

VALIDATION OF SURROGATE ENDPOINTS IN ADVANCED SOLID TUMORS: SYSTEMATIC REVIEW OF STATISTICAL METHODS, RESULTS, AND IMPLICATIONS FOR POLICY MAKERS

Oriana Ciani

Peninsula Technology Assessment Group, University of Exeter Medical School, Veysey Building, Salmon Pool Lane, Exeter, EX2 4SG, UK

CERGAS - Università Commerciale L. Bocconi, via Roentgen, 1 20136 Milan, IT

Sarah Davis, Paul Tappenden

Decision Support Unit, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK

Ruth Garside

European Centre for Environment and Human Health University of Exeter Medical School Knowledge Spa, Royal Cornwall Hospital, Truro, TR1 3HD, UK

Ken Stein

Peninsula Technology Assessment Group, University of Exeter Medical School, Veysey Building, Salmon Pool Lane, Exeter, EX2 4SG, UK

Anna Cantrell

Decision Support Unit, School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK

Everardo D. Saad

Dendrix Research, Rua Joaquim Floriano, 72/24, 04534-000, Sao Paulo, Brazil

Marc Buyse

International Drug Development Institute, Avenue Provinciale, 30, 1340, Louvain-la-Neuve, Belgium

Rod S. Taylor

Peninsula Technology Assessment Group, University of Exeter Medical School, Veysey Building, Salmon Pool Lane, Exeter, EX2 4SG, UK

Objectives: Licensing of, and coverage decisions on, new therapies should rely on evidence from patient-relevant endpoints such as overall survival (OS). Nevertheless, evidence from surrogate endpoints may also be useful, as it may not only expedite the regulatory approval of new therapies but also inform coverage decisions. It is, therefore, essential that candidate surrogate endpoints be properly validated. However, there is no consensus on statistical methods for such validation and on how the evidence thus derived should be applied by policy makers.

Methods: We review current statistical approaches to surrogate-endpoint validation based on meta-analysis in various advanced-tumor settings. We assessed the suitability of two surrogates (progression-free survival [PFS] and time-to-progression [TTP]) using three current validation frameworks: Elston and Taylor's framework, the German Institute of Quality and Efficiency in Health Care's (IQWiG) framework and the Biomarker-Surrogacy Evaluation Schema (BSES3).

Result: A wide variety of statistical methods have been used to assess surrogacy. The strength of the association between the two surrogates and OS was generally low. The level of evidence (observation-level versus treatment-level) available varied considerably by cancer type, by evaluation tools and was not always consistent even within one specific cancer type.

Conclusions: Not in all solid tumors the treatment-level association between PFS or TTP and OS has been investigated. According to IQWiG's framework, only PFS achieved acceptable evidence of surrogacy in metastatic colorectal and ovarian cancer treated with cytotoxic agents. Our study emphasizes the challenges of surrogate-endpoint validation and the importance of building consensus on the development of evaluation frameworks.

Keywords: surrogate outcomes, PFS, TTP, metastatic cancer, validation

Surrogate endpoints are intended to substitute for final patient-relevant endpoints that directly measure how patients feel, function or survive in clinical trials (1). Evidence from surrogate endpoints may not only expedite the regulatory approval of new health technologies but also inform coverage and reimbursement decisions. In the United Kingdom, several recommendations of the National Institute for Health and Care Excellence (NICE) have been based on cost-effectiveness analyses entirely based on treatment effects derived from clinical trials assess-

ing surrogate endpoints (2). Moreover, this type of evidence may still be relied upon even when patient-relevant endpoints are available, for example in clinical trials that have terminated prematurely or for which data on the final endpoint are not fully mature. Nevertheless, relying on evidence from surrogate endpoints poses a serious challenge for decision makers, as several failures of candidate surrogate endpoints have been reported over the last decades (3–5); such failures have arisen not only from discrepancies in the magnitude of treatment

effects between surrogate and final endpoints (6), but also in their directions (5). Hence, in order for policy makers to use a surrogate endpoint with confidence, there must be a process of “surrogate validation.”

The statistical validation of surrogate endpoints has been a major focus of research activity over the last 2 decades (7;8), but no consensus exists with respect to the standards needed to identify valid surrogates. Nevertheless, two key tenets dominate current views on the issue, namely the “correlation” and the “meta-analytic” approaches (9;10). According to these two tenets, the core goal of surrogate validation is to demonstrate a correlation between the surrogate and the final endpoint in the context of a clinical trial as well as between treatment effects on the surrogate and on the final endpoint within the context of a meta-analysis of randomized controlled trials (RCTs) (10). The uptake of surrogate validation methods in technology assessment and coverage decisions is limited (11), a potential explanation being the lack of harmonization of statistical techniques that should be used. Moreover, while decision tools have been proposed to assist policy makers in judging the strength of such validation evidence for a candidate surrogate, there has been little or no empirical testing of these decision tools to date (2;7;12).

Cancer trials are one of the areas in which surrogate endpoints have become most common (13–16). Progression-free survival (PFS), measured as the time from randomization to either documented tumor progression or death, is often used as primary endpoint in RCTs as a surrogate for overall survival (OS) (17). Tumor progression includes radiographic evidence (18–20) and, in some instances, non-radiographic criteria such as “symptomatic progression” or “clinical deterioration” determined by a clear, unequivocal worsening of the symptoms and signs of disease that are not evident on radiographic assessment (21). Some trials use time-to-progression (TTP) rather than PFS, the difference being that in TTP, patients are censored at the time of death with no prior documentation of disease progression. Other surrogate endpoints have been adopted in oncology, the most common being tumor response rate. However, TTP and PFS are more often used in phase III clinical trials (17) and cost-effectiveness analyses of treatments for metastatic solid tumors (22), whilst tumor response rate better served for this purpose in hematologic malignancies (16;23).

In this study, we review current statistical methods of surrogate-endpoint validation that use a meta-analytic framework. In addition, we assess the strength of evidence for PFS and TTP as surrogates for OS and test the application of current surrogate validation decision tools to the evidence base in several advanced solid tumors.

METHODS

Study Identification and Selection

Meta-analyses of RCTs quantifying the statistical association between PFS or TTP and OS in advanced solid tumors were

sought. Conventional literature searches of electronic bibliographic databases returned a large number (>3,000) of references, and attempts to make the search more specific resulted in the exclusion of many of the papers already known to the authors. Therefore, we used a “citation pearl-growing” approach to study identification (24), with backward and forward citation searching from an initial list of six papers known to the authors (25–30). The citation searches were conducted using Medline and the Science Citation Index in March 2012, and forward citation searching up until December 2012. Two reviewers independently screened titles and abstracts. We excluded conference abstracts, letters to the editor, papers reporting results from single trials, meta-analyses reporting treatment effects on PFS or TTP and OS without assessing an association between them, and descriptive reviews (17;31;32). Meta-analyses that focused on oncology treatments with curative intent were also excluded, as PFS and TTP are not relevant endpoints in this case (33;34).

