

BMJ Open Systematic bias between blinded independent central review and local assessment: literature review and analyses of 76 phase III randomised controlled trials in 45 688 patients with advanced solid tumour

Jianrong Zhang,^{1,2,3,4} Yiyin Zhang,^{5,6,7} Shiyan Tang,^{5,8} Long Jiang,^{1,2,3} Qihua He,^{1,2,3} Lindsey Tristine Hamblin,^{1,9} Jiaxi He,^{1,2,3} Zhiheng Xu,^{2,3} Jieyu Wu,^{10,11} Yaoqi Chen,^{1,2,3} Hengrui Liang,⁵ Difei Chen,⁵ Yu Huang,⁵ Xinyu Wang,⁵ Kexin Deng,⁵ Shuhan Jiang,⁵ Jiaqing Zhou,⁵ Jiaxuan Xu,⁵ Xuanzuo Chen,⁵ Wenhua Liang,^{1,2,3} Jianxing He^{1,2,3}

To cite: Zhang J, Zhang Y, Tang S, *et al*. Systematic bias between blinded independent central review and local assessment: literature review and analyses of 76 phase III randomised controlled trials in 45 688 patients with advanced solid tumour. *BMJ Open* 2018;**8**:e017240. doi:10.1136/bmjopen-2017-017240

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-017240>).

JZ, YZ and ST contributed equally.

Some data of this research have been presented at ESMO 2016 Annual Meeting.

Received 10 April 2017
Revised 11 April 2018
Accepted 24 July 2018



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Wenhua Liang;
liangwh1987@163.com and Dr Jianxing He;
drjianxing.he@gmail.com

ABSTRACT

Objective Unbiased assessment of tumour response is crucial in randomised controlled trials (RCTs). Blinded independent central review is usually used as a supplemental or monitor to local assessment but is costly. The aim of this study is to investigate whether systematic bias existed in RCTs by comparing the treatment effects of efficacy endpoints between central and local assessments.

Design Literature review, pooling analysis and correlation analysis.

Data sources PubMed, from 1 January 2010 to 30 June 2017.

Eligibility criteria for selecting studies Eligible articles are phase III RCTs comparing anticancer agents for advanced solid tumours. Additionally, the articles should report objective response rate (ORR), disease control rate (DCR), progression-free survival (PFS) or time to progression (TTP); the treatment effect of these endpoints, OR or HR, should be based on central and local assessments.

Results Of 76 included trials involving 45 688 patients, 17 (22%) trials reported their endpoints with statistically inconsistent inferences (p value lower/higher than the probability of type I error) between central and local assessments; among them, 9 (53%) trials had statistically significant inference based on central assessment. Pooling analysis presented no systematic bias when comparing treatment effects of both assessments (ORR: OR=1.02 (95% CI 0.97 to 1.07), p=0.42, I²=0%; DCR: OR=0.97 (95% CI 0.92 to 1.03), p=0.32, I²=0%; PFS: HR=1.01 (95% CI 0.99 to 1.02), p=0.32, I²=0%; TTP: HR=1.04 (95% CI 0.95 to 1.14), p=0.37, I²=0%), regardless of funding source, mask, region, tumour type, study design, number of enrolled patients, response assessment criteria, primary endpoint and trials with statistically consistent/inconsistent inferences. Correlation analysis also presented no sign of systematic bias between central and local assessments (ORR, DCR, PFS: r>0.90, p<0.01; TTP: r=0.90, p=0.29).

Strengths and limitations of this study

- To our knowledge, this is the largest literature review and pooling analysis comparing treatment effects between blinded independent central review and local assessment in phase III randomised controlled trials on solid tumours.
- We performed an exhaustive literature search to include all potential studies fulfilling the inclusion criteria.
- We carefully extracted the data based on the independent and double-blind principle, in order to guarantee the accuracy of the data applied for further analysis.
- Compared with our study-level analysis, the analysis using individual patients' data could be more robust.
- For using trial data of both blinded independent central review and local assessment, the findings and conclusion of this research may not be generalisable for all phase III oncological randomised controlled trials, because the situation of either assessment could be unknown when trials did not implement or report both central and local assessments.

Conclusions No systematic bias could be found between local and central assessments in phase III RCTs on solid tumours. However, statistically inconsistent inferences could be made in many trials between both assessments.

INTRODUCTION

In phase III randomised controlled trials (RCTs), response-related or progression-related endpoints like objective response rate (ORR), disease control rate (DCR), progression-free survival (PFS) and time to

progression (TTP) are key for reflecting treatment effects of the experimental arm and the control arm for patients with advanced solid tumour.¹⁻³ During trials, determination of tumour response should be assessed with accuracy, which is the prerequisite of implementation with standardised response assessment criteria (eg, Response Evaluation Criteria in Solid Tumors (RECIST) and WHO) as well.

Unlike overall survival, these endpoints assessed by local investigators are more influenced by subjective factors, including variability during tumour measurement, target lesion selection, failure to diagnose new lesions and different interpretations of non-target or immeasurable lesions.⁴ In open-label trials, the knowledge of investigators regarding treatment assignment could influence their assessment. Even in some double-blind trials, the investigators' knowledge may not be completely eliminated due to the adverse effects; for example, the investigators might be able to tell which treatments are assigned for their patients according to the different manifestations of treatments' adverse effects.⁵

Treatment effect is one of the main results considered for drug approval. If aforementioned subjective factors impact the assessment for trial endpoints, the subsequent result will overestimate or underestimate the true effect of treatments, which is called systematic bias.⁶ In order to detect potential bias from local investigators, blinded independent central review is requested by the regulatory authorities (eg, the US Food and Drug Administration (FDA)). During its implementation, all imaging examinations are reviewed by independent radiologists who are blinded to patients' treatment assignments and clinical information.⁷ However, this mechanism has some drawbacks. It increases the burden of time and expenditure on trials. Additionally, it may introduce missing data, information censoring and the neglect of symptomatic progression. These factors could result in different discrepancy rates of central and local assessments and sometimes among central reviewers themselves, which impacts treatment effects and may even cause potential bias.^{4,7,8}

Given the pros and cons of assessment by central reviewers, the FDA Oncology Drugs Advisory Committee discussed how to design a reliable assessment strategy for clinical trials with central review: if there is no strong evidence indicating systematic bias from two assessments, a sample-based central review could be considered in future usage instead of the complete assessment for all patients in the trials.⁹ This strategy may effectively reduce the complexity and implementation burden, without compromising the reliability of the RCTs.⁹

Accordingly, in order to understand the reliability of local assessment, as well as the necessity of central review, we conducted this literature review and analyses in order to investigate whether systematic bias existed in previous phase III RCTs on solid tumours.

METHOD

Search strategy and study selection

In accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement,¹⁰ a PubMed search was conducted by JRZ using the dates of 1 January 2010 to 30 June 2017. The search strategy is shown in online supplementary table 1. Inappropriate articles such as reviews, systematic reviews and/or meta analyses, guidelines and commentaries were excluded.

Eligible trials were those directly evaluating therapeutic efficacy of anticancer agents in phase III RCTs for patients with advanced solid tumour; additionally, the imaging assessment for tumour response or progression was conducted by both central reviewers and local investigators. As some authors reported their data in more than one article, we used the name and/or National Clinical Trial (NCT) number of eligible RCTs as search terms to re-search PubMed (without the time interval limitation), to find out if there were more available articles on those RCTs. Endnote X7 (Thomson Reuters, New York City, New York, USA) was used in the above process.

Data extraction

The process of data extraction was carried out independently and double-blindly by three reviewers (JRZ with YYZ and SYT; in blocks of 50 articles allocated at random; discrepancies were resolved by WHL). To ensure consistency between reviewers, we used the same data extraction form, piloted the data extraction by using a sample of 16 included trials and had discussions before and during the extraction process to confer how to properly extract and interpret the data.

The following characteristics of each trial were extracted: author, year, NCT number, funding source (pharmaceutical or academic), mask (open label, single blind or double blind), region (global or intracontinental), tumour type (eg, breast cancer, ovarian cancer, melanoma), study design (superiority, non-inferiority or hybrid; hybrid design includes the design of superiority and non-inferiority), number of enrolled patients, response assessment criteria (RECIST or WHO), primary endpoint (central assessed, local assessed or other) and the statistical inference of the primary endpoint according to whether the p value was lower than the probability of the type I error (positive, negative or indeterminate). We also extracted estimated treatment effects from both central and local assessments, including the OR of experimental arm ORR to control arm ORR, OR of experimental arm DCR to control arm DCR, HR of experimental arm PFS to control arm PFS, and HR of experimental arm TTP to control arm TTP. Regarding overlapped data from more than one article on one trial, we selected data based on primarily larger analysis or recently updated analysis. For PFS and TTP, if both intention to treat (or other methods with a larger population) and per-protocol population were available for trials' treatment effects, we preferred the former in our

research. According to characteristics, the risk of bias was evaluated in each trial (online supplementary efigure 1).

Statistical analysis

First, we investigated whether there were trials with statistically inconsistent inferences between two assessments in primary and secondary endpoints (including ORR, PFS and TTP). If these trials could be identified, we calculated the percentage of these trials among all our eligible trials. Statistically inconsistent inferences are defined as the treatment effect from one of the assessments (eg, central assessment) indicating significant difference (p value is lower than the probability of the type I error or the confidence interval of the treatment effect does not cross 1), but the treatment effect from another assessment (eg, local assessment) indicating non-significant difference (p value is higher than the probability of the type I error, or the confidence interval of the treatment effect crosses 1).

