JNCI Cancer Spectrum (2019) 3(1): pkz002

doi: 10.1093/jncics/pkz002
First published online February 6, 2019
Review

REVIEW

A Systematic Review and Recommendation for Reporting of Surrogate Endpoint Evaluation Using Meta-analyses

Wanling Xie, Susan Halabi, Jayne F. Tierney, Matthew R. Sydes, Laurence Collette, James J. Dignam, Marc Buyse, Christopher J. Sweeney*, Meredith M. Regan*

See the Notes section for the full list of authors' affiliations.
*Authors contributed equally.This paper is dedicated to the memory of Professor Douglas G Altman who was instrumental in starting the CONSORT initiative, which later led to PRISMA, STROBE, STREGA, STARD, SQUIRE, MOOSE, GNOSIS, TREND, ORION, COREQ, QUOROM, REMARK, and now ReSEEM.
Correspondence to: Wanling Xie, MS, Department of Data Sciences, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215 (e-mail: wxie@jimmy.harvard.edu).

Abstract

Background: Meta-analysis of randomized controlled trials (RCTs) has been widely conducted for the evaluation of surrogate endpoints in oncology, but little attention has been given to the adequacy of reporting and interpretation. This review evaluated the reporting quality of published meta-analyses on surrogacy evaluation and developed recommendations for future reporting.
Methods: We searched PubMed through August 2017 to identify studies that evaluated surrogate endpoints using the meta-analyses of RCTs in oncology. Both individual patient data (IPD) and aggregate data (AD) meta-analyses were included for the review.
Results: Eighty meta-analyses were identified: 22 used IPD and 58 used AD from multiple RCTs. We observed variability and reporting deficiencies in both IPD and AD meta-analyses, especially on reporting of trial selection, endpoint definition, study and patient characteristics for included RCTs, and important statistical methods and results. Based on these findings, we proposed a checklist and recommendations to improve completeness, consistency, and transparency of reports of meta-analytic surrogacy evaluation. We highlighted key aspects of the design and analysis of surrogate endpoints and presented explanations and rationale why these items should be clearly reported in surrogacy evaluation.
Conclusions: Our reporting of surrogate endpoint evaluation using meta-analyses (ReSEEM) guidelines and recommendations will improve the quality in reporting and facilitate the interpretation and reproducibility of meta-analytic surrogacy evaluation. Also, they should help promote greater methodological consistency and could also serve as an evaluation tool in the peer review process for assessing surrogacy research.

Overall survival (OS) is the gold standard primary efficacy endpoint in oncology randomized controlled trials (RCTs), but it often requires prolonged follow-up and a substantial number of patients. In addition, evaluation of OS is likely influenced by subsequent lines of therapies. As such, investigation of surrogate endpoints for OS has received increasing interest in oncology in the recent two decades in the hope of reducing the duration of trials and decreasing the cost of drug development. To establish surrogacy, investigators need to provide evidence that "a drug-induced effect on the surrogate predicts the desired effect on the clinical outcome of interest" (1,2) using robust statistical methods before it replaces the definitive endpoint.

Received: September 13, 2018; **Revised:** November 12, 2018; **Accepted:** January 3, 2019

Meta-analysis of RCTs has been widely conducted for the evaluation of surrogate endpoints in oncology, but little attention has been given to the adequacy of reporting and interpretation. The two-stage meta-analytic approach developed by Buyse et al. requires demonstration of strong correlation between the surrogate and definitive endpoints ("outcome surrogacy") as well as correlation of treatment effects on both endpoints ("trial-level or effect surrogacy") (3,4). Meta-analysis of individual patient data (IPD) remains the optimal approach to meta-analysis in general, and to surrogacy evaluation in particular, because it enables the standardization of methods across IPD sets and robust analysis at both the patient and trial levels (5). However, because IPD meta-analyses are time and resource intensive, meta-analyses of outcome correlation or trial-level associations using aggregate data (AD) are more often reported (6).

The PRISMA (7) and PRISMA-IPD (8) statements provide reporting guidelines for systematic reviews and meta-analyses. However, some requirements, for example, quantification of heterogeneity, may not apply to meta-analytic surrogacy evaluation, whereas other important design and analysis aspects that are unique for surrogacy work are not covered by the PRISMA or PRISMA-IPD. Therefore, we began by reviewing the reporting quality of published IPD and AD meta-analyses on surrogacy evaluation, using the PRISMA and PRISMA-IPD guidelines. We also identified the additional items that, if not or incompletely reported, could severely affect interpretation of surrogacy studies. We limited this review to the field of oncology, in which surrogate endpoints have been frequently investigated. On the basis of our systematic review, we provided evidence-based recommendations to improve the consistency and quality of reporting these studies in the future.

Methods

Study Selection and Identification

We conducted a systematic review of published articles that reported on surrogate endpoint evaluation in oncology using the meta-analytic approach. Articles were eligible if they evaluated surrogate endpoints using meta-analyses of RCTs in oncology and were published in English as full text. Articles were excluded if surrogacy analyses were based on a single RCT, observed retrospective studies, or single-arm phase I/II studies. Commentaries, reviews, and studies not focusing on surrogacy were also excluded. The PubMed database was searched for relevant entries up to August 31, 2017 (with no restriction on the start date), using the terms: ["surrogate end point" or "surrogate endpoint" or "surrogate outcome" or "intermediate endpoint" or "intermediate end point" or "intermediate outcome"] and ["cancer" or "neoplasms"]. Two authors performed the database search (W. Xie and M. M. Regan). In addition, reference lists of relevant review articles were manually searched by a single reviewer (W. Xie). Exclusion of ineligible articles after the full text review was determined by the two reviewers with consensus. Articles and abstracts were screened in accordance with the PRISMA guidelines (7).

Data Extraction and Analysis

Articles meeting the eligibility criteria were included for data abstraction. We collected information regarding the general characteristics of meta-analyses, study design, and statistical methods for surrogacy evaluation, reporting of information on included trials and patient characteristics, and reporting of the results for surrogacy evaluation. The initial set of items included relevant elements for the general meta-analysis from the PRISMA and PRISMA-IPD as well as particular elements required for surrogacy evaluation based on the authors' experience and pilot review of 10 eligible articles (W. Xie and M. M. Regan). A few additional items were added during the process of data extraction from full review and manuscript development. The final list comprised those items that, if not or incompletely reported, could severely affect interpretation of the results from surrogacy analysis. Detailed data elements are listed in Tables 1 and 2. Data extraction was conducted by a single reviewer (W. Xie). Descriptive methods were used for data summarization and analyses.

