REVIEW

A Systematic Review and Meta-Analysis of Bevacizumab in First-Line Metastatic Breast Cancer: Lessons for Research and Regulatory Enterprises

Spencer Phillips Hey, Bishal Gyawali ⓘD , Elvira D'Andrea, Manoj Kanagaraj, Jessica M. Franklin, Aaron S. Kesselheim

See the Notes section for the full list of authors' affiliations.
Correspondence to: Spencer Phillips Hey, PhD, Harvard Center for Bioethics, 641 Huntington Ave, Boston, MA 02115 (e-mail: heyspencer@gmail.com).

Abstract

Background: The US Food and Drug Administration's accelerated approval and later withdrawal of bevacizumab in patients with metastatic breast cancer (mBC) is a seminal case for ongoing debates about the validity of using progression-free survival (PFS) as a surrogate measure for overall survival (OS) in cancer drug approvals. We systematically reviewed and meta-analyzed the evidence around bevacizumab's regulatory approval and withdrawal in mBC.
Methods: We searched for all published phase II or III clinical trials testing bevacizumab as a first-line therapy for patients with mBC. Data were extracted on trial demographics, interventions, and outcomes. Descriptive analysis was stratified by whether the trial was initiated before, during, or after the accelerated approval. We used a cumulative random-effects meta-analysis to assess the evolution of evidence of the effect of bevacizumab on PFS and OS. We estimated the association between the trial-level PFS and OS effect using a nonlinear mixed-regression model.
Results: Fifty-two studies were included. Trial activity dramatically dropped after the accelerated approval was withdrawn. Eight clinical trials reported hazard ratios (hazard ratios) and were meta-analyzed. The cumulative hazard ratio for PFS was 0.72 (95% CI = 0.65 to 0.79), and the cumulative hazard ratio for OS was 0.90 (95% CI = 0.80 to 1.01). The regression model showed a statistically nonsignificant association between PFS benefit and OS benefit ($\beta = 0.43$, SE $= 0.81$).
Conclusion: The US Food and Drug Administration's decision-making in this case was consistent with the evolving state of evidence. However, the fact that seven clinical trials are insufficient to conclude validity (or lack thereof) for a trial-level surrogate suggests that it would be more efficient to conduct trials using the more clinically meaningful endpoints.

Progression-free survival (PFS) is a commonly used endpoint in clinical trials of metastatic breast cancer (mBC) treatments, and many investigational drugs have earned approval from the US Food and Drug Administration (FDA) for mBC on the basis of a statistically significant improvement in PFS (1). Although PFS is believed to be a reliable surrogate measure for extending overall survival (OS) for some cancers, in other cancers, treatments may simply delay the time until a cancer progresses but still fail to improve OS (2).

The relationship between PFS and OS is particularly complicated in patients with mBC. Meta-analyses of the association between PFS and OS in mBC trials have not always produced consistent results (2), and consequently, the use of PFS as a surrogate endpoint in this setting remains controversial (3). The FDA's 2008 decision to grant accelerated approval to bevacizumab (Avastin) is a seminal case for this discussion, and bevacizumab's clinical utility continues to be a focus of evidence reviews and meta-analyses (4). Bevacizumab is a humanized monoclonal antibody directed against vascular endothelial growth factor, and its accelerated approval was related to its use in combination with paclitaxel (Taxol) as a first-line treatment in patients with *HER2*-negative mBC. The approval was

based on an improvement in PFS observed in a single multicenter, open-label, randomized study comparing paclitaxel alone to paclitaxel plus bevacizumab (at the time of approval, the bevacizumab arm had not shown an improvement in OS) (5).

Bevacizumab's accelerated approval led to widespread clinical uptake (6). However, subsequent randomized trials and follow-up of the original study revealed that the drug offered no benefit to patient survival, while substantially increasing the risk of serious adverse effects (7). After a year of public discussion, the FDA ultimately removed bevacizumab's approval for this indication in November 2011 (7). But controversy continued; immediately after the withdrawal, the Centers for Medicare and Medicaid Services announced it would continue to reimburse for the use of bevacizumab as first-line treatment in patients with mBC, a policy that continues to the present day (although use of bevacizumab in this indication has substantially declined [6]). The National Comprehensive Cancer Network's (NCCN's) drug compendium also still categorizes bevacizumab as "2 A," which indicates "uniform NCCN consensus that the intervention is appropriate" for mBC.

Discordance between the regulatory action and the reimbursement policy and present-day practice guidelines underscores important and unresolved questions about how to evaluate the evolving state of scientific evidence for (or against) surrogate measures used in the accelerated approval pathway. We therefore sought to review the evolution of evidence surrounding bevacizumab as a first-line treatment for mBC, examining the level of evidence available when the FDA granted its accelerated approval, as well as exploring whether subsequent data supported greater confidence or skepticism for the utility of PFS as a surrogate measure in mBC for this drug. To elucidate these details about the research process and regulatory decision-making related to bevacizumab and mBC, we conducted a systematic review, cumulative meta-analysis, and evidence mapping of phase II and phase III bevacizumab trials in first-line treatment for mBC.

Methods

Literature Search

We searched PubMed, ClinicalTrials.gov, and the Cochrane Systematic Review Database (initially in May 2017 and updated in March 2018) for all phase II or III clinical trials testing bevacizumab as an experimental first-line treatment for mBC. Our search included terms for the condition (*metastatic* and *breast neoplasms*), the intervention (*bevacizumab* or *Avastin*), and the study type (*Clinical Trial, Phase II* or *Clinical Trial, Phase III*).

