Journal Club Comment

Prapaporn Pornsuriyasak, M.D.

6 Jan 2023

Surrogate endpoint needs to be validated

- Overall survival (OS): gold standard primary efficacy endpoint in oncology randomized controlled trials (RCTs)
- But often requires prolonged follow-up and a substantial number of patients and OS is likely influenced by subsequent lines of therapies
- If using surrogate endpoint, it needs to be associated with the disease and treatment
- “a drug-induced effect on the surrogate predicts the desired effect on the clinical outcome of interest” using robust statistical methods
- Important to note that surrogate endpoint may be valid for a particular indication, not for all cancers

REVIEW

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

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

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

- Meta-analysis of RCTs is a widely used approach for surrogacy evaluation in oncology
- ReSEEM guidelines and recommendations will improve the quality in reporting and facilitate the interpretation and reproducibility of metaanalytics surrogacy evaluation

Two-stage meta-analytic approach for surrogacy evaluation

- Stage 1 demonstration of strong correlation between the
- surrogate and definitive endpoints

("outcome surrogacy")

- Stage 2 demonstration of correlation of treatment effects on both endpoints

("trial-level or effect surrogacy")

Meta-analysis for surrogacy endpoints

- Individual patient data (IPD) meta-analysis
  - Enables the standardization of methods across IPD sets and robust analysis at both the patient and trial levels
  - Time and resource intensive

- Aggregate data (AD)

REVIEW

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

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

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

Eligible studies

- Evaluated surrogate endpoints using meta-analyses of RCTs in oncology
- published in English as full text

Excluded studies

- 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

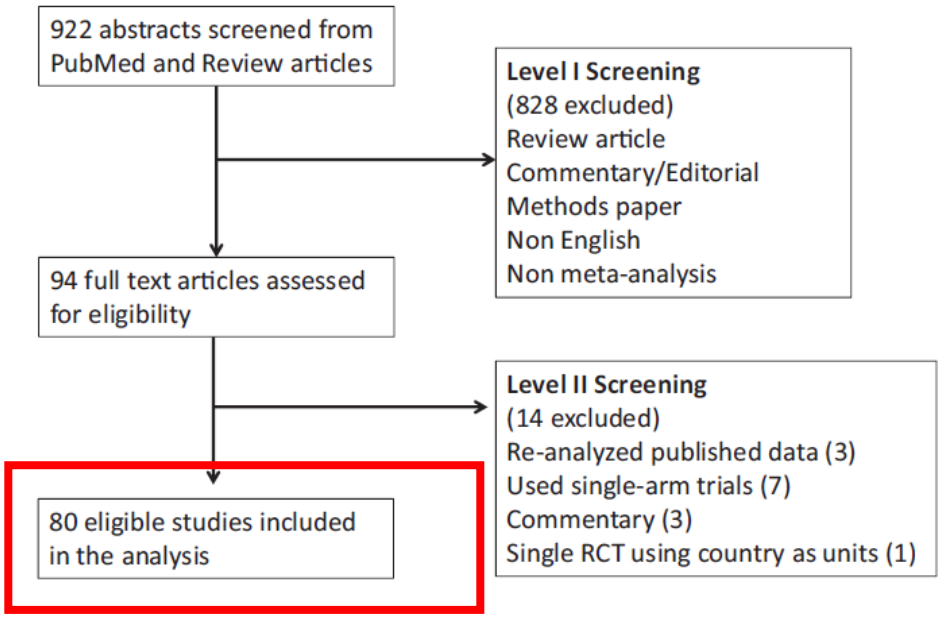
Search PubMed database: No restriction on start date up to Aug 2017

REVIEW

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

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

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



Using IPD 27%, AD 73%

Localized disease 19%:

- DFS as surrogate endpoint 87%

Advanced/metastatic disease 81%

- PFS or TTP 83%
- Tumor response 46%

OS definitive endpoint

Colorectal (17 studies), breast (15 studies), and lung cancers (14 studies)

