Original Article

Evaluation bias in objective response rate and disease control rate between blinded independent central review and local assessment: a study-level pooled analysis of phase III randomized control trials in the past seven years

Jianrong Zhang[1,2,3,4*], Yiyin Zhang[5,6,7*], Shiyan Tang[5,8*], Hengrui Liang[5], Difei Chen[5], Long Jiang[1,2,3], Qihua He[1,2,3], 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]

[1]Department of Thoracic Surgery and Oncology, First Affiliated Hospital of Guangzhou Medical University, Guangzhou 510120, China; [2]Guangzhou Institute of Respiratory Disease & China State Key Laboratory of Respiratory Disease, Guangzhou 510120, China; [3]National Clinical Research Centre of Respiratory Disease, Guangzhou 510120, China; [4]George Warren Brown School of Social Work, Washington University in St. Louis, St. Louis, USA; [5]Nanshan School, Guangzhou Medical University, Guangzhou 510182, China; [6]Department of Pancreatic Surgery/Oncology, Fudan University Shanghai Cancer Center, Shanghai 200032, China; [7]Pancreatic Cancer Institute, Fudan University, Shanghai 200032, China; [8]Department of Surgery and Cancer, Imperial College London, London, UK

Contributions: (I) Conception and design: J Zhang, W Liang, J He; (II) Administrative support: W Liang, J He; (III) Provision of study materials or patients: W Liang; (IV) Collection and assembly of data: J Zhang, Y Zhang, SY Tang, HR Liang, D Chen, Y Huang, X Wang, K Deng, S Jiang, J Zhou, J Xu, X Chen; (V) Data analysis and interpretation: J Zhang, L Jiang, Q He; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

*These authors contributed equally to this work.

Correspondence to: Prof. Wenhua Liang, MD. Department of Thoracic Surgery and Oncology, First Affiliated Hospital of Guangzhou Medical University, Guangzhou Institute of Respiratory Disease & China State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou 510120, China. Email: liangwh1987@163.com; Prof. Jianxing He, MD, PhD, FACS, FRCS. Department of Thoracic Surgery and Oncology, First Affiliated Hospital of Guangzhou Medical University, Guangzhou Institute of Respiratory Disease & China State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou 510120, China. Email: drjianxing.he@gmail.com.

Background: In previous studies, complete-case implementation of blind independent central review has been considered unnecessary based on no sign of systematic bias between central and local assessments. In order to further evaluate its value, this study investigated evaluation status between both assessments in phase III trials of anti-cancer drugs for non-hematologic solid tumors.

Methods: Eligible trials were searched in PubMed with the date of Jan 1, 2010 to Jun 30, 2017. We compared objective response rate (ORR) and disease control rate (DCR) between central and local assessments by study-level pooled analysis and correlation analysis. In pooled analysis, direct comparison was measured by the odds ratio (OR) of central-assessed response status to local-assessed response status; to investigate evaluation bias between central and local assessments, the above calculated OR between experimental (exp-) and control (con-) arms were compared, measured by the ratio of OR.

Results: A total of 28 included trials involving 17,466 patients were included (28 with ORR, 16 with DCR). Pooled analysis showed central assessment reported lower ORR and DCR than local assessment, especially in trials with open-label design, central-assessed primary endpoint, and positive primary endpoint outcome, respectively. However, this finding could be found in both experimental [exp-ORR: OR=0.81 (95% CI: 0.76–0.87), P<0.01, I^2=11%; exp-DCR: OR=0.90 (0.81–1.01), P=0.07, I^2=42%] and control arms [con-ORR: OR=0.79 (0.72–0.85), P<0.01, I^2=17%; con-DCR: OR=0.94 (0.86–1.02), P=0.14, I^2=12%]. No sign of evaluation bias between two assessments was indicated through further analysis [ORR: ratio of OR=1.02 (0.97–1.07), P=0.42, I^2=0%; DCR: ratio of OR=0.98 (0.93–1.03), P=0.37, I^2=0%], regardless of mask (open/blind), sample size, tumor type, primary endpoint (central-assessed/local-

assessed), and primary endpoint outcome (positive/negative). Correlation analysis demonstrated a high-degree concordance between central and local assessments (exp-ORR, con-ORR, exp-DCR, con-DCR: r>0.90, P<0.01).

Conclusions: Blind independent central review remained irreplaceable to monitor local assessment, but its complete-case implementation may be unnecessary.

Keywords: Blind independent central review; randomized control trials (RCTs); tumor assessment

Submitted Sep 18, 2017. Accepted for publication Nov 07, 2017.
doi: 10.21037/atm.2017.11.24
View this article at: http://dx.doi.org/10.21037/atm.2017.11.24

Introduction

The assessment for current response and progression endpoints, such as objective response rate (ORR), disease control rate (DCR), progression-free survival (PFS), and time-to-progression (TTP), has to be based on investigators' professional knowledge and experience. Namely, their assessment could be influenced by subjective factors, including failure to diagnose new lesions, variability during tumor measurement, target-lesion selection, and different interpretations on non-target or immeasurable lesions (1). In addition, the knowledge of investigators regarding treatment assignment would influence their assessment as well, especially in trials with open-labelled design (2). The above subjective factors may impact the assessment for trial endpoints, and subsequently the expected outcome will over- or underestimate the true effect of the treatments from experimental arm (exp) to control arm (con), possibly causing systematic bias (3). Therefore, blinded independent central review has been increasingly implemented in recent phase III oncological randomized control trials (RCTs). During implementation, all imaging examinations are acquired as part of protocol and reviewed by independent physicians who are blinded to treatment assignments and various information of patients (4), in order to detect and control potential bias from local investigators.

