



Published in final edited form as:

Eur J Cancer. 2020 September ; 136: 176–185. doi:10.1016/j.ejca.2020.06.015.

Progression-Free Survival is a Suboptimal Predictor for Overall Survival Among Metastatic Solid Tumor Clinical Trials

Dario Pasalic, M.D.¹, Gwendolyn J. McGinnis, M.D.¹, C. David Fuller, M.D., Ph.D.¹, Aaron J. Grossberg, M.D., Ph.D.², Vivek Verma, M.D.³, Walker Mainwaring, M.D.⁴, Austin B. Miller, B.S.⁵, Timothy A. Lin, M.D.⁶, Amit Jethanandani, M.D., M.P.H.⁷, Andres F. Espinoza, M.D.⁴, Markus Diefenhardt, M.D.^{8,9}, Prajnan Das, M.D., M.S., M.P.H.¹, Vivek Subbiah, M.D.¹, Ishwaria M. Subbiah, M.D.¹, Reshma Jagsi, M.D., D.Phil.¹⁰, Adam S. Garden, M.D.¹, Emmanouil Fokas, M.D., D.Phil.^{8,9,11,12}, Claus Rödel, M.D.^{8,9,11,12}, Charles R. Thomas Jr., M.D.², Bruce D. Minsky, M.D.¹, Ethan B. Ludmir, M.D.^{1,*}

¹The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

²Oregon Health & Science University, Portland, OR, USA.

³Allegheny Health Network Cancer Institute, Pittsburgh, PA, USA.

⁴Baylor College of Medicine, Houston, TX, USA.

⁵The University of Texas Health Science Center McGovern Medical School, Houston, TX, USA.

⁶The Johns Hopkins University School of Medicine, Baltimore, MD, USA.

⁷The University of Tennessee Health Science Center, Memphis, TN, USA.

⁸University of Frankfurt, Frankfurt, Germany.

⁹Frankfurt Cancer Institute, Frankfurt, Germany.

¹⁰University of Michigan, Ann Arbor, MI, USA.

¹¹German Cancer Research Center, Heidelberg, Germany.

¹²German Cancer Consortium, Frankfurt, Germany.

Abstract

Background: The use of overall survival (OS) as the gold standard primary endpoint (PEP) in metastatic oncologic randomized controlled trials (RCTs) has declined in favor of progression-free survival (PFS) without a complete understanding of the degree to which PFS reliably predicts for OS.

Methods: Using [ClinicalTrials.gov](https://clinicaltrials.gov), we identified 1,239 phase 3 oncologic RCTs, 260 of which were metastatic solid tumor trials with a superiority-design investigating a therapeutic intervention

*Corresponding Author: Ethan B. Ludmir, MD. 1515 Holcombe Blvd Unit 1422, Houston, TX 77030. Phone: 713-792-5863. ebludmir@mdanderson.org.

Conflicts of Interest Statement

The authors report no financial disclosures or conflicts of interests related to this work.

using either a PFS or OS PEP. Each individual trial was reviewed to quantify RCT design factors and disease-related outcomes.

Results: A total of 172,133 patients were enrolled from 1999–2015 in RCTs that utilized PFS (56.2%, 146/260) or OS (43.8%, 114/260) as the PEP. PFS trials were more likely to restrict patient eligibility using molecular criteria (15.1% vs. 4.4%, $p=0.005$), utilize targeted therapy (80.1% vs. 67.5%, $p=0.048$), accrue fewer patients (median 495 vs. 619, $p=0.03$), and successfully meet the trial PEP (66.9% vs. 33.3%, $p<0.0001$). On multiple binary logistic regression analysis, factors that predicted for PFS or OS PEP trial success included choice of PFS PEP ($p<0.0001$), molecular profile restriction ($p=0.02$) and single agent therapy ($p=0.02$). Notably, there was only a 38% (31/82) conversion rate of positive PFS-to-OS benefit; lack of industry sponsorship predicted for PFS-to-OS signal conversion (80.0% without industry sponsorship vs. 35.1% with industry sponsorship, $p=0.045$).

Conclusions: A PFS PEP has suboptimal positive predictive value for OS among phase 3 metastatic solid tumor RCTs. Regulatory agency decisions should be judicious in utilizing PFS results as the primary basis for approval.

Introduction:

Patients with metastatic malignancies are at high risk of disease-related mortality as well as suffering from poor quality of life. In this patient population, the need for treatment guidelines based on rationally-designed randomized controlled trials (RCTs) with a sound testable hypothesis cannot be understated.¹ To that end, the choice of primary endpoint (PEP) for RCTs is paramount. Historically, overall survival (OS) has been considered the “gold standard” PEP, as it provides an objective, easily measurable, and patient-centered outcome. However, the practical limitations of cost, trial design (including power analysis and crossover arm strategies), extended follow-up, and urgency of need for new treatments has led to a decline in OS as the PEP among oncologic RCTs.² Instead, contemporary trials and regulatory approvals have increasingly relied on surrogate endpoints.^{3–5}

The most commonly utilized surrogate endpoint among metastatic solid tumor trials is progression-free survival (PFS), generally defined as tumor growth beyond a certain threshold or death.⁶ Given the disease-centered nature and subjectivity commonly associated with PFS, there is debate in the oncologic community on the usefulness of PFS as a PEP in patients with advanced cancer.⁷ If PFS is to be useful, it should provide a reasonable degree of predictive value for clinically meaningful endpoints such as OS or quality of life.

Examining the disease-site-level relationship between PFS and OS, meta-analyses of RCTs for individual disease sites—thirteen trials in colorectal cancer⁸, seven in small-cell lung cancer⁹, and five in ovarian cancer¹⁰—found PFS to be a potential surrogate for OS. Meanwhile, the same cannot be said for RCTs evaluating non-small cell lung cancer¹¹, breast cancer¹², and prostate cancer.¹³ A disease-site agnostic systematic review of one hundred and ninety-three oncologic RCTs found that the majority (54%) of trials had a low correlation coefficient between surrogate endpoints and OS.¹⁴ Such a comprehensive review offers valuable insight but is limited by the heterogeneity of included surrogate endpoints (i.e. PFS, pathologic complete response, locoregional control, disease-free survival, event-

free survival, and/or time to progression) and treatment settings (i.e. locally advanced and metastatic). Moreover, a correlation coefficient provides limited clinically-meaningful information to most patients on how to interpret the relationship between PFS and OS when choosing to participate in a trial.¹⁵

To address these limitations in the current literature, we sought to determine the positive predictive value of PFS for OS through a broad analysis of phase 3 RCTs across all metastatic solid tumor disease sites.

