Method

CHALLENGES AND METHODOLOGIES IN USING PROGRESSION FREE SURVIVAL AS A SURROGATE FOR OVERALL SURVIVAL IN ONCOLOGY

Karla Hernandez-Villafuerte
German Cancer Research Center (DKFZ)
k.hernandezvillafuerte@dkfz.de

Alastair Fischer
Office of Health Economics

Nicholas Latimer
ScHARR, University of Sheffield

Objectives: A primary outcome in oncology trials is overall survival (OS). However, to estimate OS accurately requires a sufficient number of patients to have died, which may take a long time. If an alternative end point is sufficiently highly correlated with OS, it can be used as a surrogate. Progression-free survival (PFS) is the surrogate most often used in oncology, but does not always satisfy the correlation conditions for surrogacy. We analyze the methodologies used when extrapolating from PFS to OS.

Methods: Davis et al. previously reviewed the use of surrogate end points in oncology, using papers published between 2001 and 2011. We extend this, reviewing papers published between 2012 and 2016. We also examine the reporting of statistical methods to assess the strength of surrogacy.

Results: The findings from 2012 to 2016 do not differ substantially from those of 2001 to 2011: the same factors are shown to affect the relationship between PFS and OS. The proportion of papers reporting individual patient data (IPD), strongly recommended for full assessment of surrogacy, remains low: 33 percent. A wide range of methods has been used to determine the appropriateness of surrogates. While usually adhering to reporting standards, the standard of scholarship appears sometimes to be questionable and the reporting of results often haphazard.

Conclusions: Standards of analysis and reporting PFS to OS surrogate studies should be improved by increasing the rigor of statistical reporting and by agreeing to a minimum set of reporting guidelines. Moreover, the use of IPD to assess surrogacy should increase.

Keywords: Cancer, Surrogate end points, Overall survival (OS), Progression-free survival (PFS), Postprogression survival (PPS)

Outcomes from clinical and other healthcare trials of most interest to patients and health systems are usually increases in the length and quality of life as a result of treatment. This poses a problem, because to estimate overall survival (OS) sufficiently accurately often requires long-term follow-up. Direct extrapolation of OS to encompass those patients who had not died by the end of the follow-up period may be carried out. Although it is the preferred method of the United Kingdom's National Institute of Health and Care Excellence (NICE), its use is not always satisfactory (2;3). An alternative is to use progression-free survival (PFS) as a surrogate outcome. Trials with an adequate surrogate end point can be shorter and involve fewer patients, and can thus help to bring a new drug or treatment to the market sooner, or allow products to be brought to market where the costs of a trial using OS would not be justified by the expected returns. This is beneficial to patients and health systems, and improves returns to manufacturers.

However, these benefits are achieved at the expense of a less accurate measure of final outcome than would occur by waiting for data such as OS to become available. Thus there is a trade-off between time elapsed from the end of study (excluding follow-up) before information becomes available, and the accuracy of the information about the benefits of treatment. This study analyses the methodologies and challenges faced when using PFS as a surrogate for OS in oncology.

METHODS

Davis et al. (1) conducted a literature review on the use of surrogate end points in oncology up to the end of 2011. They identified 266 articles, using citation searching to identify relevant papers from an initial list of three papers already known to the authors. They said a systematic literature review was not feasible, because an exploratory search returned a very large
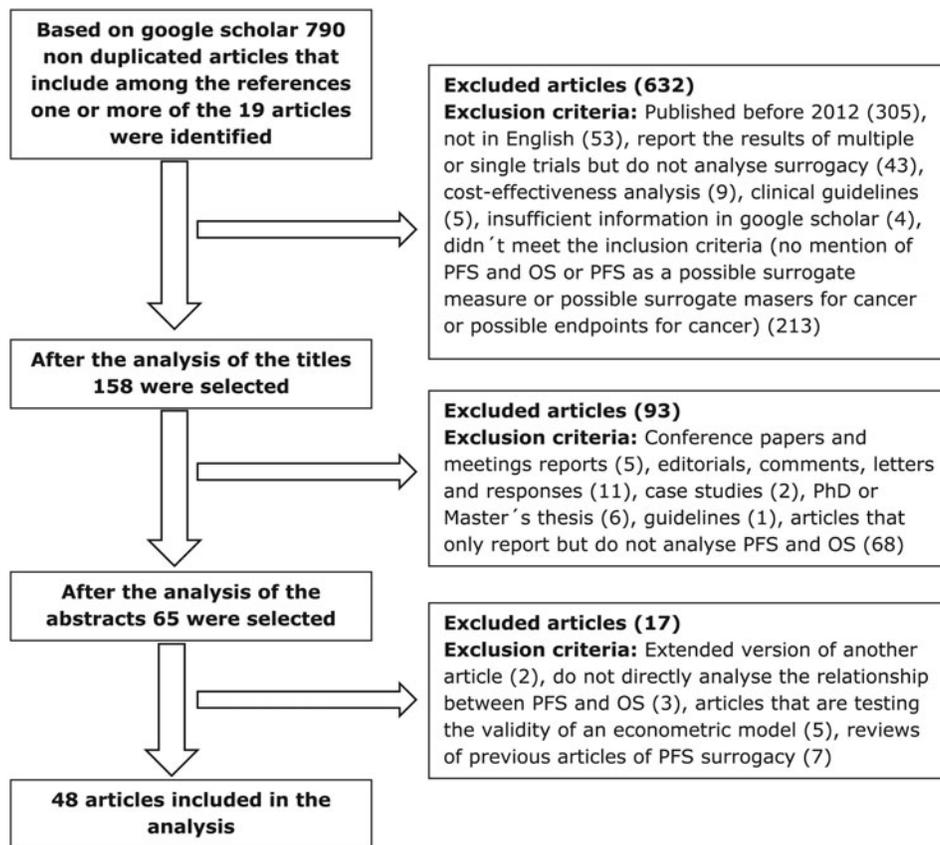
Figure 1. Literature search and selection of articles. OS, overall survival; PFS, progression-free survival; PhD, doctor of philosophy.

number of references (over 3,000), and because any attempt to make the search more specific resulted in many relevant papers being excluded.

Davis et al. (1) included all reviews that examined a statistical relationship between OS and either PFS or time to progression (TTP) and considered any form of treatment where curing the disease was not expected. Nineteen key articles concerning the relationship between PFS/TTP and OS in advanced/metastatic cancer were included.

We updated the review conducted by Davis et al. (1) to 2016 using similar search methods and selection criteria to preserve as much comparability as possible, and using the nineteen papers identified by them as our key papers upon which to base our citation search. We considered only articles in which PFS was mentioned, and excluded those articles that analyzed only TTP. Davis et al. (1) did not include radiographic progression-free survival studies, and so neither did we. A previous follow-up of Davis et al. (1) was carried out by Ciani et al. (4). Our analysis differs from that of Ciani et al. (4) as our aim has additionally been to examine the statistical methodologies most commonly applied as well as the main challenges faced by authors when assessing the validity of PFS as a surrogate.

In August 2016, we conducted a Google Scholar citation search from January 2012 to June 2016, identifying a total of 790 articles which had cited any of the original nineteen key articles identified by Davis et al. (1). We applied four inclusion criteria: (i) mentioned PFS and OS in the title, or (ii) mentioned PFS as a surrogate (including surrogate + outcome/end point/ measure), or (iii) analysis of possible surrogate measures in cancer, or (iv) analyzed end points for cancer. An additional seven articles were excluded because they are reviews of previous studies of PFS surrogacy. After applying a series of exclusion criteria (Figure 1), forty-eight articles were included in the analysis.