In current stage, however, systematic bias between central and local assessments in phase III RCTs is out of evidence from studies. In 2011, Amit *et al.* conducted the first study on systematic bias according to 27 phase III RCTs (5). This study compared the treatment effects of PFS between central and local assessments through meta-analysis, and found no systematic bias. No evidence of systematic bias has been further verified by two subsequent meta-analyses based on 28 and 61 RCTs, respectively (6,7).

Nevertheless, no evidence of systematic bias at the level of treatment effects, does not mean the evaluation concordance between central and local assessments when directly comparing response status. Further, the reliability of local assessment, as well as unnecessary implementation of central assessment could not be concluded only based on no evidence of systematic bias. For example, when comparing with central reviewers, if in experimental arm the local investigators overestimated the endpoints but in control arm they did not overestimate or even underestimated the endpoints, evaluation bias would possibly occur, even there might be no systematic bias when comparing the treatment effects between central and local assessments in the same study. Based on this assumption, central review is still valuable in clinical trials.

In order to verify the value of central assessment to local assessment, in this literature review and analyses, we investigated response status of ORR and DCR between central and local assessments among recently-published phase III RCTs on all non-hematologic solid tumors.

Methods

Search strategy and study selection

In accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement (8), a PubMed search was conducted by JRZ using the dates of Jan 1, 2010 to Jun 30, 2017. The retrieval formula was: ("neoplasms"(MeSH Terms) OR "neoplasms"(All Fields) OR "cancer"(All Fields)) AND random* AND ("Phase 3" OR "Phase III") (English) (Human) (Clinical Trial, Phase III). Articles of inappropriate publication types were excluded, including reviews, systematic reviews and/or meta-analyses, guidelines, and commentaries.

Eligible articles were regarding the therapeutic efficacy of anti-cancer agents in phase III RCTs for patients with non-hematologic solid tumors. In these articles, imaging assessment for ORR and/or DCR 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 clinicaltrials.gov identifier (NCT number) of eligible RCTs as search terms respectively to re-search PubMed (without the time interval limitation), to find out if there were more available articles of those trials. Endnote X7 (Thomson Reuters, New York, USA) was used in above process.

Data extraction

The process of data extraction was carried out independently and double-blindly by three reviewers working in pairs (Jianrong Zhang respectively blind to Yiyin Zhang and Shiyan Tang; in blocks of 17 articles allocated at random; discrepancies resolved by Wenhua Liang). To ensure consistency between reviewers, we used a same data extraction form, piloted the data extraction by using a sample of 16 included trials, and had discussion before and during the extraction process to confer how to properly extract and interpret the data.

Following characteristics of each trial were extracted: author, year, NCT number, mask (open/blind), sample size, tumor type, primary endpoint (central-assessed/local-assessed/other), and primary endpoint outcome (positive/negative/indeterminate). We also extracted ORR and DCR from both central and local assessments.

Statistical analysis

First, we directly compared response status between two assessments in both experimental and control arms by pooled analysis with Mantel-Haenszel method. The measurement was odds ratio (OR), defined as the ratio of central-assessed response status to local-assessed response status: ratio greater than 1 indicated central reviewers overestimated response status compared with local assessment; significant discrepancy between two assessments was shown if P<0.05. Second, we investigated evaluation bias between central and local assessments through pooled analysis with Inverse Variance method. In this procedure, above calculated ORs between experimental and control arms were compared, and the ratio of OR was the measure: regardless higher or lower than 1, P<0.05 indicated significant evaluation bias. During above two procedures, we also made subgroup analysis based on trial characteristics: mask, sample size (based on median value of all included trials), tumor type, and primary endpoint with its outcome. All mentioned procedures were conducted in Review Manager 5.3 (The Cochrane Collaboration, London, England), with initially fixed-effect model. If the corresponding p value for heterogeneity was less than 0.05 or the I^2 index (I^2) was over 50%, we used random-effect model, in order to reduce the heterogeneity effect.

In order to investigate the concordance between two assessments, we conducted correlation analysis by using SPSS Version 23 (SPSS Software, Chicago, USA). The test for normality was completed first, followed by correlation analysis if normal distribution was indicated, we estimated the correlation by the Pearson correlation coefficient; if not, Spearman correlation was applied. Significant correlation was indicated when p value was less than 0.05.

Results

Trial searching and characteristics

Based on article identification and selection, we totally included 28 trials from 35 articles (9-30), involving 17,466 randomly assigned patients (*Figure 1*) (31-43).

Summary and detailed characteristics are presented in *Table 1* and *Table S1*. All 28 included trials reported exp- and con-ORR from two assessments, and 16 trials reported exp- and con-DCR.

Direct comparison of response status between central and local assessments

Pooled analysis presented lower response frequency of ORR and DCR in central assessment compared with local assessment (*Table 2*, *Figures 2* and *3*), regardless in experimental [ORR: OR=0.81 (95% CI: 0.76–0.87), P<0.01, I^2=11%; DCR: OR=0.90 (0.81–1.01), P=0.07, I^2=42%] or control arm [ORR: OR=0.79 (95% CI: 0.72–0.85), P<0.01, I^2=17%; DCR: OR=0.94 (0.86–1.02), P=0.14, I^2=12%]. During above comparison, there was no significant interaction effect of therapy allocation (experimental arm versus control arm) in both ORR (P=0.56, I^2=0%) and DCR (P=0.42, I^2=0%).

