



OPEN ACCESS

Detection bias in open-label trials of anticancer drugs: a meta-epidemiological study

Satoshi Funada ^{1,2}, Yan Luo ¹, Yuki Kataoka ^{3,4,5,6}, Takashi Yoshioka,² Yusuke Fujita,⁷ Shinya Yoshida,⁸ Morihiro Katsura,^{9,10} Masafumi Tada,^{1,11} Norihiro Nishioka,¹² Yoshiaki Nakamura,^{13,14} Kentaro Ueno,¹⁵ Ryuji Uozumi,¹⁶ Toshi A Furukawa ¹

10.1136/bmjebm-2023-112332

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjebm-2023-112332>).

For numbered affiliations see end of article.

Correspondence to:

Dr Satoshi Funada,
Department of Health Promotion and Human Behavior, Kyoto University, Graduate School of Medicine and Faculty of Medicine / School of Public Health, Kyoto 606-8501, Japan; sfunada@kuhp.kyoto-u.ac.jp



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY. Published by BMJ.

To cite: Funada S, Luo Y, Kataoka Y, *et al.* *BMJ Evidence-Based Medicine* 2023;**28**:372–382.

Abstract

Objectives In anticancer clinical trials, particularly open-label trials, central reviewers are recommended to evaluate progression-free survival (PFS) and objective response rate (ORR) to avoid detection bias of local investigators. However, it is not clear whether the bias has been adequately identified, or to what extent it consistently distorts the results. Therefore, the objective of this study was to evaluate the detection bias in oncological open-label trials by confirming whether local investigators overestimate the PFS and ORR compared with the findings of central reviewers.

Design Meta-epidemiological study.

Data sources MEDLINE via PubMed from 1 January 2010 to 30 June 2021.

Eligibility criteria for selecting studies Open-label, parallel-group superiority, randomised trials of anticancer drugs that adjudicated PFS or ORR by both central reviewers and local investigators.

Review methods We assessed the values for the same outcome (PFS and ORR) adjudicated by both central reviewers and local investigators. A random-effects model was used to estimate the ratio of HR (RHR) for PFS and the ratio of OR (ROR) for ORR between central reviewers and local investigators. An RHR lower than 1 and an ROR higher than 1 indicated an overestimation of the effect estimated by local investigators.

Results We retrieved 1197 records of oncological open-label trials after full-text screening. We identified 171 records (PFS: 149 records, ORR: 136 records) in which both central reviewers and local investigators were used, and included 114 records (PFS: 92 records, ORR: 74 records) for meta-analyses. While the RHR for PFS was 0.95 (95% CI 0.91 to 0.98), the ROR of ORR was 1.00 (95% CI 0.91 to 1.09). The results remained unchanged in the prespecified sensitivity analysis.

Conclusions This meta-epidemiological study found that overestimation of local investigators has a small impact on evaluating PFS and ORR in oncological open-label trials. However, a limitation of this study is that it did not include data from all trials; hence, the results may not fully evaluate detection bias. The necessity of central reviewers in oncological open-label trials needs to be assessed by further studies that overcome this limitation.

Trial registration number CTR-UMIN000044623.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The US Food and Drug Administration and European Medicines Agency recommend the use of central reviewers in oncological open-label trials to avoid detection bias; however, this recommendation is not based on evidence. Previous meta-epidemiological studies have compared progression-free survival (PFS) or objective response rate (ORR) adjudicated by central reviewers and local investigators, but have not clearly identified an overestimation by local investigators. However, these studies did not specifically focus on open-label trials, and there is a concern that an adequate number of studies were not included through an appropriate search strategy.

WHAT THIS STUDY ADDS

⇒ This meta-epidemiological study found that only a small fraction of the studies on anticancer drugs adjudicated PFS or ORR using both central and local investigators, and only half of these studies reported both outcomes. A meta-analysis of these studies showed that PFS may have been slightly overestimated by local investigators, while ORR was not. A sensitivity analysis which used a range of assumptions did not change these results. Our study suggests that the impact of the differences between central and local adjudications is not substantial. However, as this study did not access data from all trials, the results may not fully evaluate detection bias. Among the studies that claimed both assessors were used, half of them only reported the results from one assessor, indicating potential selective outcome reporting bias. Until the findings of this study

ARE VALIDATED BY STUDIES THAT OVERCOME THIS LIMITATION, IT IS DESIRABLE TO ESTABLISH CENTRAL REVIEWERS IN ONCOLOGICAL OPEN-LABEL TRIALS.