Titles and abstracts of the database results were screened by two reviewers (SPH and MK) to exclude nonhuman studies, nonfirst-line trials, nonmetastatic breast cancer trials, and nonprimary trial reports (eg, evidence reviews). After obtaining full-text versions of the remaining records, we further excluded any published studies that were not breast cancer, were not metastatic disease, were not first-line therapy trials, did not include bevacizumab, did not report either PFS or OS data, or did not report primary data. For unpublished studies, exclusion criteria 1–4 were used. We then conducted a recursive manual search for additional published studies based on the references in eligible reports.

Data Extraction

For studies meeting our inclusion criteria, two authors (SPH, BG, or MK) independently extracted the following elements (any discrepancies were resolved through consultation with a third author): NCT registration number (if available), trial phase, study design (eg, single-arm or randomized), tumor subtype, treatment interventions, trial status (according to registration record), primary study completion date (according to registration record), availability of results on ClinicalTrials.gov, study start and completion dates, date of publication (if published, or date on which results were uploaded to ClinicalTrials.gov), sample size, primary study endpoint, median PFS and OS, hazard ratios (hazard ratios) for PFS and OS, and whether quality of life (QoL) measures were included as trial outcomes.

Descriptive Analysis

We sought to investigate the effect that the FDA's accelerated approval may have had on bevacizumab trial characteristics. Therefore, we stratified our descriptive analysis into three periods: trials initiated during the preapproval period (before February 2008), trials initiated during approval (March 2008 to November 2011), and trials initiated postwithdrawal (after November 2011). We used AERO graphing to examine the patterns of research activity (8). This method graphically represents each study as a node arranged in time along the *x*-axis and stratified by study design properties along the *y*-axis. Node shape and color are then used to represent qualitative or categorical properties of the study outcome.

Statistical Analysis

We analyzed the evidence on the effect of bevacizumab on PFS and OS over time using a cumulative random-effects meta-analysis (9). Trials were sequentially added by the year their results on PFS or OS became available, either through journal publication or reporting on ClinicalTrials.gov. Between-study heterogeneity was assessed using the I^2 statistic, which describes the percentage of total variation across studies that is the result of heterogeneity rather than chance (10).

To estimate the association between the trial-level PFS hazard ratio and the trial-level OS hazard ratio, we fit a joint nonlinear mixed-effects model (11). The model of surrogacy described a linear relationship between the true log hazard ratio for OS and the true log hazard ratio for PFS, accounting for the uncertainty in the hazard ratio estimates. If PFS is a reliable trial-level surrogate for OS, β (the slope of the linear relationship) should be positive and large in absolute value.

All statistical analyses were performed using STATA 15 (STATA Corp, College Station, TX) and R statistical software. The trial-level surrogate analysis was performed using the R source package developed by Korn et al. (Available from https://brb.nci.nih.gov/programdownload/pCRsoftware.html, accessed April 19, 2019.) Results were considered statistically significant when the 95% confidence interval did not cross 1.00.

Results

Sample Characteristics

Fifty-two studies met eligibility criteria (see Supplementary Figure 1, available online, for PRISMA diagram). These trials evaluated 44 different bevacizumab combination regimens and collectively enrolled 11 897 participants. Table 1 lists the properties of the total sample, as well as the properties of the three time-stratified subsets. At the time of this analysis, 48 (92.3%) of

Table 1. Characteristics of sample

Characteristic	Total No. (%)	Before Feb. 2008 No. (%)	Mar. 2008–Nov. 2011 No. (%)	After Nov. 2011 No. (%)
Trials	52 (100)	29 (53.8)	19 (36.5)	4 (7.7)
Completed	48 (92.3)	28 (96.5)	19 (100.0)	1 (25.0)
Design				
Phase 3	14 (26.9)	8 (27.6)	5 (26.3)	1 (25.0)
Randomized	23 (44.2)	13 (44.8)	6 (31.6)	4 (100.0)
Double-blind	7 (13.5)	5 (17.2)	1 (5.3)	1 (25.0)
Single arm	29 (55.8)	16 (55.2)	13 (68.4)	0 (0.0)
Arms > 2	8 (15.4)	6 (20.7)	1 (5.3)	1 (25.0)
BVZ in all arms	3 (5.8)	0 (0.0)	3 (15.8)	0 (0.0)
Primary endpoint				
ORR	13 (25.0)	7 (24.1)	6 (31.6)	0 (0.0)
PFS	33 (63.5)	19 (65.5)	11 (57.9)	3 (75.0)
OS	1 (1.9)	0 (0.0)	1 (5.3)	0 (0.0)
Dose/safety	4 (7.7)	3 (10.3)	1 (5.3)	0 (0.0)
Other	1 (1.9)	0 (0.0)	0 (0.0)	1 (25.0)
Tumor type				
HER2-	27 (51.9)	15 (51.7)	10 (52.6)	2 (50.0)
TNBC	7 (13.5)	1 (3.4)	5 (26.3)	1 (25.0)
HER2+	5 (9.6)	4 (13.8)	1 (5.3)	0 (0.0)
Data availability*				
Published	27 (56.3)	16 (57.1)	10 (55.6)	1 (100.0)
Results on ClinicalTrials.gov	26 (54.2)	19 (67.9)	6 (33.3)	1 (100.0)
Outcome†				
Favorable	6 (13%)	3 (10.7)	3 (20.0)	0 (0.0)
Unfavorable	13 (29%)	10 (35.7)	2 (13.3)	1 (100.0)
Mixed	1 (2%)	0 (0.0)	1 (6.7)	0 (0.0)
Terminated	10 (21%)	6 (21.4)	4 (22.2)	0 (0.0)
Unknown	14 (29%)	9 (32.1)	5 (27.8)	0 (0.0)

*For calculating the percentage of trials with available data, the denominator excludes active trials. Abbreviations: ORR = objective response rate; PFS = progression-free survival; OS = overall survival; TNBC = triple-negative breast cancer.
†For calculating the percentage of favorable, unfavorable, and mixed outcomes, the denominator excludes active trials and trials that contained BVZ in all treatment arms. For calculating the percentage of terminated trials or trials whose outcome is unknown, the denominator excludes active trials.

the trials were completed. Of the 48 completed trials, 27 (56.3%) have published their results, and 34 (70.8%) have either published or reported results on ClinicalTrials.gov (5,12–36). In total, data from 8354 (67.5%) patient subjects are available.