Furthermore, to statistically investigate whether systematic bias existed, we made a comparison of treatment effects between central and local assessments, by conducting a pooling analysis with the inverse variance method and fixed-effect model in Review Manager 5.3 (The Cochrane Collaboration, London, England). In this process, if the corresponding p value for heterogeneity was less than 0.05 or the I^2 index was over 50%, we used a random-effect model instead of the fixed-effect model in order to reduce the effect of heterogeneity. The pooled OR and HR were the measure of this comparison, expressed as the ratio of central-assessed treatment effects (eg, OR of ORR, OR of DCR, HR of PFS, HR of TTP) to local-assessed treatment effects.¹¹ The OR (of ORR or DCR) greater than 1 indicated that central review overestimated the efficacy of the therapeutic strategy in the experimental arm; while a HR (of PFS or TTP) greater than 1 indicated that central review underestimated the therapeutic efficacy of the experimental arm (compared with local assessment). Regardless of whether the ratio was higher or lower than 1, we concluded no sign of a significant systematic bias if: (1) the corresponding p value was higher than 0.05, which means the 95% CI of the pooled ratio (HR, OR) crossed 1; (2) the 95% CI of the pooled ratio was extremely tight (<5%) if the first consideration was not met. For the above summary synthesis of ORR, DCR, PFS and TTP, a funnel plot was used to estimate publication bias (online supplementary efigure 2). Furthermore, we conducted subgroup analysis based on the trial characteristics: funding source, mask, region, trial design, number of enrolled patients (based on median value of all included trials), tumour type, response assessment criteria, primary endpoint and its outcome, as well as statistical inferences between central and local assessments (consistent/inconsistent).

In order to verify the result of the pooling analysis, we conducted correlation analysis for the treatment effects between central and local assessments, by using SPSS V.23 (SPSS, Chicago, Illinois, USA). The test for normality was completed first, followed by correlation analysis with a

bivariate model: if normal distribution was indicated, we estimated the correlation by the Pearson correlation coefficient; if not, the Spearman's correlation was applied. Significant correlation was indicated when the p value was less than 0.05. The correlation between two assessments was also demonstrated in scatterplots, constructed by using Excel 2011 (Microsoft, Seattle, Washington, USA).

Patient and public involvement

Due to the nature of the literature review, we do not have patient and public involvement in this research.

RESULTS

Trial searching and characteristics

Based on article identification and selection (figure 1), we included a total of 76 trials from 100 articles, involving 45 688 randomly assigned patients.^{12–111}

Summary and detailed characteristics are presented in table 1 and in online supplementary efigure 2. A majority of the 100 articles were published in high-impact journals: *Journal of Clinical Oncology* (29), *Lancet Oncology* (24), *New England Journal of Medicine* (18), *Lancet* (10), *European Journal of Cancer* (4), *Gynecologic Oncology* (4), *Annals of Oncology* (3), *Oncologist* (3) and so on. In all 76 included trials, 15 trials^{13–17 26 27 30 31 41 48 64 67 68 90 97 101 105 109 110} reported both central-assessed and local-assessed treatment effects of ORR and DCR; among them, 14 trials^{13–17 26 27 30 31 41 64 67 68 90 97 101 105 109 110} had those of ORR, DCR and PFS, including one trial⁶⁸ with those of ORR, DCR, PFS and TTP. Another 12 trials^{18 28 29 33 37 51 57 65 79 84 85 91 92 103} with both central and local assessments only contained treatment effects of ORR and PFS.

Statistical analysis

Statistically inconsistent inferences of central and local assessments

From a total of 76 included trials, 17 trials (22%) had statistically inconsistent inferences (significant difference/non-significant difference) of ORR, PFS and/or TTP between central and local assessments.^{17 29 33 48 57 66 68 69 79 87 97 105 110} Among these 17 trials, 2 trials^{29 33} had inconsistent inferences in both of the primary endpoint and secondary endpoint simultaneously. In total, there were 9 of 17 trials (53%) with significant difference based on central assessment; 5 (56%) of these 9 trials were on open-label design (table 2).

Systematic bias between central and local assessments

All comparison results of pooling analysis are presented at table 3. There was no significant difference in the treatment effects of ORR between central and local assessments (OR: 1.02 (95% CI 0.97 to 1.07), p=0.42; heterogeneity: p=0.91, $I^2=0\%$; online supplementary efigure 3). Similarly, no sign of significant difference was in DCR (OR: 0.97 (95% CI 0.92 to 1.03), p=0.32; heterogeneity: p=0.93, $I^2=0\%$; online supplementary efigure 4), PFS (HR: 1.01 (95% CI 0.99 to 1.02), p=0.32; heterogeneity: p=1.00,

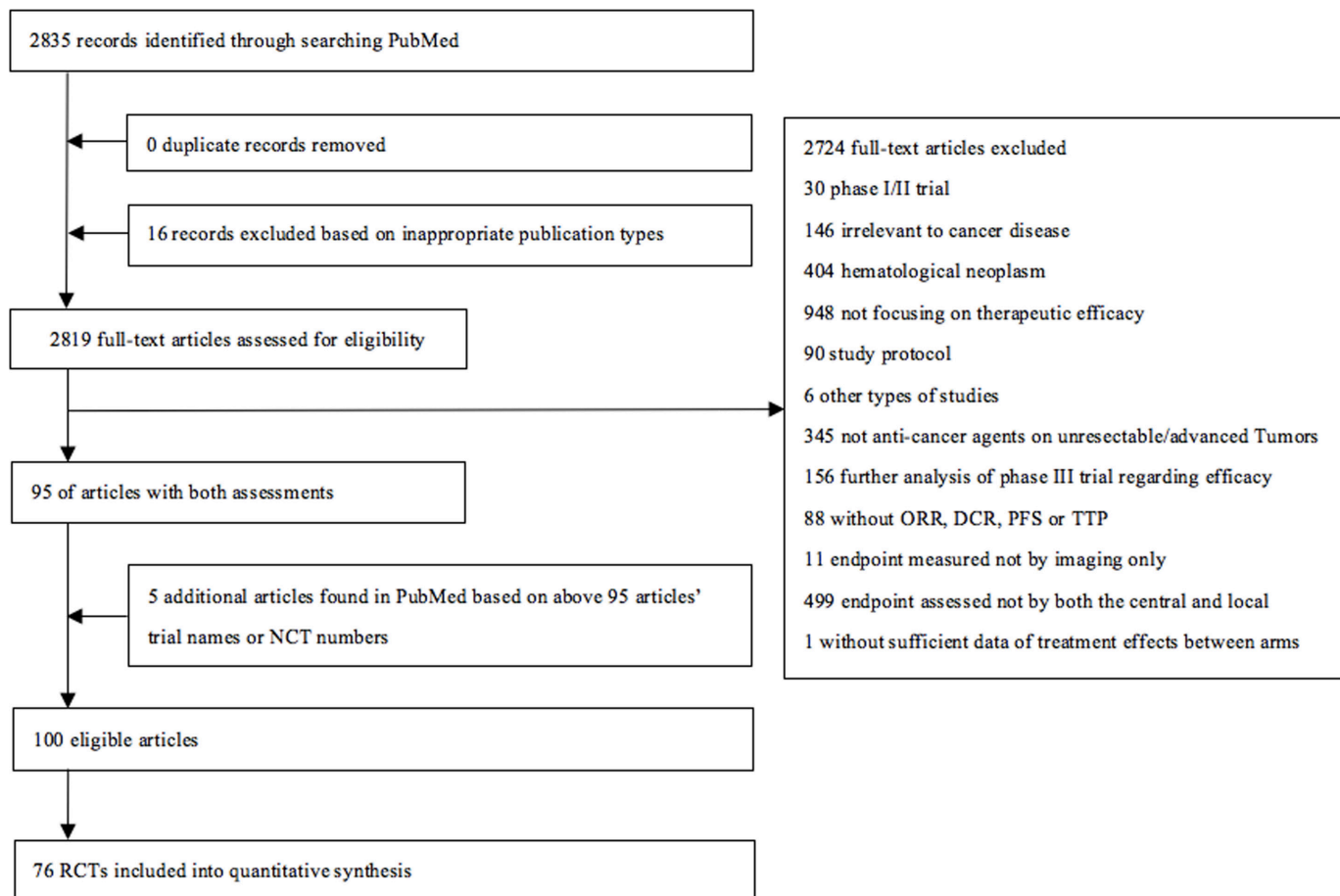


Figure 1 Flow chart of study identification and selection. DCR, disease control rate; NCT, National Clinical Trial; ORR, objective response rate; PFS, progression-free survival; RCT, randomised controlled trial; TTP, time to progression.

$I^2=0\%$; online supplementary efigure 5) and TTP (HR: 1.04 (95% CI 0.95 to 1.14), $p=0.37$; heterogeneity: $p=0.59$, $I^2=0\%$; online supplementary efigure 6). Subgroup analysis also presented no significant difference between central and local assessments, and no significant interaction effect between different elements of subgroup factors, including open label or blind design (table 3).

The strength of the correlation between central and local assessments regarding treatment effect of ORR, DCR, PFS and TTP was 0.91 ($p<0.01$), 0.93 ($p<0.01$), 0.94 ($p<0.01$) and 0.90 ($p=0.29$), respectively (figure 2).

DISCUSSION

To our knowledge, this is the largest literature review with data analyses investigating blinded independent central review and local assessment in phase III RCTs on solid tumours. Also, it is the first research article to report the statistically inconsistent inferences (significant difference or not) of primary and secondary endpoints assessed by central reviewers and local investigators. We found 22% of trials (17/76) with inconsistent inferences between central and local assessments. However, our subsequent pooling analysis and correlation analysis based on all 76 trials confirmed no sign of systematic bias between central and local assessments, regardless of funding source, mask,

region, tumour type, study design, number of enrolled patients, response assessment criteria, primary endpoint and outcome, as well as trials with statistically consistent/inconsistent inferences.