Results of the Systematic Review

Characteristics of Eligible Meta-Analysis Studies

Over 900 abstracts were initially identified for screening from PubMed and review articles. A total of 828 ineligible abstracts were excluded (non-meta-analysis, methodology papers, review, commentary, or editorial articles, non-English articles; Figure 1). We conducted the full text review for the 94 potentially eligible articles. Fourteen of 94 articles were excluded from the review for the following reasons: reanalyzed published data (n = 3 articles), used single-arm trials (n = 7), commentary (n = 3), single RCT using countries as analysis units (n = 1). The remaining 80 articles met the inclusion criteria (Figure 1; Supplementary Table S1, available online).

Notably, the number of meta-analyses on surrogacy evaluation in oncology has grown markedly in recent years, with 60 (75%) of the relevant articles published since 2010 (Figure 2). The most frequent tumor types examined were colorectal, breast, and lung cancers (17, 15, and 14 meta-analyses, respectively). Twenty-two of 80 (27%) meta-analyses used IPD and 58 (73%) used AD (Table 1). Fifteen (19%) meta-analyses evaluated surrogate endpoints in a localized (or locally advanced) disease setting, with disease-free survival (DFS) as the most frequently evaluated surrogate endpoint (13 of 15, 87%). For metastatic or advanced diseases (65 studies), progression-free survival (PFS) or time to progression (TTP) were the major surrogate endpoints examined (54 of 65, 83%) followed by tumor response (30 of 65, 46%). OS was the definitive endpoint of interest in all but four studies (Table 1).

Reporting of Surrogacy Evaluation Study Design

Some studies reported a preexisting protocol (7 of 58, 12%; and 7 of 22, 32% in AD and IPD meta-analyses, respectively). For trial selection, the majority (57 of 58, 98%) of AD meta-analyses reported using a systematic literature review to identify included RCTs; 42 (72%) provided a trial selection flowchart. For the 22 meta-analyses using IPD, 15 (68%) reported systematic literature review and 9 (41%) presented the flowchart (Table 2). The scope of the reported literature search criteria included study population, type of intervention, and other trial characteristics.

Harmonized endpoint definitions across included trials were more frequently provided in IPD meta-analyses (22 of 22, 100%) than in AD meta-analyses (23 of 58, 40%). Meanwhile, heterogeneity in defining the endpoints across included trials was reported in 41 studies: nine (41%) in IPD and 32 (55%) in AD

Table 1. Characteristics of 80 meta-analysis study articles reviewed (up to August 31, 2017)

Characteristic	No. (%)
Type of meta-analysis data	
AD	58 (73)
IPD	22 (27)
Type of cancer*	
Colorectal	17 (21)
Breast	15 (19)
Lung	14 (18)
Other	38 (48)
Disease stage and surrogate endpoints*	
Localized or locally advanced	15 (19)
Tumor response	3 (20)
DFS	13 (87)
MFS	2 (13)
Other	2 (13)
Metastatic or advanced†	65 (81)
Tumor response	30 (46)
TTP or PFS	54 (83)
Other	11 (17)
Definitive endpoint	
OS	69 (86)
OS and DFS or EFS or PFS	7 (9)
PFS	1 (1)
PPS	3 (4)
No. of trials included, median (range)	24 (3–191)
No. of patients in the included trials	9223 (870–67158)

*Total is greater than 100% because some studies examined multiple diseases or multiple endpoints. AD = aggregate data; DFS = disease-free survival; EFS = event-free survival; IPD = individual patient data; MFS = metastasis-free survival; OS = overall survival; PFS = progression-free survival; PPS = postprogression survival; TTP = time to progression.
†Included four hematologic cancers, which evaluated PFS or response as surrogates for OS.

meta-analyses. The reported variations in endpoint definitions included type of failure events for DFS or PFS endpoints, disease evaluation criteria (eg, World Health Organization [WHO] vs Response Evaluation Criteria in Solid Tumors [RECIST]), endpoint assessment schedule, and censoring rules for time-to-event endpoints (Table 2).

In terms of the surrogacy criteria, 19 (24%) meta-analyses predefined the rule for declaring surrogacy, of which 11 meta-analyses specified an R-squared of 0.7 or higher as the cut point to determine a statistically acceptable correlation.

Reporting of Included Trial and Patient Characteristics

Among the AD meta-analyses, 13 of 58 (22%) reported years of enrollment for included RCTs, and reporting of this variable was 17 of 22 (77%) for IPD meta-analyses. Patients' ages, disease stages or characteristics, trial follow-up duration, and numbers of events (for time-to-event endpoints) from included trials were also more frequently reported in IPD meta-analyses than in AD meta-analyses (Table 2).

Reporting of Outcome Surrogacy

Fifty studies (63%) reported outcome surrogacy (ie, correlation between the surrogate and definitive endpoints irrespective of treatment effect). Fourteen IPD-based studies reported using a copula model to estimate individual-level correlation coefficient for failure-time endpoints. Copula is a modeling analysis that describes the dependence of two endpoints through modeling their distributions jointly. There are a series of candidate models (eg, Clayton's, Hougaard's, and Plackett's models) for time-to-event endpoints (9); hence, application of the copula methodology consists of selecting an appropriate copula model from among these candidates. Although nine (64%) studies specified the type of copula, only three studies clearly stated the method (eg, based on the Akaike information criteron) used to select a copula model. Correlation was expressed through the Spearman or Kendall's tau correlation coefficients in 12 (86%) studies, but without explicit justification for the choice of measure (Table 2; Supplementary Table S2, available online). Nine IPD meta-analyses also analyzed a binary surrogate (eg, tumor response), whose correlation with OS was quantified as hazard ratios from Cox's regression model, log-rank test of statistical significance, or survival odds ratios from a joint copula.

Outcome surrogacy was also reported using AD, which examined the correlation between endpoints on estimated summary metrics (such as median time to event, event rate at selected timepoints, or response rate) with trial and treatment arm as the analysis unit. Twenty-nine studies reported use of a correlation coefficient to measure the strength of correlation, including two studies using aggregated IPD. Although the majority of studies explicitly specified the type of correlation coefficients (Pearson, Spearman, or Kendall's tau) or whether weights were used in the calculation, only about 50% studies provided confidence intervals for these estimates. The coefficient of determination (ie, R-squared) from a linear regression (unweighted, weighted, or error-in-variable adjusted; Table 2) was reported in 20 studies, but the R-squared confidence interval and the regression equation were not fully reported (5 of 20, 25%; and 9 of 20, 45%, respectively).