The majorities of trials in our sample were phase II (73.1%), nonrandomized (55.8%), and open-label (86.5%). The most common primary trial endpoint was PFS (63.5%), followed by objective response rate (25.0%). Only one trial used OS as its primary endpoint. Most trials restricted enrollment to patients with either *HER2*–negative (n = 27, 51.9%) or triple-negative (n = 7, 13.5%) tumors. Three trials (5.8%) involved administering bevacizumab in all study arms (23,30,35).

Six trials (13.3%) achieved statistically significant outcomes favorable to bevacizumab on the primary study endpoint (classified as having a "favorable" outcome in Table 1) (12,21,22,26,32,34). All seven trials that reported an hazard ratio for OS found no survival benefit with addition of bevacizumab, but each was small and not well powered to identify moderate effect sizes (5,13,15,24,29,36). Ten trials (20.8%) were terminated, and the results of 14 completed trials (29.2%) remain unknown. One trial had a "mixed" result: a single-arm study of docetaxel+epirubicin+bevacizumab that found activity with the regimen but also substantial toxicity (25).

Trial characteristics were similar across our three time periods. Our evidence mapping (see Supplementary Figure 2, available online) found that 55.8% (29 of 52) of trials were initiated before accelerated approval. Only four (7.7%) were initiated following the FDA's withdrawal of bevacizumab's mBC indication. The three trials that had bevacizumab across all study arms were initiated during the approval period (2008–2011). The most vigorous period of research activity occurred in the period leading up to and immediately following bevacizumab's initial approval. Thirty-seven (71.1%) trials were initiated between 2006 and 2009.

Eight (15.4%) of the completed studies, which included nine comparisons of a bevacizumab-containing regimen to a control regimen, reported sufficient data to be included in the quantitative analysis. Seven of these presented their results in publications, and one unpublished study (NCT00520975) has results available on ClinicalTrials.gov (5,13,15,24,29). Table 2 presents the trial design, tumor types, treatment regimens, sample sizes, and outcome data extracted from these reports. For the three trials (Miller et al. [5], NCT00520975, and Gianni et al. [24]) that included QoL assessments along with PFS or OS outcomes, none found a statistically significant difference between the bevacizumab and control arms.

In total, 36 different bevacizumab-containing regimens were evaluated across this portfolio. Only seven regimens (19.4%) were evaluated in more than one trial. The most frequently tested regimen was bevacizumab+paclitaxel (eight trials). When regimens were retested, the results were largely consistent. In one instance, completed trials of the same regimen

Table 2. PFS and OS outcomes for randomized controlled trials of BVZ as a first-line treatment in patients with metastatic breast cancer*

Study	Masking	Tumor type	Exp	Ctrl	No. exp	Med. PFS	Med. OS	No. ctrl	Med. PFS	Med. OS	Delta PFS	Delta OS	PFS HR (95% CI)	OS HR (95% CI)	Included QoL endpoint?
Miller et al. 2007 (5)	Open label	Any	BVZ+paclitaxel	Paclitaxel	347	11.8	26.7	326	5.9	25.2	5.9	1.5	0.6 (0.51 to 0.70)	0.87 (0.72 to 1.05)†	Yes
Miles et al. 2010 (13)	Double-blind	HER2–	BVZ+docetaxel	Docetaxel	247	10.1	30.2	241	8.2	31.9	1.9	–1.7	0.77 (0.64 to 0.93)	1.03 (0.70 to 1.33)	No
Robert et al. 2011 (15)‡	Double-blind	HER2–	BVZ+capecitabine	Capecitabine	409	8.6	25.7	206	5.7	22.8	2.9	2.9	0.69 (0.56 to 0.84)	0.85 (0.63 to 1.14)	No
			BVZ+anthra/Tax	Anthra/tax	415	9.2	27.5	207	8.0	NR	1.2	–	NR	NR	No
Martin et al. 2011 (17)	Double-blind	HER2–	BVZ+paclitaxel	Paclitaxel	97	11.5	NR	94	9.0	NR	2.5	–	0.79 (0.53 to 1.17)	NR	No
Gianni et al. 2013 (24)	Open label	HER2+	BVZ+docetaxel+ trastuzumab	Docetaxel+ trastuzumab	216	16.5	38.5	208	13.7	38.3	2.8	0.2	0.82 (0.65 to 1.02)	1.01 (0.74 to 1.38)	Yes
Martin et al. 2015 (29)	Open label	HER2–	BVZ+letrozole or fulvestrant	Letrozole or fulvestrant	190	19.3	52.1	184	14.4	51.8	4.9	0.3	0.83 (0.65 to 1.06)	0.87 (0.58 to 1.32)	No
NCT00520975	Double-blind	HER2+	BVZ+carboplatin+ paclitaxel+ trastuzumab	Carboplatin+ paclitaxel+ trastuzumab	48	13.8	63.0	48	11.1	49.1	2.7	13.9	0.73 (0.43 to 1.23)	1.09 (0.61 to 1.97)	Yes
Miles et al. 2016 (36)	Double-blind	HER2–	BVZ+paclitaxel	Paclitaxel	238	11.0	NR	233	8.8	NR	2.2	–	0.68 (0.51 to 0.91)	0.81 (0.61 to 1.08)	No

*Individual trial properties and median outcome data for the eight randomized controlled trials in our sample that reported PFS or OS outcome data. Anthra/Tax = anthracycline-taxane combination therapy; BVZ = bevacizumab; Exp = experimental arm; Ctrl = control arm; Med = median; NR = not reported; OS = overall survival; PFS = progression-free survival; QoL = quality of life.