In subgroup analysis (*Table 2*), the discrepancy between two assessments in trials with open-label design, positive

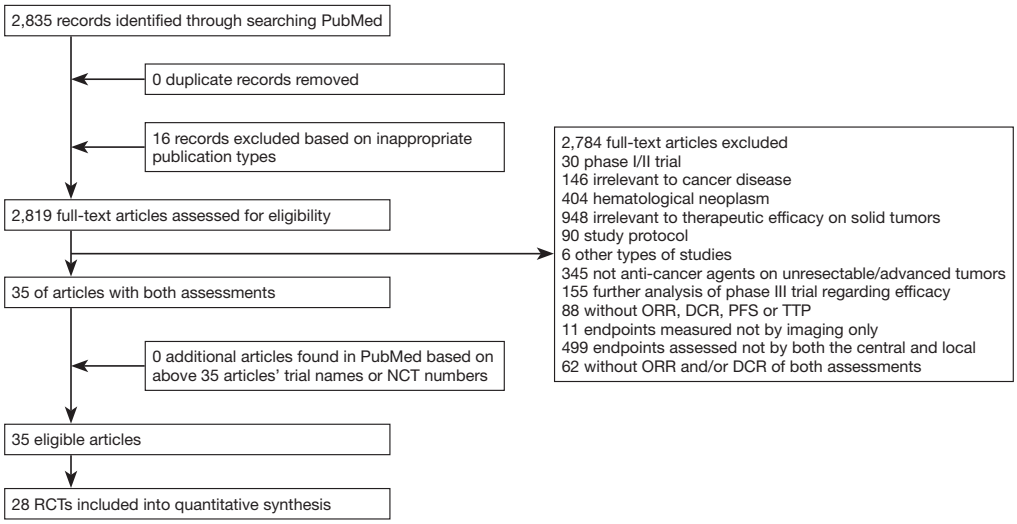

Figure 1 Flow chart of study identification and selection.

Table 1 Summary characteristics of included trials

Trial-level characteristics	Trial (N=28)	Patient (N=17,466)
Mask		
Open	22	14,133
Blind[†]	6	3,333
Sample size		
Max	–	1,110
Median	–	618.5
Min	–	185
Tumor type		
NSCLC	8	5,172
Breast	7	4,452
Renal-cell	6	3,917
Ovarian	3	1,985
Other[‡]	4	1,940
Primary endpoint		
Central-assessed[§]	16	10,668
Other[¶]	7	4,465
Local-assessed[a]	5	2,333
Outcome		
Positive	18	11,328
Indeterminate[b]	1	185
Negative	9	5,953

[†], 5 double blind, and 1 single blind; [‡], 2 melanoma, 1 gastrointestinal stromal tumor, 1 pancreatic tumor; [§], 1 central-assessed objective response rate, and 15 central-assessed progression-free survival; [¶], 6 overall survival and 1 unknown-assessed objective response rate; [a], 5 local-assessed progression-free survival; [b], in one study, objective response rate (ORR) was the primary endpoint: significant difference in central review (P=0.03), not in local assessment (P=0.05). We consider "indeterminate" because we are unable to determine whether central- or local-assessed ORR was the primary endpoint (24). NSCLC, non-small cell lung cancer; Breast, breast cancer; Renal-cell, renal-cell carcinoma; Ovarian, ovarian cancer.

Table 2 Direct comparison of response status between central versus local assessments

Summary/subgroup	Exp-ORR (central/local)				Con-ORR (central/local)				Exp-DCR (central/local)				Con-DCR (central/local)			
	S (n)	OR (95% CI)	P[†]	$I^{2‡§}$ (%)	S (n)	OR (95% CI)	P[a]	$I^{2‡§}$ (%)	S (n)	OR (95% CI)	P[a]	$I^{2‡§}$ (%)	S (n)	OR (95% CI)	P[a]	$I^{2‡§}$ (%)
Summary	28	0.81 (0.76–0.87)	<0.01	11	28	0.79 (0.72–0.85)	<0.01	17	16	0.90 (0.81–1.01)	0.07	42	16	0.94 (0.86–1.02)	0.14	12
Mask																
Open	22	0.80 (0.74–0.87)	<0.01	0	22	0.76 (0.69–0.83)	<0.01	75*	12	0.88 (0.77–1.02)	0.08	0	12	0.93 (0.85–1.02)	0.15	0
Blind	6	0.86 (0.73–1.01)	0.07		6	0.95 (0.78–1.17)	0.64		4	0.98 (0.83–1.15)	0.77		4	0.96 (0.80–1.16)	0.70	
Sample size																
> median (618.5)	14	0.81 (0.74–0.88)	<0.01	0	14	0.81 (0.73–0.90)	<0.01	0	10	0.89 (0.78–1.02)	0.10	0	10	0.96 (0.87–1.05)	0.34	0
< median (618.5)	14	0.82 (0.73–0.91)	<0.01		14	0.75 (0.65–0.86)	<0.01		6	0.93 (0.76–1.15)	0.52		6	0.88 (0.73–1.06)	0.17	
Tumor type																
NSCLC	8	0.76 (0.67–0.86)	<0.01	0	8	0.78 (0.67–0.91)	<0.01	19	4	0.95 (0.82–1.11)	0.53	62*	4	0.95 (0.80–1.13)	0.57	61*
Breast	7	0.88 (0.76–1.01)	0.07		7	0.70 (0.59–0.83)	<0.01		5	0.97 (0.75–1.27)	0.85		5	0.96 (0.82–1.12)	0.58	
Renal-cell	6	0.82 (0.71–0.95)	<0.01		6	0.81 (0.68–0.97)	0.02		4	0.72 (0.61–0.85)	<0.01		4	0.81 (0.70–0.94)	<0.01	
Ovarian	3	0.73 (0.59–0.90)	<0.01		3	0.78 (0.62–0.97)	0.03		1	1.09 (0.83–1.44)	0.52		1	1.34 (1.02–1.75)	0.04	
Others	4	0.88 (0.73-1.06)	0.17		4	0.98 (0.77-1.26)	0.90		2	1.10 (0.73-1.64)	0.65		2	1.00 (0.69-1.46)	1.00	
Primary endpoint																
Central-assessed	16	0.80 (0.73–0.87)	<0.01	53	16	0.75 (0.68–0.82)	<0.01	55	10	0.86 (0.73–1.02)	0.08	22	10	0.88 (0.80–0.98)	0.02	49
Others	7	0.77 (0.67–0.89)	<0.01		7	0.94 (0.77–1.15)	0.58		5	0.95 (0.83–1.08)	0.43		5	1.02 (0.87–1.19)	0.82	
Local-assessed	5	0.96 (0.81–1.15)	0.68		5	0.82 (0.65–1.03)	0.08		1	1.00 (0.97–1.03)	0.96		1	1.20 (0.83–1.73)	0.34	
Primary outcome																
Positive	18	0.80 (0.73–0.87)	<0.01	0	18	0.74 (0.66–0.82)	<0.01	37	10	0.88 (0.77–1.00)	0.05	0	10	0.89 (0.79–0.99)	0.03	31
Indeterminate	1	0.79 (0.36–1.72)	0.55		1	0.64 (0.22–1.89)	0.42		1	0.83 (0.46–1.51)	0.65		1	0.87 (0.48–1.58)	0.65	
Negative	9	0.84 (0.74–0.94)	<0.01		9	0.85 (0.75–0.97)	0.01		5	0.97 (0.77–1.22)	0.77		5	1.03 (0.90–1.18)	0.68	

