Surrogate endpoints in advanced sarcoma trials: a meta-analysis

Marion Savina[1,2,3,4], Saskia Litière[5], Antoine Italiano[6], Tomasz Burzykowski[7], Franck Bonnetain[8], Sophie Gourgou[9], Virginie Rondeau[3,10], Jean-Yves Blay[11,12], Sophie Cousin[6], Florence Duffaud[13], Hans Gelderblom[14], Alessandro Gronchi[15], Ian Judson[16], Axel Le Cesne[17], Paul Lorigan[18], Joan Maurel[19], Winette van der Graaf[20,21,23], Jaap Verweij[22], Simone Mathoulin-Pélissier[1,2,3,4] and Carine Bellera[1,2,3,4]

[1]Clinical and Epidemiological Research Unit, Institut Bergonié, Comprehensive Cancer Center, Bordeaux cedex 33076, France

[2]INSERM CIC-EC 14.01 (Clinical Epidemiology), Bordeaux 33000, France

[3]INSERM, ISPED, Centre INSERM U1219 Bordeaux Population Health Center, Epicene Team, Bordeaux 33000, France

[4]University of Bordeaux, ISPED, Centre INSER M U1219 Bordeaux Population Health, Epicene Team, Bordeaux 33000, France

[5]European Organisation for Research and Treatment of Cancer (EORTC), Brussels 1200, Belgium

[6]Medical Oncology Unit, Institut Bergonié, Comprehensive Cancer Center, Bordeaux cedex 33076, France

[7]Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Hasselt University, Hasselt 3500, Belgium

[8]Methodology and Quality of life in Oncology Unit, Besançon EA3181, France

[9]Biometrics Unit, Institut du Cancer de Montpellier, Univ. Montpellier, Montpellier 34298, France

[10]INSERM, ISPED, Centre INSERM U1219 Bordeaux Population Health Center, Biostatistic Team, Bordeaux 33000, France

[11]Centre Léon Bérard, Comprehensive Cancer Center, Lyon 69008, France

[12]University Claude Bernard Lyon I, Lyon 69000, France

[13]Medical Oncology Unit, University Hospital La Timone and University of Aix-Marseille, Marseille 13005, France

[14]Department of Medical Oncology, Leiden University Medical Center, Leiden 2300RC, The Netherlands

[15]Sarcoma Service, Department of Surgery, Fondazione IRCCS Istituto Nazionale dei Tumori, Milano, Italy

[16]Institute of Cancer Research, Sutton, Surrey, United Kingdom

[17]Medicine Department, Institut Gustave Roussy, Comprehensive Cancer Center, Villejuif 94800, France

[18]University of Manchester and Christie NHS Foundation Trust, Manchester M20 4BX, UK

[19]Department of Medical Oncology, Hospital Clinic, CIBERehd, Translational Genomics and Targeted Therapeutics in Solid Tumors (IDIBAPS), Barcelona 08036, Spain

[20]The Institute of Cancer Research, Sutton, London SM2 5NG, United Kingdom

[21]Radboud University Medical Centre, Department of Medical Oncology, GA Nijmegen 6525, The Netherlands

[22]Department of Medical Oncology, Erasmus University Medical Center, CE Rotterdam 3015, The Netherlands

[23]Royal Marsden NHS Foundation Trust, Chelsea, London, United Kingdom

Correspondence to: Carine Bellera, **email:** C.Bellera@bordeaux.unicancer.fr

ABSTRACT

Background: Alternative endpoints to overall survival (OS) are frequently used to assess treatment efficacy in randomized controlled trials (RCT). Their properties in terms of surrogate outcomes for OS need to be assessed. We evaluated the surrogate properties of progression-free survival (PFS), time-to-progression (TTP) and time-to-treatment failure (TTF) in advanced soft tissue sarcomas (STS).

Results: A total of 21 trials originally met the selection criteria and 14 RCTs (*N* = 2846) were included in the analysis. Individual-level associations were moderate (highest for 12-month PFS: Spearman's rho = 0.66; 95% CI [0.63; 0.68]). Trial-level associations were ranked as low for the three endpoints as per the IQWiG criterion.

Materials and Methods: We performed a meta-analysis using individual-patient data (IPD). Phase II/III RCTs evaluating therapies for adults with advanced STS were eligible. We estimated the individual- and the trial-level associations between then candidate surrogates and OS. Statistical methods included weighted linear regression and the two-stage model introduced by Buyse and Burzykowski. The strength of the trial-level association was ranked according to the German Institute for Quality and Efficiency in Health Care (IQWiG) guidelines.

Conclusions: Our results do not support strong surrogate properties of PFS, TTP and TTF for OS in advanced STS.

INTRODUCTION

The choice of the primary endpoint is a key step when designing a randomized controlled trial (RCT). In oncology, the most commonly used endpoint to assess the efficacy of a new treatment in RCTs is overall survival (OS) which is easily measurable, objectively defined as the time from randomization to death and validated by health regulatory authorities [1]. Alternative time-to-event endpoints are commonly used in practice in phase II trials and increasingly being used instead of OS in phase III trials [2]. These composite endpoints include biological and clinical events, such as disease progression or treatment toxicity. Their development is driven by the need to reduce the number of patients, the trial duration, the delay to reach conclusions and ultimately the cost of the trials. However, it is essential to rigorously assess their surrogate properties for OS, and as such whether or not they can be used as primary endpoints for assessing the benefit of new therapies. This approach does not preclude their intrinsic value as parameters of patient benefit of a treatment.

As of today, the meta-analytic surrogacy evaluation scheme proposed by Buyse and Burzykowski *et al*. [3, 4] is considered as the most statistically rigorous method for the validation of surrogate endpoints [5, 6]. This approach requires individual-patient data (IPD) from multiple RCTs with similar design and treatment to address surrogacy from a multi-level framework. At the patient level, the surrogate endpoint should be correlated and predictive of the final endpoint regardless of the treatment (individual-level association). At the trial level, the treatment effect on the surrogate endpoint should be correlated and predictive of the treatment effect on the final endpoint (trial-level association).

Soft-tissue sarcomas (STS) are a heterogeneous group of diseases that account for 1% of all malignancies in adults [7]. Despite adequate locoregional treatment, up to 40% of patients with STS develop metastatic disease [7]. When metastases are detected, the standard of care is palliative chemotherapy. Due to their rarity, conducting large RCTs to evaluate the benefit of new treatment for metastatic STS is complex. The identification of valid surrogate endpoints for OS that would reduce the number of included patients would be of a great advantage for clinical research. To our knowledge, only one meta-analysis evaluating response rate and PFS as surrogates for OS in metastatic STS was conducted [8]. The study was however limited to the analysis of aggregated data.

