ELSEVIER

CrossMark

Journal of Clinical Epidemiology 68 (2015) 833–842

Journal of Clinical Epidemiology

Meta-analyses of randomized controlled trials show suboptimal validity of surrogate outcomes for overall survival in advanced colorectal cancer

Oriana Ciani[a,b,*], Marc Buyse[c,d], Ruth Garside[e], Jaime Peters[a], Everardo D. Saad[f], Ken Stein[a], Rod S. Taylor[a]

[a]*Institute of Health Research, University of Exeter Medical School, Veysey Building, Salmon Pool Lane, Exeter EX2 4SG, UK*
[b]*Centre for Research on Health and Social Care Management, Bocconi University, via Rontgen, 1, Milan 20141, Italy*
[c]*Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Hasselt University, Universitaire Campus, 3590 Diepenbeek, Belgium*
[d]*International Drug Development Institute (IDDI), 185 Alewife Brook Parkway, Suite 410, Cambridge, MA 02138, USA*
[e]*European Centre for Environment and Human Health University of Exeter Medical School Knowledge Spa, Royal Cornwall Hospital, Truro TR1 3HD, UK*
[f]*Dendrix Research, Rua Joaquim Floriano, 72/24, Sao Paulo 04534-000, Brazil*

Accepted 25 February 2015; Published online 6 March 2015

Abstract

Objectives: To quantify and compare the treatment effects on three surrogate end points, progression-free survival (PFS), time to progression (TTP), and tumor response rate (TR) vs. overall survival (OS) based on a meta-analysis of randomized controlled trials (RCTs) of drug interventions in advanced colorectal cancer (aCRC).

Study Design and Setting: We systematically searched for RCTs of pharmacologic therapies in aCRC between 2003 and 2013. Trial characteristics, risk of bias, and outcomes were recorded based on a predefined form. Univariate and multivariate random-effects meta-analyses were used to estimate pooled summary treatment effects. The ratio of hazard ratios (HRs)/odds ratios (ORs) and difference in medians were used to quantify the degree of difference in treatment effects on the surrogate end points and OS. Spearman ρ, surrogate threshold effect (STE), and R^2 were also estimated across predefined trial-level covariates.

Results: We included 101 RCTs. In univariate and multivariate meta-analyses, we found larger treatment effects for the surrogates than for OS. Compared with OS, treatment effects were on average 13% higher when HRs were measured and 3% to 45% higher when ORs were considered; differences in median PFS/TTP were higher than on OS by an average of 0.5 month. Spearman ρ ranged from 0.39 to 0.80, mean R^2 from 0.06 to 0.65, and STE was 0.8 for HR_{PFS}, 0.64 for HR_{TTP}, or 0.28 for OR_{TR}. The stratified analyses revealed high variability across all strata.

Conclusion: None of the end points in this study were found to achieve the level of evidence (ie, mean $R^2_{trial} > 0.60$) that has been set to select high or excellent correlation levels by common surrogate evaluation tools. Previous surrogacy relationships observed between PFS and TTP vs. OS in selected settings may not apply across other classes or lines of therapy.

Keywords: Surrogate outcome; Colorectal cancer; PFS; TTP; Tumor response; Health technology assessment

1. Introduction

Surrogate end points have been defined as biomarkers or intermediate outcomes that can substitute for a final patient-relevant end point to successfully measure the effect of health interventions [1]. In colorectal cancer, the second commonest cause of cancer-related mortality in high-income countries [2], predictive end points for overall survival (OS) are needed to accelerate the availability of promising new therapies for patients. A number of surrogate end points for OS in clinical oncology trials have been proposed, including progression-free survival (PFS), time to progression (TTP), and tumor response rate (TR) [3–5]. However, to ensure that these surrogate end points provide the same answer as the final end point (OS) about the experimental therapy, they should undergo a process of surrogate validation [6]. Several authors have dealt with the validation of PFS [7–12], TTP [8,10,13], or TR [13,14] as surrogate end points for OS in advanced colorectal cancer

Conflict of interest: O.C. received a PhD scholarship from the Peninsula College of Medicine and Dentistry. Other coauthors have no conflict of interest to declare.

* Corresponding author. Evidence Synthesis & Modelling for Health Improvement, Institute of Health Research, University of Exeter Medical School, Veysey Building, Salmon Pool Lane, Exeter EX2 4SG, UK. Tel.: 44 7575918306; fax: +39 02 5836.2598.

E-mail address: O.Ciani@exeter.ac.uk (O. Ciani).

http://dx.doi.org/10.1016/j.jclinepi.2015.02.016
0895-4356/

What is new?

- The meta-analyses showed that treatment effect sizes were always larger for the surrogate end points than for overall survival (OS). The stratified analyses revealed high variability across all strata.
- Progression-free survival (PFS), time to progression (TTP), and tumor response rate (TR) have been proposed as surrogate for OS in advanced colorectal cancer (aCRC); however, previous surrogacy relationship observed in selected aCRC therapies may not directly apply across other classes or lines of therapy.
- None of the end points in this study were found to achieve the level of evidence that has been set to select high or excellent correlation levels by common surrogate evaluation tools. Where PFS and TTP are deemed acceptable surrogates for OS, policy makers still need to consider that the anticipated treatment effect on OS is likely to be smaller than that observed on the surrogate measure when weighing up the evidence in their licensing and coverage decisions. TR should not be used as a surrogate end point for OS when evaluating the efficacy of drug interventions in aCRC.

(aCRC) over the last decade. Although most of the studies are of high quality, some are not based on systematic review of the available evidence, either because they were based on opportunistically available individual-patient data (IPD) [7,9,11,12,14] or focused on subgroups of trials and therapies [8,13] and did not, therefore, provide a comprehensive examination of the issue. The present study seeks to overcome these limitations by systematically looking at all available randomized controlled trials (RCTs), across drug classes and lines of therapy, and considering different approaches to surrogate validation, with the primary aim of quantifying and comparing treatment effects on surrogates and on OS.

2. Methods

We conducted and reported this systematic review in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement [15].