Data Extraction and Analysis

Three levels of data were extracted from included meta-analyses using standardized *pro-formas*: information on the general characteristics of each meta-analysis (authors and date of publication, criteria for inclusion of studies, number and nature of studies included, number of patients included, and type of tumor and interventions considered); details of the statistical methods reported to assess the association between surrogate and final endpoints, the results of these analyses, and each study authors’ conclusions based on the results; and details of the literature search performed to identify included studies. Data were extracted by a single reviewer and checked by a second. Finally, we sought to analyze the suitability of PFS and TTP as surrogates for OS using established surrogate validation frameworks. Three surrogate validation frameworks identified by a recent review of surrogate-endpoint methods were applied to each meta-analysis (35); they are outlined briefly below. To ensure consistency, they were applied to each meta-analysis by a single reviewer and checked by a second reviewer, and discrepancies resolved with involvement of a third reviewer.

Elston and Taylor’s Framework

In 1999, Bucher and colleagues (36) proposed a set of validity criteria to inform the use of an article measuring the effect of an intervention on surrogate endpoints in clinical practice. These criteria were adapted by Elston and Taylor (2) into a three-level evidence hierarchy: Level 1, evidence demonstrating that treatment effects on the surrogate (i.e., the change on the surrogate endpoint of treatment versus control arm) correspond to treatment effects on the final patient-relevant endpoint (from RCTs); Level 2, evidence demonstrating a consistent association between surrogate outcome and final patient-relevant outcome (from at least epidemiological/observational studies); Level 3, evidence of biological plausibility of a relationship

between surrogate and final patient-relevant outcomes (from pathophysiologic studies and/or understanding of the disease process).

German Institute of Quality and Efficiency in Health Care (IQWiG) framework

In 2011, the Institute of Quality and Efficiency in Health Care (IQWiG), an independent health technology assessment (HTA) agency that assesses the benefits and harms of drug and non-drug technologies on behalf of the German Federal Joint Committee and the Federal Ministry of Health, published a framework for the validation of surrogate endpoints in oncology (12). The IQWiG framework proposes that two levels of consideration are required to judge the suitability of a surrogate endpoint in the assessment of cancer therapy: the reliability of the evidence and the strength of evidence for surrogate validation. Reliability is measured as high, limited, moderate, or low on the basis of the following aspects: (i) application of a recognized approach described in the specialized statistical literature, (ii) conduct of analyses to test the robustness and generalizability of results, (iii) systematic compilation of data, (iv) sufficient restriction of indications or degrees of disease severity and of interventions, and (v) clear definitions of the endpoints investigated. The strength-of-evidence criterion considers the degree of correlation of effects on the surrogate and the patient-relevant endpoint according to predefined thresholds (i.e., high correlation, when the lower limit of the 95 percent confidence interval for $R \geq 0.85$; low correlation, when the upper limit of the 95 percent confidence interval for $R \leq 0.7$; and medium correlation otherwise). Depending on the categorization produced by an algorithm that takes into account both levels of consideration, a conclusion about the validity of the surrogate endpoint is drawn and expressed as proof, indication, hint or no proof of an effect on the patient-relevant endpoints as derived from an observed effect on the surrogate endpoint. While the IQWiG framework provides a list of elements that contribute to “reliability”, we needed to introduce a system of scoring that enabled us to categorize this dimension in a reproducible manner (e.g., a “high” score required all contributing elements to be met).

Biomarker-Surrogacy Evaluation Schema (BSES3)

The Biomarker-Surrogacy Evaluation Schema (BSES3) (37) is a revised version of a previous scheme (BSES) (38), proposed in 2010. The BSES3 validation framework consists of four domains: study design, target endpoint, statistical evaluation, and generalizability. Details of the elements that comprise these domains are shown in the online data supplement (Supplementary Table 1 which can be viewed online at <http://dx.doi.org/10.1017/S0266462314000300>). Each domain is ranked from 0 to 3 and combined to determine an overall score (ranging from 0 to 12). A hierarchical scale of validity is attached to the overall score, with “A” corresponding to highest validity (i.e., overall score 12) and “F-” to lowest. The developers suggest that an overall

Table 1. Summary of the Characteristics of Included Meta-analyses, $N = 31$

	Frequency (%)
<i>Type of meta-analysis</i>	
Aggregate data	18 (58)
IPD	13 (42)
<i>Number of trials, Median (IQR)</i>	
Aggregate data	28 (9 – 51)
IPD	39 (31 – 67)
<i>Number of patients, Median (IQR)</i>	
Aggregate data	4138 (1167 – 15262)
IPD	15850 (10714 – 23492)
<i>Advanced tumor types</i>	
Colorectal cancer	12 (39)
Ovarian cancer	7 (23)
Breast cancer	8 (26)
Lung cancer	9 (29)
Renal cell carcinoma	3 (10)
Prostate cancer	2 (6)
Glioblastoma multiforme	2 (6)
Gastric cancer	1 (3)
Head and neck cancer	1 (3)
Pancreatic cancer	1 (3)
<i>Number of tumor types examined</i>	
1	24 (77)
2	4 (13)
> 2	3 (10)
<i>Surrogate endpoint considered</i>	
PFS	15 (48)
TTP	6 (19)
PFS and TTP ^a	3 (10)
PFS/TTP ^b	5 (16)
PFS and TTP and PFS/TTP	2 (6)

IPD, individual patient data; NSCLC, non-small-cell lung cancer; PFS, progression-free survival; TTP, time to progression.

^aPFS and TTP analyzed as two distinct endpoints.

^bPFS and TTP analyzed as single endpoint.

score of 9 or above, equivalent to a category of “A” or “B,” is required to identify a good level of evidence of surrogate validation.

RESULTS

Characteristics of Included Meta-analyses

Of the 758 papers identified by citation searching, thirty-one publications were included. Figure 1 summarizes the selection process, whereas Table 1 presents a summary of the characteristics of the included meta-analyses. Details for each meta-analysis are provided in Supplementary