Methods:

We queried the [ClinicalTrials.gov](https://clinicaltrials.gov) database to identify superiority-design oncologic RCTs for patients with metastatic solid tumors. The following search parameters were utilized: Terms: “cancer”; Study Type: “All Studies”; Status: excluded “Not yet recruiting”; Phase: “3”; and Study Results: “With Results.” This search yielded 1,239 trials, which were then screened for metastatic-solid-tumor-specific phase 3 randomized multi-arm trials assessing a therapeutic intervention (Figure 1). Trials were eligible for analysis if the primary endpoint (PEP) evaluated a disease-related outcome (DRO) such as disease control and/or survival. Trials with a DRO PEP that was not PFS or OS (e.g. overall response rate, durable response rate) were also deemed ineligible to limit trial PEP heterogeneity. If a trial had a PFS and OS co-primary PEP, these trials were assorted to the PFS PEP group. Trial screening and abstraction of trial variables were independently performed by a minimum of two individuals. For each trial, the peer-reviewed literature was reviewed through [PubMed.gov](https://pubmed.ncbi.nlm.nih.gov/) to identify publication of trial results.

Data abstraction was finalized December 2019. Statistical analyses included Pearson’s *Chi*-squared testing to compare proportions between groups; independent-samples Mann-Whitney *U*-tests and multiple binary logistic regression modeling were utilized for continuous variables. A multiple binary logistic regression analysis was performed to determine factors associated with trial PEP success; it was decided, *a priori*, that the model would only include factors that met the $p < 0.05$ threshold to limit variable inputs, spurious associations, and dilution of true associations. All statistical tests were performed using SPSS (Version 22.0).⁹

Results:

Two hundred and sixty trials met inclusion criteria as superiority-design trials assessing a therapeutic intervention for patients with a metastatic solid tumor (Figure 1). A minority of trials (128/388, 33.0%) were deemed ineligible due to evaluation of a non-DRO PEP (64/388, 16.5%), non-inferiority trial design (42/388, 10.8%), and lack of OS or PFS DRO PEP (22/388, 5.7%; Figure 1). The combined total enrollment for the 260 eligible trials using either PFS (146/282, 56.2%) or OS (43.8%, 114/260) as the PEP was 172,133 patients, with years of enrollment initiation ranging from 1999 to 2015.

Comparative characteristics of PFS and OS PEP trials are detailed in Table 1. PFS trials enrolled significantly fewer patients than OS trials with a median enrollment of 495 patients compared to 619 patients, respectively ($p=0.03$). Among the 114 OS PEP trials, the median

time from start of accrual to primary completion date, defined as the final data collection date for primary outcome measure, was 39 months (interquartile range: 32–54 months); comparatively, among the 146 PFS PEP trials, the median time from start of accrual to primary completion date was 36 months (interquartile range: 27–58 months, $p=0.48$). PFS trials were no more likely to be sponsored by industry than OS trials (89.0% vs 92.1%, $p=0.41$). By disease site, breast cancer trials were more likely to utilize PFS as the PEP, whereas genitourinary and gastrointestinal trials were more likely to utilize OS ($p<0.0001$). PFS trials were more likely to include molecular profile restriction criteria, defined by those trials that restricted eligibility to patients with specific tumor molecular characteristics (i.e. – *EGFR* mutation, *ALK* fusion, *BRAF* mutation) (15.1% vs 4.4%, $p=0.005$). Similarly, PFS trials were more likely to test a targeted therapy (such as a monoclonal antibody or small molecular inhibitor) versus a cytotoxic chemotherapy than OS trials (80.1% vs 67.5%, $p=0.048$). There was no difference between PFS and OS trials testing a first-line therapy or second-line and-higher therapy ($p=0.66$). Notably for trials that reported PEP results ($n=235$), PFS trials succeeded in reaching statistically-significant benefit for the PEP (as defined by each trial) markedly more often than OS trials (66.9% vs 33.3%, $p<0.0001$).

Among those trials with reported PEP results ($n=235$), these were pooled to determine predictors of trial success as detailed in Table 2. We found that cooperative group trials were less likely to succeed compared to studies not sponsored by cooperative groups (30.6% vs 56.3%, $p=0.004$), while the opposite trend was seen for industry-sponsored studies (53.8% vs 31.3%, $p=0.08$). Molecular profile-restricted trials, on the other hand, were more likely to succeed than unrestricted trials (84.6% vs 48.3%, $p<0.0001$). Trials testing an investigational agent alone rather than in combination with other therapies were also more likely to succeed (64.5% vs 43.9%, $p=0.002$). Considering only PFS PEP trials, the method of PFS assessment was determined for each trial; we found a trend toward higher rates of success amongst PFS trials that defined PFS based on local investigator assessment versus independent central review (81.0% vs 65.5%, $p=0.09$). Similarly, total trial enrollment, disease site, treatment order, and co-primary endpoint were not found to predict trial success. Multiple binary logistic regression demonstrated that trial PFS PEP (vs. OS PEP; odds ratio [OR] of 3.6, 95% confidence interval [CI] 3.0–4.2, $p<0.0001$), use of molecular profile restriction (yes vs. no; OR of 4.1, 95% CI 2.9–5.3, $p=0.02$), and use of combination oncotherapeutics (vs. single agent; OR of 2.1, 95% CI 1.5–2.7, $p=0.02$) remained independently associated with trial success, whereas cooperative group sponsorship status did not (yes vs. no; OR of 2.1, 95% CI 1.2–2.9, $p=0.08$).