HOW MIGHT THIS AFFECT RESEARCH, PRACTICE OR POLICY

⇒ It is expected that the results of this study will lead to a trend towards reporting outcomes adjudicated by both central and local investigators. Additional meta-epidemiological studies would then be conducted to validate this study and accumulate knowledge on detection bias in open-label trials of anticancer drug.

Introduction

Detection bias can systematically distort the results of randomised controlled trials.¹ This bias, also known as observer bias or measurement bias, can occur when the outcomes are subjective and the adjudicators are not blinded (or masked), leading them to evaluate the results optimistically. In general medicine, several meta-epidemiological studies have assessed the magnitude of detection bias by comparing the same outcomes between blinded and non-blinded adjudicators, with conflicting findings and a lack of consensus.^{2–5} Recently, the MetaBLIND study found no difference in the estimated treatment effect between trials with and without blinded outcome adjudicators; however, since they compared the outcomes in different trials, the study had a risk of confounding.⁶ Furthermore, these meta-epidemiological studies have diverse specialities and outcomes, and whether they can be generalised to specific outcomes in oncology requires validation.

Open-label trials are more common in oncological clinical trials than in non-oncological trials,⁷ and there has been a trend towards using progression-free survival (PFS) and objective response rate (ORR) as primary outcomes in oncological trials rather than overall survival (OS).⁸ This is due to the abbreviated time required to evaluate efficacy, smaller sample size requirements and lack of subsequent treatments. However, PFS and ORR involve more subjective judgement than OS, particularly in open-label settings. Given recent trends, detection bias is a particular concern in oncological clinical trials.

The US Food and Drug Administration recommends the use of central reviewers blinded to study treatment to verify tumour assessment to minimise bias in oncological clinical trials when the primary endpoint is PFS or ORR.⁹ The European Medicines Agency also emphasises the importance of blinded independent central reviewers, especially when the majority of events are based on imaging rather than clinical progression.¹⁰ However, previous oncological meta-epidemiological studies have yielded inconsistent findings and do not necessarily demonstrate over-estimation by local investigators.^{11–15} Nevertheless, these studies did not focus on open-label trials, and there is a concern that an adequate number of studies were not included through an appropriate search strategy. Therefore, while the importance of central reviewers is widely recognised in oncology, there is limited evidence to support their importance.

This meta-epidemiological study focused on open-label trials of anticancer drugs and aimed to evaluate detection bias by confirming whether local investigators overestimate the PFS and ORR compared with the estimates by central reviewers.

Methods

Study design

This was a meta-epidemiological study of randomised controlled trials registered in the UMIN Clinical Trial Registry (CTR-UMIN000044623). The protocol was published as a preprint in the Open Science Foundation¹⁶ and descriptive summary results of the identified studies have been reported elsewhere.¹⁷ The study was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.¹⁸

Eligibility criteria

Types of studies

We included open-label, parallel-group superiority and randomised trials that investigated the efficacy of anticancer drugs. Non-inferiority and equivalence trials were excluded because the null hypothesis was different from that of the superiority trials. We also excluded records that were not in English language.¹⁹

Types of participants

We focused on solid tumours and excluded haematological diseases, including leukaemia, lymphoma and multiple myeloma. Because haematological cancers are evaluated biologically, their evaluations are less likely to be influenced by adjudicators than solid tumours. Therefore, we only included solid cancers of all histological types and stages.

Types of interventions

Eligible interventions included molecularly targeted therapy, immune checkpoint inhibitors, immune therapy, chemotherapy or hormone therapy; and eligible comparisons included standard therapy, supportive care or no treatment. Combination therapies were also included (eg, therapy A plus standard therapy vs standard therapy). Owing to heterogeneity, we excluded neoadjuvant and adjuvant interventions.

Types of outcomes

We included trials that used PFS or ORR as measurements of treatment efficacy.

Information sources

We identified relevant trials from the MEDLINE database.

Search strategy

We searched MEDLINE from 1 January 2010 to 30 June 2021. Online supplemental appendix lists the search terms used on 1 July 2021. Furthermore, we conducted a manual search of the reference lists attached to relevant articles. We restricted the search period to 2010 to allow the possibility of contacting the corresponding authors.