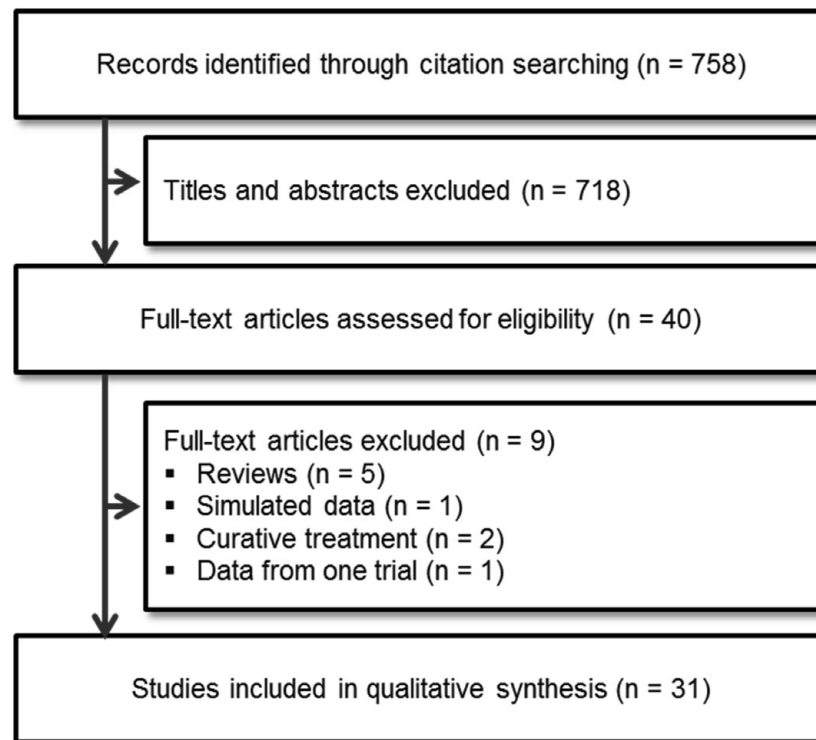


Figure 1. Process of screening and identification of included meta-analyses.

Table 2, which can be viewed online at <http://dx.doi.org/10.1017/S0266462314000300>. The majority of them ($N = 24$, 77 percent) restricted their analyses to a single tumor type, although some reported separate analyses for two (26;29;39;40), or more tumor types (25;41;42). Two meta-analyses (43;44) of patients with glioblastoma multiforme were included; the poor median survival and the fact that metastases are seldom found in this disease suggest that PFS and OS would be important endpoints. The most frequent tumor types examined were colorectal cancer (25–29;39–42;45–47) non–small-cell lung cancer (NSCLC) (25;29;41;48–52), breast cancer (25,30,39,41,42,53–55), and ovarian cancer (9,26,40–42,56,57).

Eighteen meta-analyses were based on aggregate data, while thirteen used individual patient data (IPD). In the aggregate-data meta-analyses, the number of included trials per meta-analysis ranged from 13 to 191 (median, 39) and the number of patients per meta-analysis ranged from approximately 4,300 to 44,000 (median, 15,850). For IPD meta-analyses, these numbers were lower, ranging from two to 27 trials (median, four) and 193 to 3,953 patients (median 1,158). Aggregate-data meta-analyses frequently reported using a systematic literature search to identify included studies (15/18, 84 percent) whereas none of the IPD meta-analyses stated so. The criteria used to select included trials varied markedly across meta-analyses. The scope of meta-analyses were determined by type of intervention (e.g., gefitinib or erlotinib monotherapy) (49), line of therapy (e.g., first-line) (29;45;46;48;52), or other trial characteristics (e.g., sample size) (27;45;46).

Statistical Methods to Assess the Association between Surrogate and Final Endpoints

A wide variety of differing methods to examine the association between surrogate and final endpoints were used across the thirty-one meta-analyses. Two broad criteria may be used to summarize these statistical methods. The first criterion is the type of meta-analysis, as noted above (meta-analyses using aggregate data and those using IPD). The second criterion is the level of association reported: ten meta-analyses (32 percent) reported on the “observation-level association” or Level-2 evidence (2) or “individual-level surrogacy” (10), i.e., the association between surrogate and final endpoints regardless of the treatment effect on each of the endpoints; twelve meta-analyses (39 percent) reported the “treatment-level association” or Level-1 evidence (2) or “trial-level surrogacy” (10), that is, the association between the treatment effect on the surrogate and the treatment effect on the final endpoint; and nine studies (29 percent) reported both levels of association. Combining these two criteria allowed for four core categorizations of the assessment and reporting of the association between PFS/TTP and OS: (i) meta-analyses that reported an observation-level association based on aggregate data (27;45;46;49;51;52;58), for example, single-arm median PFS/TTP versus median OS; (ii) meta-analyses that reported an observation-level association based on IPD (9;28;40;43;44;47;50;53;56;59–61), for example, patients’ TTP versus survival time; (iii) meta-analyses that reported a treatment-level association using aggregate data (25;27;29;30;39;41;42;46;48;54;55;57;58;62), for

example, hazard ratio (HR) for PFS/TTP versus HR for OS; and (iv) meta-analyses that reported a treatment-level association using IPD (9;26;28;40;47;53;59). An overview of the statistical methods used presented according to these four categorizations is provided in a supplementary technical note, with further details shown in Supplementary Tables 2 and 3, which can be viewed online at <http://dx.doi.org/10.1017/S0266462314000300>.

Assessment of the Validity of PFS and TTP as Surrogates for OS

The main results of meta-analyses on the potential role of PFS or TTP as surrogates for OS are presented in Tables 2 and 3, respectively. The validity of these candidate surrogates was assessed according to the Elston and Taylor's, IQWiG, and BSES3 frameworks applied to each meta-analysis, grouped according to the tumor type. An extract of the original authors' conclusions on the surrogacy of PFS or TTP is also presented for each meta-analysis. The four most frequently evaluated advanced solid tumors were colorectal cancer, NSCLC, breast cancer, and ovarian cancer. While the available evidence consistently shows an association between treatment effects on PFS or TTP and treatment effect on OS (i.e., Level 1 evidence according to Elston and Taylor's framework) in metastatic colorectal cancer, the validity of these surrogate measures appear relatively low when rated by both the IQWiG and BSES3 frameworks (Tables 2 and 3). However, four studies (26–28;47) provide an “indication” of an effect on the final endpoint given the effect observed on PFS, according to the IQWiG framework (Table 2). Nevertheless, as these analyses were limited to trials within a specific treatment setting (i.e., the comparison of fluorouracil (FU) plus leucovorin with either FU alone or with raltitrexed) (47) and did not provide evidence across different risk populations and drug-class mechanisms, they were scored down on the BSES3 framework. For advanced lung cancer, three meta-analyses (49,50,61) only reported observation-level association between PFS and OS in NSCLC; in small-cell lung cancer, Foster et al. (59) reported high correlation ($R^2_{\text{trial}} = 0.79$) between HR observed on PFS and OS (on the log scale), thus providing an “indication” for an effect on OS having observed an effect on PFS according to the IQWiG framework (Table 2). TTP does not appear to be a good surrogate measure in advanced lung cancer according to any of the three frameworks (Table 3). In metastatic breast cancer, despite the moderate to high quality of the meta-analyses assessed (42;53;55), PFS is not judged to be a valid surrogate for OS according to the three evaluation frameworks adopted (Table 2). However, Hackshaw and colleagues (54) reported a medium association between TTP and OS ($R^2 = 0.56$) in trials of first-line chemotherapy, which provided a “hint” for an effect on the final endpoint according to the IQWiG framework (Table 3). In metastatic ovarian cancer, three IPD meta-analyses, two related to PFS (Table 2) (26;40), and one to TTP (9) (Table 3) show an indication of an effect on OS drawn on the observation of an

effect on the two surrogate endpoints, with R^2_{trial} ranging from 0.83 to 0.95. Nonetheless, as according to the BSES3 criteria (see Supplementary Table 1) they lack generalizability, these studies were scored down. The remaining six solid tumor types (renal, prostate, brain, gastric, head and neck, and pancreatic) were each assessed in one or two meta-analyses (Tables 2 and 3). Across these indications, the level of evidence was mixed and the strength was poor; moreover, the endpoints were not always clearly specified, therefore, all scores for strength of surrogacy relationship were low in both the IQWiG and BSES3 frameworks (in brain and gastric cancer Level 2 was the highest level of evidence according to Elston and Taylor's framework).