Consequently, we evaluated the association of successful PFS PEP trials and subsequent positive OS signal. Out of 146 trials with PFS as the PEP, 133 trials (91.1%) had PEP results that were evaluable; 13 trials were excluded because they either did not have peer-reviewed results published at the time of analysis or their results were published as part of a pooled-analysis rather than the individual trial-level data (Figure 2). Among the 133 trials, 82 trials (61.7%) successfully met the PFS PEP and also reported on any OS results either as a co-PEP or secondary endpoint (Figure 2). Out of 82 trials, 26 trials (31.7%) reported a statistically significant OS benefit at initial publication, and 31 trials (37.8%) reported a statistically significant OS benefit at any time point (Figure 2). Among the 133 trials, 43 trials (32.3%) did not successfully meet the PFS PEP and also reported on any OS results

(Figure 2). Out of 43 trials, 2 trials (4.7%) reported a statistically significant OS benefit (Figure 2).

We next examined whether any trial-specific factors might predict for an OS benefit among the 82 PFS-positive trials (Table 3). Of all trial-specific factors, we only found that industry sponsorship was less likely to predict for OS benefit as compared to studies not sponsored by industry (35.1% vs. 80.0%, $p=0.045$). We additionally evaluated the role of crossover following disease progression and the effect of PFS hazard ratio (HR) magnitude. When evaluating crossover, 40 trials (53.3%) allowed crossover from one study arm to another following disease progression; studies allowing crossover had comparable rates of subsequent OS benefit as those that did not allow crossover (40.0% vs 28.6%, $p=0.30$). On an analysis of PFS HR, there was no association with a subsequent OS benefit (median PFS HR 0.58) vs. no OS benefit (median PFS HR 0.60, $p=0.70$).

Finally, given the possible effects of insufficient trial power, an alternative method of evaluating OS benefit among PFS PEP trials was performed by analyzing OS HRs and PFS HRs. When plotting log of PFS HRs vs. log of OS HRs among PFS-positive PEP trials, linear regression revealed a weak correlation (correlation coefficient $r=0.44$). An additional analysis of trials that met an OS HR threshold of 0.80 was observed in 34% ($n=27/79$) of PFS-positive PEP trials vs. 17% ($n=6/36$) of PFS-negative PEP trials ($p=0.052$). Among PFS-positive PEP trials, binary logistic regression revealed that PFS HR was associated with OS HR 0.80 ($p=0.01$).

Discussion:

Among phase 3 oncologic RCTs for patients with metastatic solid tumors, PFS is a suboptimal predictor for OS. A positive PFS signal with a subsequent OS benefit was observed in only 31 of 82 trials (positive predictive value of 38%). This is comparable to the rate of success for all trials with OS as the PEP, where an OS benefit was observed for 33% of trials. Apart from lack of industry sponsorship, we identified no discernable pattern to which PFS-positive trials have a higher likelihood of OS benefit. Altogether, it appears PFS functions best as a screen for an OS signal (sensitivity 94%, specificity 45%, negative predictive value 95%), which is logical given its development as a tool to detect therapeutic activity in early phase 1/2 trials rather than a clinically-meaningful endpoint in and of itself. To our knowledge, this represents the first report comprehensively assessing trial factors and predictive value of PFS for OS in the metastatic cancer setting.

Notably, the present study provides new insights into the characteristics and predictors of trial success in phase 3 RCTs addressing patients with metastatic disease. Particularly when evaluating factors between PFS PEP trials and OS PEP trials, we observed that PFS PEP trials were more likely to restrict patient eligibility using molecular criteria, utilize targeted therapy, accrue lower numbers of patients, successfully meet the PEP, and enroll in certain disease sites. In the era of precision medicine, patient screening and selection are fundamental in identifying these subgroups of patients who may derive the most benefit. The reasons for the observed association between PFS PEP trials and both molecular profile restriction as well as assessment of targeted agents are unclear. One explanation for this

association is that PFS allows for quantification of disease response to targeted therapy or in a molecular-subtype-restricted population while reserving cytotoxic therapy and/or crossover as a salvage treatment. Alternatively, there may be a component of convenience or of commercial interest given PFS trials enroll fewer patients and are more likely to successfully meet their PEP, likely a result of the shorter median time to an event and larger number of events that occur when compared to OS PEP trials.¹⁶ As such, novel targeted agents may be set up for success through the use of PFS as the PEP. This is further supported by our multivariable analysis that demonstrated PFS, molecular profile restriction, and the use of single agent therapy were independently associated with PEP success. The observation that patient selection through molecular profile restriction enriches for PEP success may carry implications for continued pursuit of precision-medicine-based efforts. Similarly, there was an association between PEP success and lack of cooperative group sponsorship. Cooperative group clinical trials, especially those in the metastatic setting, are at high risk of low/poor accrual and subsequent lack of successful PEP achievement, secondary to a host of factors including greater trial complexity.¹⁷ Of note among PFS and OS trials with a positive, successful PEP, there was a trend with industry-sponsorship; meanwhile, the only association for PFS-positive trials with a subsequent OS-positive endpoint was lack of industry sponsorship. While the signal is admittedly faint ($p=0.045$), it suggests that industry trials do well to achieve a PFS-positive signal but fail to demonstrate an OS benefit, perhaps due to study design or lack of confirmatory studies.¹⁸ Finally, we observed variability between disease site and utilization of PFS as the PEP—breast trials were more likely to utilize PFS while gastrointestinal, genitourinary, and thoracic trials were more likely to utilize OS. The observed association in genitourinary and thoracic disease sites is expected as there is a growing body of evidence that PFS is a poor predictor of OS in certain malignancies within these broad disease sites.^{11,13} However, we were surprised to note that breast trials were associated with PFS PEP despite only a moderate correlation coefficient (0.69) to OS in previous studies; in the same vein, gastrointestinal cancer trials were less associated with PFS PEP despite of a strong correlation coefficient (0.82) to OS in previous studies evaluating colorectal cancer.^{8,12}

It has previously been suggested that the role of patient crossover in cancer trial design may explain observed trials where PFS improvements do not translate into OS improvements.^{19–21} One prior analysis, which focused exclusively on non-small-cell lung cancer trials, observed that the correlation coefficient between PFS and OS HRs became stronger following exclusion of trials that permitted crossover.²¹ Yet, the absolute correlation coefficient even after exclusion of crossover trials remains at best a ‘low-strength’ correlation ($r = 0.7$) as previously defined by the Institute of Quality and Efficiency in Health Care.²² Other than this disease-site-limited analysis, no prior studies to our knowledge have systemically examined the effect of crossover on the role of oncologic surrogate endpoints. In our series, half of PFS PEP trials that demonstrated a positive PFS signal allowed crossover. The rate of subsequent OS improvements was in fact higher among those studies that allowed crossover (40%) than those that did not (29%), although this difference was not statistically significant. This observation runs counter to the hypothesis that a crossover design may decrease detection of a subsequent OS signal among PFS-positive trials.^{14,21} While it remains possible that at the individual-trial level crossover may

affect conversion of a PFS signal into an OS signal, this pooled trial-level analysis does not demonstrate an effect of crossover on the PFS-OS relationship.