The nineteen papers reviewed by Davis et al. (1) and the forty-eight in this study were mostly summaries of many studies. For each study within a study, there was a single aggregate data point pair, the average/median PFS and the average/ median OS. This pair of points was then used as one observation in an analysis that pooled all such pairs from the many studies considered in the study. The sample size for the estimated correlations or regression parameters equals the number of studies considered by the paper. Such data are known as aggregated clinical trial data (ACTD), in which large amounts of data (all the data points of individuals within each trial) have been ignored. In both Davis et al. (1) and our paper, a minority of analyses used unaggregated source data. That is, each patient within the study has a pair of observations which is brought together in a sample, the size of

which is the number of patients, for estimating correlations and regression parameters. Such data are known as individual patient data (IPD). Information regarding methodology, data, and factors affecting the relationship between PFS and OS were extracted. Author affiliation and publication journal were also collected by K.H.V. The quality and accuracy of the extraction was verified by A.F.

RESULTS

The 2012–16 results from this review are very similar to those of Davis et al. (1). Davis et al. (1) usually found a positive correlation between PFS/TTP and OS for individual patients, individual trial arms and the treatment effect between trial arms: only 10.5 percent (2/19) of articles did not support the idea that PFS/TTP could be a useful surrogate for OS. However, the size of the correlation and its statistical significance varied considerably across studies, particularly between cancer types. The authors attributed this variation to the dissimilarities in patient characteristics from study to study, such as tumor type, line of therapy, and diversity of treatment methods.

We classified the results into three groups: (i) papers that explicitly mentioned that PFS is a good surrogate for OS, (ii) those that indicated that PFS could be a good surrogate only under certain conditions, and (iii) those that concluded that PFS is not a good surrogate for OS. Davis et al. (1) (7/12) and our (17/32) analysis both indicate that around 55 percent of the articles using ACTD from multiple trials support PFS surrogacy. Eight of the remaining fifteen articles in our study do not support surrogacy, while seven articles support PFS surrogacy only under particular conditions (e.g., treatment line).

The lack of IPD is evident if we consider that around 35 percent of the articles found both by Davis et al. (1) (7/19) and by this review (16/48) include IPD data (Table 1). Among the ten articles that used solely IPD in our review, four supported surrogacy, five did not and one supported surrogacy for treatments that have a major impact on PFS. Moreover, six of the ten IPD articles in our review were based on information collected in a single Japanese institution, which indicates that the conclusions should be viewed with caution (5–10).

Methodologies and Statistical Results

In analyzing the relationship between PFS and OS, the most usual preference has been for the use of correlation such as Spearman, Pearson, or Kendall's τ (71 percent-34/48) and weighted or unweighted linear regression (73 percent-35/48) (Table 2), comparable to the findings of Davis et al. (1). Moreover, like Davis et al. (1), we found that many different variations in methodology have been applied which makes it difficult to compare the results of studies (e.g., Aboshi et al. (11) and Bria et al. (12), Table 2).

Additionally, of the nine articles that used Pearson correlation, seven support surrogacy and two support surrogacy only under particular factors. However, from the twenty-one articles that used Spearman correlation only seven support surrogacy. This suggests an effect of the correlation test selected. Pearson correlations are affected by outliers more strongly than Spearman correlations; the use of the Pearson correlation without considering outliers could lead to misleading conclusions.

A common practice for testing whether the surrogate is capable of predicting the clinical end point is to analyze the relationship between the actual values of the PFS and OS. Those articles that include ACTD mostly use median PFS and median OS, but in some cases the estimation uses a logarithmic transformation of the variables (27). In the case of IPD articles, the relationship between the actual values per patient of PFS and OS is used to estimate the predictive capacity of PFS.

The effect of treatment in changing PFS to predict the effect of treatment in changing OS is explored throughout the analysis of ACTD. Here it is common to compare hazard ratios (HRs); however, we also identified articles in which the differences in median PFS and OS were examined (13;33;41).

By comparing the distribution of the observations by Surrogacy (Appropriate surrogate [AP]; Depends on particular factors [DPF]; PFS is not an appropriate surrogate [NoAP]) (Table 1) with the distribution of the observations split into two depending on the type of data (ACTD, IPD, or Both), we observe a significant relationship (Fisher's Exact Test 0.039; Pearson's Chi-squared 0.043) between Surrogacy and Type of data, which suggests that the availability of IPD data affects the final conclusion.

Surrogate Threshold Effect

Surrogate threshold effect (STE) is defined as the minimum treatment effect on the surrogate necessary to predict a nonzero effect on the true end point (47). It is normally presented as a HR, for instance, a STE equal to 0.8 means that a PFS HR smaller than 0.8 would need to be observed to predict a less than 1.0 HR for OS. This concept has the advantage of not being a yes or no answer to the question of surrogacy, but a lower bound for PFS that, if achieved, would indicate that a statistically significant effect of the treatment on OS can be predicted. STE has been used relatively more frequently since 2012. In Davis et al. (1), STE was reported in only two of nineteen papers; we identified eleven of forty-eight articles (including five of the six articles that included both IPD and ACTD). However, it is not clear whether STE is affecting authors' final recommendations. For instance, Foster et al. (49) and Shi et al. (50) supported PFS surrogacy with an STE for PFS HR smaller than 0.70 while Ciani et al. (41) rejected surrogacy with an STE of 0.80 (Table 3).

Table 1. Comparison between This Study and Davis et al. (1) Review

	Davis et al. (1) (2001–2011)				Current review (2012–2016)[a]			
	All	ACTD	IPD	Both	All	ACTD	IPD	Both
No. of articles	19	12 (63%)	4 (21%)	3 (16%)	48	32 (67%)	10 (21%)	6 (12%)
No. of patients	193–44,125	4,323–44,125	193–1,296	870–3,953	35–43,459	2,148–43,459	35–2,331	689–16,762
No. of clinical trials	3–191	13–191	3–9	9–11	1–153	6–153	1–7	8–29
Cancer types[b]								
Breast cancer	6	5	0	1	3	2	0	1
Colorectal cancer	7	6	0	1	7	6	0	1
NSCLC	5	4	1	0	10	6	4	0
SCLC/lung	1	0	0	1	4	0	2	2
Renal cell carcinoma	1	0	0	1	6	4	2	0
Gastric cancer	0	0	0	0	6	4	1	1
Multiple myeloma	0	0	0	0	2	2	0	0
Urothelial	0	0	0	0	2	0	1	1
Others	3	1	2	0	7	7	0	0
No particular cancer type	0	1	0	0	1	1	0	0
Surrogacy								
AP	13	7	4	2	25	17	4	4
DPF/no clear	3	3	0	0	11	10	1	0
NoAP	3	2	0	1	12	5	5	2

Note. Source: Authors' search.

[a]Excluding the seven articles that are reviews of previous studies.

[b]Three of the Davis et al. (1) studies include more than one cancer type.

[c]Correlation between PFS and OS or between PFS/TTP when the first was not reported.