†The 95% CI for OS HR was not reported in the publication; however, the US Food and Drug Administration does report this information in its 2010 memorandum on the decision to withdraw the indication.

‡The trial reported in Robert, 2011 (15) was a four-arm clinical trial comparing the effects of adding BVZ to two different background regimens.

REVIEW

found discordant results: a favorable single-arm phase II trial of docetaxel+trastuzumab+bevacizumab that showed promising PFS (26) was followed by an unfavorable phase III trial that found no PFS or OS benefit of docetaxel+trastuzumab+bevacizumab over docetaxel+trastuzumab alone (24).

Relationship Between PFS and OS

Figure 1 shows the results of the cumulative meta-analyses for PFS and OS hazard ratios across the sample. The final pooled estimate for PFS benefit was 0.72 (95% CI = 0.65 to 0.79), which regressed from the initial estimate of 0.60 (95% CI = 0.51 to 0.70). By contrast, the pooled hazard ratio for OS remained relatively stable over time, ranging between 0.87 and 0.92. The FDA's decision to withdraw the indication was based on the first three trials (5,13,15) in the sample (7), and our analysis accords with the Agency's assessment that the evidence from those trials did not show a statistically significant OS benefit. The final pooled estimate for OS, which includes results from four additional trials, remains statistically nonsignificant (hazard ratio = 0.90, 95% CI = 0.80 to 1.01) but does show a stable 10% OS benefit.

Figure 2 shows the association between trial-level hazard ratio for PFS effects and trial-level hazard ratio for OS effects observed in the seven trials with data on both outcomes. The solid line represents equality between OS and PFS effects, whereas the dashed line represents the estimated linear association from the random-effects model. The pattern of points clustering around the dashed line suggests a positive slope ($\beta = 0.43$, SD = 0.81), meaning that a 1-unit increase in log-hazard ratio for PFS was associated with a 0.43-unit increase in log-hazard ratio for OS. However, our sample size (seven studies) was small, the standard error was large, and these data would be consistent with a negative slope as well. The model also indicates that nearly all of the heterogeneity in OS treatment effects between studies can be explained by PFS or random variability, as there was little residual heterogeneity ($g = 0$). All estimates of the parameters from the model of surrogacy are presented in Table 3.

We conducted sensitivity analyses for all of the statistical analyses, removing the one study whose results were only available on ClinicalTrials.gov. This did not qualitatively change the results (see Supplementary Figures 3–5 and Supplementary Table 1, available online).

Discussion

The portfolio of clinical trials testing bevacizumab as a first-line treatment for mBC showed that at the time of the FDA's decision to withdraw bevacizumab's accelerated approval, the total body of evidence showed a statistically significant 32% improvement in PFS and a non-statistically significant 10% improvement in OS. As additional evidence has accumulated since that time, these estimates have remained stable, validating the FDA's decision.

We also observed extensive exploration of bevacizumab-containing combination regimens, amounting to trials of 36 different regimens initiated within a period of 12 years. To date, none of these 36 regimens has been shown to offer a statistically significantly improvement in patient survival or QoL over a non-bevacizumab comparator. Although this does not preclude the possibility that an effective bevacizumab combination therapy for mBC could still be found, the FDA's withdrawal of