PFS is a heterogeneously-defined endpoint, with progression being assessed by local investigators or by an independent central review, depending on the trial. Prior studies have sought to characterize the differences between investigator-assessed and centrally-determined PFS, with variable results.^{23–26} The degree of concordance between these two forms of PFS remains debated; similarly, whether one form of PFS assessment will systematically exaggerate results relative to the other form of PFS is unclear.^{25,26} Our data demonstrate that PFS trials with investigator-assessed PFS as the PEP have higher rates of trial success (81%). This success rate decreases to 66% among trials with centrally-determined PFS as the PEP. However, trials using either method of PFS assessment had similar predictive values for OS; an OS signal was seen for 31% of investigator-assessed PFS-positive trials and 35% of centrally-assessed PFS-positive trials. Therefore, the basis for defining progression among PFS trials does not appear to impact the likelihood of PFS-to-OS signal conversion. The variable categorization of death in respect to PFS may also introduce heterogeneity that is difficult to account for when comparing trials, including in this effort.

Overall, our results call into question the broad application of PFS at the regulatory level which is timely considering that regulatory agencies are increasingly utilizing surrogate endpoints for decision-making in oncologic RCTs. By some estimates, 57% (39/68) of drug approvals at the European Medicines Agency from 2009 to 2013, and 67% (36/54) of drug approvals at the United States Food and Drug Administration from 2008 to 2012 were based on a surrogate endpoint.^{4,27} Similarly, PFS as a PEP among oncologic RCTs has also increased over several decades, from 2% between 1984 and 1994 to 26% between 2005 and 2009.⁷ The hazard of relying on surrogate endpoints is the lack of evidence that they provide a clinically meaningful outcome as validation, and post-marketing studies are rarely initiated.³ Even when these additional studies are performed and fail to demonstrate an improvement in quality of life or OS, regulatory approval is rarely revoked, with the exception of a few high-profile cases.^{3,28,29}

There are also several shortcomings to using PFS as a surrogate endpoint when considering the patient perspective, especially in non-biomarker driven trials. Patients benefit if material for decision-making is easy to understand. However, PFS is often difficult to interpret in a clinically meaningful way, with widely disparate definitions among clinical research studies, making informed consent potentially challenging.¹⁵ It is also difficult to counsel patients with correlation coefficients alone, on which prior attempts to quantify the association between surrogate endpoints and OS have relied. The present study offers a simpler comparison: only about one-third (38%) of trials showing benefit to PFS result in improvement in OS. As 33% of all trials designed with OS as the PEP succeed, PFS does not appear to add significant value to our interpretation of the meaningfulness of trial interventions.

There are multiple limitations to the current study. First, the analysis includes RCTs that are heterogeneous in nature, from trial design to disease site, thereby making direct comparisons

difficult. With regards to disease site heterogeneity, we chose to do this analysis only among patients with metastatic disease as the clinical course and prognosis tend to be similar (broadly-speaking) in this patient population, thereby limiting the effect of disease site.³⁰ Second, we were unable to accurately characterize the burden of disease as there is a wide range of metastases, including limited soft-tissue oligometastasis versus multiple visceral organ metastases. The third limitation relates to post-progression survival (PPS), which is an important consideration in metastatic cancer trials, as patients with a longer PPS could potentially dilute the PFS treatment effect, whereas patients with a shorter PPS could demonstrate a strong correlation between PFS and OS.³¹ As we did not have direct access to patient-level data, our study cannot directly answer that question. Instead, we provide an indirect measure of first-line therapy versus second-line or higher therapy which did not predict for PEP success or PFS-to-OS conversion. The final limitation to consider is the observation that PFS PEP trials enroll fewer patients than OS PEP trials, and therefore PFS PEP trials may have insufficient power to detect an OS difference. As such, the trials that demonstrate a PFS-to-OS signal may be underpowered to make such a conclusion, while ones that do not demonstrate a signal may be underpowered to detect it in the first place. However, these trials should show preliminary differences in HR signal if not significant, a finding which our analysis did not support given that there was no difference in PFS HRs in trials with or without a statistical OS benefit (HR 0.58 vs 0.60). Similarly, there was a weak association between PFS HR and OS HR (correlation coefficient $r=0.44$) for PFS-positive PEP trials.²² On further analysis of trials meeting an OS HR of 0.80, a threshold that was set *a priori* in keeping with a prior study that examined varying degrees of trial power,³¹ showed a non-significant trend among PFS-positive PEP trials vs. PFS-negative PEP trials (34% vs. 17%). We acknowledge that PFS may serve as a potential signal for OS, especially as PFS HR is associated with achieving an OS HR threshold of 0.80. However, the strength of that signal appears to be weak and should not replace OS as a PEP unless there is robust evidence that PFS may be a reliable surrogate endpoint.

Taken together, our results demonstrate suboptimal predictive value of PFS for OS metastatic solid tumor oncologic RCTs. Combining our findings with recent data suggesting a weak association between PFS benefit and patient-reported quality of life all call into question the predictive value of PFS for improvements to either patient quality or quantity of life.⁸

Acknowledgement:

Dr. Jagsi reports stock ownership and an advisory role in Equity Quotient; personal fees from Amgen and Vizient; grants from the National Institutes of Health, the Doris Duke Foundation, the Greenwall Foundation, and Blue Cross Blue Shield of Michigan.

Dr. Das reports personal fees from Adlai Nortye.

Dr. Fuller reports honoraria from the University of Texas Health Science Center San Antonio and Elekta AB; royalties from Demos Medical Publishing; travel expenses / accommodations from Oregon Health & Science University, Great Baltimore Medical Center, University of Illinois Chicago, Elekta AB, the Translational Research Institute of Australia; grants from the National Science Foundation and National Institutes of Health.

Dr. Thomas reports receiving a stipend as deputy editor of JAMA Oncology.