We performed a meta-analysis of RCTs using IPD to assess the surrogate properties for OS of three commonly used endpoints in advanced STS: progression-free survival (PFS), time-to-progression (TTP) and time-to-treatment failure (TTF). This manuscript follows the international recommendations of the PRISMA guidelines for reporting meta-analyses [9].

RESULTS

Data

After screening 231 abstracts, we identified 21 eligible trials and obtained the trial sponsor's agreement for 19 RCTs (Figure 1). IPD were available for 14 RCTs [10–23]. Trials characteristics are presented in Table 1. Three trials had two experimental arms evaluating different administration schedules for the same drug [10, 15, 23]. We combined the two experimental arms into one for these studies. One trial was designed as two parallel randomized trials [14], it was included in the meta-analysis as two distinct trials so that we considered a total of 15 trials. Aside from one trial [18], RCTs evaluated chemotherapy-based regimens. Most trials compared an experimental chemotherapy to a doxorubicin or ifosfamide-based chemotherapy regimen as first-line treatment (Table 1). One trial focused on leiomyosarcomas [20] and one trial excluded liposarcomas [23], all 13 other trials presented similar histological subtypes. Out of the 14 trials included, 11 relied on radiological central review at study entry and two trials used histology review (local or in a specialized center).

IPD from the 2846 patients included in the trials were analyzed. Median follow-up duration ranged from 9 to 93 months (median: 36 months). Figure 2 displays forest plots for the treatment effects estimated by hazard ratios (HR) on two-year OS, and one-year PFS, TTP and TTF for each trial. Among the 2165 patients who died during the total follow-up, 1704 (79%) died during the first 18 months. For each of the three candidate surrogate endpoints, the number of events observed at 6 and 12 months is provided in Supplementary Table 1.

Correlation between the candidate surrogate endpoints and OS (individual-level surrogacy)

We relied on a one-parameter Clayton copula model, considered the best fitting model compared to Plackett or Hougaard copula. Considering a six-month follow-up for the surrogate endpoints, the individual-level correlations with 18-month OS were moderate, with PFS showing the highest correlation (0.62; 95% CI [0.59; 0.65]) (Table 2). Correlations obtained when using a one-year follow-up for
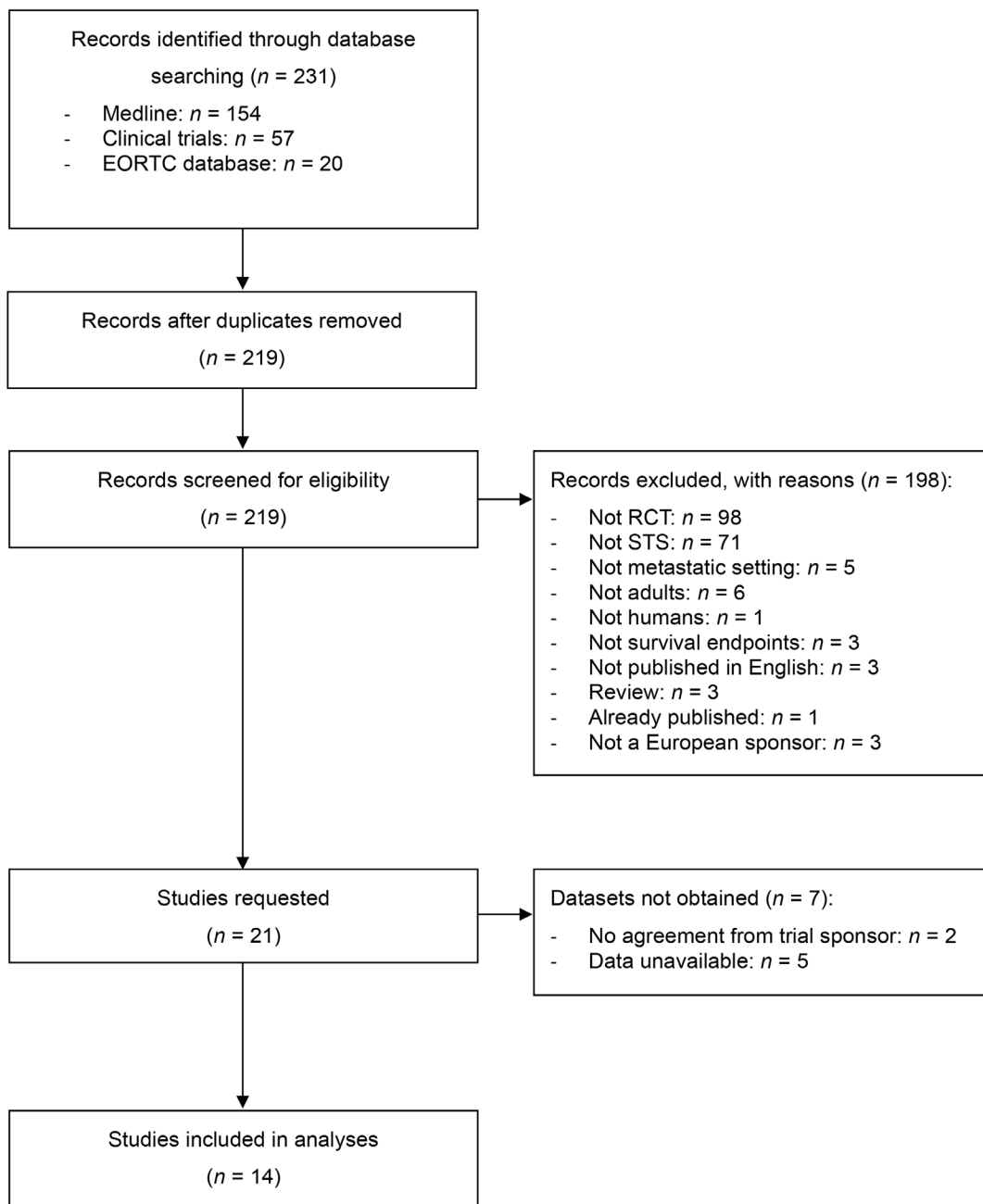
Figure 1: Flow of information through the different phases of the study selection, as per PRISMA guidelines [9][*].
[*]EORTC = European Organisation for Research and Treatment of Cancer; RCT = randomized controlled trial; STS = soft-tissue sarcoma.