ACTD, aggregated clinical trial data; AP, Appropriate surrogate; DPF, Depends on particular factors; IPD, individual patient data; NoAP, PFS is not an appropriate surrogate; NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer

Weaknesses of the Current Approaches

Lack of Rigor in Applying Methodology. Weighted linear regression, the most frequently used method of analysis, is based on assumptions that are not tested in the majority of articles when analyzing surrogacy. In only a few such cases was the type of model mentioned (14). Exceptions are Félix et al. (16) that use the generalized method of moments to control for heteroscedasticity; and Johnson et al. (35) whose results showed unsatisfactory diagnostics with non-normality and heteroscedasticity in the residuals.

Although linear regression assumptions should be so widely known as to be considered irrelevant to report them, their absence leaves many analyses open to the suggestion of failure to handle complications, such as the presence of outliers. Clear outliers are shown in Yoshino et al. (9) and Moriwaki et al. (38). Only 23 percent (11/48) of studies consider or mention outliers. In five of the eleven cases, authors test the sensitivity of the results by applying a "leave-one-out" strategy (26;47;49;52). The six remaining articles test the sensitivity of the results by excluding those trials that are considered outliers (13;16;23;24;32;36).

Publication bias could also have a significant impact on the results, particularly for ACTD. Of the thirty-one articles that include systematic literature reviews, eleven mentioned publication bias as a possible limitation of the study while an additional six articles included a step to overcome the possible bias. Among the six, some articles considered both published and unpublished clinical trials (15;30;40); the others analyzed the extent to which bias represents a problem using Egger's regression test (13;24;41).

Apparently Inconsistent Conclusions. The German Institute of Quality and Efficiency in Health Care (IQWiG) framework considers "low correlation" to be present when the upper limit (95 percent confidence interval) of the correlation is under 0.70 (Pearson R^2 smaller than 0.49) (53). Two of the five articles concerning colorectal cancer that support surrogacy have correlation

Table 2. Studies Included: Methodology and Estimations

Author	Cancer type	Surrogacy	Patients	Clinical trials	Treatment effect corr.	Actual values corr.	Treatment effect R^{2a}	Actual value R^{2a}	Selected particular features of the methodology and the final conclusion
Articles based on ACTD									
Delea et al. (13)	Renal cell carcinoma	AP	10,943	31	For the difference in median: 0.54 (Pe)/For lnHRs: 0.80(Pe)		For difference in median: 0.28/ For lnHRs: 0.63		1) If HRs are not reported, estimated using data from Kaplan–Meier curves/numbers of events/log-rank statistics 2) Median survival estimated by fitting Weibull survival functions to Kaplan–Meier curves
Li et al. (14)	NSCLC	AP	9,903	60				Simple reg.: 0.70/ multivariate reg.: 0.74	1) Multivariate linear regression models 2) Significance of surrogate as a survival marker compared: Area under their ROC curves 3) Discrimination ability tested
Shitara et al. (15)	Gastric	AP	10,484	36	0.80(Sp)	0.70(Sp)			
Félix et al. (16)	Multiple myeloma	AP	22,696	153		0.75(Sp)			1) Two-step approach to a simultaneous Tobit model. Censored normal-weighted regression 2) Heteroscedasticity: generalized method of moments
Giessen et al. (17)	Colorectal	AP	22,736	50			0.76	0.74	
Han et al. (18)	Glioblastoma	AP	7,125	91			0.92	0.70	1) Lead-time that could be gained by using PFS instead of OS
Petrelli and Barni (19)	Colorectal	AP	16,408	34	0.59(Sp)	0.64(Sp)		0.43	
Petrelli et al. (20)	NSCLC	AP	4,176	10	0.64(Sp)	0.37(Sp)[b]		0.00[b]	
Sidhu et al. (21)	Colorectal	AP	20,438	24	0.86(Pe)		0.73		
Beauchemin et al. (22)	Breast	AP	43,459	1,44	0.43(Sp)	0.43(Sp)	0.86		1) Linear regression: only studies with statistically significant difference in both PFS/TTP and OS between treatment arms
Flaherty et al. (23)	Melanoma	AP	4,416	12	0.85(Pe)				1) Fixed and random effects
Singh et al. (24)	Neuroendocrine	AP	2,584	22			0.03	0.22	1) Information regarding clinician perceptions of disease progression was extracted from the clinicians survey

Table 2. Continued

Author	Cancer type	Surrogacy	Patients	Clinical trials	Treatment effect corr.	Actual values corr.	Treatment effect R^{2a}	Actual value R^{2a}	Selected particular features of the methodology and the final conclusion
Bria et al. (12)	Renal cell carcinoma	AP	8,791	19	0.45(Pe)[b]	Targeted treatments: 0.85(Pe)/ Immunotherapy: 0.84(Pe)		Targeted treatments: 0.73/ Immunotherapy: 0.71	1) Monthly PFS and OS rates from month 1 to month 12 extracted from Kaplan–Meier curves 2) Cumulative monthly PFS and OS rates: weighted-average approach 3) Treatment arms were merged according to 3 groups: -targeted agents, -immunotherapy, and -placebo 4) Correlation analysis between 3-month PFS and 9-month OS or 6-month PFS and 12-month OS
Cartier et al. (25)	Multiple myeloma	AP	NA	21	HRs: 0.82(Pe)/ logHRS:0.80 (Pe)		HRs: 0.67/ logHRs:0.63		
Chen et al. (26)	Nasopharyngeal	AP	5,212	21			0.90		1) Errors-in-variables linear regression model with a conservative reliability coefficient of 0.9 (weighted by trial size)
Giessen et al. (27)	Colorectal	AP	10,800	23		0.73(Pe)			
Petrelli et al. (28)	Pancreatic	AP	8,467	30	0.78(Sp)	0.75(Sp)	0.69	0.60	
Amir et al. (29)	Not appropriate	DPF	NA	26	PPS < 12: 0.64 (Pe) PPS > 12: 0.38 (Pe)				**Not appropriate when post-progression is long[c]** 1) Cutoff for median PPS with the greatest discrimination: non-parametric (spline) smooth function applied to the correlation between the ratio of OS to PFS and to PPS
Hotta et al. (30)	NSCLC	DPF	24,158	34			0.69		**Appropriate surrogate when post-study treatments were seldom employed[c]**
Kawakami et al. (31)	Gastric	DPF	11,802	43	0.55(Sp)	All:0.496(Sp)/ OlderTrials:0.689 (Sp)/ RecentTrial 0.282(Sp)			**Correlation between PFS and OS decreases over the years[c]** 1) Dichotomised (trials before and after 2006): effect of year of completion of trial enrolment.