A

B
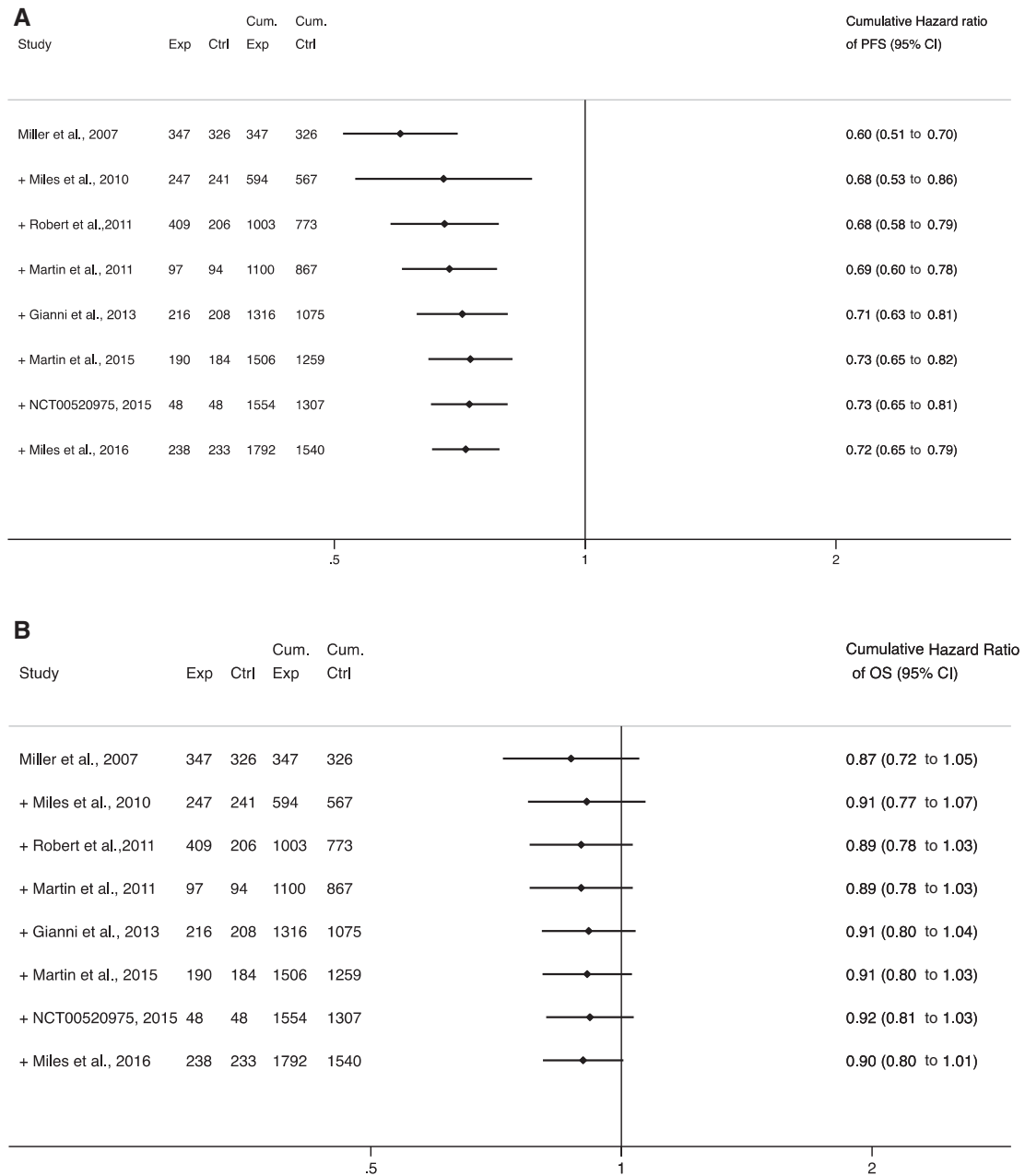
Figure 1. Forest plots of cumulative meta-analysis on the effect of bevacizumab on progression-free survival (PFS) and overall survival (OS). **A)** PFS and **B)** OS panels show cumulative meta-analyses performed by adding individual trials chronologically and summarizing the point estimate as each new trial is added. The plus symbol before each trial indicates that the point estimate (ie, hazard ratio) of that trial is pooled with the previous summary estimate. The final result is a trial-level cumulative estimate over time and corresponding 95% confidence intervals. The diamonds represent the cumulative point estimates and the horizontal segments represent the 95% confidence intervals. Ctrl = number of participants in the control group; Cum. Ctrl = cumulative number of participants in the control group; Cum. Exp = cumulative number of participants in the experimental group; Exp = number of participants in the experimental group; OS = overall survival; PFS = progression-free survival.

bevacizumab's mBC indication in 2011 (and perhaps the controversy surrounding this decision) appears to have dampened enthusiasm for this search. Only one registered trial (NCT01898117) testing a new combination has been initiated since the start of 2012, and in January 2018, this trial replaced bevacizumab with the PD-L1 inhibitor atezolizumab as the experimental intervention.

Finally, we found that the accumulated evidence covering seven randomized clinical trials that enrolled 3141 patients remains insufficient to draw definitive conclusions about the validity of PFS as a surrogate for this treatment and indication. This result challenges a common interpretation of this case, which emphasizes the dangers of relying on unvalidated surrogate measures for new drug approvals (1,4). Our analysis
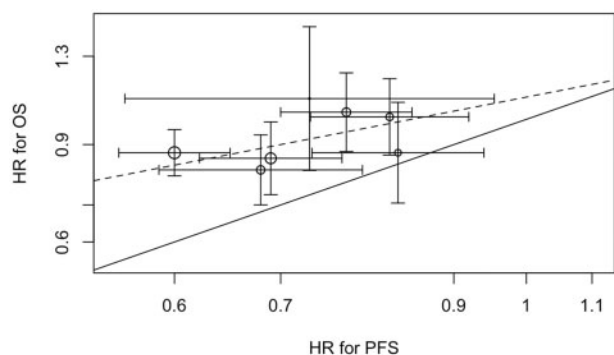
Figure 2. Association of trial-level PFS hazard ratio (HR) and OS hazard ratio. The graph shows the association of trial-level PFS and OS effects expressed as hazard ratios. The solid line represents equality between OS and PFS effects. The dashed line represents the estimated slope of the linear association 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. OS = overall survival; PFS = progression-free survival.

Table 3. Parameters ($\pm$ standard error) for a nonlinear mixed-effects model for the association of log-HR for OS and log-HR for PFS

Parameter*	Estimate	Standard error
α	0.045	0.291
β	0.425	0.805
μ	-0.332	0.054
g	0	0.001
σ^2	0.007	0.009

*α = intercept of the linear relationship between the log-HR for OS and log-HR for PFS (when the HR for PFS is 1.0, the estimated HR for OS is $e^{0.045} = 1.05$); β = coefficient of the linear relationship between the log-HR for OS and log-HR for PFS; μ = average log-HR for PFS across trials; corresponding to an HR of $e^{-0.332} = 0.72$; g = variance of log-HR for OS across trials that is not explained by PFS; σ^2 = variance of log-HR for PFS across trials. HR = hazard ratio; OS = overall survival; PFS = progression-free survival.

suggests, by contrast, that we do not yet have sufficient evidence to conclude whether PFS is a valid or invalid trial-level surrogate in this indication, and indeed, simulation studies have suggested that, when the number of trials is small, estimation of the surrogacy model may be highly unstable and the statistical power to demonstrate surrogacy is limited (37).