Dr. V. Subbiah reports funding for clinical trials from Roche/ Genentech, Novartis, Bayer, GlaxoSmithKline, Nanocarrier, Vegenics, Celgene, Northwest Biotherapeutics, Berghealth, Incyte, Fujifilm, Pharmamar, D3, Pfizer, Multivir, Amgen, Abbvie, Alfa-sigma, Agensys, Boston Biomedical, Idera Pharma, Inhibrx, Exelixis, Blueprint medicines, Loxo Oncology, Medimmune, Altum, Dragonfly Therapeutics, Takeda, National Comprehensive Cancer Network, NCI-CTEP and UT MD Anderson Cancer Center, Turning point therapeutics, Boston Pharmaceuticals; for travel from Novartis, Pharmamar, ASCO, ESMO, Helsinn, Incyte; for consultancy from Helsinn, LOXO Oncology/ Eli Lilly, R-Pharma US, INCYTE, QED pharma, Medimmune, Novartis; and for other duties from Medscape.

Role of the funding source: This research was funded in part by the National Institutes of Health / National Cancer Institute Support Grant P30 CA016672.

References

- Garrett SB, Koenig CJ, Trupin L, et al. What advanced cancer patients with limited treatment options know about clinical research: a qualitative study. *Support Care Cancer* 2017; 25(10): 3235–42. [PubMed: 28488050]
- Kay A, Higgins J, Day AG, Meyer RM, Booth CM. Randomized controlled trials in the era of molecular oncology: methodology, biomarkers, and end points. *Ann Oncol* 2012; 23(6): 1646–51. [PubMed: 22048151]
- Kemp R, Prasad V. Surrogate endpoints in oncology: when are they acceptable for regulatory and clinical decisions, and are they currently overused? *BMC Med* 2017; 15(1): 134. [PubMed: 28728605]
- 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): 1992–4. [PubMed: 26502403]
- Robinson AG, Booth CM, Eisenhauer EA. Progression-free survival as an end-point in solid tumours—perspectives from clinical trials and clinical practice. *Eur J Cancer* 2014; 50(13): 2303–8. [PubMed: 24974171]
- Davis S, Tappenden P, Cantrell A. A Review of Studies Examining the Relationship between Progression-Free Survival and Overall Survival in Advanced or Metastatic Cancer. London; 2012.
- Booth CM, Eisenhauer EA. Progression-free survival: meaningful or simply measurable? *J Clin Oncol* 2012; 30(10): 1030–3. [PubMed: 22370321]
- Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol* 2007; 25(33): 5218–24. [PubMed: 18024867]
- Foster NR, Renfro LA, Schild SE, et al. Multitrial Evaluation of Progression-Free Survival as a Surrogate End Point for Overall Survival in First-Line Extensive-Stage Small-Cell Lung Cancer. *J Thorac Oncol* 2015; 10(7): 1099–106. [PubMed: 26134227]
- Oza AM, Castonguay V, Tsoref D, et al. Progression-free survival in advanced ovarian cancer: a Canadian review and expert panel perspective. *Curr Oncol* 2011; 18 Suppl 2: S20–7. [PubMed: 21969808]
- Soria JC, Massard C, Le Chevalier T. Should progression-free survival be the primary measure of efficacy for advanced NSCLC therapy? *Ann Oncol* 2010; 21(12): 2324–32. [PubMed: 20497965]
- 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. [PubMed: 18421050]
- Armstrong AJ, Febbo PG. Using surrogate biomarkers to predict clinical benefit in men with castration-resistant prostate cancer: an update and review of the literature. *Oncologist* 2009; 14(8): 816–27. [PubMed: 19684076]
- 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. [PubMed: 30528804]
- Raphael MJ, Robinson A, Booth CM, et al. The Value of Progression-Free Survival as a Treatment End Point Among Patients With Advanced Cancer: A Systematic Review and Qualitative Assessment of the Literature. *JAMA Oncol* 2019.

16. Saad ED, Buysse M. Statistical controversies in clinical research: end points other than overall survival are vital for regulatory approval of anticancer agents. *Ann Oncol* 2016; 27(3): 373–8. [PubMed: 26578738]
17. Bennette CS, Ramsey SD, McDermott CL, Carlson JJ, Basu A, Veenstra DL. Predicting Low Accrual in the National Cancer Institute's Cooperative Group Clinical Trials. *J Natl Cancer Inst* 2016; 108(2).
18. Naci H, Smalley KR, Kesselheim AS. Characteristics of Preapproval and Postapproval Studies for Drugs Granted Accelerated Approval by the US Food and Drug Administration. *JAMA* 2017; 318(7): 626–36. [PubMed: 28810023]
19. Prasad V Double-crossed: why crossover in clinical trials may be distorting medical science. *J Natl Compr Canc Netw* 2013; 11(5): 625–7. [PubMed: 23667211]
20. Prasad V, Grady C. The misguided ethics of crossover trials. *Contemp Clin Trials* 2014; 37(2): 167–9. [PubMed: 24365533]
21. Hashim M, Pfeiffer BM, Bartsch R, Postma M, Heeg B. Do Surrogate Endpoints Better Correlate with Overall Survival in Studies That Did Not Allow for Crossover or Reported Balanced Postprogression Treatments? An Application in Advanced Non-Small Cell Lung Cancer. *Value Health* 2018; 21(1): 9–17. [PubMed: 29304946]
22. IQWiG. Validity of surrogate endpoints in oncology Executive summary of rapid report A10–05, Version 1.1 Institute for Quality and Efficiency in Health Care: Executive Summaries. Cologne, Germany; 2005.
23. Dodd LE, Korn EL, Freidlin B, et al. Blinded independent central review of progression-free survival in phase III clinical trials: important design element or unnecessary expense? *J Clin Oncol* 2008; 26(22): 3791–6. [PubMed: 18669467]
24. Jones CF, Soto Barrientos JF, Monnickendam G. Investigating discrepancies in assessments of PFS by study investigators and independent review. *Annals of Oncology* 2017; 28(suppl_5).
25. Zhang J, Zhang Y, Tang S, et al. Systematic bias between blinded independent central review and local assessment: literature review and analyses of 76 phase III randomised controlled trials in 45 688 patients with advanced solid tumour. *BMJ Open* 2018; 8(9): e017240.
26. Stone A, GebSKI V, Davidson R, Bloomfield R, Bartlett JW, Sabin A. Exaggeration of PFS by blinded, independent, central review (BICR). *Ann Oncol* 2019; 30(2): 332–8. [PubMed: 30475951]
27. Davis C, Naci H, Gurpinar E, Poplavska E, Pinto A, Aggarwal A. Availability of evidence of benefits on overall survival and quality of life of cancer drugs approved by European Medicines Agency: retrospective cohort study of drug approvals 2009–13. *BMJ* 2017; 359: j4530. [PubMed: 28978555]
28. Prasad V, Cifu A, Ioannidis JP. Reversals of established medical practices: evidence to abandon ship. *JAMA* 2012; 307(1): 37–8. [PubMed: 22215160]
29. Prasad V, Vandross A, Toomey C, et al. A decade of reversal: an analysis of 146 contradicted medical practices. *Mayo Clin Proc* 2013; 88(8): 790–8. [PubMed: 23871230]
30. Dillekas H, Rogers MS, Straume O. Are 90% of deaths from cancer caused by metastases? *Cancer Med* 2019; 8(12): 5574–6. [PubMed: 31397113]
31. Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. *J Natl Cancer Inst* 2009; 101(23): 1642–9. [PubMed: 19903805]