2.1. Data sources and selection strategy

We searched the following databases from 2003 to January 31, 2013: MEDLINE, EMBASE (via OVID), and the Cochrane Central Register of Controlled Trials. A copy of the bibliographic searches is provided in the Supplementary Material at www.jclinepi.com. We limited our searches to the last 10 years of drug interventions in metastatic colorectal cancer to limit the heterogeneity in our sample and, at the same time, to reflect current clinical practice in most developed countries. We checked citations in identified studies and systematic reviews already known to the authors [16] as additional sources of potentially eligible trials.

Trials were included if they were RCTs in advanced or metastatic colorectal cancer assessing a pharmacologic therapy against either a placebo or other drug therapy. Trials had to report OS and either PFS or TTP or TR. We excluded adjuvant setting trials and trials assessing radiotherapy, supportive-care drugs, other nonantineoplastic drugs, nondrug treatments, and trials that were stopped early, with accrual rate less than 70% of the target sample size. When multiple publications of the same RCT were available, only the most recent one reporting both surrogate and final end points was included. Titles and abstracts were screened independently by two reviewers, and disagreements were resolved by full-text retrieval and, when necessary, involvement of a third reviewer.

2.2. Data extraction

One reviewer extracted the data using a standardized form, and a second reviewer independently checked the extraction. Information collected included: general characteristics of the trial (ie, study design, sample size), patient characteristics (ie, median age, performance status), treatments under comparison, risk of bias assessment (using the Cochrane Collaboration tool [17]), and treatment effects on OS and PFS, TTP, or TR. In multiarm trials, all available between-arm comparisons were recorded. OS was defined as the time from randomization to death from any cause, with patients censored when they are last seen alive or when they are lost to follow-up [18]; PFS was defined as the time between randomization and tumor progression (however defined) or death from any cause; and TTP as the time between randomization and tumor progression (however defined), with censoring of patients who died without prior documentation of progression. Tumor response is based on objective tumor measurements by imaging methods that allow the classification of patients with a complete or partial confirmed best response as responders. Responses are usually determined according to the Response Evaluation Criteria in Solid Tumors guidelines [19] or the World Health Organization recommendations [20].

For OS, PFS, and TTP, the hazard ratio (HR) and median survival time, together with the 95% confidence intervals (CIs) for each arm, were recorded whenever available. The number of events (ie, deaths or tumor progressions or tumor responses) were also recorded to estimate odds ratios (ORs).

2.3. Statistical analyses

We derived the sample size for this present study based on a previous publication comparing the treatment effects

in RCTs ($N = 185$) assessing surrogate vs. final patient-relevant outcomes [21]. Using the subgroup of RCTs in this study assessing drug therapies in oncology, we derived a summary OR (95% CI) of 0.48 (0.39 to 0.59) and 0.68 (0.61 to 0.77) for evaluating surrogate or final primary end points, respectively. To detect this observed relative difference of 29% in the ORs for surrogate compared with the final outcomes at 80% power and α level of 5%, we estimated we would require a total of 114 trials (38 trials for each surrogate end point, that is, PFS, TTP, TR).

2.4. Estimation of pooled treatment effects

We compared treatment effects on OS and treatment effects on the surrogate end points using several analytical approaches.

For each individual trial, ORs and HRs were expressed so that a value less than 1.0 indicated beneficial effect of the intervention compared with control. To exploit all available data in trial reports, we also considered median survival time differences across arms: a positive difference in median survival time on absolute scale indicates a more favorable effect of intervention than control. Using each of these three treatment effect metrics (ORs, HRs, and median differences), separate der Simonian and Laird random-effects univariate meta-analyses were used to calculate the pooled treatment effect (95% CI) for OS and each of the three surrogate end points across included trials. Statistical heterogeneity as expressed by the I^2 statistic was examined [22], and Egger's or Harbord's tests were used to assess potential small-study effects and publication bias for all outcomes [23,24]. We also performed separate random-effects multivariate meta-analyses. Multivariate meta-analysis combines estimates of several related parameters over several studies. In this case, to estimate pooled treatment effects taking into account the within-trial relationship between OS and each of the three surrogate end points we used the Stata "mvmeta" command [25]. As we did not have access to IPD, we assumed within-study correlations varying between 0 and 1.0 and checked the likelihood of each attributed correlation value from previous IPD meta-analyses, which explored the association between surrogate end points and OS in metastatic colorectal cancer [7].

2.5. Comparison of OS and surrogate end points treatment effects

We sought to compare treatment effects on OS and on the surrogate end points by estimating the ratio of HRs (RHR), ratio of ORs (ROR), and difference in the absolute differences (DAD) in median survival times. Where an RHR (for PFS and TTP) or ROR (for tumor progression and response) is greater than 1.0 and DAD in median survival times is lower than 0, a more beneficial intervention effect for the surrogate end point than for OS is implied. We implemented a univariate meta-regression–based approach proposed [26] to calculate the ROR that was extended to calculate RHR and DAD [27]. Because this method is based on HRs, ORs, and differences in medians for OS and surrogate end points that are independent, being derived from separate trials, in our primary analysis, we calculated the within-trial RHR, ROR, and DAD (for each individual trial difference in OS and surrogates) using the indirect treatment comparison approach [28]. These within-trial estimates were then pooled across trials using random-effects univariate meta-analyses. For multiarm trials, we selected one of the available comparisons to contribute data to our primary analyses based on clinical judgment and without regard to any correlation between surrogate end points and OS. All analyses were then repeated using all available comparisons from multiarm trials. Because included RCTs did not have a common control therapy, we determined in advance for each study which arm would be the reference group and the experimental group, taking into account the innovativeness of the regimen, common oncology practice, and number of combined agents in additive comparisons.

2.6. Surrogacy metrics

In addition, we calculated commonly used indicators of surrogacy validity [16]: Spearman ρ correlation coefficient [10,29,30], the R^2_{trial} (95% CI) for the relationship between the treatment effects on the surrogate and the final outcomes variables on the log scale derived from a weighted least-squares regression [7,13,31–33], and the surrogate threshold effect (STE), that is, the intercept of the regression line with zero effect on OS [34].