Reporting of Trial-Level Surrogacy

Seventy-four studies reported trial-level surrogacy by examining correlation between the treatment effect on the surrogate endpoint and the treatment effect on the definitive endpoint. The majority of meta-analyses (71 of 74, 96%) reported using trial as the analysis unit, whereas three IPD-based studies reported using country or center as the analysis unit. Treatment effect was measured in a variety of ways, including absolute difference in medians of DFS or PFS, or response rates; ratios or percent increases in medians of DFS or PFS, or response rates; and hazard ratios of DFS or PFS, or odds ratios of response (with or without log transformation). Trial-level correlation was quantified as a correlation coefficient (Spearman or Pearson) in 37 studies, with confidence interval reported in about 50% of studies. Fifty-four studies reported R-squared from a linear regression; the types of regression included unweighted, weighted, or error-in-variable adjusted (Table 2). For 12 studies that reported copula R-squared, introduced by Burzykowski et al. (9), five studies used error-adjusted models. In addition, incomplete reporting on R-squared confidence interval or regression equation was noted in both conventional or copula model based regression analyses (Table 2). Other novel trial-level correlation measures, such as the Bayesian adjusted R-squared proposed by Renfro (10), were not reported in the meta-analyses that were reviewed.

Table 2 Reporting on surrogate endpoint evaluation using meta-analysis approach*

Reported elements	Meta-analysis of AD (n=58) No. (%)	Meta-analysis of IPD (n=22) No. (%)
Reporting of surrogacy evaluation study design (n=80)	58	22
A protocol existed for the meta-analysis	7 (12)	7 (32)
Systematic search	57 (98)	15 (68)
Specified search term(s)	55 (95)	8 (36)
Trial selection flowchart	42 (72)	9 (41)
Harmonized endpoint definition	23 (40)	22 (100)
Variation in endpoint definition across trials	32 (55)	9 (41)
Variation in time-to-event endpoint failuretypes	20 (34)	1 (5)
Variation in endpoint evaluation criteria	8 (14)	4 (18)
Variation in endpoint assessment schedule	4 (7)	3 (14)
Variation in censoring rules	0 (0)	1 (5)
Specified surrogacy criteria (eg, correlation cutoff)	13 (22)	6 (27)
Reporting of included trial and patient characteristics (n=80)	58	22
Patient enrollment period	13 (22)	17 (77)
Patient age	14 (24)	14 (64)
Patient disease characteristics	25 (43)	17 (77)
Number of events	4 (7)	11 (50)
Median follow-up duration	19 (33)	19 (86)
Reporting of outcome surrogacy (using IPD)		
Reporting of correlation between time-to-event endpoints using a copula (n=14)	n/a	14
Copula type	n/a	9 (64)
Copula selection criteria	n/a	3 (21)
Type of correlation coefficient	n/a	12 (86)
Confidence interval for correlation	n/a	13 (93)
Reporting of correlation between binary surrogates and time-to-event endpoint (n=9)	n/a	9
Type of correlation measure		
Hazard ratio from Cox regression	n/a	3 (33)
Hazard ratio from Bayesian hierarchicalanalysis	n/a	1 (11)
Log-rank test of significance	n/a	1 (11)
Survival odds ratio from Plackett copula	n/a	4 (44)
Reporting of outcome surrogacy (using trial level summary data)		
Reporting of correlation between endpoints (n=29)	27	2
Type of correlation coefficient		
Non-parametric (Kendall's tau andSpearman)	15 (56)	2 (100)
Pearson (8 weighted, 1 unweighted)	9 (33)	0 (0)
Not reported	3 (11)	0 (0)
Confidence interval for correlation	15 (56)	1 (50)
Reporting of R-squared from linear regression (n=20)	17	3
Type of linear regression specified		
Weighted by sample size	14 (82)	2 (67)
Weighted by inverse variance of surrogate	0 (0)	1 (33)
Error-in-variables adjusted	1 (6)	0 (0)
Unweighted simple linear regression	2 (12)	0 (0)
R-squared confidence interval	3 (18)	2 (67)
Regression equation	8 (47)	1 (33)
Bubble plot for regression model	14 (82)	3 (100)
Reporting of trial-level surrogacy		
Reporting of correlation of treatment effects on endpoints (n=37)	33	4
Type of correlation coefficient specified		
Pearson (11 weighted, 1 unweighted)	10 (30)	2 (50)
Spearman	20 (61)	2 (50)
AUC	1 (3)	0 (0)
Not reported	2 (6)	0 (0)
Confidence interval for correlation	15 (45)	2 (50)

(continued)

Table 2 (continued)

Reported elements	Meta-analysis of AD (n=58) No. (%)	Meta-analysis of IPD (n=22) No. (%)
Reporting of R-squared from linear regression of treatment effects on endpoints (n=54)	38	16
Type of linear regression specified		
Weighted by sample size	24 (63)	14 (88)
Weighted by other factors	4 (11)	1 (6)
Unweighted simple linear regression	4 (11)	0 (0)
Errors-in-variables adjusted	3 (8)	0 (0)
Not reported	3 (8)	1 (6)
R-squared confidence interval	12 (32)	14 (88)
Regression equation	32 (84)	6 (38)
Bubble plot for regression model	32 (84)	16 (100)
Reporting of R-squared from a copula model (n=12)	n/a	12
Type of regression specified		
Errors-in-variables adjusted	n/a	5 (42)
Weighted by sample size	n/a	1 (8)
Not reported	n/a	6 (50)
R-squared confidence interval	n/a	11 (92)
Regression equation	n/a	6 (50)
Bubble plot for regression model	n/a	12 (100)
Reporting of STE (n=21)	9	12
STE estimation method		
Using regression line prediction interval	5 (56)	10 (83)
Using regression line confidence interval	3 (33)	0 (0)
Method not specified	1 (11)	2 (17)
Reporting of additional analysis (n=80)	58	22
Any sensitivity analysis	20 (34)	11 (50)
Any subgroup analysis	40 (69)	11 (50)
Any cross-validation analysis (eg, leave-one-out)	4 (7)	10 (45)
Any external validation analysis (eg, using other trials)		5 (23)

*AD=aggregate data; AUC=area under the curve; IPD=individual patient data; n/a =not applicable; STE=surrogate threshold effect

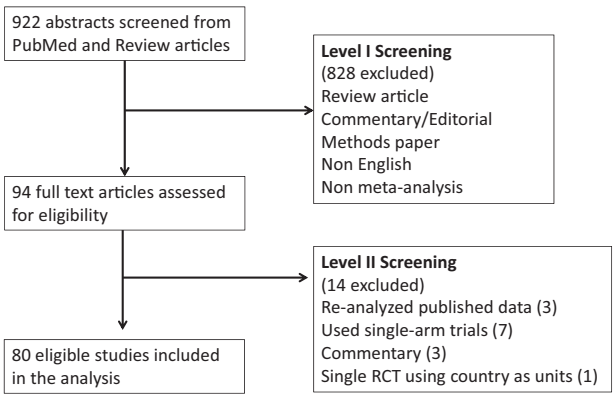

Figure 1. Meta-analysis articles inclusion flowchart. RCT = randomized controlled trial.