[†], P value for the comparison between central and local assessments; P<0.05 indicated significant discrepancy; [‡], 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; *, significant interaction effects between the elements of each subgroup factor. Exp, experimental arm; Con, control arm; ORR, objective response rate; DCR, disease control rate; OR, odds ratio; S, study.

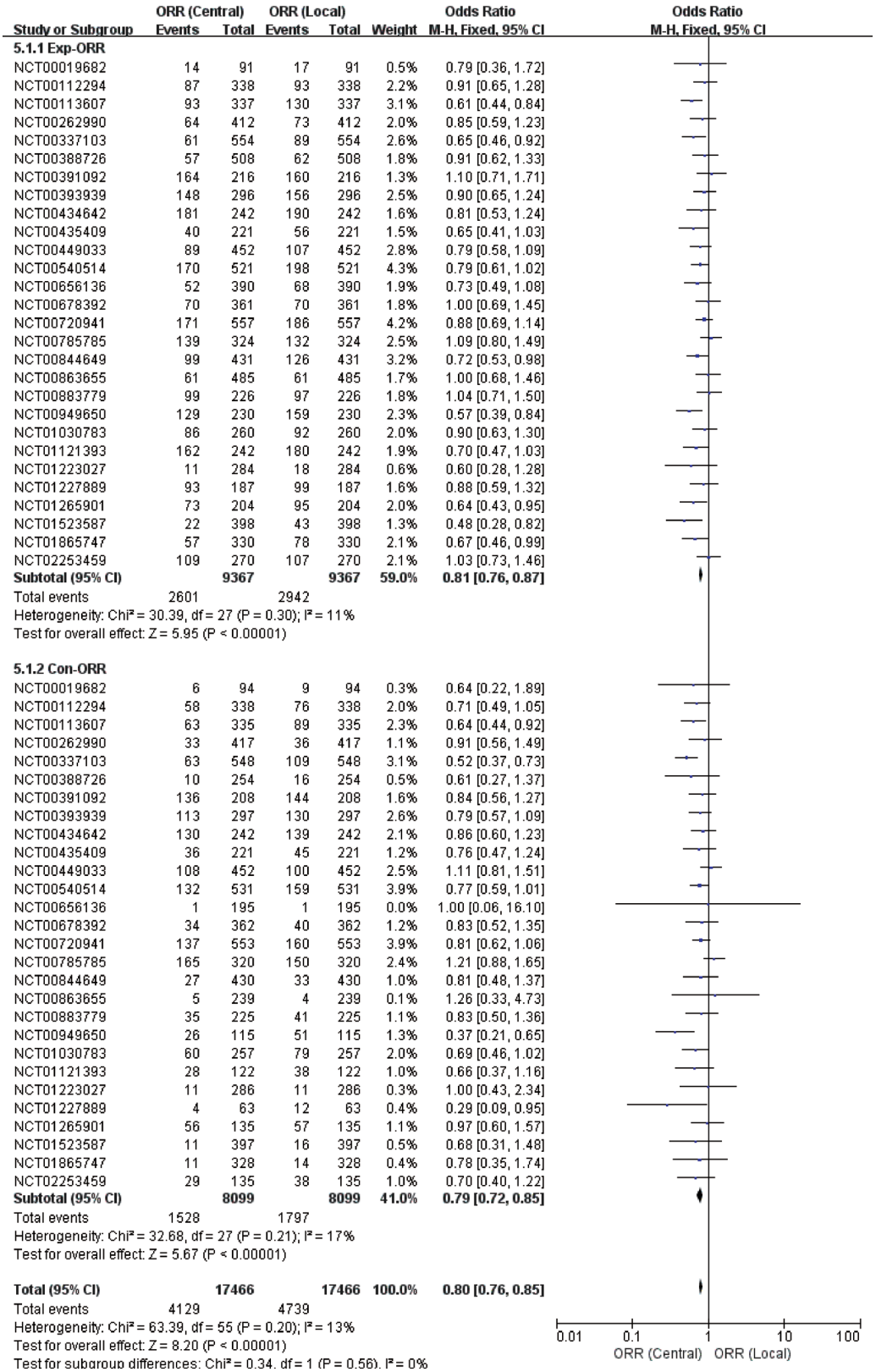

Figure 2 Forest plot for direct comparison of ORR between central and local assessments. ORR, objective response rate.