2.7. Stratification according to trial-level covariates

To assess how surrogacy indicators might vary across included trials, we stratified our analyses according to a predefined set of trial-level covariates: type of intervention (systemic chemotherapy vs. other agents), type of comparison (additive vs. other type of comparisons), and comparator (active vs. inactive, that is, placebo or best supportive care), stage of therapy (first line vs. other lines of therapy), study design (superiority vs. other), a balanced use of postrandomization therapies or crossover (low other bias factors vs. high/unclear other biases), type of primary end point (final vs. surrogate), phase of the study (III/IV vs. II), funding source (for profit vs. other), center status (multicenter vs. single center).

Data were assumed to be missing at random, and no data imputation was undertaken. Data analyses were performed in Stata 12 StataCorp LP, College Station, Texas, USA.

3. Results

3.1. Characteristics of included RCTs

In total, 101 RCTs were included that reported 117 trial arm comparisons and randomized 40,243 patients

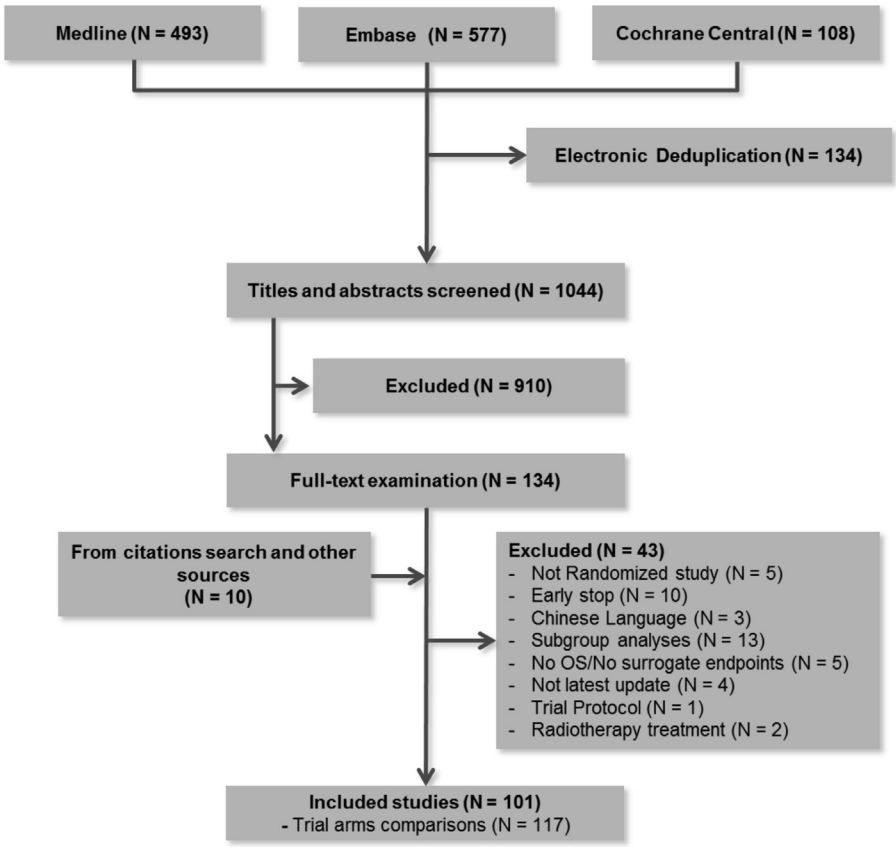

Fig. 1. Flow diagram of study selection process.

(Fig. 1, list of articles available in the Supplementary Material at www.jclinepi.com). Table 1 provides a summary of trial and patients characteristics. The publication years spanned 2003 and 2012, with a maximum of 15 publications in year 2011. On average recruitment periods lasted 29 months and publications occurred approximately 4 years after the start of recruitment. The stated primary end point was OS or survival rate in 27 (27%) of all trials, the remainder defining the primary outcome as either a surrogate (25% PFS, 32% TR, 6% TTP) or toxicity (4%) end point. As a result of the randomization, all population characteristics and performance status [35] seemed well balanced across the study arms (data not shown). The RCTs that assessed systemic chemotherapy alone contained one or more among fluoropyrimidines (fluorouracil, capecitabine, tegafur, doxifluridine), irinotecan, or oxaliplatin. A few studies had additional chemotherapeutic agents, such as raltitrexed, methotrexate, pemetrexed, cisplatin, mitomycin C, and vincristine. For hepatic intra-arterial chemotherapy, floxuridine was the agent most commonly used. The most frequent antiangiogenic compound was bevacizumab, whereas the anti-EGFR agents assessed in RCTs were cetuximab, panitumumab, or gefitinib. Finally, the targeted agents not belonging to the previous two categories were celecoxib, perifosine, tipifarnib, and the anticarcinoembryonic antigen antibody (3H1).

3.2. Risk of bias assessment

A number of trials failed to provide sufficient detail to assess their potential risk of bias (Table 2). Where details were provided, all trials reported evidence of appropriate methods for random sequence generation and allocation concealment. A number of trials ($N = 43$) were described as ‘‘open label’’ and therefore considered at risk of bias due to lack of blinding of participant and clinical personnel. However, several trials ($N = 26$) stated procedures for independent blinded assessment of tumor response or progression. Few trials ($N = 7$) reported substantive ($>20\%$) data losses at follow-up or failed to report the outcome findings for all specified outcomes ($N = 8$). Sixteen publications (16%) reported that treatment crossover across arms in the trial was allowed, whereas use of postprogression therapies was reported in 54 (53%). In 14 of these cases, there appeared to be substantive imbalance in postprogression therapy between intervention and control groups likely to result in performance bias that is a systematic difference between groups in the care that is provided.