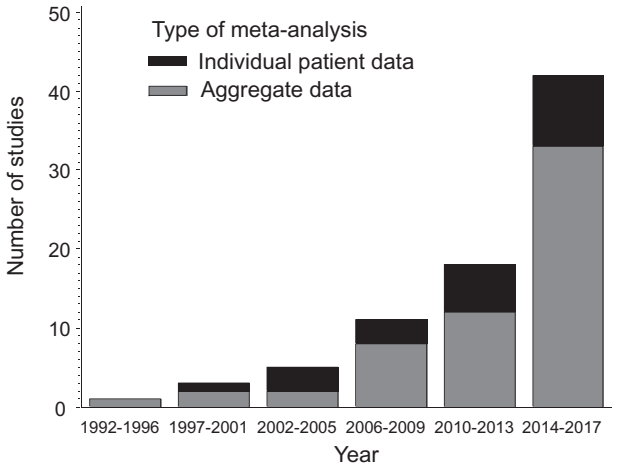

Figure 2. Publication trends for meta-analytic surrogacy evaluation in oncology.

Surrogate threshold effect (STE), the minimum treatment effect on the surrogate necessary to predict a non-zero effect on the definitive endpoint, is another metric to evaluate trial-level surrogacy (11). To obtain the STE, Burzykowski and Buyse proposed using the 95% prediction limits of the regression line for the trial-level surrogacy (11). For the meta-analysis studies we reviewed, 26% (21 of 80) reported STE: 15 studies reported using the 95% prediction limits of the regression line, three studies reported using the 95% confidence interval, and three studies did not specify which method was used to construct the STE.

Reporting Sensitivity, Subgroup, and Validation Analyses

Overall, 31 studies (39%) reported additional sensitivity analysis, such as use of various cutoffs for follow-up times, exclusion of certain trials, use of alternative endpoint definitions, use of alternative weights, or statistical models for analysis. A total of 51 (64%) studies reported subgroup analyses. The most commonly reported subgroup analyses were by type of treatment (40, 50%), by enrollment or publication years (18, 23%), and by patient characteristics (16, 20%). Eighteen studies also reported validation analysis and the majority used a leave-one-out validation (Table 2; Supplementary Table S3, available online).

Recommendations for the Future

Our systematic review identified wide variability and reporting deficiencies in the published meta-analyses on surrogate endpoint evaluation in oncology. To assist in the comprehensive and consistent reporting of meta-analytic surrogacy evaluation, we propose a checklist and recommendations (Table 3). Some design aspects are covered by the standard PRISMA (7) and PRISMA-IPD (8) guidelines (eg, items 5–13 on reporting of methods for study design and data collection). Other aspects are more specific and essential for surrogacy evaluation (such as items 14–16 on reporting of surrogacy criteria and analysis methods). We highlight below the key aspects of the design and analysis of surrogate endpoints and why they should be clearly reported in meta-analysis studies.

Develop Protocol for Surrogacy Evaluation

The PRISMA and PRIMA-IPD guidelines recommend registration and reporting of the protocol of any systematic review and meta-analysis. Compared with other meta-analyses aiming for the pooled treatment effects, meta-analysis for surrogacy evaluation usually targets broader treatment classes, more mixed study population, and multiple surrogate candidates. With the heterogeneity of trials and rich analyses of multi-level data (outcome level vs trial level), there is a greater risk of post-hoc, data-driven decision-making. The development and adherence to a protocol is therefore critical for meta-analytic surrogacy analyses.

A protocol will help investigators to choose appropriate candidate surrogates, generate hypotheses in the context of disease, and define a priori the threshold of correlation deemed necessary to meet the surrogacy criteria. As noted above, an R-squared of 0.7 or above has been the conventional degree of correlation used to define surrogacy in oncology. However, it should be realized that different types of surrogates (eg, clinical or biomarker based) may serve different roles in different phases of drug development (12). In one instance, a RCT may be designed using an intermediate clinical endpoint as a strong surrogate for OS to reduce trial duration and sample size. In other situations, lack of improvement of a very early endpoint, such as biochemical recurrence before radiographic recurrence of a prostate cancer, can be used to screen for ineffective drugs in an early-phase clinical trial, even though presence of a biochemical recurrence may not have a strong correlation with OS. Hence, the threshold of success needs to be defined in a context-dependent manner as well as ideally tied to a specific purpose.

The rigorously planned protocol should include meta-analysis study design, trial and patient selection, endpoint definition, and a statistical analysis plan that precisely defines upfront all details of the surrogacy analysis method that will be conducted, including, for example, the details method used to choose a model over another. It is best to define subgroup and sensitivity analyses as a priori in the protocol and to justify them in the context of disease. A well-developed protocol will help investigators in projecting the timeline, estimating the cost to conduct these analyses, and avoiding post-hoc data interrogation for the designed results. To increase transparency, the protocol should be explicitly reported in the meta-analyses of surrogate evaluation and made publicly accessible. Notably, several meta-analyses that we reviewed published their protocols (13,14), setting good examples of using well-developed protocols to guide analysis for surrogacy evaluation.

Sufficient Reporting on Trial Selection

In practice, very few meta-analyses were able to include all potentially eligible trials for surrogacy evaluation. A recent review article has indicated that only a subset (about 50%) of potential eligible studies were ultimately included in even the most rigorous meta-analyses for trial-level surrogacy evaluation (15). Besides the common trial selection issues in meta-analyses (eg, use of only published articles, resource limitation, data sharing obstacles for IPD collection), surrogacy evaluation requires outcome and treatment effect data for multiple endpoints, and not all eligible trials can provide such complete data needed for surrogacy evaluation. For instance, some AD meta-analyses can include only RCTs that reported both PFS and OS hazard ratios (16). The excluded trials that did not report PFS and OS hazard ratios were more likely to have negative treatment effects on these endpoints (17,18).