Nevertheless, this analysis provides several insights into the use of trial-level surrogates and the FDA's accelerated approval pathway. First, the accelerated approval standard in the statute is that a surrogate measure must be "reasonably likely" to predict clinical endpoints measuring how a patient feels, functions, or survives. However, umbrella reviews suggest the FDA routinely grants approvals for drugs treating cancer based on effects observed in surrogates that are known to have poor correlation with overall survival or for which the correlation is unproven (2). A recent analysis of the FDA–mandated confirmatory trials for new cancer therapies that received accelerated approval based on a surrogate endpoint also found that many confirmatory trials used the same surrogate endpoint for their primary outcome, and that in some cases, a negative outcome in the confirmatory trial did not lead to withdrawal of the indication (38). The apparent lack of consistency in regulatory actions suggests that greater guidance is needed about what degree of evidence is needed to satisfy the "reasonably likely" standard, as well as what endpoints should be considered clinically beneficial.

In the case of bevacizumab and mBC, seven randomized controlled trials were insufficient to determine whether PFS is a valid trial-level surrogate for this one intervention and setting. It is therefore clearly impractical to suggest that a surrogate measure could be validated for each intervention and setting. Yet, neither does it seem prudent to suggest that a surrogate validated for one drug (or mechanism) and setting will likely be valid for other drugs or mechanisms or settings. Indeed, evidence from a recent umbrella review of trial-level surrogates in cancer provides compelling evidence that even within one setting, drugs that act through a different mechanism may have a very different surrogacy relationship between PFS and OS (2). Therefore, any extrapolation of surrogate validity from other drugs or other diseases is speculative.

We therefore propose that whenever regulators approved a new drug on the basis of a surrogate measure, they should require follow-up trials using more clinically meaningful endpoints, such as OS or QoL measures. Whereas such follow-up trials were required after accelerated approval of bevacizumab for the treatment of mBC, more recent first-line treatments for mBC (eg, everolimus, palbociclib) have received standard approval from the FDA on the basis of PFS benefits only, and similar to bevacizumab, the early evidence on OS has not shown a clear benefit (39). Because these treatments received standard approvals, rather than accelerated approvals, the FDA does not have the same authority to mandate follow-up trials to establish efficacy, and these indications cannot be as easily withdrawn. Future legislation should provide authority to the FDA to require follow-up trials with patient-centered clinical endpoints when standard approvals are based on surrogate measures.

Second, our findings show how regulatory decisions can affect the clinical research enterprise. A regulatory approval not only allows a product to be marketed and enter clinical practice but also can stimulate research activity. Trials after a regulatory approval may be more likely to be modeled on the preapproval studies, since these have established the de facto standard of evidence for an approved indication. For example, the fact that PFS was the basis for bevacizumab's approval may explain why nearly every phase III trial in our sample also adopted PFS as its primary endpoint, despite the fact that OS is the more clinically meaningful and patient-centered outcome. Although patient crossover and postprogression therapy can complicate the analysis of OS as a trial endpoint (40), it would nevertheless seem to be more informative and efficient if, following an accelerated approval, at least some randomized controlled trials in the same indication were adequately powered to assess the gold-standard, clinical endpoint or QoL measures. If so, then even in cases of high clinical need, we can be assured of timely resolution of the uncertainty about the product's clinical utility.

Finally, the withdrawal of bevacizumab's approval largely halted further research into its use for this indication. This highlights another important power of the FDA. Accelerated approvals are naturally based on limited data. Requiring the manufacturer to conduct follow-up studies is thus valuable and necessary to address the remaining uncertainties about a product's benefit and risk profile. These results show that enforcing that requirement and withdrawing an indication (when warranted) can have a beneficial effect on the research enterprise by mitigating exposure of future research participants to unnecessary risks and burdens.

Limitations of our study include the following: First, we examined only first-line mBC treatment trials that included bevacizumab. It is possible that there is a more robust correlation between PFS benefit and OS benefit in mBC treatment across

the entire portfolio of experimental therapies. Therefore, additional analyses will be needed to clarify the domains of validity and utility for PFS as a surrogate measure in mBC treatment more broadly. Second, most of the trials in our analysis were not powered to detect effects on OS, and therefore, our pooled estimate for bevacizumab's effect on OS is based on a limited number of events. Third, although the trials in our meta-analysis were all testing the benefits of adding bevacizumab to a background regimen, the control arms did differ across the sample, and this heterogeneity may make our pooled estimates for PFS and OS less reliable. Fourth, some commentators argue that even if PFS does not predict OS benefit, PFS should still be considered a beneficial endpoint in its own right or may be predictive of improvements in QoL. However, prior studies have suggested that PFS is generally a poor surrogate for QoL (41). Fifth, our evidence mapping can only estimate when a result became known, disseminated, or shared with the FDA, based on study completion and publication dates. It is therefore possible that the FDA or other decision-makers may have access to more trial data than are reflected in our study. Sixth, although we included both registered and published trials in our analysis, we cannot rule out the influence of publication bias. Seventh, we did not have access to patient-level data from these trials, which would have permitted a more precise evaluation of surrogacy that could account for the fact that death is a common event between PFS and OS outcomes. Finally, our quantitative analysis was limited to hazard ratios for measures of treatment effects, even though differences in survival time may be the more clinically relevant measure. This limitation was due to the fact that measures of precision were generally unavailable for the reported survival time differences, making appropriate quantitative analysis on these endpoints impossible.