Table 1: Key characteristics of the included trials[*]

Study	Ref	Phase	Inclusion period	N	Treatment line	Control arm	Experimental arm	Median follow-up	
								All patients	Patients alive
62901	[10]	II	≥1991	334	1st line	Doxorubicin	Epirubicin	50.2 months	13.3 months
62941	[11]	II	≥1995	86	1st and 2nd line	Doxorubicin	Docetaxel	35.9 months	10.9 months
62903	[12]	III	1992–1995	315	1st line	Doxorubicin + Ifosfamide	Doxorubicin + Ifosfamide + GM-GSF	91.4 months	40.1 months
62962	[13]	II	≥1997	95	1st line	Doxorubicin	Doxorubicin pegylated liposomal	35.2 months	14.5 months
62912	[14]	II	1992–1994	78	2nd-line	Ifosfamide5 g/m²/1 day	Ifosfamide3 g/m²/3 days	30.6 months	16.2 months
			1994–1996	103	1st line	Ifosfamide5 g/m²/1 day	Ifosfamide3 g/m²/3 days	35.5 months	6.6 months
62971	[15]	III	1998–2001	326	1st line	Doxorubicin	Ifosfamide	51.7 months	43.1 months
GEIS9	[16]	II	2003–2007	132	1st line	Doxorubicin	Intensified Doxorubicin + Ifosfamide	22.5 months	15.4 months
Palsar 1	[17]	III	1994–1997	145	1st line	MAID	Intensified MAID	93.0 months	89.7 months
62072	[18]	III	2008–2010	369	2nd to 5th line	Placebo	Pazopanib	14.6 months	12.2 months
Palsar 2	[19]	III	2000–2008	87	1st line	MAID	MAID + MICE	22.3 months	21.4 months
Taxogem †	[20]	II	2006–2008	70	2nd line	Gemcitabine	Gemcitabine + Docetaxel + Lenograstime	32.5 months	24.9 months
62012	[21]	III	2003–2010	455	1st line	Doxorubicin	Intensified Doxorubicin + Ifosfamide	56.4 months	30.7 months
62061	[22]	II	2006–2008	118	1st line	Doxorubicin	Brostallicin	21.3 months	19.3 months
62091	[23]	IIb/III	≥2011	133	1st line	Doxorubicin	Trabectedin	9.4 months	8.6 months

[*]GM-GSF: Recombinant human granulocyte-macrophage colony-stimulating factor; MAID = Doxorubicin, Ifosfamide and Dacarbazine; MICE: Mesna, Ifosfamide, Carboplatin and Etoposide; Ref = reference.
†Only patients with leiomyosarcoma included.

the surrogate endpoints were also moderate, even though slightly higher (PFS: 0.66; 95% CI [0.63; 0.68]).

Correlation between treatment effects on the candidate surrogate endpoints and treatment effect on OS (trial-level surrogacy)

A total of 15 pairs of log(HR) were compared for each endpoint. When considering a six-month follow-up for the surrogates, the trial-level associations R^2_{WLR} and R^2_{2SM}, were low ($R^2_{WLR} \leq 0.60$; $R^2_{2SM} \leq 0.60$) (Table 2). When considering a one-year follow-up, the association measures remained low ($R^2_{WLR} \leq 0.60$; $R^2_{2SM} \leq 0.05$). Regression curves calculated based on the WLR models are shown in Figure 3. As per IQWiG guidelines, all trial-level associations estimated were ranked as medium.

Subgroup analyses

The first subgroup analysis focused on trials comparing systemic therapy to doxorubicin- or ifosfamide-based chemotherapies in the first-line setting (N_{trial} = 11; $N_{patient}$ = 2243). When considering a six-month follow-up, the three endpoints were moderately associated with 18-month OS at the patient level ($0.56 \leq \rho_{Spearman} \leq 0.67$). At the trial level, the association between the candidate surrogates and 18-month OS was low ($R^2_{WLR} \leq 0.60$; $R^2_{2SM} \leq 0.11$) (Table 2). When considering a 12-month follow-up, the individual-level and trial-level associations increased. Confidence intervals for trial level associations were large, irrespective of the statistical method used. The number of events observed in the subgroup of trials at 6 and 12 months is provided in Supplementary Table 2. For the second subgroup analysis focusing on

leiomyosarcomas, the treatment effects on OS and on the candidate surrogates could not be estimated for one trial due to a lack of events, it was then excluded from the subgroup analysis. Individual-level correlations slightly decreased compared to the primary analysis. At the trial level however, the correlations significantly increased (Table 2), although again, confidence interval were large.

DISCUSSION

We pooled IPD data from 2846 patients included in 14 RCTs to evaluate the surrogate properties of PFS, TTP and TTF for OS in advanced STS. At the individual-level, associations between the three endpoints and OS were moderate, with the highest correlation observed for PFS. At the trial level, associations between the treatment effects on three endpoints and treatment effect on OS were low with wide confidence intervals. The strength of the trial-level association was quantified as medium as per the IQWiG criteria, indicating that the validity of the endpoints as surrogates for OS remains unclear.

Several statistical methods are available to assess surrogacy. We relied on the two-stage approach developed by Buyse and Burzykowski based on IPD [4], considered the most rigorous statistical approach for surrogacy assessment [5, 6]. Similarly, several criteria have been proposed to assess the validity of surrogate endpoints [24–26]. Although they present differences, they all require a lower limit of the 95% CI for the trial-level correlation coefficient at least higher than 0.6 to definitely validate a surrogate endpoint. As such, there was no sufficient evidence to conclude for surrogacy.

To the best of our knowledge, this is the second meta-analysis conducted in advanced STS patients, and

the first on IPD. In the first meta-analysis, conducted on aggregated data, the authors reported a 0.61 trial-level association when assessing the surrogate properties of PFS, and concluded that PFS was an appropriate surrogate for OS [8]. However, we feel that data were insufficient to provide strong surrogacy evidence. No confidence interval for the trial-level association was reported, a key element to quantify the validity of a surrogate endpoint using appropriate criteria [24–26]. As the correlation estimate reported was derived from a smaller set of trials than in our study, it is likely that the precision was also poor.

