; breast cancer and adjuvant therapy for stage III non-small cell lung cancer	Disease-free survival (DFS)	Accelerated/ Traditional	Mechanism agnostic
Cancer: solid tumors	Patients with breast cancer; neuroblastoma	Event-free survival (EFS)	Accelerated/ Traditional	Mechanism agnostic
Cancer: solid tumors	Patients with breast cancer	Pathological complete response	Accelerated	Mechanism agnostic
Cancer: solid tumors	Patients with nonmetastatic castrate-resistant prostate cancer	Metastasis-free survival	Accelerated/ Traditional	Mechanism agnostic
Cancer: solid tumors	Patients with advanced prostate cancer	Plasma testosterone levels	Traditional	Gonadotropin-releasing hormone antagonist

in control arm) and differences in median OS times (median OS in treatment arm minus median OS in control arm).

If more than two correlation studies were available, we evaluated each. We extracted the correlation coefficient or the  $R^2$  from the correlation studies, as well as the authors' conclusion (interpretation) of the results. However, the authors' conclusions can often be quite subjective without objective criteria for surrogacy validation. Based on the results, we classified surrogate measures as strongly correlated (correlation coefficient  $\geq 0.85$ ) or not strongly correlated. This categorization was based on the criteria proposed by the Institute of Quality and Efficiency in Health Care [11] and has been used in previous studies [12]. When no such study was discovered, we classified that surrogate as an "unstudied" surrogate. For studies where only an  $R^2$  is reported, we used the cut-off value of 0.7, as it is the square of the cut-off for correlation coefficient and also seems to be the most commonly accepted cut-off in surrogacy studies in oncology [13].

#### Role of the funding source

This work was funded by the Arnold Ventures. All authors had full access to all data used in the study. The corresponding author had the final responsibility for the decision to submit for publication.

#### Results

In the FDA table, surrogate measures listed for breast cancer included: (1) pathological complete response rates (pCR); (2) event-free survival (EFS); (3) disease-free survival (DFS); (4) objective response rates (ORR); and (5) progression-free survival (PFS). Pathological complete response rate (pCR) was described as suitable only for accelerated approvals. The other four surrogates were classified as potentially suitable for accelerated or regular approval. However, because the table rows for EFS, DFS, ORR, and PFS group together

multiple solid tumor types, it is difficult to know if the FDA considered these surrogates to be suitable for accelerated or regular approval in the specific case of breast cancer.

#### Pathological complete response rates (pCR)

Pathological complete response rates (pCR) is used as a surrogate measure in the setting of neoadjuvant therapy. pCR is defined as the absence of residual invasive cancer on hematoxylin and eosin evaluation of the complete resected breast specimen and all sampled regional lymph nodes following completion of neoadjuvant systemic therapy [14]. The FDA table includes pCR as a surrogate measure only for breast cancer. We found two meta-analyses testing the validity of pCR as a surrogate measure for breast cancer included in the systematic reviews of correlation studies (Table 2) [15,16]. Both these studies tested the correlation of log HR of OS with log OR of pCR. The first study included 29 trials and found an  $R^2$  of 0.09 while the second study included 12 trials and found an  $R^2$  of 0.24 between the log HR of OS and log OR of pCR. Both the studies concluded that the treatment effects on pCR could not be considered a surrogate for treatment effects on OS in breast cancer. Thus, we categorized pCR as a "not strongly correlated" surrogate. No new correlation study for pCR was discovered in our updated search.

#### Disease-free survival (DFS)

DFS is defined as the time from randomization until disease recurrence or death from any cause, and is employed usually in the adjuvant setting [17]. We found one correlation analysis evaluating DFS as a surrogate for OS included in the two systematic reviews of correlation studies [18]. This study evaluated the correlation between difference in 2 year DFS rates with difference in 5 year OS rates among 126 trials and concluded that the "correlation was

significant but not strong for DFS to be used as a predictor for OS.” The correlation coefficient was 0.62. We found a new correlation study from our updated search assessing the correlation of treatment effects on DFS with that on OS in patients with HER2-positive early breast cancer receiving adjuvant trastuzumab up to 1 year [19]. This study reported an  $R^2$  of 0.75 between log HR OS and log HR DFS and concluded that “it is appropriate to continue to use disease-free survival as a surrogate for overall survival in trials in HER2-positive, early breast cancer.” Thus, we classified DFS as a “not strongly correlated” surrogate overall, but “strongly correlated” for trials of HER2-positive early breast cancer.