DISCUSSION

We sought to review the current statistical approaches to surrogate endpoint validation in advanced solid tumors, as well as to assess the suitability of PFS and TTP as surrogates for OS using currently available validation frameworks (35). Our review included thirty-one meta-analyses (1,363 RCTs enrolling more than 290,000 patients) and showed that a variety of statistical methods have been used to examine the relationship between PFS or TTP and OS. In addition, we observed a degree of variation in validity rating when using different validation frameworks across meta-analyses in general and even within a particular tumor type.

The various statistical methods used thus far in surrogacy research can be summarized in two broad categorizations. First, according to whether the assessment of the statistical association is made between the surrogate and final endpoint (observation-level association, which does not take treatment into account and is, therefore, an assessment of the prognostic role of the candidate surrogate), or between the treatment effects on both surrogate and final endpoints (treatment-level association, which assesses the predictive role of the candidate surrogate by taking treatment into account). Second, according to whether aggregate data or IPD were used. Observation-level association has been reported both using aggregate data and IPD, with different metrics used to quantify the correlation between endpoints (e.g., Spearman's ρ for median PFS versus median OS in the former (27;45;46;51;58) and $R^2_{\text{individual}}$ in the latter) (9). In several cancer types, such as metastatic gastric cancer and glioblastoma multiforme, only the observation-level association has been investigated so far. This is acknowledged to be insufficient evidence to establish surrogacy for putative surrogate endpoints (10). For most tumor types, including colorectal cancer, breast cancer, and NSCLC, both observational-level and treatment-level surrogacy has been investigated, with treatment-level surrogacy being assessed using both IPD and aggregate data. Although treatment-level associations were often reported using the common statistic of R^2_{trial} , this was calculated using different analytic approaches (e.g., meta-regression for aggregate

Table 2. Assessment of the Validity of PFS as Surrogate for OS: Comparison of Meta-analyses by Tumor Type across Evaluation Frameworks

Tumor type	Meta-analysis	Elston and Taylor framework ^a	IQWiG framework ^b			BSES3 ^c		Authors' conclusions
			Reliability	Correlation	Conclusion	Overall Score /12	Level of Evidence	
Colorectal cancer	Louvet 2001 ⁴³	Level 2	Low	-	No proof	7	D	"In conclusion, PFS certainly deserves further evaluation as an endpoint measure."
	Tang 2007 ⁴⁴	Level 1	Moderate	Medium	Hint	10	C	"In first-line chemotherapy trials for metastatic CRC, improvements in PFS are strongly associated with improvements in OS. In this patient population, PFS may be an appropriate surrogate for OS."
	Wilkerson 2009 ⁴⁰	Level 1	Moderate	Low	No Proof	8	D+	"We conclude that PFS is not a surrogate for OS; rather it is a straightforward measure of on-therapy benefit."
	Chirila 2012 ²⁵	Level 1	Limited	Medium	Indication	9	C-	"We have shown that the correlation of OS with PFS, either alone or aggregated with TTP, in clinical trials of patients with metastatic CRC is robust across lines of therapy and provides a useful means of predicting improvements in OS."
	Burzykowski 2001 ^{38§}	Level 1	Limited	Low	No Proof	7	D	"These results suggest that PFS is neither trial level nor individual level valid"
	Burzykowski 2006 ^{24§}	Level 1	Limited	Medium	Indication	8	D+	"This clearly illustrates that PFS would not be an acceptable, even 'potentially', surrogate for survival in the set of trials analyzed here. [However] the association between the treatment effects on both endpoints may have been dominated by random noise."
Lung cancer	Buyse 2007 ^{45§}	Level 1	Limited	High	Indication	10	C	"The analyses presented here show that, in historical trials comparing FU leucovorin with single-agent FU or with raltitrexed, PFS was an acceptable surrogate for OS"
	Green 2008 ^{26§}	Level 1	Limited	Medium	Indication	10	C-	"We conclude that there is modest evidence for surrogacy between one-year PFS and two-year OS."
	Hotta 2011 ⁵⁰	Level 2	Low	-	No Proof	8	D+	"A PFS advantage is unlikely to be associated with an OS advantage any longer due to this increasing impact of PFS on OS."
	Li 2012 ⁴⁷	Level 2	Low	-	No Proof	7	D	"Our data suggest that PFS is appropriate survival marker in the clinical trials of EGFR-TKIs for advanced NSCLC."
	Mandrekar 2010 ^{48§}	Level 2	Low	-	No Proof	6	D-	"Our present findings (based on data from phase II trials) demonstrate that PFS is a significant predictor of patient survival in advanced NSCLC."
	Foster 2011 ^{57§}	Level 1	Limited	Medium	Indication	8	D+	"PFS showed the most promise as a surrogate endpoint for OS (in SCLC) at the patient and the trial-level across all the statistical methods assessed."
Breast cancer	Miksad 2008 ⁵³	Level 1	Limited	Low	No Proof	8* 9**	D+ C-	"This meta-analysis suggests that the trial-level TE on PFS is significantly associated with the trial-level TE on OS. However, prediction of OS based on PFS is surrounded with uncertainty."
	Wilkerson 2009 ⁴⁰	Level 1	Moderate	Low	No Proof	8	D+	"We conclude that PFS is not a surrogate for OS; rather it is a straightforward measure of on-therapy benefit."
	Burzykowski 2008 ^{51§}	Level 1	High	Low	No Proof	7	D	"No end point could be demonstrated as a good surrogate for OS in these trials."