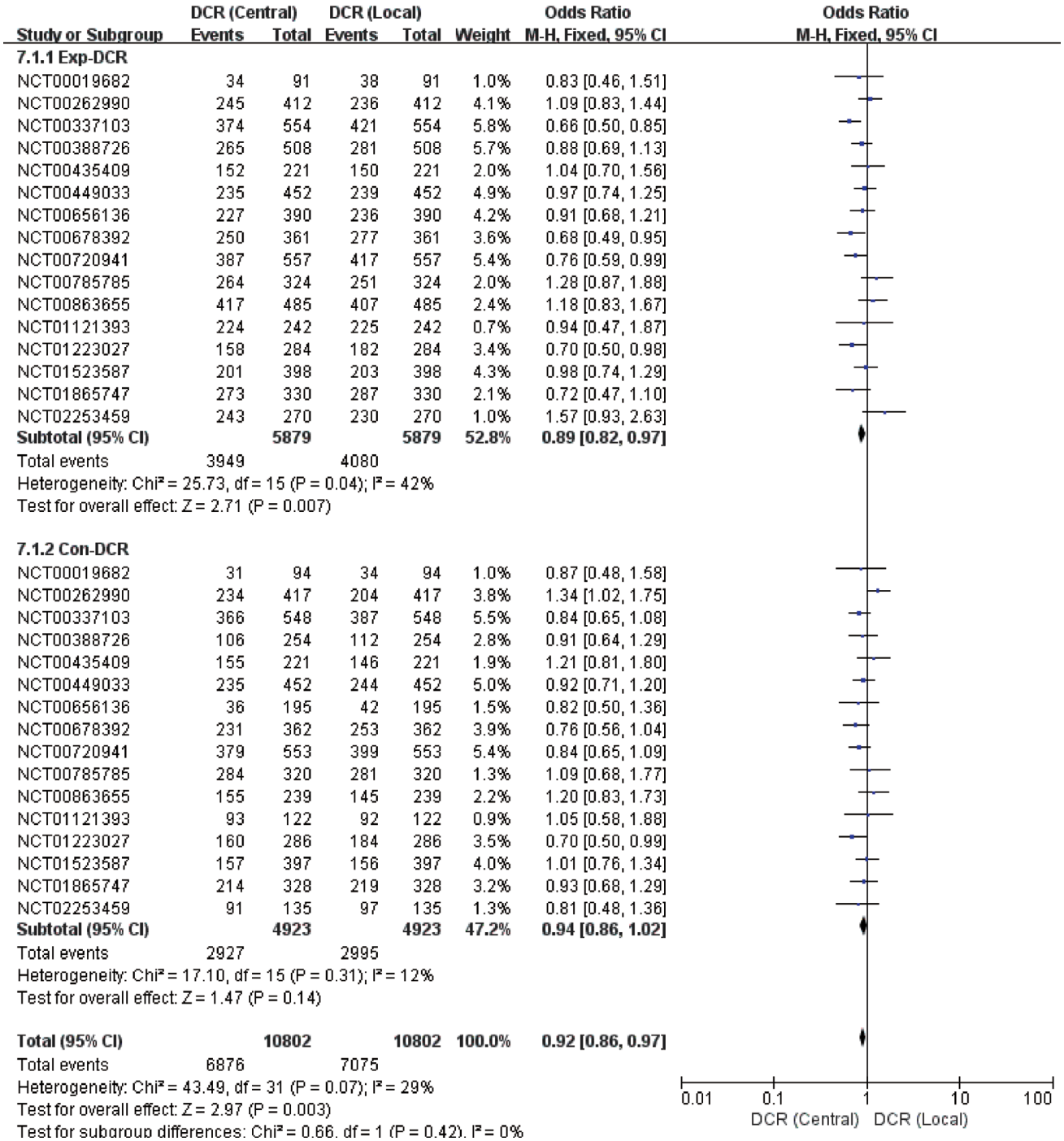

Figure 3 Forest plot for direct comparison of DCR between central and local assessments. DCR, disease control rate.

primary outcome, and central assessed primary endpoints, was larger than the discrepancy in trials with blind design, negative primary outcome, and local-assessed primary endpoints, respectively. Correspondingly significant interaction effect was only found in con-ORR (P=0.04, I^2=75%) between the mask pattern (open versus blind).

Evaluation bias between central and local assessments

No evidence of evaluation bias between central and local assessments was indicated by pooled analysis. Of ORR, the ratio of OR was 1.02 (0.97–1.07) (P=0.42); of DCR, the ratio of OR was 0.98 (0.93–1.03) (P=0.37). Subgroup analysis further verified no sign of evaluation bias, including the mask pattern and primary endpoint outcomes (*Table 3*).

Concordance between central and local assessments

Correlation analysis presented high-degree concordance between two assessments. The outcome (r) was 0.985 (P<0.01), 0.962 (P<0.01), 0.962 (P<0.01) and 0.926 (P<0.01) of exp-ORR, con-ORR, exp-DCR and con-DCR, respectively (*Figure 4*).

Discussion

As we acknowledge, this is the largest study to investigate response status between central and local assessments in recent 28 phase III oncological clinical trials on different advanced solid tumors, and it is the first study involving DCR for this topic. Based on pooled analysis, we found even though local assessment estimated higher treatment efficacy than the efficacy of central assessment, this phenomenon existed in both experimental and control arms. In other words, both response statuses of central and local assessments between two arms were concordant. This was verified by correlation analysis. More importantly, there was no sign of evaluation bias between two assessments after further pooled analysis.