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)

Two-stage meta-analytic approach for surrogacy evaluation

- Stage 1 demonstration of strong correlation between the surrogate and definitive endpoints: OS ("outcome surrogacy")

IPD

- Copula model to estimate individual-level correlation coefficient for failure-time endpoints
- Correlation between binary surrogate (tumor response) and time-to-event endpoint: Cox regression expressed as HR, log-rank test

AD

- Correlation between endpoints on estimated summary metrics with trial and treatment arm as the analysis unit: linear regression Pearson, Spearman, or Kendall's tau

- Stage 2 demonstration of correlation of treatment effects on both endpoints ("trial-level or effect surrogacy")
  - Pearson, Spearman
  - R2 confidence interval
  - Surrogate threshold effect (STE)

Results of surrogacy analysis

1. Patient level responder analysis

- Spearman rank correlation coefficient
- assess individual-level associations between PFS and OS considering the individual patient data from all the clinical trials
- patient-level responder analysis was performed to compare PFS and OS between responders and non-responders irrespective of the treatment assignment using the pooled data set

Cancer Therapy: Clinical

Clinical Cancer Research

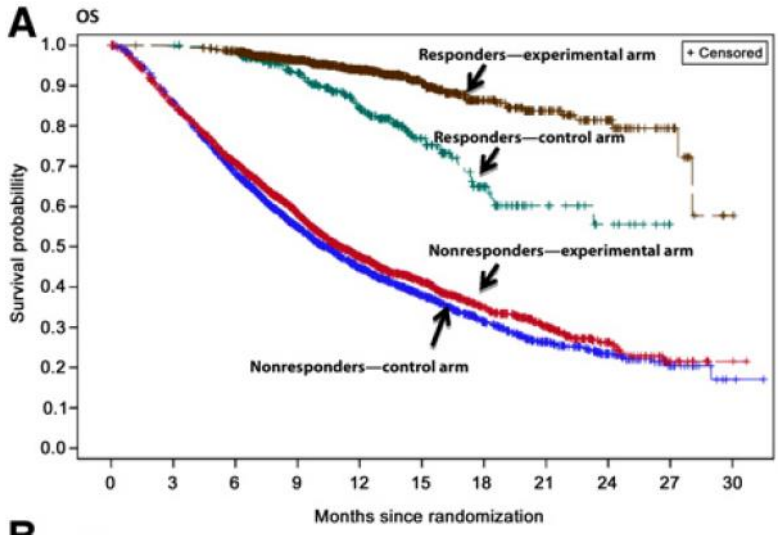
Evaluation of Overall Response Rate and Progression-Free Survival as Potential Surrogate Endpoints for Overall Survival in Immunotherapy Trials