#### *Event-free survival (EFS)*

EFS is used in the neoadjuvant setting to imply time from randomization to disease relapse, recurrence, progression or death (the only difference between EFS and DFS is that EFS is in the neoadjuvant setting while DFS is in the adjuvant setting) [17]. In past trials in the adjuvant setting, EFS was used to mean DFS. However, because the patient is not technically “disease-free” at the time of randomization in neoadjuvant setting, the term EFS is preferred. The two systematic reviews of correlation studies in cancer did not include any studies assessing the correlation of the treatment effects for EFS as a surrogate for treatment effects for OS.<sup>9,10</sup> No other such correlation studies were discovered in our updated search. We therefore categorized this surrogate marker as “unstudied.”

#### *Objective response rate (ORR)*

ORR is defined as the proportion of patients with tumor size reduction of a predefined amount and for a minimum time period [17]. We found two correlation studies evaluating the treatment effects on ORR as a surrogate for treatment effects on OS included in the two systematic reviews of correlation studies, both of them assessing the correlations between log OR (ORR) with log HR OS. One of these studies included individual patient data from 11 trials and found a correlation coefficient of 0.57. The study concluded that ORR was “not a good surrogate for OS” [20]. The second study included 42 studies and found that the  $R^2$  between log HR OS and log OR ORR was 0.34. This study concluded “only modest association with OS” [21]. No new correlation studies for ORR were discovered in our updated search. Thus, we classified ORR as a “not strongly correlated” surrogate.

#### *Progression-free survival (PFS)*

PFS is defined as the time from randomization until objective tumor progression or death, whichever occurs first [17]. We found 4 studies included in the two systematic reviews of correlation studies and three new correlation studies from our updated search, all assessing the correlation between log HR PFS and log HR OS (Table 2) [20,22–27]. One study that included individual patient data from 11 trials found a correlation coefficient of 0.48 and concluded that the treatment effects on PFS was not a good surrogate for treatment effects on OS. A second study showed an  $R^2$  of 0.49 for 16 trials involving anthracyclines and an  $R^2$  of 0.35 for 15 trials involving taxanes. This study concluded significant correlation but poor prediction. The third study involving 72 trials showed an  $R^2$  of 0.31. However, when limited to only those trials in second or later lines of therapy, the  $R^2$  improved to 0.55. This study concluded that the treatment effects on PFS cannot be recommended as a surrogate for treatment effects on OS for first-line therapy. The fourth study involving 9 trials demonstrated an  $R^2$  of 0.51. This study concluded that there was moderate correlation. The fifth study involved 37 trials and treated PFS and time to tumor progression as the same. This study reported a correlation coefficient between log HR OS and log HR

(PFS/time to tumor progression) of 0.56 and concluded moderate correlation. The sixth study included 16 trials done in patients with hormone positive, HER2-negative metastatic breast cancer and reported a correlation coefficient of 0.72. This study concluded that PFS benefit may predict OS benefit as long as the upper confidence interval for the hazard ratio for PFS is less than 0.60. Finally, the seventh study included only trials using bevacizumab in first line metastatic breast cancer and reported a statistically nonsignificant coefficient of linear relationship between log HR OS and log HR PFS of 0.43 and concluded that the evidence was insufficient to conclude the validity of PFS as a trial level surrogate in this setting.

Taking all this information together, we concluded PFS was “not strongly correlated”.

#### *Results summary*

Based on these results, we made a revised version of the surrogate table published by the FDA for breast cancer with references to the correlation studies (Table 2).