Whereas for other meta-analyses publication bias usually leads to exaggerated treatment effects, surrogacy evaluation may encounter other challenging issues. A simulation study demonstrated that surrogacy estimation via regression models is more accurate and precise in the settings of a larger number of trials, a low rate of censoring, and a wide range of treatment effects (ie, including both positive and negative trials) (19). When one or more factors deviate from the “optimal” scenarios, regression tends to underestimate the definitive surrogacy with increased variability (19). A wide range of treatment effects across included trials improves the performance of the regression model and the magnitude of R-squared, because the variance of the regression coefficient is inversely proportional to the spread of the predictor variable. Hence, heterogeneity of treatment effects (ie, including both positive and negative trials) is an advantage rather than a drawback in a meta-analysis aimed at surrogate validation.

Use of nonexhaustive sets of trials might be less of a concern compared with other meta-analyses as long as the included trials reflect the expected range of heterogeneity of treatment effects. However, it is still important for the study selection process to be inclusive and explicit for a rigorous surrogate evaluation. To reduce potential bias due to trial selection, reasons for trial exclusion should be reported at the trial selection stage. A comparison of trial characteristics and range of treatment effects across trials between included and excluded eligible trials can also provide valuable information for trial representation and generalization.

Table 3. Recommendation for ReSEEM

Section and topic	No.	Checklist items
Title		
Title	1	Identify whether this is a report of meta-analysis of individual patient or AD; specify surrogates examined in the context of disease (eg, PFS as a surrogate endpoint for OS in advanced lung cancer: meta-analyses of IPD)
Abstract		
Structured summary	2	Provide a structured summary including as applicable: • Background: state main objectives, surrogate endpoints and definitive endpoint examined, participants and interventions • Methods: report eligibility criteria, data sources (individual patient or AD), surrogacy criteria, and primary analysis method for surrogacy evaluation • Results: provide key results for patient-level (or outcome level) and trial-level surrogacy analysis • Conclusion: summarize the strength of surrogacy and implications for future research
Introduction		
Rationale	3	Provide justification for the use of a surrogate endpoint in the context of disease
Objectives	4	State the objective of meta-analysis or any prespecified hypotheses for surrogacy evaluation, including endpoints examined, participants, interventions, and study design
Methods		
Design/data collection		
Protocol and registration	5	Follow PRISMA and PRISMA-IPD statement Provide rationale for choice of endpoint, treatment, and population; generate hypotheses in a context-dependent manner; include details on study design, trial, and patient selection, endpoint definition, and a statistical analysis plan
Eligibility criteria	6	Follow PRISMA and PRISMA-IPD statement
Information sources	7	Follow PRISMA and PRISMA-IPD statement
Search	8	Follow PRISMA and PRISMA-IPD statement
Study selection	9	Follow PRISMA and PRISMA-IPD statement
Data collection process	10	Follow PRISMA and PRISMA-IPD statement
Data items	11	Follow PRISMA and PRISMA-IPD statement
Risk of bias within studies	12	Follow PRISMA and PRISMA-IPD statement
Risk of bias across studies	13	Follow PRISMA and PRISMA-IPD statement
Endpoint definitions	14	Precisely define all endpoints examined Provide description of between-trial variability in endpoint definition (eg, disease assessment criteria and schedule, type of events included in time to event endpoint, methods used for censoring endpoints)
Surrogacy criteria	15	Define surrogacy criteria and cutpoint determination in the specific context of disease; provide justification for what level of correlation would be deemed as surrogacy at individual and trial level
Statistical analyses	16	
Individual-level correlation		A Specify copula methods used to estimate individual-level correlation: choice of copula and justification, choice of correlation coefficient (eg, Spearman vs Kendall's tau), and rationale for the choice Specify other methods used for individual-level correlation if appropriate (eg, hazard ratio from Cox regression, landmark or time-dependent model, information theory, Bayesian methods)
Outcome correlation using aggregate data		B Specify the analysis unit (eg, trial, arm, country, and center) Specify type of outcome measures (eg, response rate, median time to event, event rate at selected timepoints, and rationale for timepoint selection) Specify how outcome measure is estimated for each study (eg, from Kaplan-Meier methodology or cumulative incidence function for time to event endpoints; from trial reported or extracted from Kaplan-Meier curves) State the statistical model to calculate correlation coefficient or R-squared (eg, weighted linear regression, error in variable regression, or nonparametric model; choice of weights and rationale)
Trial-level correlation		C Specify the analysis unit (eg, trial, country, and center) Specify the metrics for treatment effects (eg, hazard or odds ratio, whether logarithm transformation is used) Specify how treatment effect is estimated for each study (eg, use of Weibull or Cox regression, from marginal or joint copula model, from trial-reported or imputed)

(continued)

Table 3. (continued)

Section and topic	No.	Checklist items
		State the statistical model to calculate correlation coefficient or R-squared (eg, weighted linear regression, error in variable regression, or nonparametric model; choice of weights and rationale)
		Specify the statistical method used to calculate STE (eg, type of regression, how prediction interval is constructed)
Validation		D Specify statistical methods used to validate surrogate evaluation (eg, leave-one-out across validation, bootstrap validation, and external validation)
Sensitivity and subgroup analysis		E State type of sensitivity and subgroup analysis
		Provide justification for these additional analyses in the context of disease
Results		
Study selection	17	Follow PRISMA and PRISMA-IPD statement
Risk of bias within studies	18	Follow PRISMA and PRISMA-IPD statement
Risk of bias across studies	19	Follow PRISMA and PRISMA-IPD statement
Study characteristics	20	Summarize trial and patient characteristics for each included trial (eg, sample size, phase, interventions, disease stage, years of enrollment, follow-up period) and provide the citations
		Provide comparison of trial characteristics between included and excluded eligible trials
Endpoints summary	21	Provide summary statistics for endpoints examined (eg, number of events, median time to events, event free rate at selected timepoints, response rate), by trial and treatment arm
		Provide trial-specific hazard/odds ratio estimates (or other metrics for treatment effect if appropriate) and confidence intervals for each endpoint; a table or forest plot is recommended
Surrogacy analysis	22	Present results of each type of surrogacy analysis done, including number of patients and number of trials (or units), any exclusion from analysis, confidence interval for correlation coefficient or R-squared, regression equation if appropriate
		Provide bubble plot of regression model, with regression line and prediction interval
		Present STE in the context of disease and implication for the trial design
Validation	23	Present results from validation analyses; indicate any discrepancy between main and validation findings
Additional analysis	24	Follow PRISMA and PRISMA-IPD statement
Conclusion		
Summary of evidence	25	Summarize the strength of surrogacy in the context of prespecified hypothesis, including subjects, interventions, and trial characteristics
		Interpret results in the context of other evidence; consider its relevance, generalization, and implication for future trial design
Limitations	26	Discuss limitations at various levels (eg, risk of bias, incomplete trial inclusion, variation in endpoint definition, incomplete data, reporting bias)
Conclusions	27	Summarize the strength of surrogacy and implications for future research.
Funding		
Funding	28	Follow PRISMA and PRISMA-IPD statement