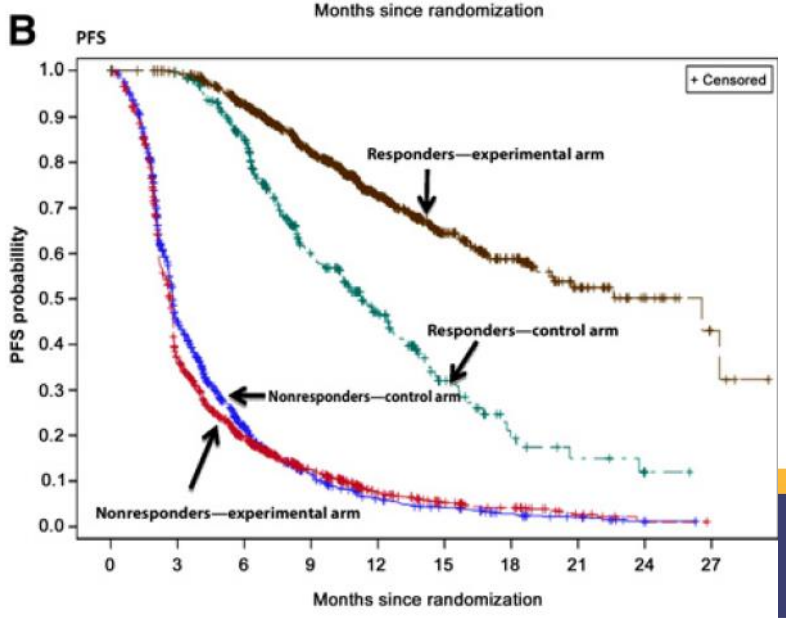
Clin Cancer Res 2018; 24(10); 2268–75

Table 1. Summary of clinical trials

Randomized comparisons	Indication	Endpoints		Number of patients	Randomization ratio	OS HR	PFS HR	ORR OR	Response rates (trt vs. control)
		Primary	Secondary						
1	Melanoma	PFS (IRC), OS	ORR (IRC)	359	1:1	0.86	0.598	6.1957	22.5 vs. 4.5%
2	Melanoma	PFS (IRC), OS	ORR (IRC)	360	1:1	0.771	0.498	8.6163	28.7 vs. 4.5%
3	Melanoma	OS, PFS (IRC)	ORR (IRC)	557	1:1	0.673	0.567	3.542	41.3 vs. 16.6%
4	Melanoma	OS, PFS (IRC)	ORR (IRC)	555	1:1	0.67	0.581	3.3401	39.9 vs. 16.6%
5	Melanoma	ORR (IRC), OS	PFS (IRC)	182[a]	2:1	0.864	0.736	4.9762	31.2 vs. 8.3%
6	Melanoma	OS	ORR, PFS (INV)	418	1:1	0.418	0.432	4.1149	40 vs. 13.9%
7	Melanoma	PFS (INV), OS	ORR (INV)	631	1:1	0.708	0.572	3.2949	43.7 vs. 19%
8	Melanoma	PFS (INV), OS	ORR (INV)	629	1:1	0.702	0.432	5.7838	57.6 vs. 19%
9	Melanoma	ORR (INV)	PFS (INV)	142	2:1	0.792	0.396	13.5655	55.8 vs. 8.5%
10	NSCLC[b]	OS, PFS (IRC)	ORR (IRC)	687	1:1	0.732	0.869	1.6951	25.5 vs. 16.8%
11	NSCLC[b]	OS, PFS (IRC)	ORR (IRC)	689	1:1	0.631	0.788	1.694	25.5 vs. 16.8%
12	NSCLC[c]	PFS (IRC)	OS, ORR (IRC)	305	1:1	0.614	0.499	2.032	47.6 vs. 30.9%
13	NSCLC	OS	PFS, ORR (INV)	287	1:1	0.675	0.912	1.0476	15.3 vs. 14.7%
14	NSCLC non-SQ	OS	ORR, PFS (INV)	582	1:1	0.75	0.909	1.6742	19.2 vs. 12.4%
15	NSCLC SQ	OS	ORR, PFS (INV)	272	1:1	0.59	0.63	2.6042	20 vs. 8.8%
16	RCC	OS	ORR, PFS (INV)	821	1:1	0.758	0.873	5.9323	25.1 vs. 5.4%
17	SCCHN	OS	PFS, ORR (INV)	361	2:1	0.691	0.882	2.5055	13.3 vs. 5.8%

Patient-level responder analysis





- Kaplan–Meier plots for OS (A) and PFS (B) for the responders and non-responders in the experimental and control arms
- The responders had longer survival times compared with non-responders

  [HR 0.14, 95% CI 0.12–0.16],

  irrespective of treatment group.
- Similarly, the responders had longer PFS

  [HR 0.12, 95% CI 0.11–0.14]

Results of surrogacy analysis

2. Trial-level analysis: defined at the population level

evaluates the ability of a surrogate endpoint to predict the effect of treatment on the true endpoint