Table 2. Continued

Author	Cancer type	Surrogacy	Patients	Clinical trials	Treatment effect corr.	Actual values corr.	Treatment effect R^{2a}	Actual value R^{2a}	Selected particular features of the methodology and the final conclusion
Petrelli and Barni (32)	Breast	DPF	10,138	20	0.78 (Sp)	0.81 (Sp)	0.73	0.61	**Not appropriate surrogate for HER2-negative disease trials where the weight of PPS is strong**[c]
Adunlin et al. (33)	Colorectal	DPF	0	72	HRs: 0.46 (Sp)/ Difference in median: 0.52 (Sp)		HRs: 0.31/ Difference in median: 0.44		**Appropriate surrogate in 2nd line and beyond**[c] 1) Multivariate regression analysis
Hotta et al. (34)	NSCLC	DPF	7,633	18			0.23		**Appropriate surrogate for molecularly selected patient trials**[c] 1) Influence of trial design (molecularly selected patients vs. all-comers) evaluated: multiple stepwise regression analysis 2) ROC: to identify the most accurate discrimination thresholds
Johnson et al. (35)	Renal cell carcinoma	DPF	10,797	30				EIV: 0.49/ OLS: 0.44	**PFS surrogacy is not generalizable across all drug classes**[c] 1) EIV regression 2) Evaluated the effect of prognostic covariates
Özer-Stillman et al. (36)	Gastrointestinal	DPF	2,189	14		0.72 (Pe)		0.52	**Not appropriate surrogate in first line therapy**[c]
Suzuki et al. (37)	NSCLC	DPF	NA	32	0.41–0.77 (Sp) Depending on PPS time	Medians = 0.37			**Appropriate surrogate with a PPS of less than 6 months**[c] 1) Optimal point of correlation of PFS-HR and OS-HR by every 1 month of PPS: by using a linear regression model
Moriwaki et al. (38)	Biliary tract	DPF	2,148	17				All: 0.66/Target therapy: 0.78/ Gemcitabine-therapies: 0.78	**Appropriate surrogate in phase II trial of a newly developed drug** 1) Trials with gemcitabine-containing therapies and with targeted agents analyzed separately
Petrelli and Barni (39)	Renal cell carcinoma	NoAP	3,188	6	0.36 (Sp)	0.87 (Sp)	0.07		

Table 2. Continued

Author	Cancer type	Surrogacy	Patients	Clinical trials	Treatment effect corr.	Actual values corr.	Treatment effect R^{2a}	Actual value R^{2a}	Selected particular features of the methodology and the final conclusion
Aboshi et al. (11)	NSCLC	NoAP	23,337	65		0.69 (Sp)		0.44	1) Covariates analysed by univariate logistic regression analysis using a fixed-effect model and multivariate logistic regression analysis
Shitara et al. (40)	Gastric	NoAP	4,286	64	0.36 (Sp)	0.56 (Sp)			
Ciani et al. (41)	Colorectal	NoAP	40,243	101	HRs: 0.75(Sp)/ Difference in median: 0.59 (Sp)		HRs: 0.34/ Difference in median: 0.52		1) der Simonian and Laird random-effects univariate meta-analyses used 2) Random-effects multivariate meta-analyses
Terashima et al. (42)	Hepatocellular carcinoma	NoAP	5,803	56				Medians = 0.30	
Articles based on IPD									
Galsky et al. (43)	Urothelial carcinoma	AP	364	7		0.86 (Fleischer)			1) Kaplan and Meier used to estimate the OS of patients stratified by disease progression at 6 or 9 months 2) Correlation estimated using the statistical model for dependence between PFS and OS (64)
Shitara et al. (10)	Gastric	AP	291	NA		0.75 (Sp)			1) Prognostic factors for PPS: uni/multivariate analyses using a Cox proportional hazards model.
Halabi et al. (44)	Renal cell carcinoma	AP	1,381	2		0.53 (Kendall's τ)			1) The Kaplan-Meier product-limit method used to estimate the OS distribution by the PFS rate at 3 months and at 6 months 2) The Cox proportional hazards model was used to assess the significance of the effect of the PFS rate at 3 months and at 6 months in predicting OS 3) Multivariable proportional hazards models 4) Adjusted association: 3.145 (considered the fact that OS is always higher than PFS)

Table 2. Continued

Author	Cancer type	Surrogacy	Patients	Clinical trials	Treatment effect corr.	Actual values corr.	Treatment effect R^{2a}	Actual value R^{2a}	Selected particular features of the methodology and the final conclusion
Négrier et al. (45)	Renal cell carcinoma	AP	750	1					1) Weibull parametric model to failure time data was fitted to determine whether longer PFS was significantly and meaningfully predictive of longer PPS. PFS was significantly predictive of longer PPS ($p < .001$) 2) In a sensitivity analysis by Kaplan–Meier non-parametric method, PPS curves for three approximately equal numbered groups of patients categorised by PFS were compared by log-rank test
Laporte et al. (46)	NSCLC	DPF	2,331	5		0.59 (Kendall's τ)	0.62-centers/ 0.72-strata		**Appropriate for treatments that have a major impact on PFS (risk reduction of at least 50%)[c]** 1) The distributions of PFS and OS were estimated using the Kaplan-Meier method. Treatment groups were compared using a Cox regression model 2) The association between PFS and OS was quantified through a bivariate copula model fitted on IPD
Imai et al. (5)	NSCLC	NoAP	39	NA		0.76 (Sp)		0.50	1) Prognostic factors for PPS: proportional hazards model with a stepwise regression procedure
Yoshino et al. (8)	NSCLC	NoAP	35	NA		0.13 (Sp)[b]		0.00[b]	
Imai et al. (6)	SCLC	NoAP	49	NA		0.58 (Sp)		0.24	
Kasahara et al. (7)	SCLC	NoAP	71	NA		0.46 (Sp)		0.38	
Yoshino et al. (9)	NSCLC	NoAP	58	NA		0.72 (Sp)		0.41	
Articles based on both IPD and ACTD									

Table 2. Continued

Author	Cancer type	Surrogacy	Patients	Clinical trials	Treatment effect corr.	Actual values corr.	Treatment effect R^{2a}	Actual value R^{2a}	Selected particular features of the methodology and the final conclusion
Mauguen et al. (47)	Lung	AP	5,211	29			Trial level: range from 0.89 to 0.97	Individual level: range from 0.77 to 0.85	1) Individual level: bivariate survival model that takes censoring into account 2) Trial level: Weighted linear regression model 3) Correlations between 2-year PFS and 5-year overall survival were assessed
Agarwal et al. (48)	Urothelial	AP	689	10		Trial level: 0.66 (Pe)		Trial level: 0.55	1) Relationship between PFS 6 months/responds rate and OS 12 months 2) Estimate of PFS 6 months: generalized linear mixed models with normal random effects for trial 3) Individual level: Pearson chi-square test with Yates continuity correction. Significantly indicates a relationship between PFS 6 and OS 12. 4) Trial level: weighted linear regression and Pearson correlation 5) A second-line phase III trial used for external validation
Foster et al. (49)	SCLC	AP	2,855	10		Individual level: 0.57 (Kendall's τ)	Trial level: Copula R^2–0.81/WLS R^2–0.77		1) Individual level: bivariate survival model constructed from a Clayton copula with Weibull marginal distributions 2) Trial level: WLS R^2 and Copula R^2 and associated standard errors 3) Use of data from seven new first-line Phase II/III ES-SCLC trials to externally validate findings

PFS as a surrogate for OS in oncology

Table 2. Continued

Author	Cancer type	Surrogacy	Patients	Clinical trials	Treatment effect corr.	Actual values corr.	Treatment effect R^{2a}	Actual value R^{2a}	Selected particular features of the methodology and the final conclusion
Shi et al. (50)	Colorectal	AP	16,762	22		Individual level: 0.51 (Rank Correlation Coef.)	Individual level: 6 months PFS - 0.69/18 months PFS - 0.51		1) Individual level: prognostic value of PFS 6 months and at 1 year assessed by the Cox model by using a land-mark approach. Correlation: bivariate Copula distribution of the two end points 2) Trial level: weighted least squares. Copula R2 was also estimated
Paoletti et al. (51)	Gastric	NoAP	4,069	20		Individual level: 0.85 (Rank Correlation Coef.)	Trial level: 0.61		
Michiels et al. (52)	Breast	NoAP	1,839	8		Individual level: 0.67 (Sp)	Trial level: 0.51	Individual level: 0.45	1) Individual level: bivariate survival model that takes censoring into account 2) Trial level: weighted linear regression model

Note. Source: authors' research.