Our study also presents some limitations. Some of the trials included date back to the 90 s. This should be regarded as a weakness, particularly with respect to the histological diagnosis. For instance, gastro-intestinal stromal tumors were considered as STS and included in some older trials. Additionally, the criteria for response assessment in solid tumors evolved in the past 20 years
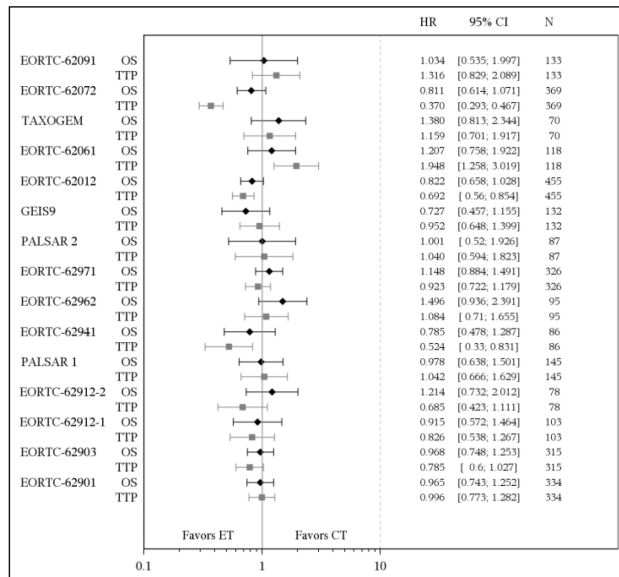

A: Progression-free survival (PFS)
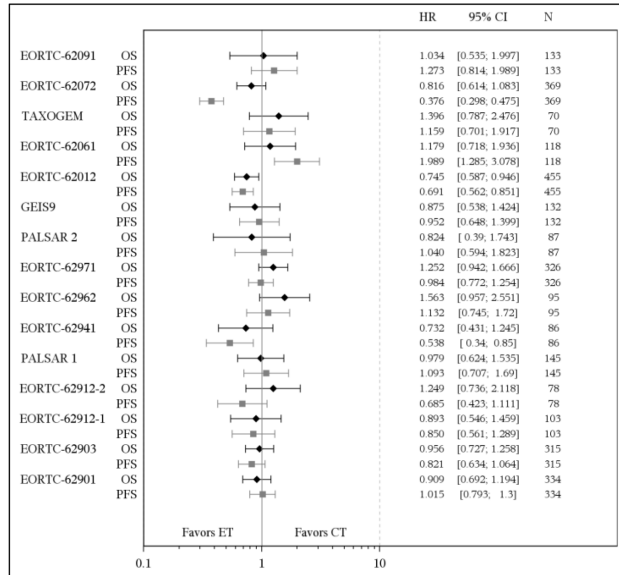
B: Time-to progression (TTP)
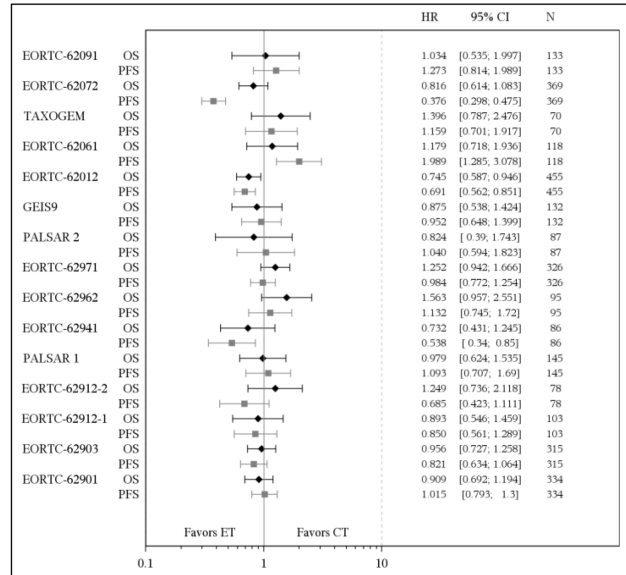
C: Time-to-treatment failure (TTF)

Figure 2: Forest plots. *Treatment effects on 12-month progression-free survival (**A**), time-to progression (**B**) and time-to-treatment failure (**C**) and on 18-month overall survival (OS) estimated by hazard ratios (HR) using separate Cox models. The first row for each trial shows the result for OS, and the second row shows the result for the candidate surrogate. The diamonds and squares represent the point estimates for OS and the candidate surrogate, respectively. The horizontal error bars show the 95% confidence interval (CI) of each hazard ratio (15 trials, 2846 patients). CT = control treatment; ET = experimental treatment.

Table 2: Individual- and trial-level associations between 6-month and 12-month progression-free survival, time-to-progression, time-to-treatment failure and 18-month overall survival[*]

Follow-up	Endpoint	Individual-level association	Trial-level association	
		$\rho_{Spearman}$[†] [95% CI]	R^2_{WLR}[‡] [95% CI]	R^2_{2SM}[φ] [95% CI]
All trials (N_{trial} = 15; $N_{patient}$ = 2846)				
6 months	PFS	0.62 [0.59; 0.65]	0.33 [0.00; 0.60]	0.04 [0.00; 0.43]
	TTP	0.59 [0.56; 0.63]	0.32 [0.00; 0.58]	0.07 [0.00; 0.60]
	TTF	0.60 [0.57; 0.63]	0.32 [0.00; 0.58]	0.06 [0.00; 0.57]
12 months	PFS	0.66 [0.63; 0.68]	0.33 [0.00; 0.60]	0.00 [0.00; 0.05]
	TTP	0.63 [0.60; 0.66]	0.30 [0.00; 0.57]	0.00 [0.00; 0.02]
	TTF	0.64 [0.61; 0.67]	0.31 [0.00; 0.58]	0.00 [0.00; 0.01]
Doxorubicin- or ifosfamide-based treatment, first-line setting (N_{trial} = 11; $N_{patient}$ = 2243)				
6 months	PFS	0.63 [0.60; 0.67]	0.30 [0.00; 0.60]	0.00 [0.00; 0.08]
	TTP	0.60 [0.56; 0.64]	0.26 [0.00; 0.58]	0.00 [0.00; 0.11]
	TTF	0.61 [0.57; 0.65]	0.27 [0.00; 0.58]	0.00 [0.00; 0.06]
12 months	PFS	0.67 [0.64; 0.70]	0.39 [0.00; 0.66]	0.08 [0.00; 0.86]
	TTP	0.64 [0.61; 0.68]	0.31 [0.00; 0.61]	0.12 [0.00; 1.00]
	TTF	0.65 [0.62; 0.68]	0.32 [0.00; 0.62]	0.10 [0.00; 1.00]
Leiomyosarcomas (N_{trial} = 14; $N_{patient}$ = 1025)				
6 months	PFS	0.57 [0.51; 0.62]	0.59 [0.15; 0.76]	0.91 [0.00; 1.00]
	TTP	0.55 [0.49; 0.60]	0.58 [0.13; 0.75]	0.97 [0.00; 1.00]
	TTF	0.53 [0.48; 0.58]	0.59 [0.14; 0.76]	0.91 [0.00; 1.00]
12 months	PFS	0.59 [0.54; 0.64]	0.59 [0.16; 0.75]	0.91 [0.00; 1.00]
	TTP	0.52 [0.47; 0.58]	0.58 [0.15; 0.75]	0.97 [0.00; 1.00]
	TTF	0.53 [0.48; 0.58]	0.58 [0.15; 0.75]	0.91 [0.00; 1.00]