Selection process

First, seven independent pairs of researchers (SF and TY; SF and YF; SF and SY; SF and NN; YL and YK; TY and MK; TY and MT) screened the titles and abstracts of the records identified by the literature search. Second, different seven pairs of researchers (SF and YL; SF and YK; SF and TY; SF and YF; SF and SY; SF and MK; SF and MT) independently screened the full texts of the records. After full-text screening, we classified the articles by outcome adjudicators into four categories: 'central and local', 'only central', 'only local' and 'unclear'. Based on this classification, we selected articles in which the outcome was adjudicated by both the central and local investigators. Any disagreement was

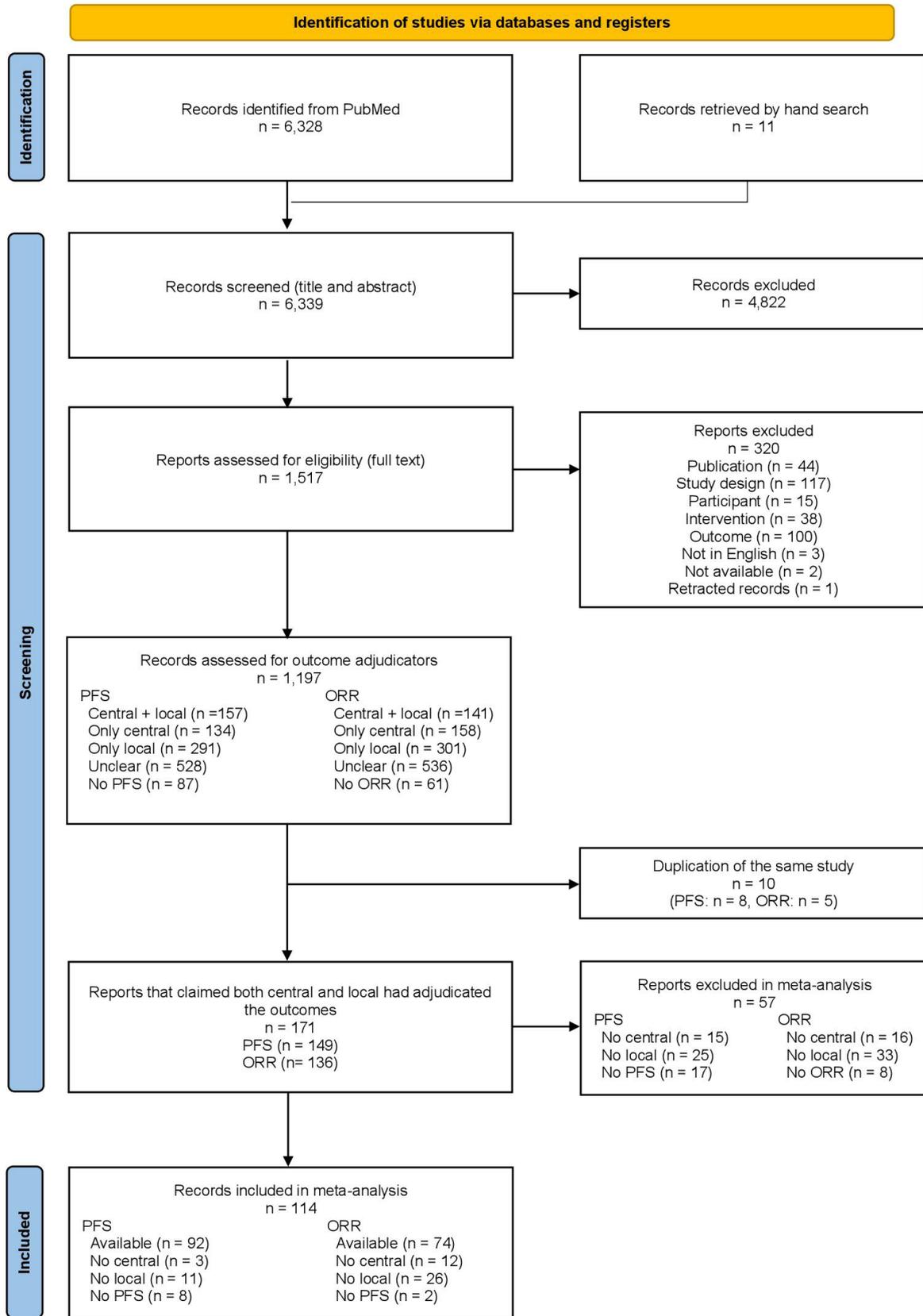


Figure 1 Flow chart of the study selection. ORR, objective response rate; PFS, progression-free survival.

resolved through discussion or consultation with another pair or TAF if necessary. The PRISMA flow chart (figure 1) summarises the reasons for excluding studies.