*AD = aggregate data; IPD = individual patient data; PFS = progression free survival; OS = overall survival; ReSEEM = reporting of surrogate endpoint evaluation using meta-analyses; STE = surrogate threshold effect.

Sufficient Reporting of Characteristics of the Included RCTs

It is important to report on all trials and patient characteristics for the included trials (eg, class of therapy, enrollment period, follow-up duration, patient characteristics). Patient characteristics, such as age and tumor staging, are important prognostic factors with respect to disease evolution and potentially affect treatment options, TTP, postprogression survival, comorbidity, and cause of death. Surrogacy results can be appropriately interpreted only if the study population is well characterized.

Furthermore, in meta-analysis for surrogacy evaluation, broad trial entry criteria increase the degree of heterogeneity between trials. However, it is possible that the strength of surrogacy may vary according to certain trial-level or patient characteristics. Analysis by subgroups is an important strategy to explore heterogeneity. Moreover, with the emergence of checkpoint inhibitor-based immunotherapy, we have seen that although most patients do not benefit, a clinically relevant minority have major benefit. Similarly, some targeted therapies may have greater benefit in a selected population subset with a mutated gene. Sufficient reporting of trial and patient characteristics provides supporting evidence and justification for the planned subgroup and sensitivity analyses.

In addition, for time-to-event endpoints, such as DFS and PFS, follow-up duration and numbers of events determine the precision of point and interval estimates. Collection and reporting of such information help investigators choose appropriate weights for the weighted regression analysis in the presence of unbalanced follow-up across trials. Unfortunately, in most studies we reviewed, authors only reported sample sizes of included RCTs and performed linear regression weighted on trial sample size without considering study follow-up duration.

Adequately Define Endpoints for Surrogacy Evaluation

For IPD meta-analyses, it is important to define endpoints accurately, with details on trial-specific disease assessment criteria and schedule, because they are integral parts of endpoint definitions. For AD meta-analyses using the protocol definition, efforts need to be made to assess and report variability in definition of endpoint. Thorough and transparent reporting does not solve issues of inconsistent endpoint definition from trials' initial designs, but it could highlight the potential problems that need to be considered in interpretation of results. For example, Colloca and Venturino (20) reported how heterogeneous criteria (WHO and RECIST) and wide ranges of timing (between 6 and 24 weeks) of assessments affect response and PFS rate estimation in first-line chemotherapy trials for advanced ovarian cancer. Several meta-analyses reported combining TTP and PFS as the same endpoint because of inconsistent endpoint definitions across the included trials and performed additional subgroup and sensitivity analyses to assure robust results (21,22). The Intermediate Clinical Endpoints in Cancer of the Prostate (ICECaP) Working Group identified important variations in local recurrence determination and in radiographic assessment methods for DFS and MFS endpoints in trials for localized prostate cancer (23). A subcommittee has been formed with the goal of harmonizing endpoints and data collection in future trials and to address incorporation of novel imaging within them. Incorporation of standardized endpoints for disease evaluation is also an important consideration for future trial designs, as has been attempted for adjuvant breast cancer studies (24) and for various diseases by the DATECAN group in Europe (25–28).

Comprehensive Reporting of Surrogacy Analysis Methods and Results

Surrogacy evaluations are subject to numerous statistical decisions. Inconsistent and incomplete reporting of the analysis methods and results leads to difficulty in assessing the appropriateness of the analyses and interpretability of study findings. For example, in the reporting of copula model-based individual-level and trial-level correlation, the choice of a copula (eg, Clayton's, Hougaard's, or Plackett's model) or a correlation coefficient (Spearman's rho vs Kendall's tau) needs to be transparent. It is possible that applying different copula models to the same dataset can lead to different conclusions about the nature and strength of association of two endpoints. Hence selecting an appropriate copula based on the goodness-of-fit (eg, goodness-of-fit tests, Akaike information criterion) (9,29,30) has been recommended to avoid post-hoc selection based on results. Empirically, we have noted Spearman's rho tended to be larger than Kendall's tau, and thus the choice of a measure of correlation is important in considering results of a study within the context of others. For studies that correlate DFS or PFS rates with OS rates at specific timepoints, it is important to justify the timepoint selection in the context of disease. The confidence intervals for the point estimate of R-squared as well as regression coefficients are essential in the reporting of trial-level correlation using linear regression. We also recommend reporting of prediction intervals (rather than confidence intervals) of regression line for the STE analysis because the interest is to predict the treatment effect on the definitive endpoint based on the observed treatment effect on the surrogate endpoint. In addition, surrogacy evaluation via regression analysis is sensitive to outliers, influential points, or nonlinear associations. The Anscombe's quartet (31) demonstrates the importance of graphing data (such as bubble plots and residual plots) along with providing numerical statistical metrics for surrogacy evaluation.

Validation and Sensitivity Analysis

It is also essential to validate a surrogacy model using well-established methods. The best situation is when trials not included in the meta-analysis can be used to validate the results of the surrogacy analyses, for instance, when IPD could be collected from a subset of trials and AD from the other trials (32). The leave-one-out cross-validation is also a commonly used approach when external datasets are limited. Validation assesses the prediction accuracy of a surrogate model. It is also an important tool to identify a potential influential outlier that greatly affects the slope of the regression line. Unfortunately, very few AD meta-analyses we reviewed reported validation for their surrogacy models.