Blinded independent central review is used to detect potential bias introduced by the assessment of local investigators. This consideration is based on a common assumption that local investigators might expect superior efficacy of experimental arm treatments compared with control arm treatments, especially in trials with open-label design. Interestingly, among the 17 trials with statistically inconsistent inferences between central and local assessments, more than half of those 17 studies (9/17; 53%) had a statistically significant difference in central assessment; in these 9 trials, 5 (56%) trials were based on open-label design. This means that central assessment seems to have more positive outcomes in favour of experimental treatments in an open-label design, which contradicts the above common assumption.

With respect to statistically inconsistent inferences between central and local assessments, we assume evaluation variability is one factor accounting for these. As we understand, variability could be impacted by many subjective factors, causing measurement errors or uncertainty.⁸ This situation occurs when one scan reviewer assesses the

Table 1 Summary characteristics of included trials

Characteristics	Trial(s) (n=76)	Patients (n=45688)
Fund source		
Pharmaceutical	73	43 557
Academic	3	2131
Mask		
Open label	37	21 455
Single blind	1	185
Double blind	38	24 048
Region		
Global	62	39 766
Intracontinental	14	5922
Design		
Superiority	71	42 213
Other*	5	3475
Number of enrolled patients		
Maximum	–	1314
Median	–	542
Minimum	–	81
Tumour type		
Breast cancer	17	11 132
NSCLC	14	9327
Renal cell carcinoma	11	6720
Ovarian cancer	6	4536
Melanoma	5	1675
Other†	23	12 298
Response assessment criteria		
RECIST	71	42 756
WHO	4	2387
Not given	1	545
Primary endpoint		
Central assessed‡	43	26 344
Other§	10	6177
Local assessed¶	23	13 167
Primary endpoint outcome		
Positive	51	29 982
Indeterminate**††	2	1106
Negative	23	14 600

*Four non-inferiority, one hybrid design combining superiority and non-inferiority.

†Four gastrointestinal stromal tumour, three pancreatic tumour, three sarcoma, three medullary thyroid cancer, two glioblastoma, two prostate cancer, two neuroendocrine tumour, one colorectal adenocarcinoma, one gastric cancer, one head and neck cancer and one hepatocellular carcinoma.

‡Forty central-assessed PFS, two central-assessed time to progression and one central-assessed ORR.

§Nine overall survival and one unknown-assessed ORR.

¶Twenty-three local-assessed PFS.

**One study used ORR as the primary endpoint, but we were unable to recognise which assessment (central or local assessment) for the ORR was considered as the primary endpoint (central-assessed ORR or local-assessed ORR?). Because a significant difference was found in central review ($p=0.03$) but not found in local assessment ($p=0.05$), we considered the outcome of the primary endpoint as indeterminate.⁴⁸

††Another study considered local-assessed PFS and OS as coprimary endpoints: a significant difference was found in PFS ($p<0.01$), but was not found in OS ($p=0.10$). We considered the outcome of the primary endpoint as indeterminate as well.⁸³

NSCLC, non-small-cell lung cancer; ORR, objective response rate; OS, overall survival; PFS, progression-free survival; RECIST, Response Evaluation Criteria in Solid Tumors.

Table 2 Trials with statistically inconsistent inferences between central and local assessments

Trial	Endpoint	Mask	Tumour type	Therapy (experimental arm vs control arm)	HR/OR and p value
Primary endpoint					
NCT00019682 ⁴⁸	ORR	Single	Melanoma	Exp: gp100:209-217(210M)+Montanide ISA-51 +interleukin 2 Con: Interleukin 2	Central: 2.86 (95% CI 1.05 to 7.82); p=0.03 Local: 2.33 (95% CI 0.98 to 5.56); p=0.05
NCT00471328 ⁶⁶	Central PFS	Open	Gastrointestinal stromal tumour	Exp: Nilotinib Con: Best supportive care/BSC+imatinib/BSC+sunitinib	Central: 0.90 (95% CI 0.65 to 1.26); p=0.56 Local: 0.58 (95% CI 0.42 to 0.80); p<0.01
NCT00112294 ³³	Central PFS	Open	Non-small-cell lung cancer	Exp: Cetuximab+taxane (paclitaxel/docetaxel)+carboplatin Con: Taxane (paclitaxel/docetaxel)+carboplatin	Central: 0.90 (95% CI 0.76 to 1.07); p=0.24 Local: 0.79 (95% CI 0.67 to 0.93); p<0.01
NCT00703326 ⁶⁹	Local PFS	Double	Breast cancer	Exp: Ramucirumab+docetaxel Con: Placebo+docetaxel	Central: 0.79 (95% CI 0.67 to 0.94); p<0.01 Local: 0.88 (95% CI 0.75 to 1.01); p=0.08
NCT00391092 ²⁹	Local PFS	Open	Breast cancer	Exp: Bevacizumab+docetaxel+trastuzumab Con: Docetaxel+trastuzumab	Central: 0.72 (95% CI 0.54 to 0.94); p=0.02 Local: 0.82 (95% CI 0.65 to 1.02); p=0.08
NCT00494299 ⁸⁷	Central TTP	Double	Hepatocellular carcinoma	Exp: Sorafenib Con: Placebo	Central: 0.87 (95% CI 0.70 to 1.09); p=0.25 Local: 0.79 (95% CI 0.62 to 1.00); p=0.049
NCT01007942 ¹²	Local PFS	Double	Breast cancer	Exp: Everolimus+trastuzumab+vinorelbine Con: Placebo+trastuzumab+vinorelbine	Central: 0.88 (95% CI 0.71 to 1.07); p=NG Local: 0.78 (95% CI 0.65 to 0.95); p<0.01
NCT01584648 ^{84 35}	Local PFS	Double	Melanoma	Exp: Trametininib+ dabrafenib Con: Placebo+ dabrafenib	Central: 0.78 (95% CI 0.59 to 1.04); p=NG Local: 0.75 (95% CI 0.57 to 0.99); p=0.03
NCT00412061 ⁴⁴	Central PFS	Double	Neuroendocrine tumour	Exp: Everolimus+octreotide Con: Placebo+octreotide	Central: 0.77 (95% CI 0.59 to 1.00); p=0.03* Local: 0.78 (95% CI 0.62 to 0.98); p=0.02*

Continued

Table 2 Continued

Trial	Endpoint	Mask	Tumour type	Therapy (experimental arm vs control arm)	HR/OR and p value
NCT00056459 ⁷⁶	Central PFS	Double	Colorectal adenocarcinoma	Exp: Vatalanib+oxaliplatin+fluorouracil+leucovorin Con: Placebo+oxaliplatin+fluorouracil+leucovorin	Central: 0.88 (95% CI 0.74 to 1.03); p=0.12 Local: 0.83 (95% CI 0.70 to 0.98); p=0.03
Secondary endpoint					
NCT00391092 ²⁹	ORR	Open	Breast cancer	Exp: Bevacizumab+docetaxel+trastuzumab Con: Docetaxel+trastuzumab	Central: 1.66 (95% CI 1.08 to 2.54); p=0.02 Local: 1.25 (95% CI 0.82 to 1.92); p=0.30
NCT00112294 ³³	ORR	Open	Non-small-cell lung cancer	Exp: Cetuximab+taxane(paclitaxel/docetaxel)+carboplatin Con: Taxane(paclitaxel/docetaxel)+carboplatin	Central: 1.67 (95% CI 1.15 to 2.43); p=0.01 Local: 1.31 (95% CI 0.92 to 1.86); p=0.13
NCT00720941 ⁵⁷	ORR	Open	Renal cell carcinoma	Exp: Pazopanib Con: Sunitinib	Central: 1.35 (95% CI 1.03 to 1.75); p=0.03 Local: 1.23 (95% CI 0.95 to 1.59); p=0.11
NCT01030783 ⁷⁹	ORR	Open	Renal cell carcinoma	Exp: Tivozanib Con: Sorafenib	Central: 1.62 (95% CI 1.10 to 2.39); p=0.01 Local: 1.23 (95% CI 0.85 to 1.78); p=0.26
NCT01523587 ⁹⁷	ORR	Open	Non-small-cell lung cancer	Exp: Afatinib Con: Erlotinib	Central: 2.05 (95% CI 0.98 to 4.29); p=0.06 Local: 2.88 (95% CI 1.60 to 5.21); p<0.01
NCT01345682 ¹⁰⁵	ORR	Open	Head and neck cancer	Exp: Afatinib Con: Methotrexate	Central: 1.90 (95% CI 0.88 to 4.14); p=0.10 Local: 3.00 (95% CI 1.3–6.9); p=0.01
NCT00785785 ¹¹⁰	ORR	Open	Gastrointestinal stromal tumour	Exp: Nilotinib Con: Imatinib	Central: 0.71 (95% CI 0.52 to 0.96); p=0.03 Local: 0.78 (95% CI 0.57 to 1.06); p=0.12
NCT00388726 ¹⁷	PFS	Open	Breast cancer	Exp: Eribulin Con: Treatment of physician's choice†	Central: 0.87 (95% CI 0.71 to 1.05); p=0.14 Local: 0.76 (95% CI 0.64 to 0.90); p<0.01

Continued

Table 2 Continued

Trial	Endpoint	Mask	Tumour type	Therapy (experimental arm vs control arm)	HR/OR and p value
NCT00449033 ⁶⁸	PFS	Double	Non-small-cell lung cancer	Exp: Sorafenib+gemcitabine+cisplatin Con: Placebo+gemcitabine+cisplatin	Central: 0.96 (95% CI 0.77 to 1.21); p=0.37* Local: 0.83 (95% CI 0.71 to 0.97); p<0.01*
	TTP				Central: 0.91 (95% CI 0.67 to 1.23); p=0.26* Local: 0.73 (95% CI 0.60 to 0.88); p<0.01*

*One side.