[a] Weighted linear regression.

[b] Value not significant at 5% confidence level.

[c] Factor which determines the validity of the PFS surrogacy.

AP, Appropriate surrogate; Copula R^2, weighted (by trial size) correlation of the joint copula effects; DPF, Depends on particular factors; EIV, errors in variables; HR, hazard ratio; NoAP, PFS is not an appropiate surrogate; NSCLC, NSCLC, non-small cell lung cancer; OS, overall survival; PFS, progression-free survival; PPS, postprogression survival; ROC, receiver operating characteristic; SCLC, small cell lung cancer; Sp, Spearman; Pe, Pearson; WLS R^2, weighted (by trial size) least squares regression of marginal Cox model effects.

Table 3. Surrogate Threshold Effect

	Cancer type	Type of study	Surrogacy	STE (HR for the effect on PFS)
Mauguen et al. (47)	Lung	Both[a]	AP	From 0.93 to 1.00 depending on therapy
Sidhu et al. (21)	Colorectal	Trial	AP	0.9
Chen et al. (26)	Nasopharyngeal	Trial	AP	PFS vs OS: 0.88/PFS at 3 years vs 5 years OS: STE = 0.84
Foster et al. (49)	SCLC	Both[a]	AP	0.67
Shi et al. (50)	Colorectal	Both[a]	AP	0.57
Laporte et al. (46)	NSCLC	Individual	DPF	0.49-centers/0.53-strata
Johnson et al. (35)	Renal cell carcinoma	Trial	DPF	STE not presented as a HR. All-trials and immunotherapy-only trials failed to demonstrate an STE. A targeted therapy trial needs a PFS difference of at least 3.7 months
Moriwaki et al. (38)	Biliary tract	Trial	DPF	0.83[b]
Paoletti et al. (51)	Gastric	Both[a]	NoAP	0.56
Ciani et al. (41)	Colorectal	Trial	NoAP	0.80
Michiels et al. (52)	Breast	Both[a]	NoAP	0.72

Note. Source: authors' research.

[a]"Both" means ACTD together with IPD.

ACTD, aggregated clinical trial data; AP = Appropriate surrogate; DPF = Depends on particular factors; HR, hazard ratio; IPD, individual patient data; NoAP = PFS is not an appropiate surrogate; NSCLC, non-small cell lung cancer; OS, overall survival; PFS, progression-free survival; SCLC, small cell lung cancer; STE, surrogate threshold effect.

[b]The ratios of the median between the control and experimental arms in each trial were used to summarise treatment effects because the HRs were not always available. STE corresponds to the upper 95% limit and a median OS ratio equal to 1.

values lower than 0.70 (19;50). For renal carcinoma, Delea et al. (13) and Halabi et al. (44) support PFS as a surrogate, despite values of association lower than 0.7 (Table 2).

Petrelli and Barni (20) stand out because of the lack of consistency between supporting PFS surrogacy and having results in which R^2 (equal to 0.00) and correlation values (treatment effect correlation 0.59 and actual values correlation 0.26) are particularly low. They observed a weak correlation between PFS and OS for NSCLC. However, they still supported PFS as a surrogate for OS, a decision that appears to be influenced by the slope of the linear regression. The slope suggests that a 1-month gain in PFS will be linked to 3 weeks' prolongation in OS. However, the reasons for concluding that the surrogacy is supported remain unclear.

Table 2 shows inconsistencies between IPD and aggregate approaches (40;42). Foster et al. (49), based on ACTD, supported surrogacy, while IPD data from the same trial were less conclusive.

Finally, it is possible that different studies analyze PFS surrogacy for a particular cancer type by including a similar list of clinical trials. However, it is not possible to observe whether similar lists of clinical trials could result in different conclusions because the references of the included trials are not mentioned in some of the papers (14).

Challenges for Analyzing PFS as a Surrogate of OS

Based on the variables included as part of sensitivity analyses or that have been included in multivariate analyses in the papers included in this review, we identify a group of factors that could affect the relationship between PFS and OS and appear in at least five studies.

Type of Treatment and/or Therapy. The literature suggests that the relationship between PFS and OS can be different within the same cancer trial depending on the treatment applied or the therapy selected (11–22;26;32;33;35;36;38;40–42;47;49;50).

Treatment Line. In some cases, the analysis cannot validate the surrogacy for first line therapy, as distinct from the second or third line therapy, mainly because postprogression survival obscures the results of the first line treatments. Petrelli and Barni (32), who analyzed twenty first-line clinical trials, proposed that the decreases in correlation between PFS HR and OS HR observed in recent years is likely to be due to the influence of postprogression treatments (14;21;22;25;26;28;33;34;36;40; 41;49;52).

Year of the Trial. The importance of the year in which the clinical trial was conducted or published was explained by the number of drugs available having increased (eleven) and because the criteria applied to measure progression have changed (e.g., RECIST published in 2000 was modified in 2010 to mRECIST (54)) (11;13;15;16;18;19;22;28;31–34;40).

Sub-group of Patients or Tumor Type. As in Davis et al. (1), the results from the validation of PFS as a surrogate point for OS vary

substantially between cancer types ([10];[11];[16];[18];[21];[26];[32];[34];[39];[52]). Six of sixteen articles for lung cancer conclude that PFS is not an appropriate surrogate for OS, and consistency does not improve when we consider the line of therapy and the phase of the clinical trial. This might be related to the fact that the criteria for supporting surrogacy differ considerably between studies. Additionally, it might not be just the observed relationship that is changing between studies. This suggests a need to standardize criteria.

Definition of PFS and Other Measures. Disease progression is often defined differently between clinical trials. This heterogeneity pertains to the period within which patients are evaluated; time intervals between radiological and clinical assessments; and what constitutes patient progression (e.g., variation of the size of the tumor, and tumor characteristics). There are many other forms of heterogeneity, some of which have been mentioned in four sub-sections above. The inclusion of Phase II trials is also likely to increase heterogeneity. This occurs because at most one arm of a dose-determining trial should be featured in a subsequent Phase III trial. Arms in which patients are over-dosed and on average die sooner than those in the control group should not feature in Phase III trials. Their inclusion in papers that measure the correlation between PFS and OS will reduce the estimated correlation.

Twenty two articles mentioned the problem of heterogeneity as a limitation, but did not adjust the methodology in response to the problem. A further seven studies adjusted the methodology ([18];[23];[26];[31]–[33];[42]), two of which included only clinical trials that have the same set of progression criteria (RECIST criteria) ([23];[42]). Three of the seven included variables such as presence of measurable lesions and tumor response in sensitivity analysis ([18];[31];[33]). Finally, two of the seven studies used established definitions to extract the information collected from the clinical trials, regardless of the terminology used by the original authors ([26];[32]).

Additionally, nineteen of the thirty-two studies based on trial data combined PFS and TTP into a single surrogate measure. In addition to progression, PFS includes death as a result of any cause while in the case of TTP the event of interest is only disease progression, although some authors consider, as part of TTP, deaths caused by the disease in question (e.g., Burzykowski et al. [55] identified by Davis et al. [1]). All-cause mortality can dilute the association between PFS/TTP and OS. Nine of the nineteen articles that include PFS and TTP analyze the sensitivity of the results by breaking down the articles into those that measure PFS and those that measure TTP ([13];[15];[22];[25];[31]–[33];[38];[42]). In contrast to what we would expect, Delea et al. ([13]), Petrelli and Barni ([32]), and Shitara et al. ([15]) found that studies that include PFS have a higher correlation with OS compared with studies that include TTP. However, Moriwaki et al.