[*]CI = confidence interval; PFS = progression-free survival; TTF = time-to-treatment failure; TTP = time-to-progression.
[†]ρSpearman represents the Spearman rank correlation coefficient between the candidate surrogates and overall survival.
[‡]R²WLR represents the coefficient of determination between treatment effect on the candidate surrogates and overall survival based on weighted linear regression models.
[φ]R²2SM represents the coefficient of determination between treatment effect on the candidate surrogates and overall survival based on the two-stage model [4].

(WHO criteria, RECIST criteria). Finally, we could not include all relevant trials, as some of the sponsors and/or PIs did not agree to share their data. This is a recurrent challenge when performing a meta-analysis [34]. Trial design for advanced sarcoma is particularly challenging due to the rarity and the heterogeneity of the disease and treatments, which may contribute to weaken the observed correlations between candidate surrogates and OS [27]. Most STS trials include different clinical phenotypes to increase their statistical power, even though specific RCTs would be required [27–29]. In the present study, the distribution of sarcoma subtypes across trials was highly variable, with proportions ranging from 0% [20] to 18% [16] for liposarcoma, 18% [19] to 100% [20] for leiomyosarcoma and 0% [20] to 14% [22] for synovial sarcoma. Locally advanced and metastatic patients have different prognoses, yet they are often conflated in trials as "advanced" sarcomas. Heterogeneity in terms of sarcoma subtypes might benefit the treatment in terms of OS, but, pre-supposing the drug is inactive, not in terms of PFS. Additionally, assuming the treatment is mainly efficient on specific histological subtypes, the treatment effects estimations might be diluted when all subtypes are combined. Heterogeneity in terms of treatment settings remains between the trials included in our study, which could also have weakened the association between the candidate surrogates and OS. Indeed, 11 trials included only first-line treatment (79% of all patients), one trial included first and second line treatments (3% of patients), one trial included 2nd line treatment only (3% of patients) and one trial included second to fifth line treatments (13% of patients). Central review at study entry is also

likely to interfere. Patients with inappropriate histological subtypes or grades could be included and thus dilute the associations. A majority of the trials included in our study performed radiological central review, and some reviewed histology locally or in a specialized center. Results from our sensitivity analysis on doxorubicin- or ifosfamide-based therapies as first-line metastatic treatment did not significantly differ from our main analysis. We estimated a higher trial-level association between the surrogate candidates and OS when focusing on patients with leiomyosarcomas. However, the individual-level associations were lower than in the primary analysis. This could be due to the limited number of patients included in the subgroup analysis. As a result, even though the results might seem promising, a larger IPD meta-analysis specific to leiomyosarcomas would be necessary to properly conclude. These factors should be accounted for when interpreting the statistically non-significant correlations observed in this meta-analysis.

Finally, absence of surrogacy could also be explained if the candidate endpoints (PFS, TTP, TTF) do not truly predict OS. Indeed, Booth and Eisenhauer, for example, have questioned the mechanisms of actions of some agents, especially those targeting cell signaling and angiogenesis, and whether with chronic administration, they could delay progression for a time but lead to evolutionary changes in tumors, producing a more aggressive phenotype after treatment, thus offsetting the earlier delay in progression [30].

In addition, absence of surrogacy evidence might be related to an absence of the treatment effect on OS.

The Accelerated Approval regulations, instituted by the Food and Drug Administration (FDA) in 1992, allow drugs for serious conditions that filled an unmet medical need to be approved based on a surrogate endpoint. Using a surrogate endpoint enabled the FDA to approve these drugs faster. As a result, an increasing number of anticancer drug product approvals by the FDA are made based on endpoints other than OS [1, 31, 32], some with no sufficient proof of their surrogate validity for OS [32]. In advanced STS for instance, FDA granted approval for pazopanib in 2012 based on proof of benefit for PFS [18], even though at the time no study had assessed trial-level association between PFS and OS. In the context of accelerated approval, the FDA guidance on the term unmet medical need is imprecise. While this is not an issue in advanced sarcoma, recent data have shown that this term can often be overused [33], and as such the use of surrogates through pathways such as accelerated approvals, may be far greater than conditions with true unmet needs [34]. This issue is well illustrated with the example of bevacizumab in metastatic breast cancer; following accelerated approval based on PFS results, this drug was subsequently withdrawn following the publication of OS data [35, 36].