Comparing with two previous meta-analyses, first, our study further confirms their results. In Lima *et al.* study based on 13 RCTs on metastatic colorectal cancer (44), local assessment had higher ORR than central assessment; this higher-estimated finding was not just in experimental arm [OR=1.16 (1.09–1.22), P<0.001], but also in control arm [OR=1.16 (1.09–1.25), P<0.001]. Parallel with our research, there was no significant interaction effect between therapy allocation (P=0.81), and also no evidence of evaluation bias by their further analysis [ratio of OR=0.97 (0.90–1.04), P=0.35]. According to the results, Lima *et al.* concluded that, the need of complete-case central assessment should be reappraised (44). In another meta-analysis based on 21 trials with different tumors, Tang *et al.* investigated the variability according to the difference of ORR and median PFS between central and local assessments (18 trials with ORR, 8 trials with PFS) (45). Comparing with central reviewers, local investigators overestimated ORR [estimated mean difference =4.57% (2.95–6.19%)], but did not overestimate median PFS [estimated mean difference =–0.19 (–0.68 to 0.29) months]. No evaluation bias was indicated by further analysis regardless of ORR (P=0.54) or PFS (P=0.31). Tang *et al.* concluded, due to the variability between central and local assessments, central review should be considered when the primary endpoint is based on response or progression assessment in oncological clinical trials (45).

The difference between our research and the above two studies is, we included larger RCTs which are recently published from 2010 to 2017, and involved DCR for analysis. Additionally, we found lower treatment response of central assessments especially in trials with open-label design, central-assessed primary endpoint, and positive primary endpoint outcome. Namely, assessment by central reviewers seemed more "conservative" in open-label studies, or in trials whose primary endpoint was based on central assessment. However, no evaluation bias could be found regardless of summary synthesis or subgroup analysis, including above-mentioned subgroup circumstances of open-labelled versus blind design, central-assessed primary endpoint versus local-assessed primary endpoint, and positive primary endpoint outcome versus negative primary endpoint outcome.

According to no evidence of systematic bias (5,6), as well as high-degree concordance without evaluation bias between central and local assessment in previous and our meta-analysis, we consider, the implementation of central assessment for all enrolled patients is unnecessary in clinical trials. Instead, we are looking forward to understanding the usage of sample-based central review as an audit strategy in future trials (1,2,5,6,46). Its value deserves further investigation.

Our research has several limitations. First, the finding of our research may be not completely generalizable to all phase III clinical trials on advanced solid tumors, in that our included trials were implemented with both assessments' response endpoints. Accordingly, we are

Table 3 Evaluation bias between central versus local assessments

Summary/subgroup	ORR (Exp/Con)				DCR (Exp/Con)			
	S (n)	OR (95% CI)	P[†]	$I^{2‡§}$ (%)	S (n)	OR (95% CI)	P[a]	$I^{2‡§}$ (%)
Summary	28	1.02 (0.97–1.07)	0.42	0	16	0.98 (0.93–1.03)	0.37	0
Mask								
Open	22	1.03 (0.98–1.09)	0.23	30	12	0.97 (0.91–1.02)	0.25	0
Blind	6	0.95 (0.84–1.08)	0.43		4	1.01 (0.91–1.13)	0.79	
Sample size								
> median (618.5)	14	1.00 (0.94–1.06)	0.99	1	10	0.96 (0.91–1.02)	0.21	0
< median (618.5)	14	1.05 (0.97–1.14)	0.20		6	1.03 (0.92–1.15)	0.63	
Tumor type								
NSCLC	8	1.01 (0.93–1.10)	0.81	0	4	1.01 (0.91–1.12)	0.85	0
Breast	7	1.08 (0.98–1.20)	0.13		5	0.98 (0.89–1.07)	0.62	
Renal-cell	6	1.01 (0.91–1.12)	0.86		4	0.95 (0.87–1.05)	0.35	
Ovarian	3	0.98 (0.85–1.12)	0.72		1	0.91 (0.77–1.09)	0.32	
Others	4	1.00 (0.86–1.15)	0.99		2	1.04 (0.83–1.30)	0.73	
Primary endpoint								
Central-assessed	16	1.04 (0.98–1.10)	0.22	50	10	0.97 (0.91–1.04)	0.37	0
Others	7	0.92 (0.83–1.03)	0.16		5	0.99 (0.90–1.08)	0.76	
Local-assessed	5	1.08 (0.94–1.24)	0.28		1	0.99 (0.80–1.23)	0.93	
Primary outcome								
Positive	18	1.04 (0.98–1.11)	0.23	0	10	0.99 (0.93–1.06)	0.76	0
Indeterminate	1	1.10 (0.61–1.95)	0.76		1	0.98 (0.68–1.42)	0.92	
Negative	9	0.99 (0.92–1.07)	0.84		5	0.96 (0.88–1.04)	0.30	

[†], P value for the comparison between central and local assessments; P<0.05 indicated significant discrepancy; [‡], 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. Exp, experimental arm; Con, control arm; ORR, objective response rate; DCR, disease control rate; OR, odds ratio; S, study.

unclear for the reliability of either assessment in trials which were only implemented or reported according to the result of one assessment. Second, even though we included trials with different solid tumors, heterogeneity analysis indicated, trials with different tumor types could have inconsistent findings when directly comparing DCR between two assessments. For example, in trials on non-small cell lung cancer, breast cancer and renal-cell carcinoma, central assessment underestimated treatment benefit on DCR compared with local assessment, but in trials on ovarian cancer, central assessment did not underestimate. However, regardless of above higher or lower treatment benefit, the OR of trials on different tumors represented same direction in both experimental and control arms. Given further analysis on evaluation bias, as well as following correlation analysis, these findings represented evaluation concordance between both assessments. Third, this meta-analysis was conducted based on study-level analysis, instead of individual-level analysis.