([38]) found a slightly lower correlation when TTP trials were excluded.

Geographical Context ([11];[14];[15];[28];[33];[40]). A reason given to explain geographical differences in trial results is the variation in comparator (i.e., standard) treatments between Asian and occidental countries. In addition, in advanced gastric cancer, Shitara et al. ([15]) pointed out several differences in tumor characteristics and practice patterns (e.g., surgery and chemotherapy) that have been identified between Asian and occidental countries.

7) Crossover ([13];[23];[30];[33];[41]). Some clinical trials allow crossover to the experimental regimen upon disease progression. This hinders the analysis of the treatment effect on OS. Eighteen of the thirty-two articles that use ACTD mentioned crossover, while six articles considered it during the estimation. For renal cell carcinoma, Delea et al. ([13]) indicate that the link between the effect of the treatment on PFS and the effect of the treatment on OS was stronger in studies that did not allow crossover. In melanoma, Flaherty et al. ([23]) suggested that correlation coefficients for the nine trials without crossover were significant and more than 7 percentage points higher than with crossover. Hotta et al. ([30]), in studying NSCLC, suggest that for clinical trials in which the median proportion of crossover was lower than 1 percent, the association between the HRs of PFS and OS was strong.

Kim and Prasad ([56]), identified by Davis et al. ([1]), evaluated previous publications to assess the strength of the surrogate-survival correlation among cancer drugs approved. They found no significant differences in survival benefit between clinical trials with or without crossover. They suggest that the results are opposed to the commonly shared idea that crossover masks OS benefits, possibly because crossover prevents observation of late toxicity. Contrary to other studies, in an analysis of colorectal cancer trials, Adunlin et al. ([33]) found that among crossover trials the strength of the association between PFS and OS was higher.

Characteristics of Postprogression Survival

Characteristics of postprogression survival (treatment line 1st/2nd/3rd, year of the clinical trial, crossover between control and treatment arms, newly diagnosed vs recurrent, and subgroup of patients) and the fact that an important number of articles analysed postprogression survival (PPS) together with PFS suggest that PPS has a role in the discussion of the validation of PFS as a surrogate of OS. Amir et al. ([29]) indicate that when PPS is short, the correlation between OS and PFS is higher than when PPS is long.

For patients with advanced NSCLC, Suzuki et al. ([37]) identified the optimal point of correlation of the HR for PFS and the HR for OS by analyzing every 1 month of PPS. They found that the correlation between the HR for PFS and for OS increases for

a PPS of less than 6 months and then decreases (<4 months 0.70; <6 months 0.77; <9 months 0.46). From the sixteen articles that analysed PPS, thirteen suggest that the relationship between OS and PPS is stronger than between OS and PFS and one of the remaining three pointed out a high correlation between PFS and PPS.

A group of Japanese researchers specializing in a study of factors that affect the relationship between PPS and OS (5–10) suggests that the significant factors to explain the effect of PPS on OS are: (i) number of regimens used after progression, (ii) response to the second or third-line treatment, (iii) performance status at progression, (iv) PFS of first line chemotherapy, (v) tumor stage after initial treatment, (vi) presence of distant metastases at recurrence.

DISCUSSION

The percentage of articles that conclude that PFS is an appropriate surrogate for OS (52 percent) is higher than the percentage of those that do not support surrogacy (25 percent). An additional 23 percent of the samples suggest that surrogacy depends on factors such as the length of the PPS and whether the treatment was first or subsequent line. In such a complicated area, it is no wonder that simple rules of thumb, to determine whether a surrogate end point can replace OS, will not work in all situations. Additionally, it seems that different investigators use different rules of thumb in the same circumstances.

The first set of criteria to establish whether a surrogate would be an adequate replacement for OS was proposed by Prentice (57). His criteria have been amended and elaborated since then. Ciani et al. (4) summarized three different frameworks that are currently applied to validate the strength of the evidence (IQWiG, Biomarker-Surrogacy Evaluation Schema, and Elston and Taylor's frameworks). All include the Prentice (57) criteria, but also analyze factors that could influence the strength of the relationship, such as the quality of the data and characteristics of the clinical trial.

Ciani et al. (58) highlight three further conditions. First, the strength of the association between the surrogate end point and the final outcome should be measured through approaches such as regression and meta-analysis. Second, it is necessary that the effect on the final outcome can be predicted and quantified based on the effect on the surrogate. The effect of the treatment on PFS must be large enough to predict an improvement in OS. Third, the level of evidence supporting the relationship between the surrogate end point and the desired outcome needs to be considered. A strong correlation should be observed between the surrogate and the end point based on individual patient data as well as between the treatment effect on the surrogate end point and the final outcomes across multiple randomized trials.

Similarly, Buyse et al. (59) propose that to validate a surrogate end point, it is necessary to analyze both individual and trial level data. ACTD is important for testing the relationship between the treatment effect on PFS and the treatment effect on OS while the IPD allows the analysis of the relationship between the actual value of PFS and the actual value of OS. Thus, despite the existence of recognized surrogate validation criteria, which have been developed over time, no consistent application of these criteria was observed in the studies we reviewed; different authors made different, often arbitrary assessments of surrogate adequacy.

The lack of any substantial proportionate increase in the number of articles including IPD data between Davis et al. (1) study and our analysis suggests the rate of progress in this field is being hampered by an unwillingness by most pharmaceutical companies to provide IPD data to independent and well-qualified researchers for analysis or even to report analyses based on IPD when those data have routinely been collected. Nevertheless, some progress in the topic has been observed, for example, firms have joined recent initiatives such as clinicalstudydatarequest.com or project data sphere that allows researchers to analyze pooled IPD data sets.

The existence of heterogeneity in the definition of progression among clinical trials and a lack of clear information in the clinical trial reports as to how disease progression was evaluated (15;17;40) indicate that there is a need to standardise clinical trial protocols to provide comparability between trials for the same cancer type.

Finally, our search process found additional evidence that went beyond the scope of Davis et al. (1) analysis. Stevens et al. (60) outlined an economic approach which bears on both clinical effectiveness and on cost-effectiveness, suggesting a framework for factoring the use of surrogates into the decision-making process. Perhaps surprisingly, the benefits (or costs) of earlier adoption of a new technology that the use of a surrogate end point will usually allow are not taken into consideration when assessing new treatments. The longer the lag between the results of a trial using a surrogate end point rather than OS, the greater the additional benefits of using a surrogate should be, provided the surrogate is valid and that subsequent treatments do not act as confounders. However, as Davis et al. (1) explain, even when strong consistent evidence supporting a correlation between the treatment effects is available, it is unclear how that should be converted into a quantified relationship between PFS and OS treatment effects within a cost-effectiveness model.

CONCLUSION

The analysis strongly suggests that the use of IPD to assess surrogacy should increase. A case could be made for release of all IPD as a condition of publication. As in Davis et al. (1), our findings show that the availability of such information has been limited, though recent data-sharing initiatives may be changing that.

There is a high variation in the characteristics of the methodologies and little apparent consistency in what should be considered appropriate statistical estimation methodology. Thus the need for standardization that allows for more consistent results. Standardization, in the form of adhering to common definitions, statistical techniques and a checklist of necessary items in reporting results, would often be virtually costless. This could facilitate the use of PFS by policy makers, if it were deemed appropriate, based upon standardized validation methodology, and could increase both the speed and accuracy of their decision making.