†Any single-agent chemotherapy or hormonal or biological treatment approved for the treatment of cancer.

BSC, best supportive care; central, central assessed; Con, control arm; Double, double blind; Exp, experimental arm; Local, local assessed; NG, not given; Open, open label; ORR, objective response rate; PFS, progression-free survival; Single, single blind; TTP, time to progression.

response status of different individual patients, as well as when several reviewers conduct the scan assessment for one trial, regardless of whether this is a central or local assessment. In this situation, the evaluation variability attenuates the treatment effect and reduces the statistical power of the clinical trials.⁶⁸ This understanding has been verified based on 21 phase III cancer trials, demonstrating large variability but no sign of systematic bias between two assessments.¹¹²

Missing data could be another factor. It occurs when some patients do not have complete follow-up to determine progression or death, or when patients stop receiving randomised treatments or use alternative treatments before they have progression.¹¹³ In oncological clinical trials, missing data are regarded as censoring. Similar to evaluation variability, the effect of censoring would not contribute to systematic bias but could attenuate the treatment effect.¹¹³

In the trials included in our study, we consider that evaluation variability, censoring and other unmentioned factors simultaneously played a role in attenuating the treatment effects, resulting in statistically inconsistent inferences between two assessments in 17 of the 76 trials. Whereas, regardless of what causes statistically inconsistent inferences, the robustness of the trial efficacy outcome needs to be carefully considered when two assessments present statistically inconsistent inferences, especially in primary endpoint. Even though this inconsistency is unnecessary to reflect a systematic bias, it would be interesting to know how policy-makers consider the approval process for corresponding anticancer agents to the specific patients with cancer.

Considering statistically inconsistent inferences, we believe that blinded independent central review is still a useful method for controlling the risk of bias from local assessment. However, we also question the necessity of central assessment as a routine assessment method for all patients (complete-case fashion) in clinical trials. According to our research, there was no sign of systematic bias: (1) the 95% CIs of all pooled ratios in ORR, DCR, PFS and TTP crossed 1, indicating non-significant difference of the treatment effects between central and local assessments; (2) the 95% CIs were tight as well (especially in PFS), representing quite a precise estimate of the bias that should be negligible. These findings could be further confirmed by our subgroup analysis, even though a small number of the intervals are too wide to be informative due to a limited number of the trials (eg, only one trial used single blind, the OR of ORR was 1.09 (95% CI 0.61 to 1.95)).

When questioning the necessity of the complete central assessment, its drawbacks should be considered as well. First, its implementation in the complete-case fashion is very costly. Second, technically it is hard to conduct a real-time central assessment along with local assessment, to determine disease progression independently. In other words, the decision of central reviewers could be impacted by local investigators when the local

Table 3 Summary results of comparing treatment effects between central and local assessments

Summary/subgroup	Objective response rate (ORR)				Disease control rate				Progression-free survival						
	Study Patient		Patient		Study Patient		Patient		Study Patient		Patient				
	(n)	(n)	OR (95%CI)	P values* I ² †	(n)	(n)	OR (95%CI)	P values* I ² †	(n)	(n)	HR (95%CI)	P values* I ² †			
Summary	29	17949	1.02 (0.97 to 1.07)	0.42	0%	15	9590	0.97 (0.92 to 1.03)	0.32	0%	72	43695	1.01 (0.99 to 1.02)	0.32	0%
Funding source															
Pharmaceutical	28	17764	1.02 (0.97 to 1.07)	0.43	0%	14	9405	0.97 (0.92 to 1.03)	0.33	0%	69	41749	1.01 (0.99 to 1.02)	0.38	0%
Academic	1	185	1.09 (0.61 to 1.95)	0.76		1	185	0.98 (0.67 to 1.42)	0.91		3	1946	1.03 (0.95 to 1.11)	0.50	
Mask															
Open label	23	14616	1.03 (0.98 to 1.09)	0.23	0%	12	7777	0.96 (0.91 to 1.02)	0.23	0%	36	20403	1.01 (0.99 to 1.04)	0.13	0%
Single blind	1	185	1.09 (0.61 to 1.95)	0.76		1	185	0.98 (0.67 to 1.42)	0.91		-	-	-	-	
Double blind	5	3148	0.95 (0.83 to 1.08)	0.41		2	1628	1.02 (0.89 to 1.17)	0.77		36	23292	1.00 (0.98 to 1.02)	0.93	
Region															
Global	23	15384	1.01 (0.95 to 1.06)	0.81	24%	12	8636	0.96 (0.91 to 1.02)	0.20	37%	61	38714	1.01 (0.99 to 1.02)	0.37	0%
Intracontinental	6	2565	1.08 (0.97 to 1.21)	0.17		3	954	1.11 (0.90 to 1.36)	0.34		11	4981	1.01 (0.97 to 1.05)	0.68	
Trial design															
Superiority	27	15787	1.02 (0.97 to 1.08)	0.48	0%	15	9590	0.97 (0.92 to 1.03)	0.32	-	69	41570	1.01 (0.99 to 1.02)	0.34	-
Other§	2	2162	1.03 (0.91 to 1.15)	0.67		-	-	-	-		3	2125	1.01 (0.95 to 1.07)	0.79	
Number of enrolled patients															
>542	17	13260	1.00 (0.94 to 1.06)	0.94	11%	10	7711	0.97 (0.91 to 1.03)	0.26	0%	37	30493	1.01 (0.99 to 1.03)	0.40	0%
<542	12	4689	1.06 (0.97 to 1.16)	0.18		5	1879	1.01 (0.88 to 1.15)	0.93		35	13202	1.01 (0.98 to 1.04)	0.60	
Tumour type															
Breast cancer	6	4028	1.08 (0.97 to 1.21)	0.18	0%	5	3435	0.97 (0.88 to 1.07)	0.58	0%	15	10410	1.00 (0.97 to 1.03)	0.88	0%
NSCLC	8	5172	1.01 (0.93 to 1.11)	0.80		3	2063	1.01 (0.89 to 1.13)	0.91		13	8275	1.02 (0.99 to 1.06)	0.19	
Renal cell carcinoma	6	3917	1.01 (0.91 to 1.12)	0.81		3	1951	0.95 (0.84 to 1.08)	0.43		11	6720	1.00 (0.96 to 1.04)	0.87	
Ovarian cancer	3	1985	0.98 (0.85 to 1.12)	0.72		1	829	0.92 (0.78 to 1.08)	0.31		6	4536	1.03 (0.99 to 1.09)	0.17	
Melanoma	2	435	1.35 (0.91 to 2.00)	0.14		1	185	0.98 (0.67 to 1.42)	0.91		4	1490	1.02 (0.92 to 1.12)	0.74	
Others¶	4	2412	0.99 (0.87 to 1.12)	0.83		2	1127	1.00 (0.84 to 1.19)	1.00		23	12264	1.00 (0.97 to 1.03)	0.94	
Response assessment criteria															
RECIST	27	17088	1.02 (0.97 to 1.07)	0.55	0%	14	9405	0.97 (0.92 to 1.03)	0.33	0%	68	40948	1.01 (0.99 to 1.02)	0.41	0%
WHO	2	861	1.11 (0.90 to 1.37)	0.33		1	185	0.98 (0.67 to 1.42)	0.91		3	2202	1.02 (0.97 to 1.08)	0.44	
Not given	-	-	-	-		-	-	-	-		1	545	1.00 (0.88 to 1.33)	0.94	
Primary endpoint															
Central assessed	17	11151	1.04 (0.98 to 1.10)	0.23	47%	10	6186	0.97 (0.90 to 1.04)	0.37	0%	40	24536	1.01 (0.99 to 1.03)	0.44	0%
Others**	7	4465	0.93 (0.83 to 1.04)	0.19		4	2680	0.98 (0.88 to 1.08)	0.65		9	5992	1.02 (0.98 to 1.07)	0.26	
Local assessed	5	2333	1.08 (0.94 to 1.24)	0.27		1	724	0.99 (0.79 to 1.24)	0.95		23	13167	1.00 (0.97 to 1.03)	0.96	

Continued

Table 3 Continued

Summary/subgroup	Objective response rate (ORR)			Disease control rate			Progression-free survival			
	Study (n)	Patient (n)	OR (95% CI)	Study (n)	Patient (n)	OR (95% CI)	Study (n)	Patient (n)	HR (95% CI)	
Primary endpoint outcome										
Positive	19	11 811	1.04 (0.98 to 1.11)	9	5484	0.99 (0.91 to 1.06)	50	28 930	1.00 (0.99 to 1.02)	0.63
Indeterminate	1	185	1.09 (0.61 to 1.95)	1	185	0.98 (0.67 to 1.42)	1	921	0.98 (0.89 to 1.07)	0.65
Negative	9	5953	0.99 (0.92 to 1.07)	5	3921	0.95 (0.88 to 1.04)	21	13 844	1.02 (0.99 to 1.04)	0.24
Statistical inferences between central and local assessments										
Consistent	18	10 726	1.02 (0.96 to 1.09)	8	5094	0.95 (0.88 to 1.03)	56	32 676	1.00 (0.99 to 1.02)	0.69
Inconsistent	11	7223	1.02 (0.95 to 1.10)	7	4496	0.99 (0.92 to 1.08)	16	11 019	1.02 (0.99 to 1.05)	0.20

*P value for the comparison between central and local assessments.