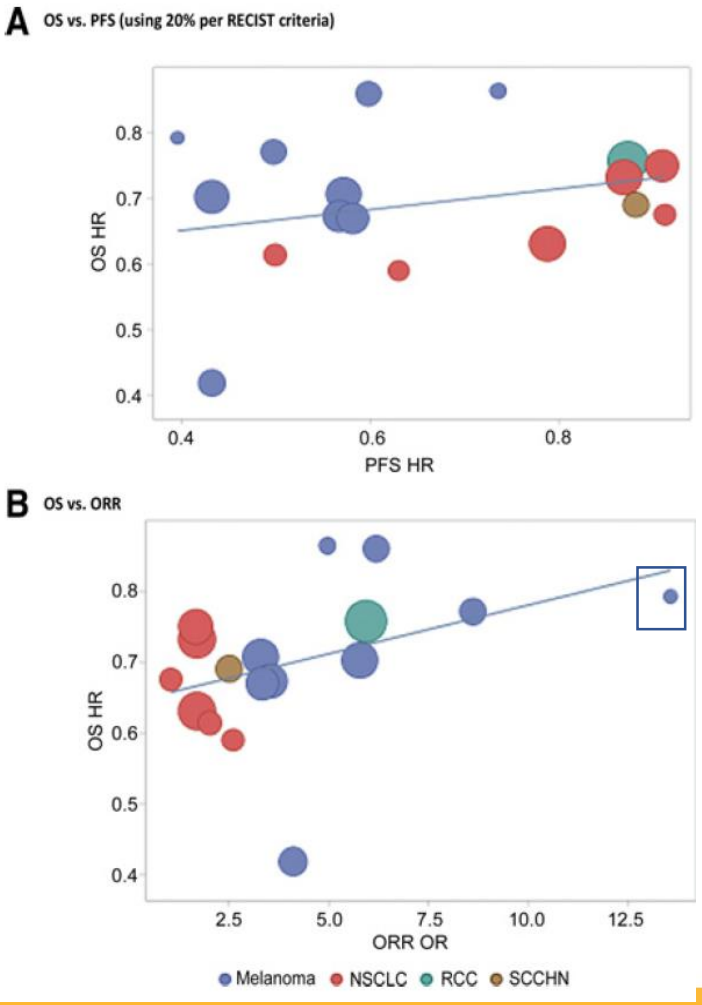
- Associations between treatment effect based on ORR, PFS, and OS were evaluated using weighted linear regression model with weights equal to the sample size of each randomized comparison
- The treatment effects based on PFS and OS were measured by the log of HR estimated using unstratified Cox proportional hazards model with treatment as the covariate

Trial-level analysis

- The coefficient of determination (R2) value

  computed from the weighted linear regression model: strength of association between PFS and OS
- R2 close to 1: perfect surrogate endpoint
- R2 =0: no association between the surrogate endpoint and the true benefit endpoint

Trial level association



- R2= 0.13
- R2= 0.127

This suggests that, although responders lived longer without disease progression and death, both PFS and ORR are not good surrogate endpoints for OS

Trial-level surrogacy

Surrogate threshold effect (STE)

- The minimum treatment effect on the surrogate necessary to predict a non-zero effect on the definitive endpoint

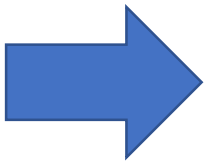
- Using the 95% prediction limits of the regression line

Reporting Sensitivity, Subgroup Analyses

- To assure robust results
  - Use of various cutoffs for follow-up times
  - Exclusion of certain trials
  - Use of alternative endpoint definitions
  - Use of alternative weights
  - Statistical models for analysis

Recommendations for reporting of meta-analytic surrogacy evaluation

- Study design and data collection
- Protocol and registration
- Eligibility criteria
- Information sources
- Search
- Study selection
- Data collection process
- Risk of bias within studies/across studies



Follow PRISMA and PRISMA-IPD statement

Recommendations for reporting of meta-analytic surrogacy evaluation

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)
		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)