In conclusion, according to the finding that local assessment estimated higher treatment efficacy than the efficacy of central assessment in our direct comparison, we believe blind independent central review remains an

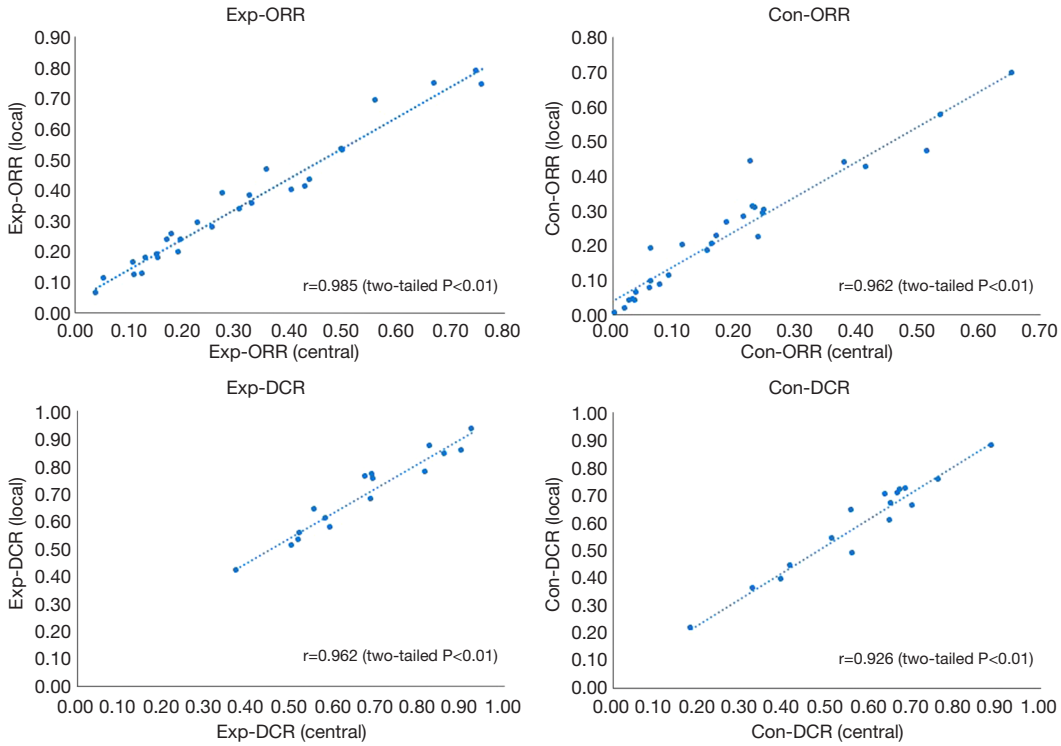

Figure 4 Scatterplot for the correlation between central and local assessments. ORR, objective response rate; DCR, disease control rate.

irreplaceable method to monitor local assessment. However, we don't believe its implementation for all patients is necessary in all trials, due to no evaluation bias between central and local assessments, as well as their high-degree evaluation concordance.

Acknowledgements

This article is the second 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, implementation to final report through systematic review and/or further analysis. We thank Ms. Xiaoru Deng (Guangzhou Medical University, Guangzhou, China) for advising the process of article searching, Ms. Elizabeth Burke (Washington University in St. Louis, St. Louis, USA) and Dr. Jieyu Wu (Karolinska Institutet, Stockholm, Sweden) for providing suggestions on manuscript revision. We also sincerely thank all authors, investigators, sponsors and patients for their effort and participation in our included studies.

Footnote

Conflicts of Interest: Some data of this research has been presented at ESMO 2016 Annual Meeting.

References

1. Amit O, Bushnell W, Dodd L, et al. Blinded independent central review of the progression-free survival endpoint. Oncologist 2010;15:492-5.
2. 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.
3. 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.
4. 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.
5. Amit O, Mannino F, Stone AM, et al. Blinded independent

central review of progression in cancer clinical trials: results from a meta-analysis. Eur J Cancer 2011;47:1772-8.
6. Zhang JJ, Chen H, He K, et al. Evaluation of Blinded Independent Central Review of Tumor Progression in Oncology Clinical Trials: A Meta-analysis. Ther Innov Regul Sci 2013;47:167-74.
7. Liang W, Zhang J, He Q, et al. Comparison of assessments by blinded independent central reviewers and local investigators: An analysis of phase III randomized control trials on solid cancers (2010-2015). Ann Oncol 2016;27:abstr 316P. doi:10.1093/annonc/mdw366.1
8. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535.
9. 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.
10. 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-2dagger. Ann Oncol 2014;25:2357-62.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. Miller VA, Hirsh V, Cadranel 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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.
25. 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.
26. 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.
27. Motzer RJ, Hutson TE, Cella D, et al. Pazopanib versus sunitinib in metastatic renal-cell carcinoma. N Engl J Med

2013;369:722-31.
28. 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.
29. 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.
30. Crown JP, Dieras 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. 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.
36. 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.
37. 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.
38. Monk BJ, Herzog TJ, Kaye SB, et al. Trabectedin plus pegylated liposomal Doxorubicin in recurrent ovarian cancer. J Clin Oncol 2010;28:3107-14.
39. 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.
40. Rini BI, Stenzl A, Zdrojowy R, et al. IMA901, a multipeptide 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.
41. 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.
42. 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.
43 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.
44. Lima JP, de Souza FH, de Andrade DA, et al. Independent radiologic review in metastatic colorectal cancer: systematic review and meta-analysis. Radiology 2012;263:86-95.
45. 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.
46. Zhang JJ, Zhang L, Chen H, et al. Assessment of audit methodologies for bias evaluation of tumor progression in oncology clinical trials. Clin Cancer Res 2013;19:2637-45.