Several conditions have to be met to ensure adequate validation of a surrogate endpoint: (i) a significant quantity of high-quality data, both in terms of trials and patients, (ii) homogeneity, in terms of disease, settings, and mechanisms of action of the drugs, and (iii) strong statistical thresholds. Our meta-analysis did not lead to the validation of a surrogate in advanced sarcoma, as
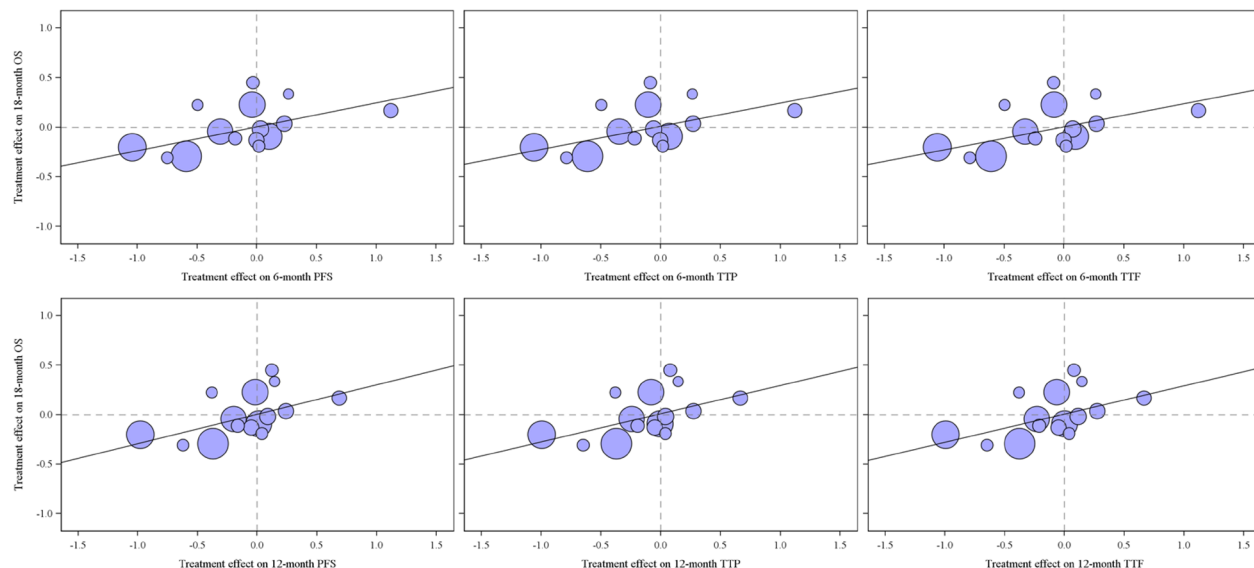
Figure 3: Trial-level association between treatment effects on the candidate surrogates and overall survival*. *Treatment effects estimated by the logarithm of hazard ratios (log[HR]) using the weighted linear regression approach. Each circle represents a trial, and the surface area of the circle is proportional to the size of the corresponding trial. PFS = progression-free survival; TTF = time-to-treatment failure; TTP = time-to-progression; OS = overall survival.

such OS should remain the primary endpoint in phase III trials evaluating systemic treatments in metastatic STS. Nevertheless, we highlighted moderate individual-level associations between the surrogate candidates and OS. These alternative endpoints thus remain useful in testing new treatments [37] in earlier drug development stages, such as phase II trials, or in futility assessment [38, 39].

MATERIALS AND METHODS

This study is registered on the clinical trial registry clinicaltrials.gov (identifier: NCT02873923).

Study selection

We identified trials by using a computerized search on MEDLINE with the following search algorithm: "sarcoma" [MeSH] AND "randomized controlled trial" [Text Word] AND trial [Text Word]. We limited our research to trials published before April the 7th, 2016. We also searched for trials on ClinicalTrials.gov and by contacting European sponsoring groups (EORTC, UNICANCER). Trials were eligible if they met the following criteria: (i) phase II or III randomized trials on humans, (ii) evaluating therapies for adults with advanced (i.e. locally advanced or metastatic) STS, (iii) at least one time-to-event endpoint other than OS as outcome, (iv) published or soon to be published in French or English, (v) signed agreement from the principal investigator and the sponsor, and (vi) available IPD.

Patients, data and outcomes

We gathered IPD from all eligible trials. We assessed the surrogate properties of PFS, TTP and TTF evaluated at six and twelve months for 18-month OS. Outcomes were defined following the international DATECAN guidelines [40]. When none of the events included in the definition was observed, 6- and 12-month PFS, TTP and TTF were censored at the date of last follow-up or 6 months, respectively 12 months, of follow-up whichever came first.

Surrogacy measures

The individual-level surrogacy was assessed following a copula-based approach [4]. The individual-level associations were estimated by the Spearman rank correlation coefficient ($\rho_{Spearman}$) calculated from the copula parameter.

The trial-level surrogacy - the association between the treatment effects - was evaluated with two frameworks. Using a weighted linear regression model (WLR), treatment effects on OS and PFS/TTP/TTF were estimated separately for each trial, based on the logarithm of the hazard ratios

(log[HR]) using Cox proportional hazard models. We assessed the association between the treatment effects using the coefficient of determination (R^2_{WLR}) of a linear regression model weighted by the trial size. The second method follows the two-stage model (2SM) adapted to time-to-event endpoints introduced by Burzykowski et al. [4]. We first simultaneously estimated the treatment effects on OS and on the candidate surrogate endpoints in each trial using a bivariate survival model based on a one-parameter Clayton copula. This approach enables taking into account the correlation between the endpoints in the estimation of the HR. We then estimated the association between the treatment effects (Weibull-distribution-based log[HR]) using an error-in-variable model, which allows taking into account the estimation errors. We assessed the trial-level association using the coefficient of determination (R^2_{2SM}).

All analyses were on an intention-to-treat basis. We reported confidence intervals for a 95% two-sided confidence-level (95% CI). All analyses were performed using SAS software v9.3 following Burzykowski et al. [4].

Strength of association

The strength of the trial-level association was ranked according to the Institute for Quality and Efficiency in Health Care (IQWiG) guidelines [24]: high association (lower limit of the 95% CI of $R^2 \geq 0.72$), low association (higher limit of the 95% CI of $R^2 \leq 0.49$) or medium association (neither low nor high), meaning that the validity of the surrogate remains unclear.

Subgroup analyses

To control for trials' heterogeneity, we performed two subgroup analyses. Firstly, we retained only trials focusing on first line treatment that included doxorubicin- or ifosfamide-based therapies in the control arm. In the second analysis, we included only patients with leiomyosarcomas.

CONCLUSIONS

Even though we highlighted moderate individual-level associations between the surrogate candidates and OS, our meta-analysis did not lead to significant evidence to validate PFS, TTP or TTF as surrogate markers for OS when assessing systemic treatment in advanced STS. PFS, TTP and TTF can be used as primary endpoints in phase II trials or as futility endpoints in phase III trials. However, OS should remain the primary endpoint in phase III trials until sufficient proof of surrogacy is provided. To achieve that goal, improvement in the conduct of sarcoma trials, particularly regarding the selection of histological subtypes, is necessary.

Ethics approval and consent to participate

Not applicable.

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions

Marion Savina: Methodology, formal analysis, software, and writing - original draft, and writing - review and editing.