REVIEW

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

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

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

Recommendations for the future

1. Develop protocol for surrogacy evaluation
2. Sufficient reporting on trial selection
3. Sufficient reporting of characteristics of the included RCTs
4. Adequately define endpoints for surrogacy evaluation
5. Comprehensive reporting of surrogacy analysis methods and results
6. Validation and sensitivity analysis

Difference of meta-analysis for surrogacy evaluation from other meta-analyses

Meta-analysis

- Aim for the pooled treatment effects

Meta-analysis for surrogacy evaluation

- targets broader treatment classes
- more mixed study population
- multiple surrogate candidates

1. Develop protocol for surrogacy evaluation

- Help investigators to choose appropriate candidate surrogates, generate hypotheses in the context of disease, and define a priori the threshold of correlation
- $R2 \geq 0.7$ used as a conventional degree of correlation to define surrogacy in oncology
- Define subgroup and sensitivity analyses as a priori in the protocol and justify them in the context of disease

2. Sufficient reporting on trial selection

- Only about 50% of potential eligible studies were ultimately included in even the most rigorous meta-analyses for trial-level surrogacy evaluation

- Surrogacy evaluation requires outcome and treatment effect data for multiple endpoints, and not all eligible trials can provide such complete data needed for evaluation

  - Some AD meta-analyses can include only RCTs that reported both PFS and OS hazard ratios

  - Excluded trials that did not report PFS and OS hazard ratios were more likely to have negative treatment effects on these endpoints

2. Sufficient reporting on trial selection

- A simulation study demonstrated that surrogacy estimation via regression models is more accurate and precise in the settings of
  - larger number of trials
  - low rate of censoring
  - wide range of treatment effects

    (i.e, including both positive and negative trials)

- When one or more factors deviate from the “optimal” scenarios, regression tends to **underestimate** the definitive surrogacy with increased variability

2. Sufficient reporting on trial selection

- A wide range of treatment effects across included trials **improves** performance of the regression model and the magnitude of R2

(because variance of the regression coefficient is inversely proportional to the spread of the predictor variable)

- Heterogeneity of treatment effects (i.e, including both positive and negative trials) is an advantage rather than a drawback in a meta-analysis aimed at surrogate validation

- Quantification of heterogeneity may not apply to meta-analytic surrogacy evaluation

2. Sufficient reporting on trial selection

- Use of non-extensive sets of trials might be less concern (compared with other meta-analyses) as long as the included trials reflect the expected range of heterogeneity of treatment effects
- Study selection process: broad and clear
- To reduce potential bias due to trial selection, reasons for trial exclusion should be reported
- Comparison of trial characteristics and range of treatment effects across trials between included and excluded eligible trials provide valuable information for trial representation and generalization

3. Sufficient reporting of characteristics of the included RCTs

- Important to report patient characteristics for the included trials (e.g, class of therapy, enrollment period, follow-up duration, patient characteristics: age and tumor staging), important prognostic factors with respect to disease evolution and potentially affect treatment options, post-progression survival), and cause of death
- Surrogacy results can be appropriately interpreted only if the study population is well characterized
- As the strength of surrogacy may vary according to certain trial-level or patient characteristics, **subgroup analysis** is an important strategy to explore the heterogeneity

3. Sufficient reporting of characteristics of the included RCTs

- For time-to-event endpoints (DFS and PFS),
  - follow-up duration
  - numbers of events

1. Determine the precision of point and interval estimates

2. Help to choose appropriate weights for the weighted regression analysis in the presence of unbalanced follow-up across trials

- Unfortunately, in most studies authors only reported sample sizes of included RCTs and performed linear regression weighted on trial sample size without considering study follow-up duration

4. Adequately define endpoints for surrogacy evaluation

Definition of endpoint

IPD meta-analyses

- important to define endpoints accurately, with details on trial-specific **disease assessment criteria** and **schedule**