Table 1. Characteristics of 101 included RCTs

Study characteristics	*N* (%)
Phase:	
II	34 (34)
III–IV	67 (66)
Study design:	
Superiority	60 (59)
Equivalence or Noninferiority	14 (14)
Noncomparative studies	27 (27)
N arms:	
2	84 (83)
3	13 (13)
>3	4 (4)
Funding source:	
For profit	47 (46)
Not for profit	20 (20)
Mixed or not reported	34 (34)
Setting	
International	29 (29)
Europe	15 (15)
Uninational	57 (56)
Center status:	
Multicenter	95 (94)
Single center	6 (6)
Median follow-up (mo)	24 (7.5–60.6)[a]
Patients randomized	290 (35–2,035)[a]
Median age	62 (50–75)[a]
Male proportions	0.61 (0.08–0.79)[a]
Line of therapy:	
First	76 (75)
Second or mix first–second	20 (20)
Third or mix second–third	5 (5)
Treatment:	
Systemic chemotherapy	56 (55)
Anti-EGFR antibodies	15 (15)
Angiogenesis inhibitors	20 (20)
Other molecular-targeted agents	5 (5)
IHA chemotherapy	5 (5)
Comparator:	
Active	87 (86)
Inactive	14 (14)
Type of comparison:	
Additive[b]	47 (47)
Substitutive	31 (31)
Other	23 (23)

Abbreviations: RCT, randomized controlled trial; anti-EGFR, anti-epidermal growth factor receptor; IHA, intrahepatic arterial.

Values are numbers (percentages) unless otherwise stated.

[a] Median (range).

[b] Additive comparison when treatment is compared with the same one plus an additional agent, substitutive comparison where one of the agents in the association is substituted by another one, and other comparisons for the remainder, including different schedules or ways of administration.

Table 2. Risk of bias assessment

Domain	Low	Unclear	High
Random sequence generation	30 (30)	71 (70)	—
Allocation concealment	34 (34)	67 (66)	—
Blinding of participants and personnel	11 (11)	47 (46)	43 (43)
Blinding of outcome assessment	26 (26)	58 (57)	17 (17)
Incomplete outcome data	79 (78)	15 (15)	7 (7)
Selective reporting	82 (81)	11 (11)	8 (8)
Use of postprogression therapies[a]	28 (28)	59 (58)	14 (14)

Number of trials (percentages) reported.

[a] Low risk: no postprogression treatment allowed when patients failed to respond to allocated therapy or subsequent treatment allowed and shown to be similar across arm; high risk: postprogression treatment allowed and shown to be different across arms.

3.3. Treatment effects on OS and surrogate end points

Depending on type of end point and metrics (HRs, ORs, or medians) used, 9 to 58 trials contributed to each pooled treatment-effect estimation. Univariate and multivariate random-effects meta-analyses demonstrate a significant benefit of the intervention relative to control for OS and each of the surrogate end points (Table 3 and Table 4). Consistently across HR, OR, and difference of median survival times, treatment effect sizes were larger for the surrogate end points than for OS. The multivariate meta-analyses gave broadly similar estimates to univariate meta-analyses (Table 4). Estimates of between-study covariance decrease as the assumed within-study correlation increases, suggesting between-study variance compensates for lack of correlation between the end points of interest in univariate meta-analyses.

Moderate to substantial levels of heterogeneity (I^2 statistic > 30%) were seen. There was no evidence of small-study effect bias, except for HR_{PFS} (Egger's test bias = −2.68; P = 0.022) and difference in median PFS time (Egger's test bias = 1.02; P = 0.027; Table 3).

3.4. Measures of surrogacy validity

There was evidence of moderate to high correlation [36] between the treatment effects on OS and on each of the three surrogate end points (ρ between 0.39 and 0.80). The coefficients of determination for the linear relationships (R^2_{trial}) and related 95% CIs are shown in Table 5. Estimated STEs were consistently lower than 1.0 for HR or OR and greater than 0 for the difference in medians. For example, a $HR_{PFS} < 0.8$ would need to be observed to predict a less than 1.0 HR_{OS} (Table 5).

3.5. Comparison of treatment effects on OS and surrogate end points

The within-study paired comparison analysis showed for PFS, TTP, and TR that the estimated RHR and ROR exceeded 1.0 and the DAD was less than 0.0, indicating a larger treatment effect on the surrogate end points than on OS (Table 5). Compared with OS, relative treatment effects on the surrogate end points were on average 13% higher when HRs were measured and 3% to 45% higher when ORs were considered. On the absolute scale, the treatment effects on PFS or TTP were higher than on OS by an average of 0.5 months. The random-effects logistic metaregression models provided similar results. In a sensitivity analysis, 16 additional arm comparisons

Table 3. Summary treatment effects[a] on OS, PFS, TTP, or TR

Study endpoints	OS/Deaths	PFS	TTP	TR
HR	0.95 (0.90 to 0.99) $P = 0.030$, $I^2 = 64.3\%$ Egger's test $P = 0.526$, $N = 49$	0.84 (0.78 to 0.91) $P = 0.000$, $I^2 = 83.9\%$ Egger's test $P = 0.022$, $N = 39$	0.75 (0.62 to 0.89) $P = 0.002$, $I^2 = 84.1\%$ Egger's test $P = 0.983$, $N = 9$	NA
OR	0.95 (0.87 to 1.04) $P = 0.270$, $I^2 = 27.1\%$ Harbord's test $P = 0.602$, $N = 34$	NA	0.88 (0.73 to 1.07) $P = 0.203$, $I^2 = 56.9\%$ Harbord's test $P = 0.535$, $N = 17$	0.71 (0.64 to 0.80) $P = 0.000$, $I^2 = 77.7\%$ Harbord's test $P = 0.129$, $N = 95$
MD	0.23 (−0.44 to 0.90) $P = 0.504$, $I^2 = 51\%$ Egger's test $P = 0.630$, $N = 58$	0.76 (0.45 to 1.06) $P = 0.000$, $I^2 = 84.9\%$ Egger's test $P = 0.027$, $N = 41$	0.56 (−0.16 to 1.29) $P = 0.130$, $I^2 = 71\%$ Egger's test $P = 0.114$, $N = 24$	NA

Abbreviations: OS, overall survival; PFS, progression-free survival; TTP, time to progression; TR, tumor response; HR, hazard ratio; NA, not applicable; OR, odds ratio; MD, median difference.