#### **Discussion**

Our study aim was to evaluate the evidence underlying the FDA’s Table of Surrogate Endpoints for breast cancer. We found that none of the FDA listed surrogates were strongly correlated with OS: the treatment effects for pCR was not correlated with the treatment effects for OS, while the treatment effects for ORR, DFS, and PFS each showed some degree of correlation in formal validation studies but none concluded the surrogate was strongly validated as a surrogate for treatment effects for OS. The only exception was a strong correlation between treatment effects in DFS and that in OS for HER2 positive early breast cancer. We also found that EFS was listed on the FDA table despite no published studies formally assessing the validity of this surrogate.

How strongly validated a surrogate measure is to a clinical endpoint can help determine which FDA regulatory pathway should be used for approval. The accelerated approval system was specifically designed for surrogates that lack strong validation, but may still be reasonably likely to be a surrogate for clinical benefit. Our review suggests that surrogate measures such as ORR, DFS, and PFS may be most appropriate for accelerated approval since they demonstrate some level of correlation. Some recent studies assessing FDA approval for cancer drugs have, however, highlighted that more drugs are receiving regular approval in recent years based on improvement in surrogate measures, not accelerated approval [28,29]. Accelerated approval drugs must be subsequently validated in follow-up testing, or else the FDA can withdraw the drug’s approval. While most of these follow-up validation studies have shown a positive effect, one survey found that confirmatory trials for accelerated approval drugs continue to use surrogate measures as endpoints—and for some of them even the same surrogate measure that was used in the preapproval pivotal study [4].

Despite none of the surrogate measures highlighted by the FDA in its table having trial-level correlations for overall survival in breast cancer, numerous drugs have been approved for use in treating breast cancer using surrogate measures in the past two decades via the regular—not accelerated—approval pathway. For example, everolimus (Afinitor) and alpelisib (Piqray) were granted full approval for advanced breast cancer although they improved only PFS with OS data immature at the time of approval [30,31]. Long-term follow up data for everolimus confirmed that the drug does not improve OS [32]. Because the regular approval pathway does not have the same confirmatory trial requirement associated with accelerated approvals, manufacturers should be required to provide evidence to the FDA—and communicate that evidence to patients—of the validation of their surrogate measures. Our current analysis shows that none of

**Table 2**

Revised table providing additional details on evidence supporting surrogate measures.