[†] I^2 in summary outcome was for heterogeneity of data synthesis.

[‡] I^2 in subgroup was for subgroup difference, representing the interaction effects between the elements of each subgroup factor.

[§]For non-inferiority, one hybrid design combining superiority and non-inferiority.

[¶]Four gastrointestinal stromal tumour, three pancreatic tumour, three sarcoma, three medullary thyroid cancer, two glioblastoma, two prostate cancer, two neuroendocrine tumour, one colorectal adenocarcinoma, one gastric cancer, one head and neck cancer and one hepatocellular carcinoma.

**Nine overall survival and one unknown-assessed ORR.

RECIST, Response Evaluation Criteria in Solid Tumors.

investigators declare progression, and ‘progressed’ patients may start to receive subsequent-line treatments. Therefore, the progression time of these specific patients is unknown for central reviewers, which is called informative censoring.^{5 6 9 11 112} Third, based only on imaging information, central reviewers could not conclude progression when patients have symptomatic deterioration. Both information censoring and withdrawal of patients with symptomatic progression (because of no radiological progression in central assessment) may potentially cause bias when the final treatment effects of the experimental arm to the control arm in RCTs are calculated.^{5 8 114}

Fourth, similar to local assessment, central assessment also shares some drawbacks, such as evaluation variability, target-lesion selection and different interpretations on non-target or immeasurable lesions.⁴⁷

In fact, the continuous implementation of the present response assessment criteria, the RECIST and the WHO criteria, has become controversial in the new era of medicine with biomarker-driven therapies, no matter whether for central or local assessment. For instance, when patients are treated with immunotherapies, some tumour lesions might manifest a sign of tumour ‘progression’ based on the RECIST/WHO criteria before manifesting a sign of tumour shrinkage, which is called pseudoprogression.¹¹⁵ Pseudoprogression was initially reported by Wolchok et al. They found that by using the immune-related response criteria (irRC), at least 10% of ipilimumab-treated patients whose response status was characterised as progression disease (PD) based on the WHO criteria could have favourable survival.¹¹⁶ The increased lesion in one case of the study was shown by histopathology as T-cell infiltration instead of tumour proliferation when PD was considered according to the WHO criteria.¹¹⁶ Similar findings have been proved by another two studies that compared the assessment of irRC with RECIST V.1.1, and immune-modified RECIST with RECIST V.1.1, respectively.^{117 118} Even though in our subgroup analysis the comparison result of central versus local assessments did not present significant difference regardless of the RECIST and WHO criteria, these criteria deserve an improvement for biomarker-driven therapies.

Our research has several limitations. First, due to using data from RCTs with both assessments, our outcome may not perfectly match all phase III trials, especially when the trials are implemented by only one type of assessment. Another situation that needs to be considered is trials evaluating two radiological assessment methods, but eventually reporting the outcomes based only on one assessment in published articles. In this situation, a statistically positive outcome may be reported in one assessment; whereas, the ‘not-yet-reported’ outcome of another assessment might be negative. Second, we included trials covering all solid tumours instead of focusing on one specific tumour type, in that we assumed that our research outcome could not be strongly impacted by tumours’ biological characteristics when comparing specific trial processes (eg, central and local assessments)

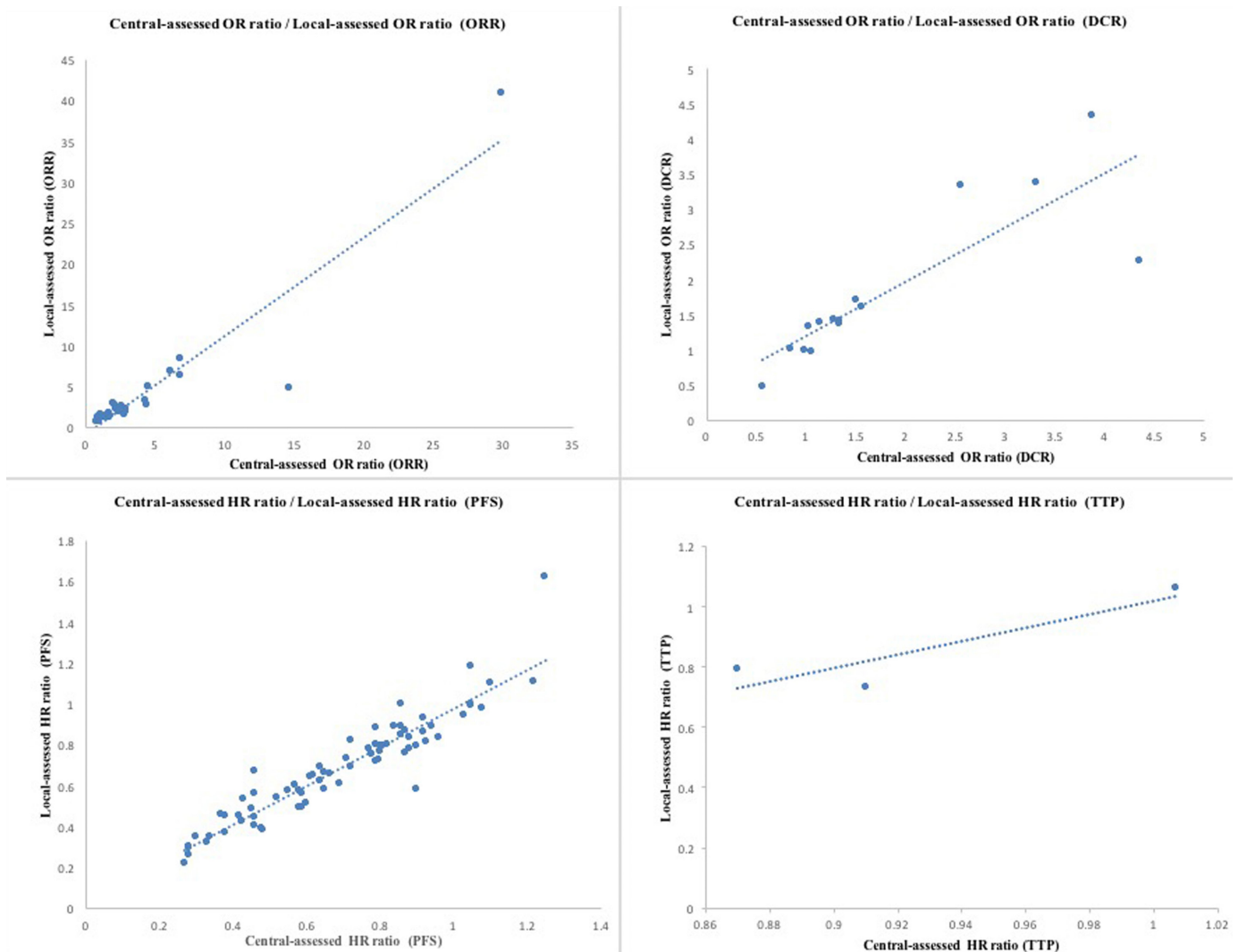


Figure 2 Scatterplot for the correlation of treatment effects between central and local assessments. DCR, disease control rate; ORR, objective response rate; PFS, progression-free survival; TTP, time to progression.

based on study-level strategy. Our subgroup analysis based on different tumour types verified our assumption.

Furthermore, individual-level data would have been the best option for our research, but we did not have access yet. However, we consider that using study-level data reported in each published article is still a good option because the aim of our research type is to investigate study-level issues. Moreover, given that the effect of informative censoring might exist on the treatment effects of PFS and TTP, we also included another important endpoint, ORR, in order to acquire a more exact understanding about whether the treatment effects of both assessments are consistent or not. In this circumstance, the effect of informative censoring could be eliminated because when assessing ORR, central reviewers and local investigators worked independently before local investigators declared progression. Lastly, even though we have done our best to minimise inconsistency during the process of data extraction, it is possible that potential errors may have accrued. Nevertheless, all reviewers have tried to ensure consistency for data interpretation.

In conclusion, we estimate that there was essentially no systematic bias between local and central assessments, as evidenced by our precisely estimated pooled ratios of OR in ORR and DCR, as well as estimated pooled ratios of HR in PFS and TTP. Despite this, we found that statistically inconsistent inferences could be made in many trials depending on whether central or local assessment was used. Considering these, we think blinded independent central review is still an irreplaceable method for controlling the risk of bias from local assessment, but its routine usage for all patients may be unnecessary in oncological randomised controlled trials.

Author affiliations

¹Department of Thoracic Surgery and Oncology, First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

²Guangzhou Institute of Respiratory Disease & China State Key Laboratory of Respiratory Disease, Guangzhou, China

³National Clinical Research Centre of Respiratory Disease, Guangzhou, China

⁴George Warren Brown School, Washington University in St. Louis, St. Louis, Missouri, USA

⁵Nanshan School, Guangzhou Medical University, Guangzhou, China

⁶Department of Pancreatic Surgery/Oncology, Fudan University Shanghai Cancer Center, Shanghai, China

⁷Pancreatic Cancer Institute, Fudan University, Shanghai, China

⁸Department of Surgery and Cancer, Imperial College London, London, UK

⁹Institute of International Education, Guangdong University of Foreign Studies, Guangzhou, China

¹⁰Department of Pathology, First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

¹¹Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden

Acknowledgements This study is the first academic output of the Design, Implementation & Report of Oncological Randomized Controlled Trials (DIRORCT) Research Project, which aims to investigate the quality of RCTs from study design and implementation to final report through systematic review and/or further analysis. We thank Ms Xiaoru Deng (Guangzhou Medical University), Dr Peter Coogan, Professor Joseph T. Steensma, Mr Natalicio Serrano (Brown School at Washington University in St. Louis), Ms Carolyn Smith (The Writing Center at Washington University in St. Louis) for advising article searching and manuscript revision. We also sincerely thank all authors, investigators, sponsors and patients for their effort and participation in our included studies.