[a] Der Simonian and Laird random-effects method. Treatment OR and HR lower than 1.0 correspond to beneficial effects, whereas positive weighted (by inverse variance) differences in treatment vs. control median survival times correspond to beneficial effects. Effect estimates (95%CI), *P*-value for significance of effect size, I^2statistic, number of trials, and Egger's or Harbord's modified tests for small-study effect reported.

from 17 multiarm trials were considered, and all the analyses were replicated without any apparent discrepancy with the primary analyses reported above (see Supplementary Material/eTable 1 at www.jclinepi.com). There was some evidence of a difference in the RHRs or RORs for OS and the surrogate end points across predefined trial-level covariate strata. Higher RHRs were seen for OS vs. PFS for trials with additive comparisons and for OS vs. TTP for trials with profit funding and inactive comparator. Higher ROR for OS vs. TR was observed for trials of second or later line therapy, superiority trials, and trials reporting OS as the primary end

Table 4. Summary treatment effects – results of the multivariate meta-analyses

Treatment effect – results of bivariate meta-analysis[a]			
Study endpoints	**OS vs. PFS**	**OS vs. TTP**	**OS vs. tumor response**
HR vs. HR			NA
($\rho = 0$)	0.94 (0.90 to 0.99) vs. 0.84 (0.77 to 0.91) [0.99]	0.95 (0.90 to 0.99) vs. 0.82 (0.70 to 0.95) [0.78]	
($\rho = 0.5$)	0.95 (0.91 to 0.99) vs. 0.83 (0.76 to 0.91) [0.86]	0.95 (0.91 to 0.99) vs. 0.82 (0.72 to 0.94) [0.70]	
($\rho = 0.75$)	0.95 (0.91 to 0.99) vs. 0.83 (0.76 to 0.90) [0.78]	0.95 (0.91 to 0.99) vs. 0.82 (0.72 to 0.93) [0.66]	
($\rho = 0.95$)	0.95 (0.91 to 0.99) vs. 0.83 (0.76 to 0.90) [0.71]	0.95 (0.91 to 0.99) vs. 0.82 (0.73 to 0.93) [0.61]	
OR vs. OR	NA		
($\rho = 0$)		0.95 (0.87 to 1.04) vs. 0.94 (0.78 to 1.13) [1.00]	0.97 (0.88 to 1.06) vs. 0.71 (0.63 to 0.80) [0.50]
($\rho = 0.5$)		0.96 (0.87 to 1.05) vs. 0.93 (0.76 to 1.14) [0.60]	0.97 (0.89 to 1.06) vs. 0.72 (0.64 to 0.80) [0.21]
($\rho = 0.75$)		0.96 (0.88 to 1.05) vs. 0.92 (0.74 to 1.13) [0.32]	0.98 (0.89 to 1.07) vs. 0.72 (0.64 to 0.80) [0.05]
($\rho = 0.95$)		0.96 (0.87 to 1.05) vs. 0.90 (0.72 to 1.12) [−0.16]	0.99 (0.90 to 1.08) vs. 0.72 (0.64 to 0.80) [−0.09]
MD vs. MD			NA
($\rho = 0$)	0.03 (−0.63 to 0.69) vs. 0.68 (0.27 to 1.10) [1.00]	0.15 (−0.55 to 0.86) vs. 0.74 (0.02 to 1.45) [0.78]	
($\rho = 0.5$)	0.10 (−0.58 to 0.78) vs. 0.67 (0.26 to 1.08) [0.90]	0.18 (−0.52 to 0.88) vs. 0.70 (0.05 to 1.35) [0.60]	
($\rho = 0.75$)	0.21 (−0.49 to 0.91) vs. 0.65 (0.24 to 1.06) [0.74]	0.19 (−0.50 to 0.89) to 0.68 (0.07 to 1.28) [0.50]	
($\rho = 0.95$)	0.31 (−0.41 to 1.02) vs. 0.65 (0.22 to 1.08) [0.54]	0.20 (−0.50 to 0.90) vs. 0.63 (0.08 to 1.17) [0.34]	

Abbreviations: OS, overall survival; PFS, progression-free survival; TTP, time to progression; HR, hazard ratio; NA, not applicable; OR, odds ratio; MD, median difference.

[a] Bivariate meta-analyses with residual maximum likelihood estimation of outcomes with common assumed within-study correlations (ρ). Between-study covariance reported in squared brackets.

Table 5. Surrogacy validity and comparison of treatment effects between OS/deaths and PFS, TTP, or TR

OS vs. surrogate treatment effect contrasts			
Study endpoints	**OS vs. PFS**	**OS vs. TTP**	**OS vs. tumor response**
HR vs. HR within-study paired analysis[a] metaregression[b]	1.13 (1.06 to 1.20) $P < 0.001$ 1.13 (1.02 to 1.25) $P = 0.026$ $N = 36$ STE = 0.8 $\rho = 0.75$*** R^2 (95%CI) = 0.34 (0.10, 0.59)	1.13 (1.02 to 1.25) $P = 0.020$ 1.14 (0.83 to 1.57) $P = 0.389$ $N = 9$ STE = 0.61 $\rho = 0.80$*** R^2 (95%CI) = 0.65 (0.09, 0.92)	NA
OR vs. OR within-study paired analysis[a] metaregression[b]	NA	1.03 (0.85 to 1.24) $P = 0.754$ 1.03 (0.76 to 1.39) $P = 0.850$ $N = 13$ STE = 0.64 $\rho = 0.39$ R^2 (95%CI) = 0.25 (0, 0.68)	1.45 (1.19 to 1.77) $P < 0.001$ 1.44 (1.14 to 1.83) $P = 0.003$ $N = 32$ STE < 0.28 $\rho = 0.53$*** R^2 (95%CI) = 0.06 (0.01, 0.29)
DAD vs. DAD within-study paired analysis[a] metaregression[b]	−0.50 (−0.85 to −0.16) $P < 0.001$ −0.19 (−1.14 to 0.75) $P = 0.689$ $N = 36$ STE = 2 $\rho = 0.59$*** R^2 (95%CI) = 0.52 (0.26, 0.72)	−0.52 (−1.14 to 0.10) $P = 0.098$ −1.09 (−2.43 to 0.26) $P = 0.110$ $N = 21$ STE > 4.0 $\rho = 0.54$** R^2 (95%CI) = 0.43 (0.10, 0.72)	NA

Abbreviations: OS, overall survival; PFS, progression-free survival; TTP, time to progression; TR, tumor response; HR, hazard ratio; STE surrogate threshold effect; NA, not applicable; OR, odds ratio; DAD, difference in absolute median difference.