Table 2. Continued

Tumor type	Meta-analysis	Elston and Taylor framework ^a	IQWiG framework ^b			BSES3 ^c		Authors' conclusions
			Reliability	Correlation	Conclusion	Overall Score /12	Level of Evidence	
Ovarian cancer	Wilkerson 2009 ⁴⁰	Level 1	Moderate	Low	No Proof	8	D+	"We conclude that PFS is not a surrogate for OS; rather it is a straightforward measure of on-therapy benefit."
	Rose 2010 ^{54§}	Level 2	Low	–	No Proof	5	E+	"We studied the correlation between PFS at six months and survival and found measures of PFS at six months correlated better than response rate to OS."
	Burzykowski 2001 ³⁸	Level 1	Limited	High	Indication	8	D+	"It seems plausible to conclude that PFS is a valid surrogate for survival in advanced ovarian cancer for treatments of the type used in the trials analyzed."
	Burzykowski 2006 ²⁴	Level 1	Limited	High	Indication	8	D+	"Consequently, we suggest a better validity of the surrogate (PFS) [than in CRC]."
Renal cell carcinoma	Heng 2011 ^{59§}	Level 2	Low	–	No Proof	5	E+	"PFS may be a meaningful intermediate endpoint for OS in patients with metastatic RCC who receive treatment with novel agents."
Prostate cancer	Halabi 2009 ^{58§}	Level 2	Low	–	No Proof	6	D-	"PFS seems to be associated with OS. These data need to be validated prospectively before it can be used routinely as an intermediate end point in phase II trials in CRPC."
GBM	Ballman 2007 ^{42§}	Level 2	Low	–	No Proof	6	D-	"In light of our assessment of the relationship between PFS and OS, it appears that PFS provides only a moderately reliable estimate of survival."
	Polley 2010 ^{41§}	Level 2	Low	–	No Proof	5	E+	"Our analysis suggested that PFS at 6 months may be an appropriate primary endpoint in the context of phase II trials evaluating treatment regimen in newly diagnosed GBM patients. Future research is needed to validate our findings in a larger population."

Note. § indicate meta-analyses using individual patient data.

CRC, colorectal cancer; CRPC, castrate-resistant prostate cancer; EGFR-TKIs, epidermal growth factor receptor tyrosine-kinase inhibitors; FU, fluorouracil; GBM, glioblastoma multiforme; NSCLC, non-small-cell lung cancer; OS, overall survival; PFS, progression-free survival; PPS, post-progression survival; RCC, renal cell carcinoma; RCT, randomized controlled trial; SCLC, small-cell lung cancer; TE, treatment effect, TTP, time to progression. *Taxanes; ** Anthracyclines.

^a Level 1 corresponds to treatment-level association, i.e. evidence showing treatment effects on the surrogate correspond to treatment effects on the final patient-relevant endpoint. Level 2 corresponds to evidence showing association between the two endpoints.

^b Reliability is assessed according to (i) use of appropriate statistical approach, (ii) robustness and generalizability of results, (iii) systematic compilation of data, (iv) sufficient restriction of indications, degrees of disease severity, interventions and (v) clear definitions of endpoints. Low, moderate, limited and high indicate growing level of reliability. High correlation corresponds to $R \geq 0.85$ whilst low correlation to $R \leq 0.70$. Correlation is not even assessed if the study is of low reliability. The conclusion about the effect on the final endpoint drawn from the effect observed on the surrogate can be a no proof, hint, indication or proof according to increasing level of validity of the surrogate endpoint.

^c Overall score sums up scores from 0 to 3 obtained in each of the four domain (i.e., study design, target endpoint, statistical evaluation and generalizability). Category A and B of level of evidence correspond to good evidence for validity of the surrogate endpoint. If the score is lower than 2 in any domain, the level of evidence drops by one alphabetic category.

Table 3. Assessment of the Validity of TTP as Surrogate for OS: Comparison of Meta-analyses by Tumor Type across Evaluation Frameworks

Tumor type	Meta-analysis	Elston and Taylor framework ^a	IQWiG framework ^b			BSES3 ^c		Authors' conclusions
			Reliability	Correlation	Conclusion	Overall Score /12	Level of Evidence	
Colorectal cancer	Johnson 2006 ²⁷	Level 1	Limited	Low	No proof	8	D+	"Our findings support the use of time to progression as a surrogate for survival in metastatic lung cancer and colorectal cancer."
	Tang 2007 ⁴⁴	Level 1	Moderate	Medium	No Proof	7	D	"Our analysis showed that improvements in PFS, TTP, and RR were all strongly associated with an improvement in OS in randomized control trials of first-line chemotherapy for metastatic CRC. [. . .] The overlapping definitions of PFS and OS may account for the superiority of PFS as a surrogate for OS, as compared with TTP or RR."
	Bowater 2008 ²³	Level 1	Low	–	No proof	8	D+	"The relationship between PFS and PPS in cancer treatment that have been examined in this study are worthy of further investigation."
	Bowater 2011 ³⁷	Level 1	Low	–	No proof	7	D	"It would appear that drugs for metastatic breast or CRC that extend, by a given amount, the TTP have a strong tendency to extend, by roughly the same amount, the OS."
	Chirila 2012 ²⁵	Level 1	Limited	Low	No proof	8	D+	"The weighted correlation value did not change for PFS and it was somewhat lower for TTP (although with confidence limits that overlap those of PFS/TTP) [. . .] the correlation of OS with PFS, either alone or aggregated with TTP, in clinical trials of patients with metastatic CRC is [. . .] a useful means of predicting improvements in OS."
NSCLC	Johnson 2006 ²⁷	Level 1	Limited	Low	No proof	8	D+	"Our findings support the use of time to progression as a surrogate for survival in metastatic lung cancer and colorectal cancer."
	Bowater 2008 ²³	Level 1	Low	–	No proof	8	D+	"The relationship between PFS and PPS in cancer treatment that have been examined in this study are worthy of further investigation."

Table 3. Continued

Tumor type	Meta-analysis	Elston and Taylor framework ^a	IQWiG framework ^b			BSES3 ^c		Authors' conclusions
			Reliability	Correlation	Conclusion	Overall Score /12	Level of Evidence	
Breast cancer	Hotta 2009 ⁴⁶	Level 1	Limited	Low	No Proof	8	D+	"TTP potentially acts as a surrogate marker, but may not be still a definitive alternative in the first-line setting."
	Li 2012 ⁴⁷	Level 2	Low	—	No Proof	7	D	"Our data suggest that PFS is appropriate survival marker in the clinical trials of EGFR-TKIs for advanced NSCLC."
	Hackshaw 2005 ⁵²	Level 1	Moderate	Medium	Hint	9	C-	"TTP may be a useful surrogate marker for predicting survival in women receiving first-line anthracycline chemotherapy and could be used to estimate the survival benefit in future trials of first-line chemotherapy."
	Bowater 2008 ²³	Level 1	Low	—	No proof	8	D+	"The relationship between PFS and PPS in cancer treatment that have been examined in this study are worthy of further investigation."
	Bowater 2011 ³⁷	Level 1	Low	—	No proof	7	D	"It would appear that drugs for metastatic breast or CRC that extend, by a given amount, the TTP have a strong tendency to extend, by roughly the same amount, the OS."
Ovarian cancer	Burzykowski 2008 ^{51§}	Level 1	Limited	Low	No Proof	7	D	"No end point could be demonstrated as a good surrogate for OS in these trials."
	Buyse 2000 ^{9§}	Level 1	Limited	High	Indication	8	D+	"We conclude that TTP can be used as a surrogate for survival in advanced ovarian cancer."
Prostate cancer	Bowater 2008 ²³	Level 1	Low	—	No Proof	8	D+	"The relationship between PFS and PPS in cancer treatment that have been examined in this study are worthy of further investigation."