S Litière, A Italiano, T Burzykowski, F Bonnetain, S Gourgou, V Rondeau, JY Blay, S Cousin, F Duffaud, H Gelderblom, A Gronchi, I Judson, A Le Cesne, P Lorigan, J Maurel, WT van der Graaf, J Verweij: writing - review and editing.

Carine Bellera: Conceptualization, methodology, funding acquisition, project administration, and writing - review and editing.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. Sridhara R, Johnson JR, Justice R, Keegan P, Chakravarty A, Pazdur R. 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; 102:230–43.

2. Mathoulin-Pelissier S, Gourgou-Bourgade S, Bonnetain F, Kramar A. Survival end point reporting in randomized cancer clinical trials: a review of major journals. J Clin Oncol. 2008; 26:3721–6.

3. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostat Oxf Engl. 2000; 1:49–67.

4. 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; 50:405–22.

5. Green E, Yothers G, Sargent DJ. Surrogate endpoint validation: statistical elegance versus clinical relevance. Stat Methods Med Res. 2008; 17:477–86.

6. Renfro LA, Shi Q, Sargent DJ. Surrogate End Points in Soft Tissue Sarcoma: Methodologic Challenges. J Clin Oncol. 2016;

7. Coindre JM, Terrier P, Guillou L, Le Doussal V, Collin F, Ranchère D, Sastre X, Vilain MO, Bonichon F, N'Guyen Bui B. Predictive value of grade for metastasis development in the main histologic types of adult soft tissue sarcomas: a study of 1240 patients from the French Federation of Cancer Centers Sarcoma Group. Cancer. 2001; 91:1914–26.

8. Zer A, Prince RM, Amir E, Abdul Razak A. Evolution of Randomized Trials in Advanced/Metastatic Soft Tissue Sarcoma: End Point Selection, Surrogacy, and Quality of Reporting. J Clin Oncol. 2016; 34:1469–75.

9. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009; 6:e1000097.

10. Nielsen OS, Dombernowsky P, Mouridsen H, Crowther D, Verweij J, Buesa J, Steward W, Daugaard S, van Glabbeke M, Kirkpatrick A, Tursz T. High-dose epirubicin is not an alternative to standard-dose doxorubicin in the treatment of advanced soft tissue sarcomas. A study of the EORTC soft tissue and bone sarcoma group. Br J Cancer. 1998; 78:1634–9.

11. Verweij J, Lee SM, Ruka W, Buesa J, Coleman R, van Hoessel R, Seynaeve C, di Paola ED, van Glabbeke M, Tonelli D, Judson IR. Randomized phase II study of docetaxel versus doxorubicin in first- and second-line chemotherapy for locally advanced or metastatic soft tissue sarcomas in adults: a study of the European organization

for research and treatment of cancer soft tissue and bone sarcoma group. J Clin Oncol. 2000; 18:2081–6.

12. Le Cesne A, Judson I, Crowther D, Rodenhuis S, Keizer HJ, Van Hoesel Q, Blay JY, Frisch J, Van Glabbeke M, Hermans C, Van Oosterom A, Tursz T, Verweij J. Randomized phase III study comparing conventional-dose doxorubicin plus ifosfamide versus high-dose doxorubicin plus ifosfamide plus recombinant human granulocyte-macrophage colony-stimulating factor in advanced soft tissue sarcomas: A trial of the European Organization for Research and Treatment of Cancer/Soft Tissue and Bone Sarcoma Group. J Clin Oncol. 2000; 18:2676–84.

13. Judson I, Radford JA, Harris M, Blay JY, van Hoesel Q, le Cesne A, van Oosterom AT, Clemons MJ, Kamby C, Hermans C, Whittaker J, Donato di Paola E, Verweij J, et al. Randomised phase II trial of pegylated liposomal doxorubicin (DOXIL/CAELYX) versus doxorubicin in the treatment of advanced or metastatic soft tissue sarcoma: a study by the EORTC Soft Tissue and Bone Sarcoma Group. Eur J Cancer. 2001; 37:870–7.

14. van Oosterom AT, Mouridsen HT, Nielsen OS, Dombernowsky P, Krzemieniecki K, Judson I, Svancarova L, Spooner D, Hermans C, Van Glabbeke M, Verweij J; EORTC Soft Tissue and Bone Sarcoma Group. Results of randomised studies of the EORTC Soft Tissue and Bone Sarcoma Group (STBSG) with two different ifosfamide regimens in first- and second-line chemotherapy in advanced soft tissue sarcoma patients. Eur J Cancer. 2002; 38:2397–406.

15. Lorigan P, Verweij J, Papai Z, Rodenhuis S, Le Cesne A, Leahy MG, Radford JA, Van Glabbeke MM, Kirkpatrick A, Hogendoorn PC, Blay JY; European Organisation for Research and Treatment of Cancer Soft Tissue and Bone Sarcoma Group Study. Phase III Trial of Two Investigational Schedules of Ifosfamide Compared With Standard-Dose Doxorubicin in Advanced or Metastatic Soft Tissue Sarcoma: A European Organisation for Research and Treatment of Cancer Soft Tissue and Bone Sarcoma Group Study. J Clin Oncol. 2007; 25:3144–50.

16. Maurel J, López-Pousa A, de Las Peñas R, Fra J, Martín J, Cruz J, Casado A, Poveda A, Martínez-Trufero J, Balañá C, Gómez MA, Cubedo R, Gallego O, et al. Efficacy of Sequential High-Dose Doxorubicin and Ifosfamide Compared With Standard-Dose Doxorubicin in Patients With Advanced Soft Tissue Sarcoma: An Open-Label Randomized Phase II Study of the Spanish Group for Research on Sarcomas. J Clin Oncol. 2009; 27:1893–8.

17. Fayette J, Penel N, Chevreau C, Blay JY, Cupissol D, Thyss A, Guillemet C, Rios M, Rolland F, Fargeot P, Bay JO, Mathoulin-Pelissier S, Coindre JM, Bui-Nguyen B. Phase III trial of standard versus dose-intensified doxorubicin, ifosfamide and dacarbazine (MAID) in the first-line treatment of metastatic and locally advanced soft tissue sarcoma. Invest New Drugs. 2009; 27:482–9.

18. van der Graaf WT, Blay JY, Chawla SP, Kim DW, Bui-Nguyen B, Casali PG, Schöffski P, Aglietta M, Staddon AP, Beppu Y, Le Cesne A, Gelderblom H, Judson IR, et al; EORTC Soft Tissue and Bone Sarcoma Group; PALETTE study group. Pazopanib for metastatic soft-tissue sarcoma (PALETTE): a randomised, double-blind, placebo-controlled phase 3 trial. The Lancet. 2012; 379:1879–86.