Tumor type	Surrogate measure	Validation study Name, year, PMID	Validation study results	Conclusion of the Study	Inference				
			Number of studies included	Population studied	Correlation assessed	Correlation Results			
Breast	Pathological complete response (pCR)	pCR Validation study 1	Berruti 2014 (PMID: 25349292)	29	RCTs of patients receiving neoadjuvant chemotherapy or targeted therapy, excluding endocrine therapy	logHR(OS)-logOR (pCR)	$R^2 = 0.09$	Does not support the use of pCR as a surrogate for OS	pCR not strongly correlated with OS
		pCR validation study 2	Cortazar 2014 (PMID: 24529560)	12	RCTs of patients receiving neoadjuvant chemotherapy or targeted therapy, excluding endocrine therapy	logHR(OS)-logOR (pCR)	$R^2 = 0.24$	pCR cannot be validated as a surrogate for OS	
	Event-free survival (EFS)	Validation study doesn't exist in the literature							EFS yet to be validated as a surrogate endpoint
	Disease-free survival(DFS)	DFS Validation study 1	Ng 2008 (PMID: 18029973)	126	Randomized adjuvant breast cancer trials	2 year $\Delta$ DFS- 5 year $\Delta$ OS	$r = 0.62$ ; $R^2 = 0.38$	Correlation significant but not sufficiently strong for DFS to be used as a predictor for OS	DFS not strongly correlated with OS overall, but strongly correlated for HER <sup>2</sup> positive trials
		DFS validation study 2	Saad 2019 (PMID: 30709633)	8	Adjuvant trials of HER <sup>2</sup> positive early breast cancer	Log HR OS-log HR DFS	$R^2 = 0.75$	Appropriate as a surrogate for OS	
	Objective Response Rates (ORR)	ORR Validation study 1	Burzykowski 2008 (PMID: 18421050)	$N = 11$ (IPD analysis)	RCTs of chemotherapy in advanced breast cancer	log HR OS-log OR (ORR)	$r = 0.57$	Not a good surrogate for OS	ORR not strongly correlated with OS
		ORR Validation study 2	Hackshaw 2005 (PMID: 16278665)	$N = 42$	RCTs of first line anthracycline chemo for advanced breast cancer	log HR OS-log OR (ORR)	$R^2 = 0.34$	Modest association with OS	
	Progression-free survival (PFS)	PFS Validation study 1	Burzykowski 2008 (PMID: 18421050)	$N = 11$ (IPD analysis)	RCTs of chemotherapy in advanced breast cancer	log HR OS- log HR PFS	$r = 0.48$	Not a good surrogate for OS	PFS not strongly correlated with OS.
		PFS Validation study 2	Miksad 2008 (PMID: 18828930)	$N = 16$ (Anthracyclines), 15 (Taxanes)	RCTs of anthracycline and taxane-based chemotherapy in advanced breast cancer	logHR OS-log HR PFS	$R^2 = 0.49$ (Anthracyclines), 0.35 (Taxanes)	Correlation significant but prediction poor	
		PFS Validation study 3	Adunlin 2015 (PMID: 26596731)	$N = 72$	RCTs of chemotherapy and targeted therapy in advanced breast cancer	log HR OS-log HR PFS	$R^2 = 0.31$ , $R^2 = 0.55$ for >2L	PFS can be a surrogate for >2 L, not recommended for 1L	
PFS Validation study 4		Michiels 2016 (PMID: 26961151)	$N = 9$	RCTs of trastuzumab or lapatinib in HER <sup>2</sup> positive advanced breast cancer	log HR OS-log HR PFS	$R^2 = 0.51$	Moderate correlation		
PFS validation study 5		Li 2018 (PMID: 28818493)	$N = 37$		log HR OS – log HR (PFS/TTP)	$r = 0.56$	Moderate correlation		

(continued on next page)



Table 2 (Continued)

Tumor type	Surrogate measure	Validation study Name, year, PMID	Validation study results	Conclusion of the Study	Inference	
			Phase 3 RCTs of advanced breast cancer			
PFS validation study 6		Lux 2019 (PMID: 31065873)	N = 16 RCTs of HR positive HER2 negative metastatic breast cancer	log HR OS – log HR PFS	$r = 0.72$	Significant effect is upper confidence limit of HR for PFS is $< 0.60$
PFS validation study 7		Hey 2018 (PMID: 31651981)	N = 8 RCTs of bevacizumab in first line metastatic breast cancer	log HR OS – log HR PFS	$B = 0.43$	Evidence insufficient to determine if PFS is a valid trial level surrogate

the surrogate measures for breast cancer meet such a requirement with the exception of DFS in HER2 positive early breast cancer, at least with respect to published systematic reviews. However, some drugs such as ribociclib (Kisqali) in first-line treatment of breast cancer approved on the basis of changes to a surrogate measure have demonstrated improved OS later [33].

Our analysis demonstrates ways in which FDA and other regulatory agencies could improve the existing Table of Surrogate Endpoints to make it more useful for industry, physicians, policy-makers, and patients. Currently, the FDA table groups all solid tumors together for given surrogates. However, the strength of surrogacy for the same surrogate may differ by tumor types. A strong surrogate for one solid tumor maybe a poor surrogate for a different solid tumor. Rather than group all tumors under “solid tumor,” we decided to investigate a single tumor-surrogate pair separately and provide evidence on any surrogate studies confirming or rejecting the validity. Repeating this analysis to provide the strength of evidence for each surrogate-tumor pair and folding this information into the table would further increase its relevance and importance for knowledge-users.