CRC, colorectal cancer; EGFR-TKIs, epidermal growth factor receptor tyrosine-kinase inhibitors; NSCLC, non-small-cell lung cancer; OS, overall survival; PFS, progression-free survival; PPS, post-progression survival; RR, response rate; TTP, time to progression.

§ indicate meta-analyses using individual patient data.

^aSee Table 2.

^bSee Table 2.

^cSee Table 2.

data (27;29;30;41;42;45;46;55;62) and hierarchical regression methods for IPD (9;26;28;40;47;53;59).

There is little literature directly comparing statistical validation of surrogates using IPD compared with aggregate level data (28). Buyse and colleagues have proposed IPD meta-analysis and calculation of both the $R^2_{\text{individual}}$ and R^2_{trial} to be the gold standard approach to the statistical surrogate validation (10). However, only 22 percent of the meta-analyses in this review met this criterion. Such a low proportion is in large part due to the practical challenges of conducting an IPD meta-analysis. Gathering, cleaning and formatting patient data from across clinical trial centers involves substantial resources, as due to commercial or academic restrictions, IPD for some trials are not immediately available in the public domain. While regulatory agencies can require companies to make such data available, this is often not the case for HTA organizations or agencies with a coverage or reimbursement mandate. Hence, while an IPD meta-analytic approach remains the optimal statistical approach to surrogate validation, it is likely that meta-analyses of treatment-level associations reporting the R^2_{trial} or equivalent statistics will continue to be undertaken. There is often a lack of appreciation that the use of aggregate data entails a loss of information that may have a profound impact on the analyses performed, and their interpretation. For instance, several meta-analyses included in our review used the ratio of medians as a measure of treatment effects (27;32;48;54;55). Such an approach could be seriously misleading if the time to event distributions were not exponential and, even if they were, the medians usually have wide confidence intervals and so their ratio is likely to be extremely unstable (63). Few regression analyses make proper allowance for the estimation error (9;26;28;40;53) other than through a weighting of the trials by their sample size. Regression analyses that ignore estimation errors are likely to underestimate the true relationship between the treatment effects on the surrogate and the final endpoint. The availability of IPD allows the association between the surrogate and the final endpoints to be modeled, which is theoretically preferable to looking only at the marginal association between the treatment effects on the two endpoints (9).

Limitations of This Study

To the best of our knowledge, this is the first study to empirically test the application of current surrogate validation frameworks across a sample of meta-analyses in a disease area. On the other hand, our study has some limitations. First, as we were unable to use a conventional search strategy, we cannot claim to have identified all relevant meta-analyses. A more exhaustive search might have been feasible if we had narrowed our scope to a single tumor type. However, our aim was to keep the scope of the study broad and to identify a sufficient number of meta-analyses to assess a variety of statistical methods. Our list of included meta-analyses appears indeed to be comprehen-

sive when compared with recent reviews in the field (12;32). Second, we have not formally appraised the overall quality of each meta-analysis. Given that the focus of the study was not to determine an unbiased estimate of the efficacy or safety of interventions, we believe that this decision was justified. However, to assess potential selection or publication bias, we noted if each included meta-analysis reported undertaking a formal literature search strategy to identify studies. In line with the findings of previous studies, we found that meta-analyses of aggregate data were more likely to undertake a literature search and include more studies than IPD meta-analyses (51). Third, we have not attempted to replicate any of the analyses presented in the included meta-analyses. This might have been useful, as it would allow us to examine whether all of the assumptions made in the presented analyses are supported by the primary data and whether the conclusions change when all relevant trials are considered in a single analysis and after updating for more recently published trials. Fourth, the application of both the IQWiG and BSES3 evaluation frameworks involved an element of subjective judgment. To minimize potential assessment bias, the application of the frameworks undertaken was independently checked by a second reviewer and a third reviewer used to resolve disagreements in judgment of these two reviewers. All of them were HTA analysts with experience in the field of oncology. Finally, although survival is a definitive patient-relevant outcome in the case of most solid tumors, there may be problems with using of OS in the context of surrogate validation. Despite its primacy, OS has been claimed to be unsuitable in detecting treatment benefit in settings for which effective therapy is available after trial participation (64). Because patients in oncology trials are often permitted to cross over from the control arm to the treatment arm or switch to other therapies due to lack of response or symptoms, the attribution of OS gain to initial treatment allocation may be confounded by these subsequent lines of therapy (65).

Implications for Policy and Practice

Surrogate validation studies also have important relevance for the assessment of the cost effectiveness of new treatments (66). Using a reported relationship between OS and the surrogate, decision analysts can estimate the incremental cost per quality-adjusted life-year (QALY) based on the observed treatment effect on the surrogate (2).

Our study has important implications for the use of surrogate outcomes in HTA and coverage/reimbursement policy decisions. To appropriately apply evidence of surrogate validation, policy makers need decision frameworks that help them do so. While the IQWiG and BSES3 frameworks are potentially useful tools for clinicians and healthcare decision makers, there are problems in their practical application. Both have elements that require subjective judgment. In addition, they require a high level of association to demonstrate surrogacy, that is, $R^2_{\text{treatment}}$

≥ 0.60 or $R_{\text{treatment}} \geq 0.85$, raising a query on the origins of such thresholds. With a small number of exceptions, we found the strength of the association between PFS or TTP and OS across meta-analyses to be consistently low (i.e., $R_{\text{treatment}} < 0.7$) across tumor types. Indeed, according to the IQWiG and BSES3 validation frameworks, the evidence available about surrogacy of PFS and TTP in metastatic cancer is still insufficient to guide policy. Moreover, we noted a degree of variation in validity rating of IQWiG and BSES3 frameworks across meta-analyses within a particular tumor type. For example, for PFS in colorectal cancer, four meta-analyses (26–28;47) showed an “indication” of an effect on OS given an effect observed on the surrogates, however the highest level of evidence achieved according to BSES3 is C, well below the minimum acceptable level for a good surrogate. We believe that this variation probably reflects differences in the evidence within each meta-analysis due to differences in the precise patient population, definition and assessment of progression, drug therapy and comparator of included trials. Moreover, variation may also be due to differences in the statistical methods applied by each meta-analysis. Finally, the criteria considered by the two evaluation frameworks are different and in some cases opposite; for instance, BSES3 favors generalizability across populations and drug-class mechanisms while the IQWiG framework gives precedence to restricted indications and therapies. Nonetheless, within each indication, different meta-analyses deal with overlapping evidence, and the underlying redundancy may have accounted for similar conclusions. When considering conclusions across indications, our results support the need for a disease-specific approach to the validation of surrogate endpoints, with careful consideration of transferability of results from one disease to the other.