For surrogacy evaluation, sensitivity and subgroup analysis provide further reassurance that the results are robust. For example, in the assessment of time to metastasis as a surrogate for prostate cancer-specific survival, a large proportion of these endpoints was censored due to non-prostate cancer deaths. As such, the ICECaP study team performed a sensitivity analysis by using cumulative incidence estimates of endpoints and subdistribution treatment effect hazard ratio estimates from competing risk models (23).

Discussion

Based on the systematic review of the reporting of the 80 meta-analyses, the conduct and reporting of surrogacy evaluation is inadequate. We have proposed a set of guidelines to improve the quality of the reporting of surrogate endpoint evaluation using meta-analyses (ReSEEM; Table 3), which can be used alongside the existing PRISMA and PRISMA-IPD guidelines for reporting meta-analyses.

The proposed guidelines apply to both IPD and AD meta-analytic surrogacy evaluation in general. Both types of surrogacy evaluation require adequate reporting to provide valuable information. When possible, IPD meta-analyses are preferred. As previously mentioned (5,6,33), the availability of IPD allows harmonizing the endpoint, estimating patient-level correlation, applying the copula model-based approach for trial-level correlation, and conducting more comprehensive subgroup and sensitivity analyses. Hence, formal surrogate evaluation should be based on a rigorous IPD meta-analysis at both the patient and trial levels. However, as a more expeditious approach, AD meta-analyses may provide some preliminary evidence before a more thorough but resource-intensive IPD meta-analysis is conducted. It can also serve for validation purposes if a plausible surrogate has been confirmed from a subset of trials with IPD.

The main focus of our review was to assess the quality of reporting of surrogate evaluation. Several previous studies examined the strength of surrogacy, suggesting that the relationship of DFS and PFS with OS vary considerably across tumor types (15,33–35). The challenges of identifying a valid surrogate for OS in cancer RCTs were attributed to multifaceted technical and clinical issues, such as the use of nonexhaustive sets of trials, RCTs allowing crossover after progression, and effective

salvage or subsequent lines of therapies (33,35). A recent simulation study by Broglio and Berry (36) showed that correlation between PFS and OS hazard ratios weakens as the median post-progression survival increases. Therefore, effective subsequent treatment may dilute the PFS and OS association. These studies provided important insights on the proper interpretation of surrogacy evidence in the context of the disease, treatment, trial designs, and chronological time.

In line with previous studies (6,33), our review revealed the heterogeneity of statistical methods for surrogacy evaluation, further highlighting the importance of building consensus on appropriate statistical techniques for surrogacy evaluation. Shi et al. (19) made head-to-head comparisons of several trial-level surrogacy measures in a large-scale simulation-based study, but no optimal approach was recommended. The authors, however, identified important trial characteristics that most affected the performance of these metrics and made practical recommendations for real applications (19). Efforts are also underway to develop sophisticated and integrated software for surrogacy evaluation (4,37). In the absence of uniformly accepted methods for surrogacy evaluation, we advocate transparency in the reporting of statistical decisions.

Notably, the contemporary meta-analytic surrogacy evaluation mainly focuses on radiographically based clinical endpoints such as DFS, PFS, and tumor response in the context of chemo- and targeted-therapies (Supplementary Table S1, available online). In the new era of immuno-oncology, reevaluation of the surrogacy of these endpoints is warranted. The surrogacy of novel endpoints in immune-oncology, such as immune-related response and correlative immune endpoints, also needs investigation (38). Moreover, there may be a difference in disease setting such as progression metastatic disease in metastatic disease vs progression to metastatic disease in the adjuvant or curative setting. Molecular, genetic, or immunoprofiling markers have great potential for use in clinical oncology, but few biomarkers have been robustly validated or confirmed as the surrogate for clinical endpoints through meta-analytic approach. The "outcome surrogacy" can be easily demonstrated by proving a biomarker's prognostic value through individual studies; however, there are greater challenges to demonstrate trial-level surrogacy because the latter requires high-quality biologic samples and standardized assays across multiple trials (39,40). With the increasing interest in exploring surrogate biomarkers, we expect formal validation of biomarkers through a robust meta-analytic approach will evolve in the future. The current reporting guidelines can be extended in the future to address additional aspects from the meta-analysis of a surrogate biomarker.

Meta-analysis of RCTs is a widely used approach for surrogacy evaluation in oncology. Our ReSEEM guidelines and recommendations will improve the quality in reporting and facilitate the interpretation and reproducibility of meta-analytics surrogacy evaluation. In an era with emerging novel therapies and enhanced understanding of cancer biology and genetics, identification of novel, more rapid clinical endpoints are critically needed to expedite drug development and to derive benefit for subpopulations for a personalized medicine. Given the rapid growth in this area, our ReSEEM guidelines should help promote greater methodological consistency and could also serve as an evaluation tool in the peer review process for assessing surrogacy research. These recommendations have high potential to foster more efficient trial design and conduct in the future.

Funding

This work was supported by the Dana Foundation Funds (to MMR, WX, CJS); and by the United States Army Medical Research W81XWH-15-1-0467 (to SH).

Notes

Affiliations of authors: Dana-Farber Cancer Institute, Boston, MA (WX, CJS, MMR); Duke University Medical Center, Durham, NC (SH); MRC Clinical Trials Unit at UCL, London, UK (JFT, MRS); European Organization for Research and Treatment of Cancer, Brussels, Belgium (LC); University of Chicago, Chicago, IL (JJD); International Drug Development Institute Inc, San Francisco, CA (MB); Harvard Medical School, Boston, MA (CJS, MMR).

This work arose from the ICECaP Working Group collaboration (14), which was supported by the Prostate Cancer Foundation Challenge Award and grants from Astellas Pharma, Bayer, Medivation, Janssen Pharmaceuticals, Millennium Pharmaceuticals, Sotio, and Sanofi.

We thank Victoria Wong, Frontier Science & Technology Research Foundation, for assistance with computer programming.