In particular, making the available evidence explicit in the table would draw increased attention to places where systematic assessments of surrogacy do not appear to exist—as we identified in the case of EFS in breast cancer. This would be valuable for identifying important avenues for future research. The NIH or the FDA could support a systematic surrogacy study since a positive result should be needed to support the continued inclusion of this surrogate endpoint for breast cancer. In fact, one of the surrogacy studies assessing pCR as a surrogate measure was partly funded by the FDA [16]. Such studies should be conducted for each surrogate-tumor pair for accepting the measure as a valid surrogate. We also recommend that such surrogacy studies be continually updated as new trials with information on treatment effects on surrogate measures and OS become available. Finally, quality of life information is lacking from most trials and to our knowledge, no studies assessing correlation between surrogate measures and quality of life specifically in patients with breast cancer have been performed [34]. If such data do become available, they should also be added to the table.

Our analysis also highlights the fact that the strength of correlation for the same surrogate-tumor pair may vary based on the treatment type or line of therapy. For example, in Miksad et al.’s 2008 analysis, PFS had a better correlation for trials testing anthracyclines than it had for trials testing taxanes [22]. In Adunlin et al.’s 2015 study, PFS was found to have a better correlation when tested as a second-line treatment, rather than first-line [23]. Another analysis also suggests that the correlation maybe stronger for targeted therapies than chemotherapies in breast cancer [25]. For DFS, although the correlation was not strong overall, the correlation was strong among trials of HER2-positive breast cancers alone. This variability across mechanism or setting is broadly consistent with the structure of the FDA’s table, since it already includes a column to specify the mechanism for which a surrogate may be useful. However, based on the available evidence, it may not seem entirely accurate to characterize PFS or DFS’s utility in breast cancer as “mechanism agnostic.” Indeed, the FDA may accept different levels of evidence for first approval versus subsequent approvals as well as for treatments as first-line therapy with multiple subsequent treatments versus last-line therapy with no subsequent treatments. For example, if a drug improved OS as third-line treatment, it could be reasonable to grant subsequent approvals for first- or second-line treatment based on improvement in surrogates alone.

Our study has some limitations. First, we relied on systematic reviews and meta-analyses for assessing the strength of validation of the surrogate measures in this study; in the case of pCR, DFS, ORR, and PFS, the data were drawn from two published systematic reviews of surrogate studies emerging from two independent groups of

researchers. We also conducted our own systematic search for newer studies published between 2018 and March 1, 2020. Nevertheless, it is possible that unpublished work—perhaps submitted to the FDA as part of the drug approval process—has demonstrated stronger validation for these surrogates. Second, we based our categorization as strongly correlated versus not strongly correlated based on the criteria proposed by the Institute of Quality and Efficiency in Health Care. The FDA may use different criteria to consider whether a correlation is strong enough to serve as a surrogate, however those criteria have not been described to the public. Furthermore, even though the criteria proposed by the Institute of Quality and Efficiency in Health Care categorizes only correlation coefficients  $\geq 0.85$  as strong and those between 0.7 and 0.85 as unclear, the FDA may accept the correlations between 0.7 and 0.85 as good enough for accelerated approval but not regular approval. Many correlation studies have methodological limitations and make subjective judgments as to what strength of correlation would make a strong surrogate [35]. To this point, it's worth noting that our study investigated surrogacy claims only in relation to the treatment effects. For example, a long PFS may very well correlate with a long OS but our study specifically studied whether a drug that improves PFS is also likely to improve OS. Third, EFS is a newer surrogate used in neoadjuvant trials, which may explain the lack of any surrogacy studies. Finally, although we explore the correlation between surrogate endpoints and OS, improvement in quality of life is also an important clinical endpoint for cancer patients. However, FDA approval of cancer drugs based on improvements in quality of life has been rare. There is also no consensus as to proper measurement of quality of life in trials, and a previous analysis has shown poor correlation between improvement in PFS and improvement in quality of life [34]. Regulatory authorities, including the FDA, have recognized this and are working towards forming a better guidance and policy to incorporate patient reported outcomes and quality of life information in the drug approval process [36].

Striking an optimal balance between timely access to new cancer therapies and confirmation of efficacy is of vital importance to safeguarding patient interests. Using breast cancer as an example, we evaluated the underlying evidence for the surrogate measures for solid tumors listed in the FDA's Table of Surrogate Endpoints. We used these data to provide an expanded, more informative table summarizing the limited evidence from trial-level correlation studies assessing treatment effects for the surrogates listed for breast cancer. For EFS, no published studies were available and is therefore an area for future research.