The three evaluation tools used herein were developed through different processes: Elston and Taylor’s framework was based on a guide for clinicians proposed by the Journal of the American Medical Association (JAMA) Evidence-Based Medicine Working Group; the algorithm for surrogate endpoints validation in oncology was developed at IQWiG after a systematic search of the literature by the Agency; whereas the BSES3’s initial version originated from a literature review followed by a stakeholder workshop that evaluated it for applications in rheumatology (38). We believe that the development of future surrogate validation tools would benefit from formal consensus methods. Further research is needed to examine the application of surrogate validation frameworks in the context of candidate surrogates both in oncology and other disease areas.

In conclusion, we found that the level of evidence available supporting a relationship between PFS or TTP and OS varies considerably by tumor type and is not always consistent even within one specific type. Overall, the strength of the association between PFS or TTP and OS was relatively low and only PFS in advanced colorectal and ovarian cancers treated with cytotoxic agents was found to be a valid surrogate endpoint according to one of the evaluation frameworks used. Our study emphasizes

the importance of building consensus on appropriate statistical techniques to examine surrogacy and on development of evaluation frameworks, not only in oncology but across all areas of medicine, across jurisdictions and scientific communities.

SUPPLEMENTARY MATERIAL

- | | | |
|---------------------|----|---|
| Supplementary Table | 1: | http://dx.doi.org/10.1017/S0266462314000300 |
| Supplementary Table | 2: | http://dx.doi.org/10.1017/S0266462314000300 |
| Supplementary Table | 3: | http://dx.doi.org/10.1017/S0266462314000300 |

CONTACT INFORMATION

Oriana Ciani, PhD (o.ciani@exeter.ac.uk), University of Exeter Medical School, Veysey Building, Salmon Pool Lane, Exeter, EX2 4SG, United Kingdom

CONFLICTS OF INTEREST

OC was funded through a Peninsula College of Medicine and Dentistry studentship. A preliminary version of this paper was based on a report which was funded by the National Institute for Health and Care Excellence through its Decision Support Unit. The views expressed in the paper are those of the authors. MB declares an association with the International Drug Development Institute and EDS with Dendrix Ltd. All other co-authors declare no conflict of interest.

REFERENCES

- De Gruttola VG, Clax P, DeMets DL, et al. Considerations in the evaluation of surrogate endpoints in clinical trials. Summary of a National Institutes of Health workshop. *Control Clin Trials*. 2001;22:485-502.
- Elston J, Taylor RS. Use of surrogate outcomes in cost-effectiveness models: A review of United Kingdom health technology assessment reports. *Int J Technol Assess Health Care*. 2009;25:6-13.
- Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. *BMJ*. 2011;343:d7995.
- Messerli FH, Bangalore S. ALTITUDE Trial and Dual RAS Blockade: The alluring but soft science of the surrogate end point. *Am J Med*. 2013;126:e1-e3.
- Fleming TR, DeMets DL. Surrogate end points in clinical trials: Are we being misled? *Ann Intern Med*. 1996;125:605-613.
- Ciani O, Buyse M, Garside R, et al. Comparison of treatment effect sizes associated with surrogate and final patient relevant outcomes in randomised controlled trials: Meta-epidemiological study. *BMJ*. 2013;346:f457.
- Lassere MN. The Biomarker-Surrogacy Evaluation Schema: A review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. *Stat Methods Med Res*. 2008;17:303-340.
- Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: A literature review. *Stat Med*. 2006;25:183-203.