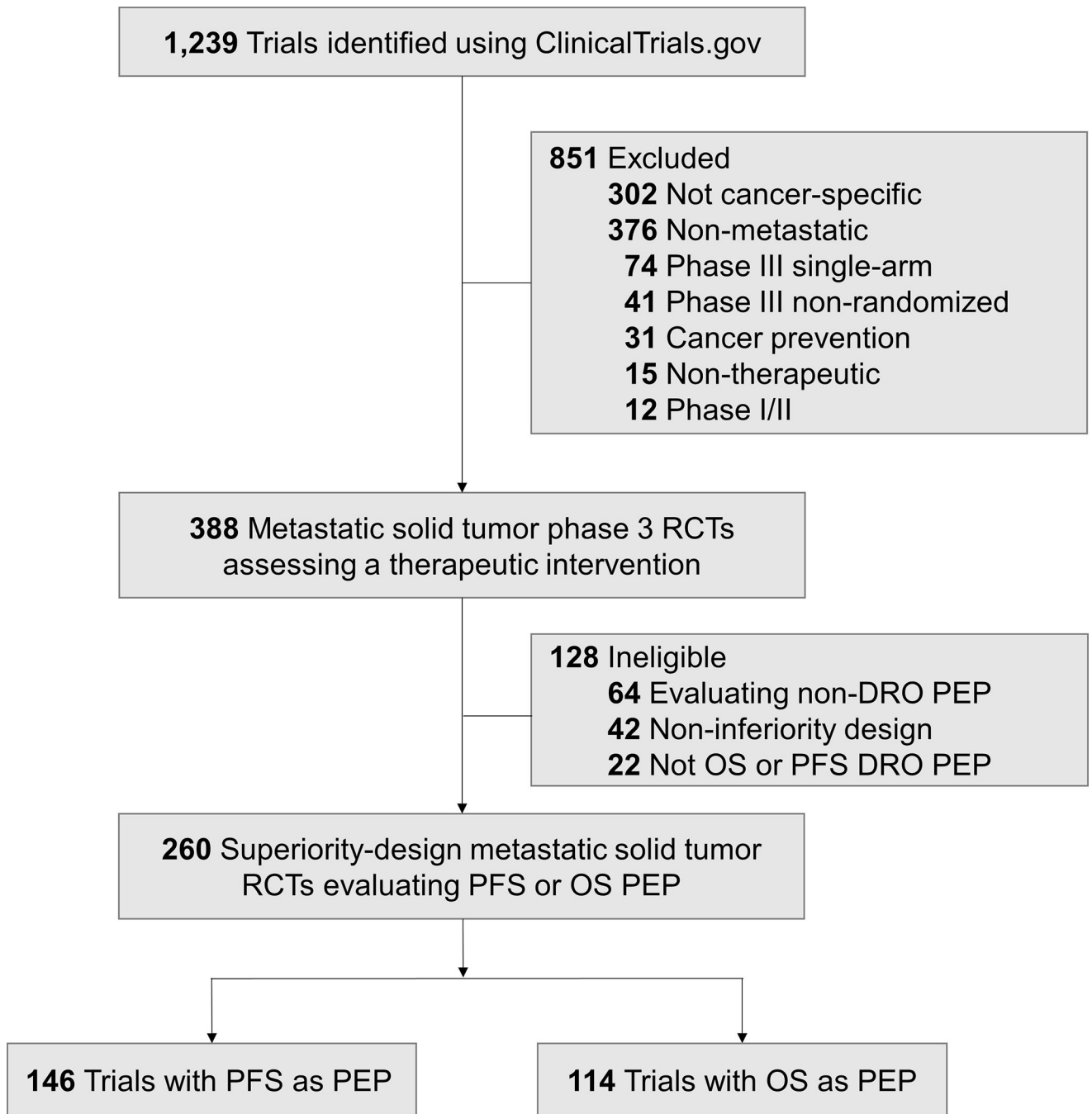


Figure 1. Flowchart of trial screening, eligibility, and inclusion. Disease-related outcomes (DROs) includes outcomes related to disease control and/or survival; non-DRO endpoints include those assessing disease- or treatment-related toxicities, supportive measures, and similar. Abbreviations: RCT, randomized controlled trials; DRO, disease-related outcome; PEP, primary endpoint; PFS, progression-free survival; OS, overall survival.