Many of the factors that affect the validation of surrogacy are related to the length and characteristics of postprogression survival. Procedures for gathering information on factors affecting the postprogression management of a disease should be described in protocols for following-up clinical trial patients, making it possible to derive stronger conclusions from statistical analysis.

Some limitations of the study need to be mentioned. First, it is not a full literature review. We conducted a citation search based on Davis et al. (1) nineteen studies that we assume captured all relevant articles. Discussions with experts and comparisons with previous systematic reviews suggest that no relevant article has been excluded from the analysis. Second, this is also not a systematic literature review of any particular cancer type. Therefore, analyzing whether PFS should or should not be used in any particular case was outside of the scope of this analysis. It is recommended that the factors that affect the relationship between PFS and OS by cancer type should be analyzed to understand the particular challenges faced in each case. Third, for pragmatic reasons, our exclusion of TTP ignores the possibility that the names of TTP and PFS have in error been used interchangeably (61).

Finally, in addition to using one of the frameworks promoted by Ciani et al. (4) to ensure a higher standard of validation of the strength of evidence, both researchers and policy makers in an area that makes use of surrogate end points need to be aware that the statistical methodology must be properly understood and documented. The importance that validating PFS as a surrogate for OS may have on allowing patients to access new health technologies more quickly should not be undermined by a poor knowledge of the methodology applied. The results of this study are broadly in line with those of Kemp and Prasad (62) who have concluded that the use of surrogate outcomes should be limited to situations where a surrogate has demonstrated robust ability to predict meaningful benefits, or where cases are dire, rare, or with few treatment options.

CONFLICTS OF INTEREST

REFERENCES

1. 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: National Institute for Health and Care Excellence (NICE), 2012.
2. Latimer NR. Survival analysis for economic evaluations alongside clinical trials: Extrapolation with patient-level data. *Med Decis Making*. 2013;33:743-754.
3. Latimer NR, Abrams KR, Lambert PC, et al. Adjusting survival time estimates to account for treatment switching in randomized controlled trials: An economic evaluation context. *Med Decis Making*. 2014;34:387-402.
4. Ciani O, Davis S, Tappenden P, et al. Validation of surrogate endpoints in advanced solid tumors: Systematic review of statistical methods, results, and implications for policy makers. *Int J Technol Assess Health Care*. 2014;30:312-324.
5. Imai H, Mori K, Ono A, et al. Individual-level data on the relationships of progression-free survival and post-progression survival with overall survival in patients with advanced non-squamous non-small cell lung cancer patients who received second-line chemotherapy. *Med Oncol*. 2014;31:1-7.
6. Imai H, Mori K, Wakuda K, et al. Progression-free survival, post-progression survival, and tumor response as surrogate markers for overall survival in patients with extensive small cell lung cancer. *Ann Thorac Med*. 2015;10:61-66.
7. Kasahara N, Imai H, Kaira K, et al. Clinical impact of post-progression survival on overall survival in patients with limited-stage disease small cell lung cancer after first-line chemoradiotherapy. *Radiol Oncol*. 2015;49:409-415.
8. Yoshino R, Imai H, Mori K, et al. Surrogate endpoints for overall survival in advanced non-small-cell lung cancer patients with mutations of the epidermal growth factor receptor gene. *Mol Clin Oncol*. 2014;2: 731-736.
9. Yoshino R, Imai H, Mori K, et al. Clinical impact of postprogression survival for overall survival in elderly patients (aged 75 years or older) with advanced nonsmall cell lung cancer. *J Cancer Res Ther*. 2015;11:606.
10. Shitara K, Matsuo K, Muro K, Doi T, Ohtsu A. Progression-free survival and post-progression survival in patients with advanced gastric cancer treated with first-line chemotherapy. *J Cancer Res Clin Oncol*. 2013;139:1383138-9.
11. Aboshi M, Kaneko M, Narukawa M. Factors affecting the association between overall survival and progression-free survival in clinical trials of first-line treatment for patients with advanced non-small cell lung cancer. *J Cancer Res Clin Oncol*. 2014;140:839-848.
12. Bria E, Massari F, Maines F, et al. Progression-free survival as primary endpoint in randomized clinical trials of targeted agents for advanced renal cell carcinoma: Correlation with overall survival, benchmarking and power analysis. *Crit Rev Oncol Hematol*. 2015;93:50-59.
13. Delea TE, Khuu A, Heng DY, Haas T, Soulieres D. Association between treatment effects on disease progression end points and overall survival in clinical studies of patients with metastatic renal cell carcinoma. *Br J Cancer*. 2012;107:1059-1068.

14. Li X, Liu S, Gu H, Wang D. Surrogate end points for survival in the target treatment of advanced non-small-cell lung cancer with gefitinib or erlotinib. *J Cancer Res Clin Oncol*. 2012;138:1963-1969.

15. Shitara K, Ikeda J, Yokota T, et al. Progression-free survival and time to progression as surrogate markers of overall survival in patients with advanced gastric cancer: Analysis of 36 randomized trials. *Invest New Drugs*. 2012;30:1224-1231.

16. Félix J, Aragão F, Almeida JM, et al. Time-dependent endpoints as predictors of overall survival in multiple myeloma. *BMC Cancer*. 2013;13:1-12.

17. Giessen C, Laubender RP, Ankerst DP, et al. Progression-free survival as a surrogate endpoint for median overall survival in metastatic colorectal cancer: Literature-based analysis from 50 randomized first-line trials. *Clin Cancer Res*. 2013;19:225-235.

18. Han K, Ren M, Wick W, et al. Progression-free survival as a surrogate endpoint for overall survival in glioblastoma: A literature-based meta-analysis from 91 trials. *Neuro Oncol*. 2013;16:696-706.

19. Petrelli F, Barni S. Correlation of progression-free and post-progression survival with overall survival in advanced colorectal cancer. *Ann Oncol*. 2013;24:186-192.

20. Petrelli F, Barni S. Is overall survival still the primary endpoint in maintenance non-small-cell lung cancer studies? An analysis of phase III randomised trials. *Transl Lung Cancer Res*. 2013;2:6-13.

21. Sidhu R, Rong A, Dahlberg S. Evaluation of progression-free survival as a surrogate endpoint for survival in chemotherapy and targeted agent metastatic colorectal cancer trials. *Clin Cancer Res*. 2013;19:969-976.

22. Beauchemin C, Cooper D, Lapierre M-È, Yelle L, Lachaine J. Progression-free survival as a potential surrogate for overall survival in metastatic breast cancer. *Onco Targets Ther*. 2014;7:1101-10.

23. Flaherty KT, Hennig M, Lee SJ, et al. Surrogate endpoints for overall survival in metastatic melanoma: A meta-analysis of randomised controlled trials. *Lancet Oncol*. 2014;15:297-304.

24. Singh S, Wang X, Law C. Association between time to disease progression end points and overall survival in patients with neuroendocrine tumors. *Gastrointest Cancer*. 2014;4:103-113.

25. Cartier S, Zhang B, Rosen VM, et al. Relationship between treatment effects on progression-free survival and overall survival in multiple myeloma: A systematic review and meta-analysis of published clinical trial data. *Oncol Res Treat*. 2015;38:88-94.