Contributors JRZ, WHL and JXH conceived and designed the study. JRZ conducted article researching and selection, and quality assessment. JRZ, YYZ and SYT took the lead for data extraction, which was checked by YQC, HRL, DFC, YH, XYW, KXD, SHJ, JQZ, JXX and XZC. JRZ, YYZ, SYT, LJ, QHH, LTH, JXH, ZHX and JYW analysed and interpreted the data. JRZ drafted the manuscript, which was critically revised for important intellectual content by all authors. WHL and JXH supervised the study. All authors, including JRZ, YYZ, SYT, LJ, QHH, LTH, JXH, ZHX, JYW, YQC, HRL, DFC, YH, XYW, KXD, SHJ, JQZ, JXX, XZC, WHL and JXH, have read and approved the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Requests for additional details regarding the study protocol may be made by contacting the author JRZ (jianrong.zhang@wustl.edu).

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- Johnson JR, Williams G, Pazdur R. End points and united states food and drug administration approval of oncology drugs. *J Clin Oncol* 2003;21:1404–11.
- Freidlin B, Korn EL, Hunsberger S, et al. Proposal for the use of progression-free survival in unblinded randomized trials. *J Clin Oncol* 2007;25:2122–6.
- Savers S. Guidance for industry: clinical trial endpoints for the approval of cancer drugs and biologics. *Biotechnology Law Report* 2007;26:375–86.
- Amit O, Bushnell W, Dodd L, et al. Blinded independent central review of the progression-free survival endpoint. *Oncologist* 2010;15:492–5.
- Dodd LE, Korn EL, Freidlin B, et al. Blinded independent central review of progression-free survival in phase III clinical trials: important design element or unnecessary expense? *J Clin Oncol* 2008;26:3791–6.
- Stone AM, Bushnell W, Denne J, et al. Research outcomes and recommendations for the assessment of progression in cancer clinical trials from a PhRMA working group. *Eur J Cancer* 2011;47:1763–71.
- Ford R, Schwartz L, Dancey J, et al. Lessons learned from independent central review. *Eur J Cancer* 2009;45:268–74.
- Dancey JE, Dodd LE, Ford R, et al. Recommendations for the assessment of progression in randomised cancer treatment trials. *Eur J Cancer* 2009;45:281–9.
- US Food and Drug Administration. FDA briefing document oncologic drugs advisory committee: evaluation of radiologic review of progression-free survival in non-hematologic malignancies. 2012 <http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/drugs/oncologicdrugsadvisorycommittee/ucm311141.pdf>
- Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
- Zhang JJ, Chen H, He K, et al. Evaluation of blinded independent central review of tumor progression in oncology clinical trials. *Ther Innov Regul Sci* 2013;47:167–74.
- André F, O'Regan R, Ozguroglu M, et al. Everolimus for women with trastuzumab-resistant, HER2-positive, advanced breast cancer (BOLERO-3): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet Oncol* 2014;15:580–91.
- Baselga J, Campone M, Piccart M, et al. Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. *N Engl J Med* 2012;366:520–9.
- Beaver JA, Park BH. The BOLERO-2 trial: the addition of everolimus to exemestane in the treatment of postmenopausal hormone receptor-positive advanced breast cancer. *Future Oncol* 2012;8:651–7.
- Yardley DA, Noguchi S, Pritchard KI, et al. Everolimus plus exemestane in postmenopausal patients with HR(+) breast cancer: BOLERO-2 final progression-free survival analysis. *Adv Ther* 2013;30:870–84.
- Piccart M, Hortobagyi GN, Campone M, et al. Everolimus plus exemestane for hormone-receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: overall survival results from BOLERO-2†. *Ann Oncol* 2014;25:2357–62.
- Cortes J, O'Shaughnessy J, Loesch D, et al. Eribulin monotherapy versus treatment of physician's choice in patients with metastatic breast cancer (EMBRACE): a phase 3 open-label randomised study. *Lancet* 2011;377:914–23.
- Bergh J, Bondarenko IM, Lichinitser MR, et al. First-line treatment of advanced breast cancer with sunitinib in combination with docetaxel versus docetaxel alone: results of a prospective, randomized phase III study. *J Clin Oncol* 2012;30:921–9.
- Blackwell KL, Burstein HJ, Storniolo AM, et al. Overall survival benefit with lapatinib in combination with trastuzumab for patients with human epidermal growth factor receptor 2-positive metastatic breast cancer: final results from the EGF104900 Study. *J Clin Oncol* 2012;30:2585–92.
- Blackwell KL, Burstein HJ, Storniolo AM, et al. Randomized study of Lapatinib alone or in combination with trastuzumab in women with ErbB2-positive, trastuzumab-refractory metastatic breast cancer. *J Clin Oncol* 2010;28:1124–30.
- Blay JY, Leahy MG, Nguyen BB, et al. Randomised phase III trial of trabectedin versus doxorubicin-based chemotherapy as first-line therapy in translocation-related sarcomas. *Eur J Cancer* 2014;50:1137–47.
- Blumenthal GM, Cortazar P, Zhang JJ, et al. FDA approval summary: sunitinib for the treatment of progressive well-differentiated locally advanced or metastatic pancreatic neuroendocrine tumors. *Oncologist* 2012;17:1108–13.
- Raymond E, Dahan L, Raoul JL, et al. Sunitinib malate for the treatment of pancreatic neuroendocrine tumors. *N Engl J Med* 2011;364:501–13.
- Burger RA, Brady MF, Bookman MA, et al. Incorporation of bevacizumab in the primary treatment of ovarian cancer. *N Engl J Med* 2011;365:2473–83.
- Burger RA, Brady MF, Rhee J, et al. Independent radiologic review of the Gynecologic Oncology Group Study 0218, a phase III trial of bevacizumab in the primary treatment of advanced epithelial ovarian, primary peritoneal, or fallopian tube cancer. *Gynecol Oncol* 2013;131:21–6.
- Wu YL, Zhou C, Hu CP, et al. Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. *Lancet Oncol* 2014;15:213–22.
- Yang JC, Wu YL, Schuler M, et al. Afatinib versus cisplatin-based chemotherapy for EGFR mutation-positive lung adenocarcinoma (LUX-Lung 3 and LUX-Lung 6): analysis of overall survival data from two randomised, phase 3 trials. *Lancet Oncol* 2015;16:141–51.
- Miller VA, Hirsh V, Cadranet J, et al. Afatinib versus placebo for patients with advanced, metastatic non-small-cell lung cancer after failure of erlotinib, gefitinib, or both, and one or two lines of chemotherapy (LUX-Lung 1): a phase 2b/3 randomised trial. *Lancet Oncol* 2012;13:528–38.