References

1. Katz R. Biomarkers and surrogate markers: an FDA perspective. *Neurorx*. 2004;1(2):189–195.
2. FDA-NIH Biomarker Working Group. *BEST (Biomarkers, EndpointS, and Other Tools) Resource*. Silver Spring (MD): Food and Drug Administration (US); Bethesda (MD): National Institutes of Health (US); 2016.
3. Buyse M, Molenberghs G, Burzykowski T, et al. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*. 2000; 1(1):49–67.
4. Alonso A, Bigirumurame T, Burzykowski T, et al. *Applied Surrogate Endpoint Evaluation Methods with SAS and R (Chapman & Hall/CRC Biostatistics Series)*. 1st ed. CRC Press; 2016.
5. Shi Q, Sargent DJ. Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. *Int J Clin Oncol*. 2009;14(2):102–111.
6. Ciani O, Davis S, Tappenden P, et al. Validation of surrogate endpoints in advanced solid tumors: systematic review of statistical methods, results, and implications for policy makers. *Int J Technol Assess Health Care*. 2014;30(03): 312–324.
7. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009; 6(7):e1000097.
8. Stewart LA, Clarke M, Rovers M, et al. Preferred reporting items for systematic review and meta-analyses of individual participant data: the PRISMA-IPD statement. *JAMA*. 2015;313(16):1657–1665.
9. Burzykowski T, Molenberghs G, Buyse M, et al. Validation of surrogate end points in multiple randomized clinical trials with failure time endpoints. *J Royal Statistical Soc C*. 2001;50(4):405–4222.
10. Renfro LA, Shi Q, Sargent DJ, et al. Bayesian adjusted R2 for the meta-analytic evaluation of surrogate time-to-event endpoints in clinical trials. *Stat Med*. 2012;31(8):743–761.
11. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharmaceut Statist*. 2006;5(3): 173–186.
12. Korn EL, Freidlin B. Surrogate and intermediate endpoints in randomized trials: what's the goal? *Clin Cancer Res*. 2018;24(10):2239–2240.
13. Rotolo F, Pignon JP, Bourhis J, et al. Surrogate end points for overall survival in loco-regionally advanced nasopharyngeal carcinoma: an individual patient data meta-analysis. *J Natl Cancer Inst*. 2017;109(4): djw239.
14. ICECaP Working Group. The Development of Intermediate Clinical Endpoints in Cancer of the Prostate (ICECaP). *J Natl Cancer Inst*. 2015;107(12): djv261.
15. Prasad V, Kim C, Burotto M, et al. The strength of association between surrogate end points and survival in oncology: a systematic review of trial-level meta-analyses. *JAMA Intern Med*. 2015;175(8):1389–1398.
16. Petrelli F, Barni S. Surrogate endpoints in metastatic breast cancer treated with targeted therapies: an analysis of the first-line phase III trials. *Med Oncol*. 2014;31(1):776.
17. Dwan K, Gamble C, Williamson PR, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One*. 2013;8(7):e66844.

18. Chan AW, Hróbjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*. 2004;291(20):2457–2465.
19. Shi Q, Renfro LA, Bot BM, et al. Comparative assessment of trial-level surrogacy measures for candidate time-to-event surrogate endpoints in clinical trials. *Comput Stat Data Anal*. 2011;55(9):2748–2757.
20. Colloca G, Venturino A. Trial-level analysis of progression-free survival and response rate as end points of trials of first-line chemotherapy in advanced ovarian cancer. *Med Oncol*. 2017;34(5):87.
21. Chirila C, Odom D, Devercelli G, et al. Meta-analysis of the association between progression-free survival and overall survival in metastatic colorectal cancer. *Int J Colorectal Dis*. 2012;27(5):623–634.
22. Miksad RA, Zietemann V, Gothe R, et al. Progression-free survival as a surrogate endpoint in advanced breast cancer. *Int J Technol Assess Health Care*. 2008; 24(4):371–383.
23. Xie W, Regan MM, Buyse M, et al. Metastasis-free survival is a strong surrogate of overall survival in localized prostate cancer. *J Clin Oncol*. 2017;35(27): 3097–3104.
24. Hudis CA, Barlow WE, Costantino JP, et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: the STEEP system. *J Clin Oncol*. 2007;25(15):2127–2132.
25. Gourgou-Bourgade S, Cameron D, Poortmans P, et al. Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials). *Ann Oncol*. 2015;26(12):2505–2506.
26. Kramar A, Negrier S, Sylvester R, et al. Guidelines for the definition of time-to-event end points in renal cell cancer clinical trials: results of the DATECAN project. *Ann Oncol*. 2015;26(12):2392–2398.
27. Bonnetain F, Bonsing B, Conroy T, et al. Guidelines for time-to-event endpoint definitions in trials for pancreatic cancer. Results of the DATECAN initiative (Definition for the Assessment of Time-to-event End-points in CANcer trials). *Eur J Cancer*. 2014;50(17):2983–2993.
28. Bellera CA, Penel N, Ouali M, et al. 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(5):865–872.
29. Dimier N, Todd S. An investigation into the two-stage meta-analytic copula modelling approach for evaluating time-to-event surrogate endpoints which comprise of one or more events of interest. *Pharm Stat*. 2017;16(5): 322–333.
30. Renfro LA, Shang H, Sargent DJ. Impact of copula directional specification on multi-trial evaluation of surrogate endpoints. *J Biopharm Stat*. 2015;25(4): 857–877.
31. Anscombe FJ. Graphs in statistical analysis. *Am Stat*. 1973;27(1):17–21.
32. Buyse M, Molenberghs G, Paoletti X, et al. Statistical evaluation of surrogate endpoints with examples from cancer trials. *Biom J*. 2016;58(1):104–132.
33. Savina M, Gourgou S, Italiano A, et al. Meta-analyses evaluating surrogate endpoints for overall survival in cancer randomized trials: a critical review. *Crit Rev Oncol Hematol*. 2018;123:21–41.
34. 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(6):713–725.
35. Sherrill B, Kaye JA, Sandin R, et al. Review of meta-analyses evaluating surrogate endpoints for overall survival in oncology. *Onco Targets Ther*. 2012;5: 287–296.
36. Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. *J Natl Cancer Inst*. 2009;101(23): 1642–1649.
37. Rotolo F, Paoletti X, Michiels S. Surrosurv: an R package for the evaluation of failure time surrogate endpoints in individual patient data meta-analyses of randomized clinical trials. *Comput Methods Programs Biomed*. 2018;155: 189–198.
38. Anagnostou V, Yarchoan M, Hansen AR, et al. Immuno-oncology trial endpoints: capturing clinically meaningful activity. *Clin Cancer Res*. 2017;23(17): 4959–4969.
39. Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points–the challenge of statistical validation. *Nat Rev Clin Oncol*. 2010;7(6):309–317.
40. Buyse M, Burzykowski T, Molenberghs G, Alonso A, Biomarker-based surrogate endpoints. In: Matsui S, Buyse M, Simon R, eds. *Design and Analysis of Clinical Trials for Predictive Medicine*. New York: Chapman and Hall/CRC Press; 2015:333–369.