#### Authors contributions

BG and ASK conceived the study. BG did the literature search and wrote the first draft of the manuscript. SPH and ASK provided critical inputs and revised the manuscript. All authors agreed to final submission.

#### Funding

Work on this project was funded by the Arnold Ventures. Dr. Kesselheim also receives grant support from the Harvard-MIT Center for Regulatory Science. 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.

#### Conflict of Interest

ASK has received unrelated research funding from the FDA Division of Health Communication (2013–2016). No other disclosures were reported.

#### Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.eclinm.2020.100332.

#### References

- [1] FDA-NIH Biomarker Working Group. BEST (Biomarkers, endpoints, and other tools) resource [Internet]. Silver spring (MD): food and drug administration (US); 2016-. validation. 2017 Nov 14. Co-published by National Institutes of Health (US), Bethesda (MD).
- [2] Chen EY, Joshi SK, Tran A, Prasad V. Estimation of study time reduction using surrogate end points rather than overall survival in oncology clinical trials. *JAMA Intern Med* 2019;179(5):642–7.
- [3] U.S. Food and Drug Administration. FDA facts: biomarkers and surrogate end-points. Accessed at <https://www.fda.gov/about-fda/innovation-fda/fda-facts-biomarkers-and-surrogate-endpoints> on Sep12, 2019.
- [4] Gyawali B, Hey SP, Kesselheim AS. Assessment of the clinical benefit of cancer drugs receiving accelerated approval. *JAMA Intern Med* 2019.
- [5] Kim C, Prasad V. Strength of validation for surrogate end points used in the US food and drug administration's approval of oncology drugs. *Mayo Clin Proc* 2016.
- [6] U.S. Food and Drug Administration. Table of surrogate end-points that were the basis of drug approval or licensure. Accessed at <https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure>, latest on August 10, 2019.
- [7] The ASCO Post. New report finds cancer drugs account for over a quarter of all new drug approvals in the United States. Accessed at <https://www.ascopost.com/news/september-2019/cancer-drugs-account-for-over-a-quarter-of-all-new-drug-approvals-in-the-us/> on Dec 23, 2019.
- [8] Darrow JJ, Avorn J, Kesselheim AS. New FDA breakthrough-drug category—implications for patients. *Eng J Med* 2014;370(13):1252–8.
- [9] 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.
- [10] Savina M, Gourgou S, Italiano A, et al. Meta-analyses evaluating surrogate end-points for overall survival in cancer randomized trials: a critical review. *Crit. Rev. Oncol. Hematol.* 2018;123:21–41.
- [11] Institute for Quality and Efficiency in Health Care [Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen] (IQWiG). Validity of surrogate endpoints in oncology: executive summary. Accessed at [http://www.iqwig.de/download/A10-05\\_Executive\\_Summary\\_v1-1\\_Surrogate\\_endpoints\\_in\\_oncology.pdf](http://www.iqwig.de/download/A10-05_Executive_Summary_v1-1_Surrogate_endpoints_in_oncology.pdf) on August 21, 2019.
- [12] 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 (Oxford, England: 1990)* 2019;106:196–211.
- [13] Xie W, Halabi S, Tierney JF, et al. A systematic review and recommendation for reporting of surrogate endpoint evaluation using meta-analyses. *JNCI Cancer Spectr* 2019;3(1).
- [14] US Food and Drug Administration. Guidance for industry. pathological complete response in neoadjuvant treatment of high-risk early-stage breast cancer: use as an endpoint to support accelerated approval. Accessed at <https://www.fda.gov/media/83507/download> on August 1, 2019.
- [15] Berruti A, Amoroso V, Gallo F, et al. Pathologic complete response as a potential surrogate for the clinical outcome in patients with breast cancer after neoadjuvant therapy: a meta-regression of 29 randomized prospective studies. *J Clin Oncol* 2014;32(34):3883–91.
- [16] Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNEOBC pooled analysis. *Lancet* 2014;384(9938):164–72.
- [17] US Food and Drug Administration. Clinical trial endpoints for the approval of cancer drugs and biologics: guidance for industry. Accessed at <https://www.fda.gov/media/71195/download> on August 3, 2019.
- [18] Ng R, Pond GR, Tang PA, MacIntosh PW, Siu LL, Chen EX. Correlation of changes between 2-year disease-free survival and 5-year overall survival in adjuvant breast cancer trials from 1966 to 2006. *Ann Oncol* 2007;19(3):481–6.
- [19] Saad ED, Squifflet P, Burzykowski T, et al. Disease-free survival as a surrogate for overall survival in patients with HER2-positive, early breast cancer in trials of adjuvant trastuzumab for up to 1 year: a systematic review and meta-analysis. *Lancet Oncol* 2019;20(3):361–70.
- [20] Burzykowski T, Buyse M, Piccart-Gebhart MJ, et al. Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. *J Clin Oncol* 2008;26(12):1987–92.
- [21] Hackshaw A, Knight A, Barrett-Lee P, Leonard R. Surrogate markers and survival in women receiving first-line combination anthracycline chemotherapy for advanced breast cancer. *Br J Cancer* 2005;93(11):1215–21.
- [22] Miksad RA, Zietemann V, Gothe R, et al. Progression-free survival as a surrogate endpoint in advanced breast cancer. *Int J Technol Assess Health Care* 2008;24(4):371–83.
- [23] Adunlin G, Cyrus JWW, Dranitsaris G. Correlation between progression-free survival and overall survival in metastatic breast cancer patients receiving anthracyclines, taxanes, or targeted therapies: a trial-level meta-analysis. *Breast Cancer Res. Treat.* 2015;154(3):591–608.