[a] Der Simonian and Laird random-effects meta-analyses on within-study paired outcomes. For OR and HR, effect estimates >1.0 correspond to higher beneficial effects measured on the surrogate. For DAD, effect estimates <0 correspond to higher beneficial effects measured on the surrogate. STE calculated as the value of the treatment effect on the surrogate outcome at which the linear regression prediction bands cross the zero effect line for the treatment effect on OS. Spearman ρ correlation coefficient is shown on log variables. *P-value < 0.10, **P-value < 0.05, ***P-value < 0.01. R^2 derived from least-squares linear regressions on log variables weighted by the inverse of variance.

[b] Random-effects logistic meta-regressions with residual maximum likelihood estimation for the between-study variance and Knapp–Hartung variance estimator for the coefficients. Only trials contributing to the within-study paired analysis considered. Ratio of OR (95%CI) and ratio of HR (95%CI) > 1.0 and DAD < 0 correspond to higher beneficial effects measured on the surrogate.

point. An informal comparison of Spearman correlation coefficient and R^2_{trial} point estimates also suggests high variability across identified strata (see Supplementary Material/eTable 2 at www.jclinepi.com).

4. Discussion

Surrogate end points are attractive to measure as they allow the efficacy of promising treatments to be established earlier than would be possible if OS had to be observed. However, to assess whether a new intervention will provide a significant survival benefit based on the effect on a surrogate measure, a reliable and plausible prediction equation should be estimated. In our systematic review and meta-analysis of 101 RCTs of pharmacologic interventions in aCRC published between 2003 and 2013, we found moderate to high levels of correlation (ρ ranging from 0.39 to 0.80) between treatment effects on OS and treatment effects on PFS, TTP, and TR. However, the coefficients of determination for the underlying linear prediction models (R^2_{trial}) were between 0.06 and 0.65, meaning that, in general, less than half of the variability in the survival benefits could be explained by the variability in the surrogate treatment effects. Moreover, the treatment effect observed on the surrogate end point appears always to be larger than that observed on the final end point, by 3% to 45%. A relative larger treatment effect observed on the surrogate end point translates into an STE as high as 0.8 for HR_{PFS}, 0.64 for HR_{TTP}, or 0.28 for OR_{TR} to achieve a benefit in terms of OS.

4.1. Comparison with previous findings

Different techniques and approaches have been proposed for the evaluation of surrogate end points [37,38]; however, the hallmark for a valid surrogate is that differences or changes observed on it must accurately reflect changes in the final end point [39]. A theoretical verification of a small treatment effect (expressed as a binary variable) on the final outcome than on the surrogate outcome even if there is strong association between the surrogate and the patient-important outcome has been provided [40]. In a recent publication [21], treatment-effect estimates in a matched cohort of RCTs across several diseases were found to be relatively larger (up to 47%) in RCTs assessing surrogate vs. final patient-relevant primary end points. Those findings are confirmed by the results of the present study, where HR_{PFS}, HR_{TTP} and OR_{TR} were all significantly larger (by 13% to 45%) than HR_{OS} and OR_{death} derived from the same cohort of RCTs.

Our results appear generally less supportive of PFS, TTP, or TR as surrogate measures for OS in aCRC than previous meta-analyses that have examined this issue. For PFS, our coefficient of determination ($R^2_{\text{trial}} = 0.34$) was in line with that estimated by Burzykowski et al. [11]

(R^2_{trial} = 0.33) but lower than those observed in other studies, ranging from 0.53 to 0.98 [7–10,12]. Spearman correlation coefficients, both for HR (ρ = 0.75) and difference in medians (ρ = 0.59), were lower than previously reported (ρ = 0.82 and 0.74, respectively) [8,9]. However, although previous studies focused on first-line treatment, 25% of RCTs in our sample relate to further lines of therapy. The STE for HR_{PFS} (0.80) compares to that proposed by Buyse et al. [7] for fluoropyrimidine-based treatments (0.86), both higher than that estimated by Burzykowski and Buyse [12] (0.12).

Fewer studies have explored the relationship between TTP and OS [8,10,13]. Their estimated ρ (0.52) and R^2_{trial} (0.32) are comparable to those found in this study (ρ = 0.54; R^2_{trial} = 0.43). Johnson et al. [13] did estimate an STE of 3.3 months for a clinical trial of 400 patients, the average sample size in our database, whereas our estimate is above 4 months.

Although our estimated coefficient of determination for TR (R^2_{trial} = 0.06) is similar to that of Johnson et al. [13] (R^2_{trial} = 0.10), both are much lower than that reported by Buyse et al. [14] (R^2_{trial} = 0.38). Their results suggest that a 50% decrease in the odds of failure to respond corresponds to a 6% decrease in the odds of death, a result more extreme than that suggested by our ROR metric (ROR = 1.45; 95% CI: 1.19, 1.77). On the other hand, Johnson et al. [13] calculate about 30% difference in response rate, compared with 0.28 in our study, to observe a minimal benefit on OS.

The lower level of association between the treatment effect for surrogate end points and OS seen in this study is likely to reflect our inclusion of all lines of drug therapy. In other words, the extrapolation of the results from previous meta-analyses that focused on particular drug treatments, often only in first-line therapy, is likely to overestimate the ability of PFS, TTP, and TP to act as surrogate end points for OS across other classes of drug treatment and later lines of therapy in aCRC.