The FDA's approval and withdrawal of bevacizumab in mBC treatment illuminates a number of important lessons for how to test and rely on surrogate measures in drug development. Surrogate measures used in trials and by regulators to facilitate clinical translation may increase the efficiency of new drug development, but the utility of relying on surrogate endpoints requires that we understand (as much as possible) how they predict (or fail to predict) patient-centered outcomes. The fact that seven clinical trials may be insufficient to conclude validity (or lack thereof) for a trial-level surrogate highlights an important limitation in the use of surrogate measures, and it suggests that it may be far more efficient over the longer term to simply conduct trials using the more clinically meaningful endpoints.

Funding

Notes

Affiliations of authors: 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 (SPH, BG, ED, JMF, ASK); Harvard Center for Bioethics, Harvard Medical School, Boston, MA (SPH, ASK); Department of Oncology, Department of Public Health Sciences, and Division of Cancer Care and Epidemiology, Queen's University, Kingston, Ontario, Canada (BG); Harvard Medical School, Boston, MA (MK).

References

1. Kim C, Prasad V. Cancer drugs approved on the basis of a surrogate end point and subsequent overall survival: an analysis of 5 years of US Food and Drug Administration approvals. *JAMA Intern Med*. 2015;175(12):1992–1994.

2. Haslam A, Hey SP, Gill J, Prasad V. A systematic review of trial-level meta-analyses measuring the strength of association between surrogate endpoints and overall survival in oncology. *Eur J Cancer*. 2019;106:196–211.

3. Wilson MK, Karakasis K, Oza AM. Outcomes and endpoints in trials of cancer treatment: the past, present, and future. *Lancet Oncol*. 2015;16(1):e32–42.

4. Nahleh Z, Botrus G, Dwivedi A, Jennings M, Nagy S, Tfayli A. Bevacizumab in the neoadjuvant treatment of human epidermal growth factor receptor 2-negative breast cancer: a meta-analysis of randomized controlled trials. *Mol Clin Oncol*. 2019;10(3):357–365.

5. Miller K, Wang M, Gralow J, et al. Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. *N Engl J Med*. 2007;357(26):2666–2676.

6. Dusetzina SB, Ellis S, Freedman RA, et al. How do payers respond to regulatory actions? The case of Bevacizumab. *JOP*. 2015;11(4):313–318.

7. Pazdur R. *Regulatory Decision to Withdraw Avastin (Bevacizumab) First-Line Metastatic Breast Cancer Indication*. Rockville, MD: FDA Center for Drug Evaluation and Research; 2010. https://www.fda.gov/downloads/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/UCM237171.pdf. Accessed December 19, 2018.

8. Hey SP, Heilig CM, Weijer C. Accumulating Evidence and Research Organization (AERO) model: a new tool for representing, analyzing, and planning a translational research program. *Trials*. 2013;14(1):159.

9. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteler F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. *N Engl J Med*. 1992;327(4):248–254.

10. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557–560.

11. 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*. 2016;27(1):10–15.

12. Lobo C, Lopes G, Baez O, et al. Final results of a phase II study of nab-paclitaxel, bevacizumab, and gemcitabine as first-line therapy for patients with HER2-negative metastatic breast cancer. *Breast Cancer Res Treat*. 2010;123(2):427–435.

13. Miles DW, Chan A, Dirix LY, et al. Phase III study of bevacizumab plus docetaxel compared with placebo plus docetaxel for the first-line treatment of human epidermal growth factor receptor 2–negative metastatic breast cancer. *JCO*. 2010;28(20):3239–3247.

14. Rochlitz C, Ruhstaller T, Lerch S, et al. Combination of bevacizumab and 2-weekly pegylated liposomal doxorubicin as first-line therapy for locally recurrent or metastatic breast cancer. A multicenter, single-arm phase II trial (SAKK 24/06). *Ann Oncol*. 2011;22(1):80–85.

15. Robert NJ, Diéras V, Glaspy J, et al. RIBBON-1: randomized, double-blind, placebo-controlled, phase III trial of chemotherapy with or without bevacizumab for first-line treatment of human epidermal growth factor receptor 2–negative, locally recurrent or metastatic breast cancer. *JCO*. 2011;29(10):1252–1260.

16. Brufsky A, Hoelzer K, Beck T, et al. A randomized phase II study of paclitaxel and bevacizumab with and without gemcitabine as first-line treatment for metastatic breast cancer. *Clin Breast Cancer*. 2011;11(4):211–220.

17. Martin M, Roche H, Pinter T, et al. Motesanib, or open-label bevacizumab, in combination with paclitaxel, as first-line treatment for HER2-negative locally recurrent or metastatic breast cancer: a phase 2, randomised, double-blind, placebo-controlled study. *Lancet Oncol*. 2011;12(4):369–376.

18. Nahleh Z, Gupta R, Abrams J, Gartner E, Reichle L. Phase II trial of biweekly gemcitabine, paclitaxel, and bevacizumab as frontline therapy for metastatic breast cancer (MBC). *JCO*. 2011;29(15_suppl):e11527.

19. Yardley DA, Burris HA, Clark BL, et al. Hormonal therapy plus bevacizumab in postmenopausal patients who have hormone receptor–positive metastatic breast cancer: a phase II trial of the Sarah Cannon Oncology Research Consortium. *Clin Breast Cancer*. 2011;11(3):146–152.

20. Borson R, Harker G, Reeves J, et al. Phase II study of gemcitabine and bevacizumab as first-line treatment in taxane-pretreated, HER2-negative, locally recurrent or metastatic breast cancer. *Clin Breast Cancer*. 2012;12(5):322–330.