AD meta-analyses using the protocol definition

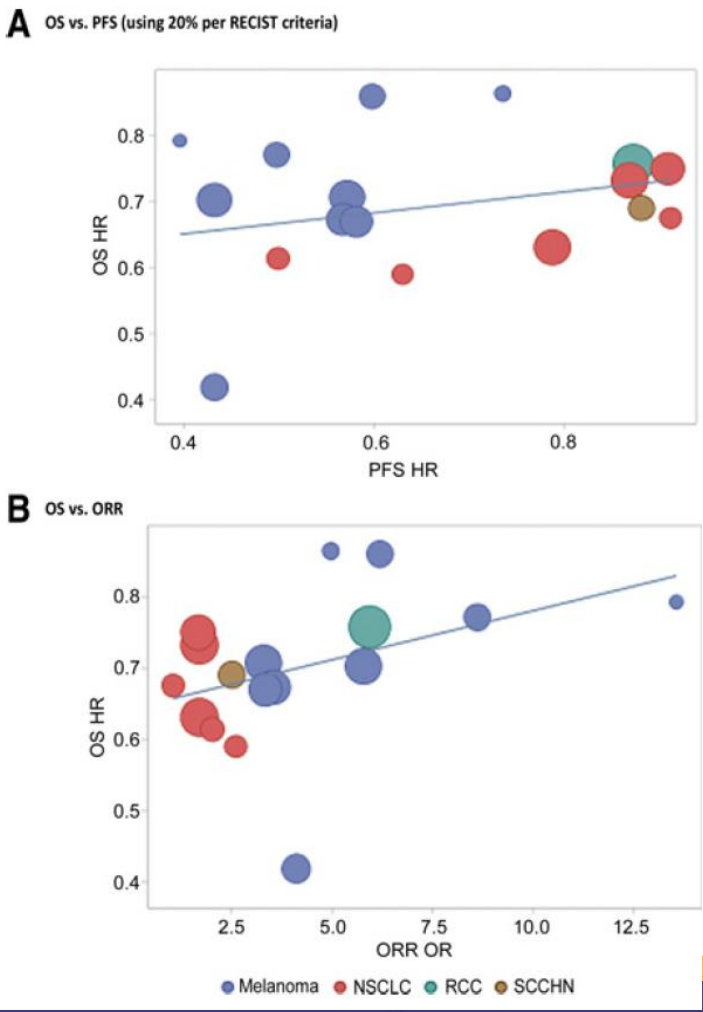
- efforts need to be made to assess and report variability in definition of endpoint

5. Comprehensive reporting of surrogacy analysis methods and results

- Applying different copula models to the same dataset can lead to different conclusions about the nature and strength of association of 2 endpoints
  - Reporting of copula model-based individual level and trial-level correlation, based on goodness-of-fit or correlation coefficient (Spearman's rho vs Kendall's tau) should be transparent
- For studies that correlate DFS or PFS rates with OS rates at specific timepoint using linear regression, important to justify the timepoint selection in the context of disease
  - Point estimate of R2 (prediction interval or 95%CI)
  - Regression coefficient (95%CI)

5. Comprehensive reporting of surrogacy analysis methods and results

- Regression analysis
- sensitive to outliers and influential points
  - Trial-level association
    - graphing data (e.g, bubble plots and residual plots) along with numerical statistical metrics



6. Validation and sensitivity analysis

Validation

- Assess the prediction accuracy of a surrogate model
- Identify potential influential outlier that greatly affects the slope of the regression line
- External dataset: IPD, AD from other trials

Sensitivity and subgroup analysis

- provide further reassurance that the results are robust
- If large proportion of endpoint (time-to-metastasis surrogate for cancer-specific survival) was censored due to non-cancer deaths

"sensitivity analysis by using cumulative incidence estimates of endpoints and subdistribution treatment effect hazard ratio estimates from competing risk model"

Summary

- IPD meta-analyses are preferred because these 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

- AD meta-analyses may provide some preliminary evidence before IPD. It can serve for validation purposes if a plausible surrogate has been confirmed from a subset of trials with IPD

- A surrogate endpoint is considered to be validated for use in phase III clinical trials when there is a strong association at both individual and trial levels

Thank you