4.2. Strengths and limitations of the study

Unlike other authors, we decided to look comprehensively at all RCTs without restrictions on treatment type [7], line of therapy, or sample size [8,29]. We believe this choice is appropriate to support any claim about PFS, TTP, or TR being good surrogate end points for drug treatments in aCRC. To reduce potential bias in the analyses, we undertook a systematic review with independent application of predefined inclusion criteria and data extraction. Moreover, we focused on all three surrogate end points in a single study, without combining PFS and TTP [41,42], and provided the most common metrics proposed for the validation of surrogate end points [16].

Our study has a number of potential limitations. First and foremost, we relied on summary data from published reports of RCTs. Potential drawbacks of aggregate meta-analyses have been highlighted elsewhere [43]. For the specific purpose of surrogate validation, a limitation of using summary statistics is that the estimation errors for the treatment effects on the surrogate and the final end point cannot be fully taken into account. Indeed, the estimation errors on the end points are generally available in published reports but not the correlation between them. We did not take account of these estimation errors in our analyses, and this may have resulted in systematic bias toward lower measures of surrogacy than if estimation errors had been taken into account as in meta-analyses based on IPD [44]. Nonetheless, given the substantial costs for sharing clinical research data [45], meta-analyses based on summary data appear to be the only feasible way to take into account most of the available evidence about efficacy of health interventions and, therefore, a useful source of data for the validation of surrogate end points. In addition, lack of IPD reduced our flexibility to handle outcome findings. For instance, not all the trials reported HR (95% CI) for time-to-event end points although they often reported median survival times. In an attempt to make use of all available data, we considered the median survival time and median survival time differences across arms as equivalent to mean values of the same distributions. Although median survival times are not always reliable substitutes for HRs [46], they are common in meta-analyses of time-to-event variables using published data [8,10,13,47–49]. Unlike previous aggregate meta-analyses, we sought to take account of the within-trial correlation of surrogate end points and survival by performing a multivariate meta-analysis. A previous IPD meta-analysis of HR_{PFS} as a surrogate for HR_{OS} in fluoropyrimidine-based treatments in aCRC [7] indicates a reasonable correlation factor lies between 0.75 and 0.95. However, despite findings being consistent across the wide range of correlation values applied (ie, 0.00 to 0.95), this level of correlation may not hold for other treatments seen in this review. The number of available trials for each surrogate-to-final outcome comparison was lower than the sample size calculated ex-ante, yet, for three of the comparisons performed (HR_{PFS} vs. HR_{OS}, HR_{TTP} vs. HR_{OS}, and OR_{TR} vs. OR_{death}), the estimated RHRs and ROR were still significantly different from 1.0.

We found several potential biases in the included trials. The methodological detail of trials was often poorly reported, in respect of random sequence generation, allocation concealment, and outcome blinding, and therefore subject to selection and assessment bias [50]. However, in our stratified analyses, we combined trials with unclear and high risk of bias, thus following a conservative approach. We observed evidence of small-study effect bias when looking at treatment effects on PFS. However, the effect disappeared when trials with PFS as primary end point and those with blinding of outcome assessment were considered. Finally, we found considerable evidence of statistical heterogeneity in the treatment effect for OS and surrogate end points across trials. As for clinical heterogeneity,

we circumscribed our study to a recent time span to account for the marked improvements in the outlook for patients with metastatic colorectal cancer seen when trials from the past 10−15 years are compared with those from earlier periods [51].

4.3. Implications for policy and research

The potential for surrogate end points to impact on health care policy making and the consequent diffusion of treatments into oncology practice is shown by the fact that 69 (68%) of all RCTs identified in our sample reported a surrogate primary outcome. By using a systematic and comprehensive approach to the identification of drug therapy trials, the findings of our meta-analysis are potentially more generalizable to the management of aCRC than previous surrogate validation studies that have focused on specific drugs or classes of drugs. None of the end points in this study were found to achieve the level of evidence (ie, mean $R^2_{\text{trial}} > 0.60$) that has been set to select high or excellent correlation levels by common surrogate evaluation tools [52,53]. Overall, our study shows that TR should not be used as a surrogate end point for OS when evaluating drug interventions in aCRC. On the other hand, depending on the statistical requirements agreed for surrogacy validity, PFS and TTP might be regarded as acceptable surrogate measures for OS in aCRC. However, our results indicate that previous good surrogacy relationship observed in selected aCRC therapies (eg, fluoruracil plus leucovorin) may not directly apply across other classes or lines of therapy. Even if applying a lower level of evidence (eg, $R^2_{\text{trial}} > 0.40$) such that PFS and TTP can be demonstrated to be valid surrogates for OS in aCRC, policy makers still need to consider that the anticipated treatment effect on OS is likely to be smaller than that observed on the surrogate measure when weighing up the evidence from use of these surrogate end points in their licensing and coverage decisions.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2015.02.016.

References

[1] De Gruttola VG, Clax P, DeMets DL, Downing GJ, Ellenberg SS, Friedman L, 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.

[2] Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2095−128.

[3] 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.

[4] Shi Q, Sargent DJ. Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. 2009. [cited 14 (Shi, Sargent) Division of biomedical statistics and informatics, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, United States]; 2: [102−111].

[5] Chakravarty A. Regulatory aspects in using surrogate markers in clinical trials. In: Burzykowsi T, Molenberghs G, Buyse M, editors. The evaluation of surrogate endpoints. New York, NY: Springer; 2005.

[6] Burzykowski T, Molenberghs G, Buyse M. The evaluation of surrogate endpoints. In: Gail M, Samet JM, Tsiatis A, Wong W, editors. Statistics for biology and health. US: Springer Science and Business Media, Inc; 2005:416.

[7] Buyse M, Burzykowski T, Carroll K, Michiels S, Sargent DJ, Miller LL, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. J Clin Oncol 2007;25:5218−24.

[8] Tang PA, Bentzen SM, Chen EX, Siu LL. Surrogate endpoints for median overall survival in metastatic colorectal cancer: literature-based analysis from 39 randomized controlled trials of first-line chemotherapy. J Clin Oncol 2007;25:4562−8.