Cite this article as: Zhang J, Zhang Y, Tang S, Liang H, Chen D, Jiang L, He Q, Huang Y, Wang X, Deng K, Jiang S, Zhou J, Xu J, Chen X, Liang W, He J. Evaluation bias in objective response rate and disease control rate between blinded independent central review and local assessment: a study-level pooled analysis of phase III randomized control trials in the past seven years. Ann Transl Med 2017;5(24):481. doi: 10.21037/atm.2017.11.24

Supplementary

Table S1 Characteristics of included trials

No.	Author & year	NCT number	P (Exp)	P (Con)	Funding source	Mask	Tumor type	Study design	Primary endpoint (outcome)
1	Baselga *et al.* 2012 (9)	NCT00863655	485	239	Pharmaceutical	Double	BC	Super	L-PFS (+)
2	Piccart *et al.* 2014 (10)								
3	Yardley *et al.* 2013 (11)								
4	Beaver *et al.* 2012 (12)								
5	Cortes *et al.* 2011 (13)	NCT00388726	508	254	Pharmaceutical	Open	BC	Super	OS (+)
6	Bergh *et al.* 2012 (14)	NCT00393939	296	297	Pharmaceutical	Open	BC	Super	C-PFS (–)
7	Wu *et al.* 2014 (15)	NCT01121393	242	122	Pharmaceutical	Open	NSCLC	Super	C-PFS (+)
8	Yang *et al.* 2015 (16)								
9	Miller *et al.* 2012 (17)	NCT00656136	390	195	Pharmaceutical	Double	NSCLC	Super	OS (+)
10	Gianni *et al.* 2013 (18)	NCT00391092	216	208	Pharmaceutical	Open	BC	Super	L-PFS (–)
11	Motzer *et al.* 2013 (19)	NCT00678392	361	362	Pharmaceutical	Open	RCC	Super	C-PFS (+)
12	Rini *et al.* 2011 (20)								
13	Lynch *et al.* 2010 (21)	NCT00112294	338	338	Pharmaceutical	Open	NSCLC	Super	C-PFS (–)
14	Hauschild *et al.* 2012 (22)	NCT01227889	187	63	Pharmaceutical	Open	M	Super	L-PFS (+)
15	Motzer *et al.* 2014 (23)	NCT01223027	284	286	Pharmaceutical	Open	RCC	Super	C-PFS (+)
16	Schwartzentruber *et al.* 2011 (24)	NCT00019682	91	94	Academic	Single	M	Super	ORR (±[†])
17	Von Hoff *et al.* 2013 (25)	NCT00844649	431	430	Pharmaceutical	Open	PT	Super	OS (+)
18	Wu *et al.* 2013 (26)	NCT00883779	226	225	Pharmaceutical	Double	NSCLC	Super	L-PFS (+)
19	Motzer *et al.* 2013 (27)	NCT00720941	557	553	Pharmaceutical	Open	RCC	Non	C-PFS (+)
20	Kaufman *et al.* 2015 (28)	NCT00337103	554	548	Pharmaceutical	Open	BC	Super	C-PFS + OS (–)
21	Sequist *et al.* 2013 (29)	NCT00949650	230	115	Pharmaceutical	Open	NSCLC	Super	C-PFS (+)
8	Yang *et al.* 2015 (16)								
22	Crown *et al.* 2013 (30)	NCT00435409	221	221	Pharmaceutical	Open	BC	Super	C-PFS (–)
23	Paz-Ares *et al.* 2012 (31)	NCT00449033	452	452	Pharmaceutical	Double	NSCLC	Super	OS (–)
24	Motzer *et al.* 2013 (32)	NCT01030783	260	257	Pharmaceutical	Open	RCC	Super	C-PFS (+)
25	Socinski *et al.* 2012 (33)	NCT00540514	521	531	Pharmaceutical	Open	NSCLC	Non	C-ORR (+)
26	Aghajanian *et al.* 2014 (34)	NCT00434642	242	242	Pharmaceutical	Double	OC	Super	L-PFS (+)
27	Aghajanian *et al.* 2012 (35)								
28	Colombo *et al.* 2012 (36)	NCT00262990	412	417	Pharmaceutical	Open	OC	Super	OS (–)
29	Monk *et al.* 2010 (37)	NCT00113607	337	335	Pharmaceutical	Open	OC	Super	C-PFS (+)
30	Monk *et al.* 2012 (38)								
31	Soria *et al.* 2015 (39)	NCT01523587	398	397	Pharmaceutical	Open	NSCLC	Super	C-PFS (+)
32	Rini *et al.* 2016 (40)	NCT01265901	204	135	Pharmaceutical	Open	RCC	Super	OS (–)
33	Choueiri *et al.* 2016 (41)	NCT01865747	330	328	Pharmaceutical	Open	RCC	Super	C-PFS (+)
34	Blay *et al.* 2015 (42)	NCT00785785	324	320	Pharmaceutical	Open	GST	Super	C-PFS (–)
35	Zhang *et al.* 2017 (43)	NCT02253459	270	135	Pharmaceutical	Open	BC	Super	C-PFS (+)

[†], in one study, objective response rate (ORR) was the primary endpoint: significant difference in central review (P=0.03), not in local assessment (P=0.05). We consider "indeterminate" because we are unable to determine whether central- or local-assessed ORR was the primary endpoint (24). P, patients; Criteria, assessment criteria for tumor response or progression; BC, breast cancer; OC, ovarian cancer; NSCLC, non-small-cell lung cancer; RCC, renal-cell cancer; M, melanoma; PT, pancreatic tumor: GST, gastrointestinal stromal tumor; Super, superiority; Non, noninferiority; C-, central-assessed; L-, local-assessed; ORR, objective response rate; PFS, progression-free survival; OS, overall survival. +, positive; –, negative.