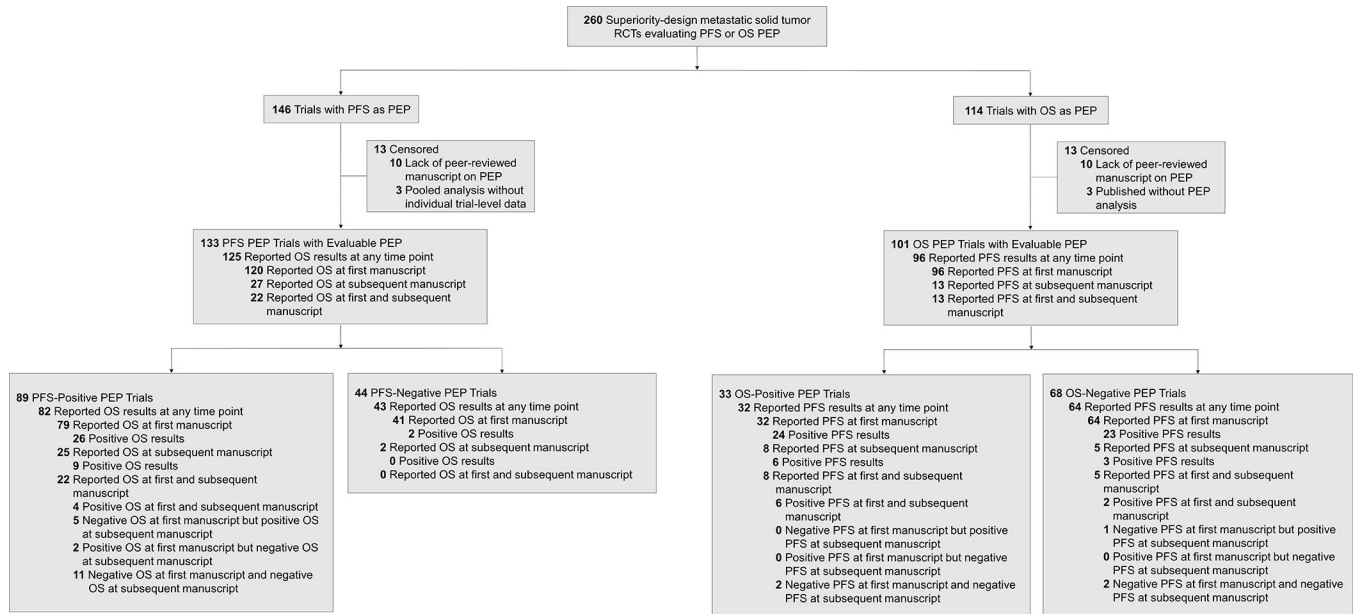


Figure 2. Flowchart of PFS and OS PEP trials with peer-reviewed manuscript reported primary and secondary endpoint results. Among PFS PEP trials, OS represents either a co-PEP or secondary endpoint. Among OS PEP trials, PFS represents a secondary endpoint. Abbreviations: RCT, randomized controlled trials; PEP, primary endpoint; PFS, progression-free survival; OS, overall survival.

Table 1.

Characteristics of PFS and OS Trials

Trial Factor	PFS Trials (N=146) N (%)	OS Trials (N=114) N (%)	p-value*
Total enrollment, median (IQR)	495 patients (302 – 751)	619 patients (416 – 896)	p=0.03
Time between accrual start and primary completion, median (IQR)	36 months (27 – 58)	39 months (32 – 54)	p=0.48
Industry funding of trial	Yes 130 (89.0%)	105 (92.1%)	p=0.41
	No 16 (11.0%)	9 (7.9%)	
Cooperative group trial	Yes 20 (13.7%)	25 (21.9%)	p=0.08
	No 126 (86.3%)	89 (78.1%)	
Disease site	Breast 39 (26.7%)	5 (4.4%)	p<0.0001
	Gastrointestinal 24 (16.4%)	30 (26.3%)	
	Genitourinary 16 (11.0%)	22 (19.3%)	
	Thoracic 36 (24.7%)	38 (33.3%)	
	Other 31 (21.2%)	19 (16.7%)	
Molecular profile restriction ^a	Yes 22 (15.1%)	5 (4.4%)	p=0.005
	No 124 (84.9%)	109 (95.6%)	
Systemic therapy ^b	Targeted therapy 117 (80.1%)	77 (67.5%)	p=0.048
	Cytotoxic chemotherapy 28 (19.2%)	33 (28.9%)	
Treatment order	First-line therapy 77 (52.7%)	57 (50.0%)	p=0.66
	Second-line or higher therapy 69 (47.3%)	57 (50.0%)	
Single-agent or combination ^c	Single-agent 61 (41.8%)	41 (36.0%)	p=0.44
	Combination 84 (57.5%)	69 (60.5%)	
Trial success (PEP met) ^d	Yes 89 (66.9%)	34 (33.3%)	p<0.0001
	No 44 (33.1%)	68 (66.7%)	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

* p-value reflects Pearson's *Chi*-squared test p-values comparing PFS and OS trials for all trial characteristics except total enrollment, for which the Mann-Whitney *U*-test p-value is provided, and time between accrual start and primary completion, for which binary logistic regression p-value is provided.

^a Molecular profile restriction refers to those trials that selected for patients with specific tumor-related mutations. This includes trials selecting for patients with *EGFR* mutation, *ALK* fusion, *BRAF* mutation, and similar.

^b Systemic therapy interventions were divided into targeted therapies (including monoclonal antibodies, small molecule inhibitors, and similar) and cytotoxic chemotherapy. 5 trials (1 PFS PEP and 4 OS PEP) investigated either a purely surgical or nuclear medicine question and were not included in this analysis.

^c Studies tested either a single-agent intervention ("Single-agent") or in combination with other oncotherapeutics ("Combination"); 5 trials (1 PFS PEP and 4 OS PEP) investigated either purely surgical or nuclear medicine questions and were not included in this analysis.

^d Trial success based on earliest peer-reviewed publication of trial PEP results; 25 trials (13 PFS PEP and 12 OS PEP) were excluded because they either did not have peer-reviewed results published at the time of analysis, results were published as part of a pooled-analysis rather than the individual trial-level data, or the publication did not analyze the PEP.

Abbreviations: PFS, progression-free survival; OS, overall survival; N, number of trials; %, percentage of trials; IQR, interquartile range; PEP, primary endpoint.

Table 2.