21. Martín M, Makhson A, Gligorov J, et al. Phase II study of bevacizumab in combination with trastuzumab and capecitabine as first-line treatment for HER-

2-positive locally recurrent or metastatic breast cancer. *Oncologist*. 2012;17(4): 469–475.

22. Ramaswamy B, Fiskus W, Cohen B, et al. Phase I–II study of vorinostat plus paclitaxel and bevacizumab in metastatic breast cancer: evidence for vorinostat-induced tubulin acetylation and Hsp90 inhibition in vivo. *Breast Cancer Res Treat*. 2012;132(3):1063–1072.

23. Rugo HS, Campone M, Amadori D, et al. A randomized, phase II, three-arm study of two schedules of ixabepilone or paclitaxel plus bevacizumab as first-line therapy for metastatic breast cancer. *Breast Cancer Res Treat*. 2013;139(2): 411–419.

24. Gianni L, Romieu GH, Lichinitser M, et al. AVEREL: a randomized phase III trial evaluating bevacizumab in combination with docetaxel and trastuzumab as first-line therapy for HER2-positive locally recurrent/metastatic breast cancer. *JCO*. 2013;31(14):1719–1725.

25. Tryfonidis K, Boukovinas I, Xenidis N, et al. A multicenter phase I–II study of docetaxel plus epirubicin plus bevacizumab as first-line treatment in women with HER2-negative metastatic breast cancer. *Breast*. 2013;22(6): 1171–1177.

26. Schwartzberg LS, Badarinath S, Keaton MR, Childs BH. Phase II multicenter study of docetaxel and bevacizumab with or without trastuzumab as first-line treatment for patients with metastatic breast cancer. *Clin Breast Cancer*. 2014;14(3):161–168.

27. Diéras V, Wildiers H, Jassem J, et al. Trebananib (AMG 386) plus weekly paclitaxel with or without bevacizumab as first-line therapy for HER2-negative locally recurrent or metastatic breast cancer: a phase 2 randomized study. *Breast*. 2015;24(3):182–190.

28. Lück HJ, Lübbe K, Reinisch M, et al. Phase III study on efficacy of taxanes plus bevacizumab with or without capecitabine as first-line chemotherapy in metastatic breast cancer. *Breast Cancer Res Treat*. 2015;149(1):141–149.

29. Martín M, Loibl S, von Minckwitz G, et al. 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. *JCO*. 2015;33(9):1045–1052.

30. Rugo HS, Barry WT, Moreno-Aspitia A, et al. Randomized phase III trial of paclitaxel once per week compared with nanoparticle albumin-bound nab-paclitaxel once per week or ixabepilone with bevacizumab as first-line chemotherapy for locally recurrent or metastatic breast cancer: CALGB 40502/NCCTG N063H (Alliance). *JCO*. 2015;33(21):2361–2369.

31. Yardley DA, Bosserman LD, O'Shaughnessy JA, et al. Paclitaxel, bevacizumab, and everolimus/placebo as first-line treatment for patients with metastatic HER2-negative breast cancer: a randomized placebo-controlled phase II trial of the Sarah Cannon Research Institute. *Breast Cancer Res Treat*. 2015;154(1):89–97.

32. Nikolaou M, Saloustros E, Polyzos A, et al. Final results of weekly paclitaxel and carboplatin plus bevacizumab as first-line treatment of triple-negative breast cancer. A multicenter phase I-II trial by the Hellenic Oncology Research Group. *Ann Oncol*. 2016;27(suppl_6). doi: 10.1186/s12885-016-2823-y.

33. Rochlitz C, Bigler M, von Moos R, et al. SAKK 24/09: safety and tolerability of bevacizumab plus paclitaxel vs. bevacizumab plus metronomic cyclophosphamide and capecitabine as first-line therapy in patients with HER2-negative advanced stage breast cancer-a multicenter, randomized phase III trial. *BMC Cancer*. 2016;16(1):780.

34. Tiainen L, Tanner M, Lahdenperä O, et al. Bevacizumab combined with docetaxel or paclitaxel as first-line treatment of HER2-negative metastatic breast cancer. *Anticancer Res*. 2016;36(12):6431–6438.

35. Zielinski C, Ling I, Inbar M, et al. Bevacizumab plus paclitaxel versus bevacizumab plus capecitabine as first-line treatment for HER2-negative metastatic breast cancer (TURANDOT): primary endpoint results of a randomised, open-label, non-inferiority, phase 3 trial. *Lancet Oncol*. 2016;17(9):1230–1239.

36. Miles D, Cameron D, Bondarenko I, et al. Bevacizumab plus paclitaxel versus placebo plus paclitaxel as first-line therapy for HER2-negative metastatic breast cancer (MERiDiAN): a double-blind placebo-controlled randomised phase III trial with prospective biomarker evaluation. *Eur J Cancer*. 2017;70: 146–155.

37. Shi Q, Renfro LA, Bot BM, Burzykowski T, Buyse M, Sargent DJ. Comparative assessment of trial-level surrogacy measures for candidate time-to-event surrogate endpoints in clinical trials. *Comput Stat Data Anal*. 2011;55(9): 2748–2757.

38. Gyawali B, Hey SP, Kesselheim AS. Assessment of the clinical benefit of cancer drugs receiving accelerated approval. *JAMA Intern Med*. 2019;179(7):906.

39. Gyawali B, Prasad V. Same data; different interpretations. *JCO*. 2016;34(31): 3729.

40. Seidman AD, Bordeleau L, Fehrenbacher L, et al. 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*. 2018;36(32):3259–3268.

41. Hwang TJ, Gyawali B. Association between progression-free survival and patients' quality of life in cancer clinical trials. *Int J Cancer*. 2018. doi: 10.1002/ijc.31957.

REVIEW