9. 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.
10. Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points—the challenge of statistical validation. *Nat Rev Clin Oncol*. 2010;7:309-317.
11. Velasco Garrido M, Mangiapane S. Surrogate outcomes in health technology assessment: An international comparison. *Int J Technol Assess Health Care*. 2009;25:315-322.
12. IQWiG. Validity of surrogate endpoints in oncology. Executive Summary. Cologne, Germany: IQWiG; 2011. https://www.iqwig.de/download/A10-05_Executive_Summary_v1-1_Surrogate_endpoints_in_oncology.pdf (accessed January, 2013).
13. Ellenberg S, Hamilton JM. Surrogate endpoints in clinical trials: Cancer. *Stat Med*. 1989;8:405-413.
14. Dunn BK, Akpa E. Biomarkers as surrogate endpoints in cancer trials. *Semin Oncol Nurs*. 2012;28:99-108.
15. Berghmans T, Pasleau F, Paesmans M, et al. Surrogate markers predicting overall survival for lung cancer: ELCWP recommendations. *Eur Respir J*. 2012;39:9-28.
16. Shi Q, Sargent DJ. Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. *Int J Clin Oncol*. 2009;14:102-111.
17. Saad ED, Katz A, Buyse M. Overall survival and post-progression survival in advanced breast cancer: A review of recent randomized clinical trials. *J Clin Oncol*. 2010;28:1958-1962.
18. Jaffe CC. Measures of response: RECIST, WHO, and new alternatives. *J Clin Oncol*. 2006;24:3245-3251.
19. Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man. Methodological consideration for using progression-free survival (PFS) or disease-free survival (DFS) in confirmatory trials. 2013. http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000406.jsp&mid=WC0b01ac0580034cf3 (accessed January, 2013).
20. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45:228-247.
21. Dancy JE, Dodd LE, Ford R, et al. Recommendations for the assessment of progression in randomised cancer treatment trials. *Eur J Cancer*. 2009;45:281-289.
22. Tappenden P, Chilcott J, Ward S, Eggington S, Hind D, Hummel S. Methodological issues in the economic analysis of cancer treatments. *Eur J Cancer*. 2006;42:2867-2875.
23. Sridhara R, Johnson JR, Justice R, et al. Review of oncology and hematology drug product approvals at the US Food and Drug Administration between July 2005 and December 2007. *J Natl Cancer Inst*. 2010;102:230-243.
24. Hartley R, Keen EM, Large J, Tedd LA. *Online searching: Principles and practice*. Epping, UK: BowkerSaur; 1990.
25. Bowater RJ, Bridge LJ, Lilford RJ. The relationship between progression-free and post-progression survival in treating four types of metastatic cancer. *Cancer Lett*. 2008;262:48-53.
26. Burzykowski T, Buyse M. Surrogate threshold effect: An alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat*. 2006;5:173-186.
27. Chirila C, Odom D, Devercelli G, et al. Meta-analysis of the association between progression-free survival and overall survival in metastatic colorectal cancer. *Int J Colorectal Dis*. 2012;27:623-634.
28. Green E, Yothers G, Sargent DJ. Surrogate endpoint validation: Statistical elegance versus clinical relevance. *Stat Methods Med Res*. 2008;17:477-486.
29. Johnson KR, Ringland C, Stokes BJ, et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: A meta-analysis. *Lancet Oncol*. 2006;7:741-746.
30. Sherrill B, Amonkar M, Wu Y, et al. Relationship between effects on time-to-disease progression and overall survival in studies of metastatic breast cancer. *Br J Cancer*. 2008;99:1572-1578.
31. Buyse M. Use of meta-analysis for the validation of surrogate endpoints and biomarkers in cancer trials. *Cancer J*. 2009;15:421-425.
32. Sherrill B, Kaye JA, Sandin R, Cappelleri JC, Chen C. Review of meta-analyses evaluating surrogate endpoints for overall survival in oncology. *Onco Targets Ther*. 2012;5:287-296.
33. Lee L, Wang L, Crump M. Identification of potential surrogate end points in randomized clinical trials of aggressive and indolent non-Hodgkin's lymphoma: Correlation of complete response, time-to-event and overall survival end points. *Ann Oncol*. 2011;22:1392-1403.
34. Tibaldi F, Barbosa FT, Molenberghs G. Modelling associations between time-to-event responses in pilot cancer clinical trials using a Plackett-Dale model. *Stat Med*. 2004;23:2173-2186.
35. EUnetHTA. *Endpoints used in REA of pharmaceuticals - Surrogate Endpoints*. 2013. <http://www.eunetha.eu/sites/5026.fedimbo.belgium.be/files/Surrogate%20Endpoints.pdf> (accessed March, 2013).
36. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. *JAMA*. 1999;282:771-778.
37. Lassere MN, Johnson KR, Schiff M, Rees D. Is blood pressure reduction a valid surrogate endpoint for stroke prevention? An analysis incorporating a systematic review of randomised controlled trials, a by-trial weighted errors-in-variables regression, the surrogate threshold effect (STE) and the Biomarker-Surrogacy (BioSurrogate) Evaluation Schema (BSES). *BMC Med Res Methodol*. 2012;12:27.
38. Lassere MN, Johnson KR, Boers M, et al. Definitions and validation criteria for biomarkers and surrogate endpoints: Development and testing of a quantitative hierarchical levels of evidence schema. *J Rheumatol*. 2007;34:607-615.
39. Bowater RJ, Lilford PE, Lilford RJ. Estimating changes in overall survival using progression-free survival in metastatic breast and colorectal cancer. *Int J Technol Assess Health Care*. 2011;27:207-214.
40. Burzykowski T, Molenberghs G, Buyse M, Geys H, Renard D. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. *J R Stat Soc Ser C Appl Stat*. 2001;50:405-422.
41. Amir E, Seruga B, Kwong R, Tannock IF, Ocana 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.
42. Wilkerson J, Fojo T. Progression-free survival is simply a measure of a drug's effect while administered and is not a surrogate for overall survival. *Cancer J*. 2009;15:379-385.
43. Polley MY, Lamborn KR, Chang SM, et al. Six-month progression-free survival as an alternative primary efficacy endpoint to overall survival in newly diagnosed glioblastoma patients receiving temozolomide. *Neuro Oncol*. 2010;12:274-282.
44. Ballman KV, Buckner JC, Brown PD, et al. The relationship between six-month progression-free survival and 12-month overall survival end points for phase II trials in patients with glioblastoma multiforme. *Neuro Oncol*. 2007;9:29-38.
45. Louvet C, de Gramont A, Tournigand C, Artru P, Maindrault-Goebel F, Krulik M. Correlation between progression free survival and response rate in patients with metastatic colorectal carcinoma. *Cancer*. 2001;91:2033-2038.
46. Tang PA, Bentzen SM, Chen EX, Siu LL. Surrogate end points for median overall survival in metastatic colorectal cancer: Literature-based analysis

- from 39 randomized controlled trials of first-line chemotherapy. *J Clin Oncol.* 2007;25:4562-4568.
47. Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol.* 2007;25:5218-5224.
 48. Hotta K, Fujiwara Y, Matsuo K, et al. Time to progression as a surrogate marker for overall survival in patients with advanced non-small cell lung cancer. *J Thorac Oncol.* 2009;4:311-317.
 49. 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.
 50. Mandrekar SJ, Qi Y, Hillman SL, et al. Endpoints in phase II trials for advanced non-small cell lung cancer. *J Thorac Oncol.* 2010;5:3-9.
 51. Hayashi H, Okamoto I, Taguri M, Morita S, Nakagawa K. Postprogression survival in patients with advanced non-small-cell lung cancer who receive second-line or third-line chemotherapy. *Clin Lung Cancer.* 2013;14:261-266.
 52. Hotta K, Kiura K, Fujiwara Y, et al. Role of survival post-progression in phase III trials of systemic chemotherapy in advanced non-small-cell lung cancer: A systematic review. *PloS One.* 2011;6:e26646.
 53. 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.
 54. Hackshaw A, Knight A, Barrett-Lee P, Leonard R. Surrogate markers and survival in women receiving first-line combination anthracycline chemotherapy for advanced breast cancer. *Br J Cancer.* 2005;93:1215-1221.
 55. Miksad RA, Zietemann V, Gothe R, et al. Progression-free survival as a surrogate endpoint in advanced breast cancer. *Int J Technol Assess Health Care.* 2008;24:371-383.
 56. Rose PG, Tian C, Bookman MA. Assessment of tumor response as a surrogate endpoint of survival in recurrent/platinum-resistant ovarian carcinoma: A Gynecologic Oncology Group study. *Gynecol Oncol.* 2010;117:324-329.
 57. Sundar S, Wu J, Hillaby K, Yap J, Lilford R. A systematic review evaluating the relationship between progression free survival and post progression survival in advanced ovarian cancer. *Gynecol Oncol.* 2012;125:493-499.
 58. 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.
 59. Foster NR, Qi Y, Shi Q, et al. Tumor response and progression-free survival as potential surrogate endpoints for overall survival in extensive stage small-cell lung cancer: Findings on the basis of North Central Cancer Treatment Group trials. *Cancer.* 2011;117:1262-1271.
 60. Halabi S, Vogelzang NJ, Ou SS, Owzar K, Archer L, Small EJ. Progression-free survival as a predictor of overall survival in men with castrate-resistant prostate cancer. *J Clin Oncol.* 2009;27:2766-2771.
 61. Heng DY, Xie W, Bjarnason GA, et al. Progression-free survival as a predictor of overall survival in metastatic renal cell carcinoma treated with contemporary targeted therapy. *Cancer.* 2011;117:2637-2642.
 62. 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.
 63. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statist Med.* 1998;17:2815-2834.
 64. Buyse M, Sargent DJ, Saad ED. Survival is not a good outcome for randomized trials with effective subsequent therapies. *J Clin Oncol.* 2011;29:4719-4720; author reply 4720-4721.
 65. Saad ED, Katz A, Hoff PM, Buyse M. Progression-free survival as surrogate and as true end point: Insights from the breast and colorectal cancer literature. *Ann Oncol.* 2010;21:7-12.
 66. Ciani O, Taylor RS. A more evidence based approach to the use of surrogate end points in policy making. *BMJ.* 2011;343:d6498.