Predictors of Trial Primary Endpoint Success

Trial Factor		Proportion of Successful Trials (%) [†]	p-value*
Total enrollment, median (IQR)	Trial Success	573 patients (372 – 828)	p=0.82
	Trial Failure	584 patients (331 – 877)	
Industry funding of trial	Yes	118/219 (53.9%)	p=0.08
	No	5/16 (31.3%)	
Cooperative group trial	Yes	11/36 (30.6%)	p=0.004
	No	112/199 (56.3%)	
Disease site	Breast	24/38 (63.2%)	p=0.23
	Gastrointestinal	27/49 (55.1%)	
	Genitourinary	16/34 (47.1%)	
	Thoracic	28/66 (42.4%)	
	Other	28/48 (58.3%)	
Molecular profile restriction ^a	Yes	22/26 (84.6%)	p<0.0001
	No	101/209 (48.3%)	
Systemic therapy ^b	Targeted therapy	100/183 (54.6%)	p=0.14
	Cytotoxic chemotherapy	21/49 (42.9%)	
Treatment order	First-line therapy	57/121 (47.1%)	p=0.10
	Second-line or higher therapy	66/114 (57.9%)	
Single-agent or combination ^c	Single-agent	60/93 (64.5%)	p=0.002
	Combination	61/139 (43.9%)	
Trial PEP	PFS	89/133 (66.9%)	p<0.0001
	OS	34/102 (33.3%)	

Trial Factor		Proportion of Successful Trials (%) [†]	p-value [*]
PFS assessment ^d	Investigator-assessed	34/42 (81.0%)	p=0.09
	Independent review	38/58 (65.5%)	
Co-primary endpoint ^e	Yes	12/18 (66.7%)	p=0.21
	No	111/217 (51.2%)	

[†] Proportion of trials where the PEP was met (trial success) over total number of trials per trial factor.

^{*} p-value reflects Pearson's *Chi*-squared test p-values comparing trial success versus trial failure for all trial characteristics except total enrollment, for which the Mann-Whitney *U*-test p-value is provided.

^a Molecular profile restriction refers to those trials that selected for patients with specific tumor-related mutations. This includes trials selecting for patients with *EGFR* mutation, *ALK* fusion, *BRAF* mutation, and similar.

^b Systemic therapy interventions were divided into targeted therapies (including monoclonal antibodies, small molecule inhibitors, and similar) and cytotoxic chemotherapy; 3 trials (1 PFS PEP and 2 OS PEP) investigated either a purely surgical or nuclear medicine question and were not included in this analysis.

^c Studies tested either a single-agent intervention ("Single-agent") or in combination with other oncotherapeutics ("Combination"); 3 trials (1 PFS PEP and 2 OS PEP) investigated either a purely surgical or nuclear medicine question and were not included in this analysis.

^d PFS assessment (investigator-assessed versus independent central review) provided for trials with PFS as PEP; 33 trials did not specify independent or investigator PFS assessment.

^e Trials which had a co-primary OS and PFS endpoint were designated as PFS primary endpoint trials.

Abbreviations: %, percentage of trials; IQR, interquartile range; PEP, primary endpoint; PFS, progression-free survival; OS, overall survival.

Table 3. Predictors of Overall Survival Benefit following Progression-Free Survival Benefit

Trial Factor	Proportion of OS+ Trials out of PFS+ Trials (%) [†]	p-value [‡]
All PFS+ Trials	31/82 (37.8%)	-
Total enrollment, median (IQR)	553 patients (314 – 766)	p=0.58
	462 patients (346 – 724)	
Industry funding of trial		p=0.045
Yes	27/77 (35.1%)	
No	4/5 (80.0%)	
Cooperative group trial		p=0.77
Yes	3/7 (42.9%)	
No	28/75 (37.3%)	
Disease site		p=0.76
Breast	7/18 (38.9%)	
Gastrointestinal	4/16 (25.0%)	
Genitourinary	3/8 (37.5%)	
Thoracic	8/21 (38.1%)	
Other	9/19 (47.4%)	
Molecular profile restriction ^d		p=0.38
Yes	8/17 (47.1%)	
No	23/65 (35.4%)	
Systemic therapy ^b		p=0.17
Targeted therapy	23/68 (33.8%)	
Cytotoxic chemotherapy	7/13 (53.8%)	
Treatment order		p=0.49
First-line therapy	17/41 (41.5%)	
Second-line or higher therapy	14/41 (34.1%)	
Single-agent or combination ^c		p=0.97
Single-agent	16/43 (37.2%)	
Combination	14/38 (36.8%)	
PFS assessment ^d		p=0.73
Investigator-assessed	9/29 (31.0%)	

Trial Factor	Proportion of OS+ Trials out of PFS+ Trials (%) [†]	p-value [‡]
Independent review	13/37 (35.1%)	
Co-primary endpoint ^e		
Yes	7/11 (63.6%)	p=0.058
No	24/71 (33.8%)	
Crossover after progression ^f		
Allowed	16/40 (40.0%)	p=0.30
Not Allowed	10/35 (28.6%)	
PFS HR, median (IQR) ^g	0.58 (0.43 – 0.72)	p=0.70
No OS benefit	0.60 (0.46 – 0.75)	

[†] Proportion of PFS PEP trials with PFS benefit (PEP met) where OS benefit was found on initial or subsequent publication. PFS benefit defined as significant PFS benefit as specified for each trial.

* p-value reflects Pearson’s *Chi*-squared test p-values except total enrollment and PFS HR, for which the Mann-Whitney *U*-test p-value is provided.

^a Molecular profile restriction refers to those trials that selected for patients with specific tumor-related mutations. This includes trials selecting for patients with *EGFR* mutation, *ALK* fusion, *BRAF* mutation, and similar.

^b Systemic therapy interventions were divided into targeted therapies (including monoclonal antibodies, small molecule inhibitors, and similar) and cytotoxic chemotherapy; 1 trial investigated a purely nuclear medicine question and was not included in this analysis.

^c Studies tested either a single-agent intervention (“Single-agent”) or in combination with other oncotherapeutics (“Combination”); 1 trial investigated a purely nuclear medicine question and was not included in this analysis.

^d PFS assessment (investigator-assessed versus independent central review) provided for trials with PFS as PEP; 16 trials did not specify independent or investigator-assessed PFS assessment and were not included in this analysis.

^e Trials which had a co-primary OS and PFS endpoint were designated as PFS primary endpoint trials.

^f Crossover between study arms following disease progression; 7 trials did not specify crossover and were not included in this analysis.

^g PFS HR reflects the HR for PFS provided at time of initial trial publication; 1 trial did not specify a HR and was not included in this analysis.

Abbreviations: OS, overall survival; PFS, progression-free survival; %, percentage of trials; IQR, interquartile range; HR, hazard ratio; PEP, primary endpoint.