29. Gianni L, Romieu GH, Lichinitser M, *et al.* AVEREL: a randomized phase III Trial evaluating bevacizumab in combination with docetaxel and trastuzumab as first-line therapy for HER2-positive locally recurrent/metastatic breast cancer. *J Clin Oncol* 2013;31:1719–25.
30. Motzer RJ, Escudier B, Tomczak P, *et al.* Axitinib versus sorafenib as second-line treatment for advanced renal cell carcinoma: overall survival analysis and updated results from a randomised phase 3 trial. *Lancet Oncol* 2013;14:552–62.
31. Rini BI, Escudier B, Tomczak P, *et al.* Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomised phase 3 trial. *Lancet* 2011;378:1931–9.
32. Elisei R, Schlumberger MJ, Müller SP, *et al.* Cabozantinib in progressive medullary thyroid cancer. *J Clin Oncol* 2013;31:3639–46.
33. Lynch TJ, Patel T, Dreisbach L, *et al.* Cetuximab and first-line taxane/carboplatin chemotherapy in advanced non-small-cell lung cancer: results of the randomized multicenter phase III trial BMS099. *J Clin Oncol* 2010;28:911–7.
34. Long GV, Stroyakovskiy D, Gogas H, *et al.* Combined BRAF and MEK inhibition versus BRAF inhibition alone in melanoma. *N Engl J Med* 2014;371:1877–88.
35. Long GV, Stroyakovskiy D, Gogas H, *et al.* Dabrafenib and trametinib versus dabrafenib and placebo for Val600 BRAF-mutant melanoma: a multicentre, double-blind, phase 3 randomised controlled trial. *Lancet* 2015;386:444–51.
36. Larkin J, Ascierto PA, Dréno B, *et al.* Combined vemurafenib and cobimetinib in BRAF-mutated melanoma. *N Engl J Med* 2014;371:1867–76.
37. Hauschild A, Grob JJ, Demidov LV, *et al.* Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial. *Lancet* 2012;380:358–65.
38. Ramalingam SS, Jänne PA, Mok T, *et al.* Dacomitinib versus erlotinib in patients with advanced-stage, previously treated non-small-cell lung cancer (ARCHER 1009): a randomised, double-blind, phase 3 trial. *Lancet Oncol* 2014;15:1369–78.
39. Kim ES, Neubauer M, Cohn A, *et al.* Docetaxel or pemetrexed with or without cetuximab in recurrent or progressive non-small-cell lung cancer after platinum-based therapy: a phase 3, open-label, randomised trial. *Lancet Oncol* 2013;14:1326–36.
40. Reck M, Kaiser R, Mellengard A, *et al.* Docetaxel plus nintedanib versus docetaxel plus placebo in patients with previously treated non-small-cell lung cancer (LUME-Lung 1): a phase 3, double-blind, randomised controlled trial. *Lancet Oncol* 2014;15:143–55.
41. Motzer RJ, Porta C, Vogelzang NJ, *et al.* Dovitinib versus sorafenib for third-line targeted treatment of patients with metastatic renal cell carcinoma: an open-label, randomised phase 3 trial. *Lancet Oncol* 2014;15:286–96.
42. Demetri GD, Reichardt P, Kang YK, *et al.* Efficacy and safety of regorafenib for advanced gastrointestinal stromal tumours after failure of imatinib and sunitinib (GRID): an international, multicentre, randomised, placebo-controlled, phase 3 trial. *Lancet* 2013;381:295–302.
43. Yao JC, Shah MH, Ito T, *et al.* Everolimus for advanced pancreatic neuroendocrine tumors. *N Engl J Med* 2011;364:514–23.
44. Pavel ME, Hainsworth JD, Baudin E, *et al.* Everolimus plus octreotide long-acting repeatable for the treatment of advanced neuroendocrine tumours associated with carcinoid syndrome (RADIANT-2): a randomised, placebo-controlled, phase 3 study. *Lancet* 2011;378:2005–12.
45. Escudier B, Bellmunt J, Négrier S, *et al.* Phase III trial of bevacizumab plus interferon alfa-2a in patients with metastatic renal cell carcinoma (AVOREN): final analysis of overall survival. *J Clin Oncol* 2010;28:2144–50.
46. Escudier B, Pluzanska A, Koralewski P, *et al.* Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. *Lancet* 2007;370:2103–11.
47. Summers J, Cohen MH, Keegan P, *et al.* FDA drug approval summary: bevacizumab plus interferon for advanced renal cell carcinoma. *Oncologist* 2010;15:104–11.
48. Schwartzentruber DJ, Lawson DH, Richards JM, *et al.* gp100 peptide vaccine and interleukin-2 in patients with advanced melanoma. *N Engl J Med* 2011;364:2119–27.
49. Shi Y, Zhang L, Liu X, *et al.* Icotinib versus gefitinib in previously treated advanced non-small-cell lung cancer (ICOGEN): a randomised, double-blind phase 3 non-inferiority trial. *Lancet Oncol* 2013;14:953–61.
50. Flaherty KT, Robert C, Hersey P, *et al.* Improved survival with MEK inhibition in BRAF-mutated melanoma. *N Engl J Med* 2012;367:107–14.
51. Von Hoff DD, Ervin T, Arena FP, *et al.* Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. *N Engl J Med* 2013;369:1691–703.
52. Wu YL, Lee JS, Thongprasert S, *et al.* Intercalated combination of chemotherapy and erlotinib for patients with advanced stage non-small-cell lung cancer (FASTACT-2): a randomised, double-blind trial. *Lancet Oncol* 2013;14:777–86.
53. Paz-Ares L, de Marinis F, Dediu M, *et al.* Maintenance therapy with pemetrexed plus best supportive care versus placebo plus best supportive care after induction therapy with pemetrexed plus cisplatin for advanced non-squamous non-small-cell lung cancer (PARAMOUNT): a double-blind, phase 3, randomised controlled trial. *Lancet Oncol* 2012;13:247–55.
54. Paz-Ares LG, de Marinis F, Dediu M, *et al.* PARAMOUNT: Final overall survival results of the phase III study of maintenance pemetrexed versus placebo immediately after induction treatment with pemetrexed plus cisplatin for advanced nonsquamous non-small-cell lung cancer. *J Clin Oncol* 2013;31:2895–902.
55. Sternberg CN, Hawkins RE, Wagstaff J, *et al.* A randomised, double-blind phase III study of pazopanib in patients with advanced and/or metastatic renal cell carcinoma: final overall survival results and safety update. *Eur J Cancer* 2013;49:1287–96.
56. Sternberg CN, Davis ID, Mardiak J, *et al.* Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial. *J Clin Oncol* 2010;28:1061–8.
57. Motzer RJ, Hutson TE, Cella D, *et al.* Pazopanib versus sunitinib in metastatic renal-cell carcinoma. *N Engl J Med* 2013;369:722–31.
58. Baselga J, Cortés J, Kim SB, *et al.* Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N Engl J Med* 2012;366:109–19.
59. Swain SM, Kim SB, Cortés J, *et al.* Pertuzumab, trastuzumab, and docetaxel for HER2-positive metastatic breast cancer (CLEOPATRA study): overall survival results from a randomised, double-blind, placebo-controlled, phase 3 study. *Lancet Oncol* 2013;14:461–71.
60. Swain SM, Baselga J, Kim SB, *et al.* Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic breast cancer. *N Engl J Med* 2015;372:724–34.
61. Swain SM, Im YH, Im SA, *et al.* Safety profile of pertuzumab with trastuzumab and docetaxel in patients from Asia with human epidermal growth factor receptor 2-positive metastatic breast cancer: results from the phase III trial CLEOPATRA. *Oncologist* 2014;19:693–701.
62. Motzer RJ, Escudier B, Oudard S, *et al.* Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *Lancet* 2008;372:449–56.
63. Motzer RJ, Escudier B, Oudard S, *et al.* Phase 3 trial of everolimus for metastatic renal cell carcinoma: final results and analysis of prognostic factors. *Cancer* 2010;116:4256–65.
64. Kaufman PA, Awada A, Twelves C, *et al.* Phase III open-label randomized study of eribulin mesylate versus capecitabine in patients with locally advanced or metastatic breast cancer previously treated with an anthracycline and a taxane. *J Clin Oncol* 2015;33:594–601.
65. Sequist LV, Yang JC, Yamamoto N, *et al.* Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J Clin Oncol* 2013;31:3327–34.
66. Reichardt P, Blay JY, Gelderblom H, *et al.* Phase III study of nilotinib versus best supportive care with or without a TKI in patients with gastrointestinal stromal tumors resistant to or intolerant of imatinib and sunitinib. *Ann Oncol* 2012;23:1680–7.
67. Crown JP, Diéras V, Staroslawska E, *et al.* Phase III trial of sunitinib in combination with capecitabine versus capecitabine monotherapy for the treatment of patients with pretreated metastatic breast cancer. *J Clin Oncol* 2013;31:2870–8.
68. Paz-Ares LG, Biesma B, Heigener D, *et al.* Phase III, randomized, double-blind, placebo-controlled trial of gemcitabine/cisplatin alone or with sorafenib for the first-line treatment of advanced, nonsquamous non-small-cell lung cancer. *J Clin Oncol* 2012;30:3084–92.
69. Mackey JR, Ramos-Vazquez M, Lipatov O, *et al.* Primary results of ROSE/TRIO-12, a randomized placebo-controlled phase III trial evaluating the addition of ramucicrumab to first-line docetaxel chemotherapy in metastatic breast cancer. *J Clin Oncol* 2015;33:141–8.
70. Ryan CJ, Smith MR, Fizazi K, *et al.* Abiraterone acetate plus prednisone versus placebo plus prednisone in chemotherapy-naïve men with metastatic castration-resistant prostate cancer (COU-AA-302): final overall survival analysis of a randomised, double-blind, placebo-controlled phase 3 study. *Lancet Oncol* 2015;16:152–60.