[9] Green E, Yothers G, Sargent DJ. Surrogate endpoint validation: statistical elegance versus clinical relevance. Stat Methods Med Res 2008;17(5):477−86.

[10] Chirila C, Odom D, Devercelli G, Khan S, Sherif BN, Kaye JA, et al. Meta-analysis of the association between progression-free survival and overall survival in metastatic colorectal cancer. Int J Colorectal Dis 2012;27(5):623−34.

[11] Burzykowski T, Molenberghs G, Buyse M, Geys H, Didier R. Validation of surrogate endpoints in multiple randomized clinical trials with failure time endpoints. J R Stat Soc Ser C (Applied Statistics) 2001;50(4):405−22.

[12] Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006;5:173−86.

[13] Johnson KR, Ringland C, Stokes BJ, Anthony DM, Freemantle N, Irs A, 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;7:741−6.

[14] Buyse M, Thirion P, Carlson RW, Burzykowski T, Molenberghs G, Piedbois P. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Meta-analysis group in cancer. Lancet 2000;356:373−8.

[15] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535.

[16] Ciani O, Davis S, Tappenden P, Garside R, Stein K, Cantrell A, et al. Validation of surrogate endpoints in advanced solid tumors: systematic review of statistical methods, results, and implications for policy makers. Int J Technol Assess Health Care 2014;30(3):312−24.

[17] Higgins J, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. New Jersey: The Cochrane Collaboration, John Wiley & Sons, Ltd; 2011.

[18] Green S, Benedetti J, Crowley J. Clinical trials in oncology. London: Chapman & Hall; 1997.

[19] Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst 2000;92:205−16.

[20] World Health Organization. WHO handbook for reporting results of cancer treatment. in WHO Offset Publication. Geneve: World Health Organization; 1979.

[21] Ciani O, Buyse M, Garside R, Pavey T, Stein K, Sterne JA, 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.
[22] Higgins JP, Green S, editors. Cochrane handbook for systematic reviews of interventions Cochrane Book Series ed. C. Collaboration. West Sussex, England: John Wiley & Sons Chichester; 2008.
[23] Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997;315:629−34.
[24] Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. Stat Med 2006;25:3443−57.
[25] Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. Stat Med 2011;30:2481−98.
[26] Sterne JA, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. Stat Med 2002;21:1513−24.
[27] Altman DG, Bland JM. Interaction revisited: the difference between two estimates. BMJ 2003;326:219.
[28] Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997;50:683−91.
[29] 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;91:2033−8.
[30] Foster NR, Qi Y, Shi Q, Krook JE, Kugler JW, Jett JR, 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;117:1262−71.
[31] 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;93:1215−21.
[32] Burzykowski T, Buyse M, Piccart-Gebhart MJ, Sledge G, Carmichael J, Lück HJ, et al. Evaluation of tumor response, disease control, progression-free survival, and time-to-progression as potential surrogate endpoints in metastatic breast cancer. J Clin Oncol 2008;26:1987−92.
[33] Hotta K, Kiura K, Fujiwara Y, Takigawa N, Hisamoto A, Ichihara E, 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.
[34] Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. 2006. 3: [173-86]. Available at http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=17080751. Accessed January 21, 2014.
[35] Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am J Clin Oncol 1982;5:649−55.
[36] Armitage P, Berry G, Matthews J. Regression and correlation. In: Statistical methods in medical research. 4th ed. Oxford: Blackwell Scientific Publications; 2002:187−204.
[37] Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. Stat Med 2006;25:183−203.
[38] 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(3):303−40.
[39] Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate endpoints—the challenge of statistical validation. Nat Rev Clin Oncol 2010;7(6):309−17.
[40] Walter SD, Sun X, Heels-Ansdell D, Guyatt G. Treatment effects on patient-important outcomes can be small, even with large effects on surrogate markers. J Clin Epidemol 2012;65:940−5.
[41] Amir E, Seruga B, Kwong R, Tannock IF, Ocaña A. Poor correlation between progression-free and overall survival in modern clinical trials: are composite endpoints the answer? Eur J Cancer 2012;48:385−8.
[42] Felix J, Aragão F, Almeida JM, Calado FJ, Ferreira D, Parreira AB, et al. Time-dependent endpoints as predictors of overall survival in multiple myeloma. BMC Cancer 2013;13:122.
[43] Buyse M, Carlson RW, Piedbois P. Meta-analyses of published results are unreliable. J Clin Oncol 1999;17:1646−7.
[44] Buyse M, Molenberghs G, Paoletti X, Oba K, Alonso A, Van der Elst W, et al. Validation of surrogate endpoints with examples from cancer clinical trials. Biom J 2015.
[45] Wilhelm EE, Oster E, Shoulson I. Approaches and costs for sharing clinical research data. JAMA 2014;311:1201−2.
[46] Michiels S, Piedbois P, Burdett S, Syz N, Stewart L, Pignon JP. Meta-analysis when only the median survival times are known: a comparison with individual-patient data results. Int J Technol Assess Health Care 2005;21(1):119−25.
[47] 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;15:379−85.
[48] Delea TE, Khuu A, Heng DY, Haas T, Soulières D. Association between treatment effects on disease progression endpoints and overall survival in clinical studies of patients with metastatic renal cell carcinoma. Br J Cancer 2012;107:1059−68.
[49] Shitara K, Ikeda J, Yokota T, Takahari D, Ura T, Muro K, 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(3):1224−31.
[50] Vale CL, Tierney JF, Burdett S. Can trial quality be reliably assessed from published reports of cancer trials: evaluation of risk of bias assessments in systematic reviews. BMJ 2013;346:f1798.
[51] Kemeny NE. Treatment of metastatic colon cancer: ''the times they are A-changing''. J Clin Oncol 2013;31:1913−6.
[52] IQWiG. Validity of surrogate endpoints in oncology. Executive summary, in IQWiG Reports. Cologne, Germany: IQWiG; 2011.
[53] 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.