- [24] Michiels S, Pugliano L, Marguet S, et al. Progression-free survival as surrogate end point for overall survival in clinical trials of HER2-targeted agents in HER2-positive metastatic breast cancer. *Ann Oncol* 2016;27(6):1029–34.
- [25] Li L, Pan Z. Progression-Free survival and time to progression as real surrogate end points for overall survival in advanced breast cancer: a meta-analysis of 37 trials. *Clin Breast Cancer* 2018;18(1):63–70.
- [26] Lux MP, Bohme S, Huchereg S, Jeratsch U, Kurschner N, Luftner D. 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. *Breast Cancer Res Treat* 2019;176(3):495–506.
- [27] 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* 2019.
- [28] Gyawali B, Sharma S, Booth CM. Is the number of cancer drug approvals a surrogate for regulatory success. *J Cancer Policy* 2019;22:100202.
- [29] Gyawali B, D'Andrea E, Franklin JM, Kesselheim AS. Response rates and durations of response for biomarker-based cancer drugs in nonrandomized versus randomized trials. *J Natl Compr Canc Netw* 2020;18(1):36–43.
- [30] André F, Ciruelos E, Rubovszky G, et al. Alpelisib for PIK3CA-Mutated, hormone receptor-positive advanced breast cancer. *Eng J Med* 2019;380(20):1929–40.
- [31] Baselga J, Campone M, Piccart M, et al. Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. *New Eng J Med* 2011;366(6):520–9.
- [32] Piccart M, Hortobagyi GN, Campone M, et al. Everolimus plus exemestane for hormone-receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: overall survival results from BOLERO-2†. *Ann Oncol* 2014;25(12):2357–62.
- [33] Im SA, Lu YS, Bardia A, et al. Overall survival with ribociclib plus endocrine therapy in breast cancer. *Eng J Med* 2019;381(4):307–16.
- [34] 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.
- [35] Hernandez-Villafuerte K, Fischer A, Latimer N. Challenges and methodologies in using progression free survival as a surrogate for overall survival in oncology. *Int J Technol Assess Health Care* 2018;34(3):300–16.
- [36] Kluetz PG, O'Connor DJ, Soltys K. Incorporating the patient experience into regulatory decision making in the USA, Europe, and Canada. *Lancet Oncol* 2018;19(5):e267–74.