26. Chen Y-P, Sun Y, Chen L, et al. Surrogate endpoints for overall survival in combined chemotherapy and radiotherapy trials in nasopharyngeal carcinoma: Meta-analysis of randomised controlled trials. *Radiother Oncol*. 2015;116:157-166.

27. Giessen C, Laubender RP, Ankerst DP, et al. Surrogate endpoints in second-line treatment for mCRC: A systematic literature-based analysis from 23 randomised trials. *Acta Oncol*. 2015;54:187-193.

28. Petrelli F, Coinu A, Borgonovo K, Cabiddu M, Barni S. Progression-free survival as surrogate endpoint in advanced pancreatic cancer: Meta-analysis of 30 randomized first-line trials. *Hepatobiliary Pancreat Dis Int*. 2015;14:124-131.

29. Amir E, Seruga B, Kwong R, Tannock IF, Ocaña A. Poor correlation between progression-free and overall survival in modern clinical trials: Are composite endpoints the answer? *Eur J Cancer*. 2012;48:385-388.

30. Hotta K, Suzuki E, Di Maio M, et al. Progression-free survival and overall survival in phase III trials of molecular-targeted agents in advanced non-small-cell lung cancer. *Lung Cancer*. 2013;79:20-26.

31. Kawakami H, Okamoto I, Hayashi H, et al. Postprogression survival for first-line chemotherapy in patients with advanced gastric cancer. *Eur J Cancer*. 2013;49:3003-3009.

32. Petrelli F, Barni S. Surrogate endpoints in metastatic breast cancer treated with targeted therapies: An analysis of the first-line phase III trials. *Med Oncol*. 2014;31:1-8.

33. 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;154:591-608.

34. Hotta K, Kato Y, Leighl N, et al. Magnitude of the benefit of progression-free survival as a potential surrogate marker in phase 3 trials assessing targeted agents in molecularly selected patients with advanced non-small cell lung cancer: Systematic review. *PLoS One*. 2015;10:e0121211.

35. Johnson KR, Liauw W, Lassere MND. Evaluating surrogacy metrics and investigating approval decisions of progression-free survival (PFS) in metastatic renal cell cancer: A systematic review. *Ann Oncol*. 2015;26:485-496.

36. Özer-Stillman I, Strand L, Chang J, Mohamed AF, Tranbarger-Freier KE. Meta-analysis for the association between overall survival and progression-free survival in gastrointestinal stromal tumor. *Clin Cancer Res*. 2015;21:295-302.

37. Suzuki H, Hirashima T, Okamoto N, et al. Relationship between progression-free survival and overall survival in patients with advanced non-small cell lung cancer treated with anticancer agents after first-line treatment failure. *Asia Pac J Clin Oncol*. 2015;11:121-128.

38. Moriwaki T, Yamamoto Y, Gosho M, et al. Correlations of survival with progression-free survival, response rate, and disease control rate in advanced biliary tract cancer: A meta-analysis of randomised trials of first-line chemotherapy. *Br J Cancer*. 2016;114:881-888.

39. Petrelli F, Barni S. Surrogate end points and postprogression survival in renal cell carcinoma: An analysis of first-line trials with targeted therapies. *Clin Genitourin Cancer*. 2013;11:385-389.

40. Shitara K, Matsuo K, Muro K, Doi T, Ohtsu A. Correlation between overall survival and other endpoints in clinical trials of second-line chemotherapy for patients with advanced gastric cancer. *Gastric Cancer*. 2014;17:362-370.

41. Ciani O, Buyse M, Garside R, Peters J, Saad ED, Stein K, et al. Meta-analyses of randomized controlled trials show suboptimal validity of surrogate outcomes for overall survival in advanced colorectal cancer. *J Clin Epidemiol*. 2015;68:833-842.

42. Terashima T, Yamashita T, Takata N, et al. Post-progression survival and progression-free survival in patients with advanced hepatocellular carcinoma treated by sorafenib. *Hepatol Res*. 2015;46:650-656.

43. Galsky MD, Krege S, Lin CC, et al. Relationship between 6-and 9-month progression-free survival and overall survival in patients with metastatic urothelial cancer treated with first-line cisplatin-based chemotherapy. *Cancer*. 2013;119:3020-3026.

44. Halabi S, Rini B, Escudier B, Stadler WM, Small EJ. Progression-free survival as a surrogate endpoint of overall survival in patients with metastatic renal cell carcinoma. *Cancer*. 2014;120:52-60.

45. Négrier S, Bushmakin AG, Cappelleri JC, et al. Assessment of progression-free survival as a surrogate end-point for overall survival in patients with metastatic renal cell carcinoma. *Eur J Cancer*. 2014;50:1766-1771.

46. Laporte S, Squifflet P, Baroux N, et al. Prediction of survival benefits from progression-free survival benefits in advanced non-small-cell lung cancer: Evidence from a meta-analysis of 2334 patients from 5 randomised trials. *BMJ Open*. 2013;3:e001802.

47. Mauguen A, Pignon J-P, Burdett S, et al. Surrogate endpoints for overall survival in chemotherapy and radiotherapy trials in operable and locally advanced lung cancer: A re-analysis of meta-analyses of individual patients' data. *Lancet Oncol*. 2013;14:619-26.

48. Agarwal N, Bellmunt J, Maughan BL, et al. Six-month progression-free survival as the primary endpoint to evaluate the activity of new agents as second-line therapy for advanced urothelial carcinoma. *Clin Genitourin Cancer*. 2014;12:130-137.

49. 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;10:1099-1106.

50. Shi Q, De Gramont A, Grothey A, et al. Individual patient data analysis of progression-free survival versus overall survival as a first-line end point for metastatic colorectal cancer in modern randomized trials: Findings from the analysis and research in cancers of the digestive system database. *J Clin Oncol*. 2015;33:22-28.

51. Paoletti X, Oba K, Bang Y-J, et al. Progression-free survival as a surrogate for overall survival in advanced/recurrent gastric cancer trials: A meta-analysis. *J Natl Cancer Inst*. 2013;105:1667-70.

52. Michiels S, Pugliano L, Marguet S, et al. Progression-free survival as surrogate endpoint for overall survival in clinical trials of HER2-targeted agents in HER2-positive metastatic breast cancer. *Ann Oncol*. 2016;27:1029-1034.

53. IQWiG. Validity of surrogate endpoints in oncology: Executive summary of rapid rep–rt A10-05, Version 1.1. Institute for Quality and Efficiency in Health Care: Executive Summaries. Cologne, Germany: Institute for Quality and Efficiency in Health Care (IQWiG); 2005.

54. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. *Semin Liver Dis*. 2010;30:52-60.

55. 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;26:1987-1992.

56. 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;91:713-725.

57. Prentice RL. Surrogate endpoints in clinical trials: Definition and operational criteria. *Stat Med*. 1989;8:431-440.

58. Ciani O, Buyse M, Drummond M, et al. Use of surrogate end points in healthcare policy: A proposal for adoption of a validation framework. *Nat Rev Drug Discov*. 2016;15:516.

59. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*. 2000;1:49-67.

60. Stevens W, Philipson T, Wu Y, Chen C, Lakdawalla D. A cost-benefit analysis of using evidence of effectiveness in terms of progression free survival in making reimbursement decisions on new cancer therapies. *Forum Health Econ Policy*. 2014;17:21-52.

61. Saad ED, Katz A. Progression-free survival and time to progression as primary end points in advanced breast cancer: Often used, sometimes loosely defined. *Ann Oncol*. 2008;20:460-464.

62. 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;15:134.