19. Bui-Nguyen B, Ray-Coquard I, Chevreau C, Penel N, Bay JO, Coindre JM, Cupissol D, Italiano A, Bonichon F, Lotz JP, Thyss A, Jimenez M, Mathoulin-Pélissier S, et al; GSF-GETO French Sarcoma Group. High-dose chemotherapy consolidation for chemosensitive advanced soft tissue sarcoma patients: an open-label, randomized controlled trial. Ann Oncol. 2012; 23:777.

20. Pautier P, Floquet A, Penel N, Piperno-Neumann S, Isambert N, Rey A, Bompas E, Cioffi A, Delcambre C, Cupissol D, Collin F, Blay JY, Jimenez M, et al. Randomized Multicenter and Stratified Phase II Study of Gemcitabine Alone Versus Gemcitabine and Docetaxel in Patients with Metastatic or Relapsed Leiomyosarcomas: A Federation Nationale des Centres de Lutte Contre le Cancer (FNCLCC) French Sarcoma Group Study (TAXOGEM study). The Oncologist. 2012; 17:1213–20.

21. Judson I, Verweij J, Gelderblom H, Hartmann JT, Schöffski P, Blay JY, Kerst JM, Sufliarsky J, Whelan J, Hohenberger P, Krarup-Hansen A, Alcindor T, Marreaud S, et al; European Organisation and Treatment of Cancer Soft Tissue and Bone Sarcoma Group. Doxorubicin alone versus intensified doxorubicin plus ifosfamide for first-line treatment of advanced or metastatic soft-tissue sarcoma: a randomised controlled phase 3 trial. Lancet Oncol. 2014; 15:415–23.

22. Gelderblom H, Blay JY, Seddon BM, Leahy M, Ray-Coquard I, Sleijfer S, Kerst JM, Rutkowski P, Bauer S, Ouali M, Marreaud S, van der Straaten RJ, Guchelaar HJ, et al. Brostallicin versus doxorubicin as first-line chemotherapy in patients with advanced or metastatic soft tissue sarcoma: an European Organisation for Research and Treatment of Cancer Soft Tissue and Bone Sarcoma Group randomised phase II and pharmacogenetic study. Eur J Cancer. 2014; 50:388–96.

23. Bui-Nguyen B, Butrynski JE, Penel N, Blay JY, Isambert N, Milhem M, Kerst JM, Reyners AK, Litière S, Marréaud S, Collin F, van der Graaf WT; European Organisation for Research and Treatment of Cancer Soft Tissue and Bone Sarcoma Group (EORTC/STBSG) and the Sarcoma Alliance for Research through Collaboration (SARC). A phase IIb multicentre study comparing the efficacy of trabectedin to doxorubicin in patients with advanced or metastatic untreated soft tissue sarcoma: the TRUSTS trial. Eur J Cancer. 2015; 51:1312–20.

24. Institute for Quality and Efficiency in Health Care (IQWiG). Validity of surrogate endpoints in oncology: Executive summary of rapid report A10-05, Version 1.1. In: Institute

for Quality and Efficiency in Health Care: Executive Summaries. Cologne, Germany: Institute for Quality and Efficiency in Health Care (IQWiG); 2005 [cited 2016 Dec 30]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK198799/.

25. Taylor RS, Elston J. The use of surrogate outcomes in model-based cost-effectiveness analyses: a survey of UK Health Technology Assessment reports. Health Technol Assess Winch Engl. 2009; 13:iii, ix–xi, 1–50.

26. 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.

27. Constantinidou A, van der Graaf WTA. The fate of new fosfamides in phase III studies in advanced soft tissue sarcoma. Eur J Cancer. 1990. 2017; 84:257–61.

28. Kummar S, Allen D, Monks A, Polley EC, Hose CD, Ivy SP, Turkbey IB, Lawrence S, Kinders RJ, Choyke P, Simon R, Steinberg SM, Doroshow JH, Helman L. Cediranib for metastatic alveolar soft part sarcoma. J Clin Oncol. 2013; 31:2296–302.

29. Constantinidou A, Miah A, Pollack S, Jones RL. New drugs and clinical trial design in advanced sarcoma: have we made any progress. Future Oncol. 2013; 9:1409–11.

30. Booth CM, Eisenhauer EA. Progression-free survival: meaningful or simply measurable. J Clin Oncol. 2012; 30:1030–3.

31. Kim C, Prasad V. Cancer Drugs Approved on the Basis of a Surrogate End Point and Subsequent Overall Survival: An Analysis of 5 Years of US Food and Drug Administration Approvals. JAMA Intern Med. 2015; 175:1992.

32. 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; 91:713–25.

33. Lu E, Shatzel J, Shin F, Prasad V. What constitutes an "unmet medical need" in oncology? An empirical evaluation of author usage in the biomedical literature. Semin Oncol. 2017; 44:8–12.

34. 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.

35. Miller K, Wang M, Gralow J, Dickler M, Cobleigh M, Perez EA, Shenkier T, Cella D, Davidson NE. Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. N Engl J Med. 2007; 357:2666–76.

36. Carpenter D, Kesselheim AS, Joffe S. Reputation and precedent in the bevacizumab decision. N Engl J Med. 2011; 365:e3.

37. LeBlanc M, Tangen C. Surrogates for Survival or Other End Points in Oncology. JAMA Oncol. 2016; 2:263–4.

38. Goldman B, LeBlanc M, Crowley J. Interim futility analysis with intermediate endpoints. Clin Trials. 2008; 5:14–22.

39. Redman MW, Goldman BH, LeBlanc M, Schott A, Baker LH. Modeling the relationship between progression-free survival and overall survival: the phase II/III trial. Clin Cancer Res. 2013; 19:2646–56.

40. Bellera CA, Penel N, Ouali M, Bonvalot S, Casali PG, Nielsen OS, Delannes M, Litière S, Bonnetain F, Dabakuyo TS, Benjamin RS, Blay JY, Bui BN, et al; Definition for the Assessment of Time-to-event Endpoints in Cancer Trials Initiative. Guidelines for time-to-event end point definitions in sarcomas and gastrointestinal stromal tumors (GIST) trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials). Ann Oncol. 2015; 26:865–72.