71. Ryan CJ, Smith MR, de Bono JS, *et al.* Abiraterone in metastatic prostate cancer without previous chemotherapy. *N Engl J Med* 2013;368:138–48.
72. Morris MJ, Molina A, Small EJ, *et al.* Radiographic progression-free survival as a response biomarker in metastatic castration-resistant prostate cancer: COU-AA-302 results. *J Clin Oncol* 2015;33:1356–63.
73. Rini BI, Bellmunt J, Clancy J, *et al.* Randomized phase III trial of temsirolimus and bevacizumab versus interferon alfa and bevacizumab in metastatic renal cell carcinoma: INTORACT trial. *J Clin Oncol* 2014;32:752–9.
74. Hutson TE, Escudier B, Esteban E, *et al.* Randomized phase III trial of temsirolimus versus sorafenib as second-line therapy after sunitinib in patients with metastatic renal cell carcinoma. *J Clin Oncol* 2014;32:760–7.
75. Iwata H, Masuda N, Ohno S, *et al.* A randomized, double-blind, controlled study of exemestane versus anastrozole for the first-line treatment of postmenopausal Japanese women with hormone-receptor-positive advanced breast cancer. *Breast Cancer Res Treat* 2013;139:441–51.
76. Hecht JR, Trabach T, Hainsworth JD, *et al.* Randomized, placebo-controlled, phase III study of first-line oxaliplatin-based chemotherapy plus PTK787/ZK 222584, an oral vascular endothelial growth factor receptor inhibitor, in patients with metastatic colorectal adenocarcinoma. *J Clin Oncol* 2011;29:1997–2003.
77. Demetri GD, Chawla SP, Ray-Coquard I, *et al.* Results of an international randomized phase III trial of the mammalian target of rapamycin inhibitor ridaforolimus versus placebo to control metastatic sarcomas in patients after benefit from prior chemotherapy. *J Clin Oncol* 2013;31:2485–92.
78. Kang YK, Ryu MH, Yoo C, *et al.* Resumption of imatinib to control metastatic or unresectable gastrointestinal stromal tumours after failure of imatinib and sunitinib (RIGHT): a randomised, placebo-controlled, phase 3 trial. *Lancet Oncol* 2013;14:1175–82.
79. Motzer RJ, Nosov D, Eisen T, *et al.* Tivozanib versus sorafenib as initial targeted therapy for patients with metastatic renal cell carcinoma: results from a phase III trial. *J Clin Oncol* 2013;31:3791–9.
80. Verma S, Miles D, Gianni L, *et al.* Trastuzumab emtansine for HER2-positive advanced breast cancer. *N Engl J Med* 2012;367:1783–91.
81. Wells SA, Robinson BG, Gagel RF, *et al.* Vandetanib in patients with locally advanced or metastatic medullary thyroid cancer: a randomized, double-blind phase III trial. *J Clin Oncol* 2012;30:134–41.
82. Socinski MA, Bondarenko I, Karaseva NA, *et al.* Weekly nab-paclitaxel in combination with carboplatin versus solvent-based paclitaxel plus carboplatin as first-line therapy in patients with advanced non-small-cell lung cancer: final results of a phase III trial. *J Clin Oncol* 2012;30:2055–62.
83. Chinot OL, Wick W, Mason W, *et al.* Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma. *N Engl J Med* 2014;370:709–22.
84. Aghajanian C, Goff B, Nycum LR, *et al.* Independent radiologic review: bevacizumab in combination with gemcitabine and carboplatin in recurrent ovarian cancer. *Gynecol Oncol* 2014;133:105–10.
85. Aghajanian C, Blank SV, Goff BA, *et al.* OCEANS: a randomized, double-blind, placebo-controlled phase III trial of chemotherapy with or without bevacizumab in patients with platinum-sensitive recurrent epithelial ovarian, primary peritoneal, or fallopian tube cancer. *J Clin Oncol* 2012;30:2039–45.
86. Stupp R, Hegi ME, Gorlia T, *et al.* Cilengitide combined with standard treatment for patients with newly diagnosed glioblastoma with methylated MGMT promoter (CENTRIC EORTC 26071-22072 study): a multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol* 2014;15:1100–8.
87. Kudo M, Imanaka K, Chida N, *et al.* Phase III study of sorafenib after transarterial chemoembolisation in Japanese and Korean patients with unresectable hepatocellular carcinoma. *Eur J Cancer* 2011;47:2117–27.
88. du Bois A, Floquet A, Kim JW, *et al.* Incorporation of pazopanib in maintenance therapy of ovarian cancer. *J Clin Oncol* 2014;32:3374–82.
89. Floquet A, Vergote I, Colombo N, *et al.* Progression-free survival by local investigator versus independent central review: comparative analysis of the AGO-OVAR16 Trial. *Gynecol Oncol* 2015;136:37–42.
90. Colombo N, Kutarska E, Dimopoulos M, *et al.* Randomized, open-label, phase III study comparing patupilone (EPO906) with pegylated liposomal doxorubicin in platinum-refractory or -resistant patients with recurrent epithelial ovarian, primary fallopian tube, or primary peritoneal cancer. *J Clin Oncol* 2012;30:3841–7.
91. Monk BJ, Herzog TJ, Kaye SB, *et al.* Trabectedin plus pegylated liposomal doxorubicin (PLD) versus PLD in recurrent ovarian cancer: overall survival analysis. *Eur J Cancer* 2012;48:2361–8.
92. Monk BJ, Herzog TJ, Kaye SB, *et al.* Trabectedin plus pegylated liposomal Doxorubicin in recurrent ovarian cancer. *J Clin Oncol* 2010;28:3107–14.
93. Yao JC, Fazio N, Singh S, *et al.* Everolimus for the treatment of advanced, non-functional neuroendocrine tumours of the lung or gastrointestinal tract (RADIANT-4): a randomised, placebo-controlled, phase 3 study. *Lancet* 2016;387:968–77.
94. Pujade-Lauraine E, Hilpert F, Weber B, *et al.* Bevacizumab combined with chemotherapy for platinum-resistant recurrent ovarian cancer: The AURELIA open-label randomized phase III trial. *J Clin Oncol* 2014;32:1302–8.
95. Husain A, Wang Y, Hanker LC, *et al.* Independent radiologic review of AURELIA, a phase 3 trial of bevacizumab plus chemotherapy for platinum-resistant recurrent ovarian cancer. *Gynecol Oncol* 2016;142:465–70.
96. Brose MS, Nutting CM, Jarzab B, *et al.* Sorafenib in radioactive iodine-refractory, locally advanced or metastatic differentiated thyroid cancer: a randomised, double-blind, phase 3 trial. *Lancet* 2014;384:319–28.
97. Soria JC, Felip E, Cobo M, *et al.* Afatinib versus erlotinib as second-line treatment of patients with advanced squamous cell carcinoma of the lung (LUX-Lung 8): an open-label randomised controlled phase 3 trial. *Lancet Oncol* 2015;16:897–907.
98. Hurvitz SA, Andre F, Jiang Z, *et al.* Combination of everolimus with trastuzumab plus paclitaxel as first-line treatment for patients with HER2-positive advanced breast cancer (BOLERO-1): a phase 3, randomised, double-blind, multicentre trial. *Lancet Oncol* 2015;16:816–29.
99. Cristofanilli M, Turner NC, Bondarenko I, *et al.* Fulvestrant plus palbociclib versus fulvestrant plus placebo for treatment of hormone-receptor-positive, HER2-negative metastatic breast cancer that progressed on previous endocrine therapy (PALOMA-3): final analysis of the multicentre, double-blind, phase 3 randomised controlled trial. *Lancet Oncol* 2016;17:425–39.
100. Turner NC, Ro J, André F, *et al.* Palbociclib in hormone-receptor-positive advanced breast cancer. *N Engl J Med* 2015;373:209–19.
101. Zhang P, Sun T, Zhang Q, *et al.* Utidelone plus capecitabine versus capecitabine alone for heavily pretreated metastatic breast cancer refractory to anthracyclines and taxanes: a multicentre, open-label, superiority, phase 3, randomised controlled trial. *Lancet Oncol* 2017;18:371–83.
102. Sternberg C, Armstrong A, Pili R, *et al.* Randomized, double-blind, placebo-controlled phase III study of tasquinimod in men with metastatic castration-resistant prostate cancer. *J Clin Oncol* 2016;34:2636–43.
103. Rini BI, Stenzl A, Zdrojowy R, *et al.* IMA901, a multi-peptide cancer vaccine, plus sunitinib versus sunitinib alone, as first-line therapy for advanced or metastatic renal cell carcinoma (IMPRINT): a multicentre, open-label, randomised, controlled, phase 3 trial. *Lancet Oncol* 2016;17:1599–611.
104. Mok TS, Wu Y-L, Ahn M-J, *et al.* Osimertinib or platinum-pemetrexid in EGFR T790M-positive lung cancer. *N Engl J Med* 2017;376:629–40.
105. Machiels JP, Haddad RI, Fayette J, *et al.* Afatinib versus methotrexate as second-line treatment in patients with recurrent or metastatic squamous-cell carcinoma of the head and neck progressing on or after platinum-based therapy (LUX-Head & Neck 1): an open-label, randomised phase 3 trial. *Lancet Oncol* 2015;16:583–94.
106. Hortobagyi GN, Stemmer SM, Burris HA, *et al.* Ribociclib as first-line therapy for hr-positive, advanced breast cancer. *N Engl J Med* 2016;375:1738–48.
107. Finn RS, Martin M, Rugo HS, *et al.* Palbociclib and Letrozole in Advanced Breast Cancer. *N Engl J Med* 2016;375:1925–36.
108. Demetri GD, von Mehren M, Jones RL, *et al.* Efficacy and safety of trabectedin or dacarbazine for metastatic liposarcoma or leiomyosarcoma after failure of conventional chemotherapy: results of a phase III randomized multicenter clinical trial. *J Clin Oncol* 2016;34:786–93.
109. Choueiri TK, Escudier B, Powles T, *et al.* Cabozantinib versus everolimus in advanced renal cell carcinoma (METEOR): final results from a randomised, open-label, phase 3 trial. *Lancet Oncol* 2016;17:917–27.
110. Blay JY, Shen L, Kang YK, *et al.* Nilotinib versus imatinib as first-line therapy for patients with unresectable or metastatic gastrointestinal stromal tumours (ENESTg1): a randomised phase 3 trial. *Lancet Oncol* 2015;16:550–60.

111. Ryu MH, Baba E, Lee KH, *et al.* Comparison of two different S-1 plus cisplatin dosing schedules as first-line chemotherapy for metastatic and/or recurrent gastric cancer: a multicenter, randomized phase III trial (SOS). *Ann Oncol* 2015;26:2097–101.
112. Tang PA, Pond GR, Chen EX. Influence of an independent review committee on assessment of response rate and progression-free survival in phase III clinical trials. *Ann Oncol* 2010;21:19–26.
113. Denne JS, Stone AM, Bailey-lacona R, *et al.* Missing data and censoring in the analysis of progression-free survival in oncology clinical trials. *J Biopharm Stat* 2013;23:951–70.
114. Campigotto F, Weller E. Impact of informative censoring on the Kaplan-Meier estimate of progression-free survival in phase II clinical trials. *J Clin Oncol* 2014;32:3068–74.
115. Litière S, Collette S, de Vries EG, *et al.* RECIST - learning from the past to build the future. *Nat Rev Clin Oncol* 2017;14:187–92.
116. Wolchok JD, Hoos A, O'Day S, *et al.* Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin Cancer Res* 2009;15:7412–20.
117. Hodi FS, Hwu WJ, Kefford R, *et al.* Evaluation of immune-related response criteria and RECIST v1.1 in patients with advanced melanoma treated with pembrolizumab. *J Clin Oncol* 2016;34:1510–7.
118. Hodi FS, Ballinger M, Lyons B, *et al.* Immune-modified response evaluation criteria in solid tumors (imRECIST): refining guidelines to assess the clinical benefit of cancer immunotherapy. *J Clin Oncol* 2018;36:850–8.