Data collection process

Eight pairs of researchers independently collected detailed information from each trial using a prepiloted form (SF and YL; SF and

Table 1 Characteristics of included studies (n=114)

	No (%)
Reported outcome	
PFS and ORR	52 (45.6)
Only PFS	40 (35.1)
Only ORR	22 (19.3)
Trial phase	
Phase III	74 (64.9)
Phase II	40 (35.1)
Cancer type	
Non-small cell lung cancer	29 (25.4)
Breast cancer	22 (19.3)
Renal cell cancer	14 (12.3)
Melanoma	9 (7.9)
Ovarian cancer	8 (7.0)
Head and neck cancer	5 (4.4)
Gastric and gastro-oesophageal junction adenocarcinoma	4 (3.5)
Sarcoma	4 (3.5)
Colorectal cancer	3 (2.6)
Gastrointestinal stromal tumour	3 (2.6)
Hepatocellular cancer	2 (1.8)
Pancreatic cancer	2 (1.8)
Others*	10 (8.8)
Drug classification of intervention	
Targeted therapy	70 (61.4)
Chemotherapy	26 (22.8)
Immune checkpoint inhibitor	15 (13.2)
Immunotherapy	3 (2.6)
Treatment line	
≥1st line	59 (51.8)
≥2nd line	47 (41.2)
≥3rd line	8 (7.0)
Primary outcome in the published paper	
PFS	82 (71.9)
ORR	13 (11.4)
OS	6 (5.3)
OS and PFS	7 (6.1)
Other†	6 (5.3)
Blinding of central reviewers	
Blinding	72 (63.2)
Unclear	42 (36.8)
Impact factor (2021)	
≥10	102 (89.5)
<10	12 (10.5)
Funding	
Industry	106 (93.0)
Public	5 (4.4)
Unclear	3 (2.6)
*Endometrial carcinoma, glioblastoma, liposarcoma or leiomyosarcoma, malignant mesothelioma, neuroendocrine tumours, pleural mesothelioma, prostate cancer, small cell lung cancer, squamous cell carcinoma of the lung.	
†OS and ORR, OS and ORR and PFS, time-to-treatment failure, tumour shrinkage, safety, incidence of cutaneous squamous cell carcinoma, PFS, response rate, duration of response and safety.	
ORR, objective response rate; OS, overall survival; PFS, progression-free survival.	

YK; SF and TY; SF and YF; SF and SY; SF and MK; SF and MT; SF and NN). When no details were reported for the data items, the corresponding authors were queried.

Data items

We collected the data items as follows:

1. General information of the study: publication journal, publication year, authors, trial registry number (National Clinical Trial number, European Union Clinical Trial number and others), trial phase (phase II, phase III and others), cancer type and number of randomised participants.
2. Intervention and comparison information: drug name, drug classification (targeted therapy, immune checkpoint inhibitor, immune therapy, chemotherapy or hormone therapy), comparison treatment (standard treatment, BSC or no treatment) and treatment line.
3. Outcome information: primary outcome, response criteria (response evaluation criteria in solid tumours (RECIST) 1.0,²⁰ RECIST V.1.1,²¹ WHO criteria²² and others), point estimates and 95% CIs of the HR of PFS assessed by central reviewers (HR_{central}), HR of PFS assessed by local investigators (HR_{local}), OR of ORR assessed by central reviewers (OR_{central}) and OR of ORR assessed by local investigators (OR_{local}). If the study did not report the 95% CIs, we collected the p value or the other CI (eg, the 90% CIs) to obtain the SE.²³

Statistical analysis

The characteristics of the included studies were summarised as numbers and relative frequencies for categorical variables. Our primary outcomes were the ratio of HR (RHR) and the ratio of OR (ROR) between HR_{central} (or OR_{central}) and HR_{local} (or OR_{local}). We performed a meta-analysis of RHR using the two-step process proposed by Sterne *et al.*²⁴ Within each trial, we divided HR_{local} by HR_{central} to calculate the RHR and divided OR_{local} by OR_{central} to calculate the ROR. An RHR<1 indicates that the local investigators overestimate the effect compared with the findings of central reviewers, which suggests the existence of a detection bias. In order to calculate the SE of RHR in each trial, we transformed RHR to log RHR and calculated the SE of the logRHR (SE(logRHR)) using the following equation:

$$SE(\log RHR) = \sqrt{\frac{\text{Var}(\log HR_{\text{central}}) + \text{Var}(\log HR_{\text{local}}) - 2 \times \rho \times SE(\log HR_{\text{central}}) \times SE(\log HR_{\text{local}})}{2}}$$

where ρ is correlation coefficients between HR_{central} and HR_{local} in each trial. Then, we performed a random-effects meta-analysis using the inverse variance method. We were unable to calculate ρ without individual patient data. We, therefore, assumed no dependency between HR_{central} and HR_{local} in each trial ($\rho=0$) for the main analysis, as it is the most conservative approach (ie, it generates the largest SE estimation and hence widest CIs).⁴ To test robustness, we also performed sensitivity analyses assuming $\rho=0.25, 0.50, 0.75$ and 0.95 . We presented a forest plot to visualise the RHR in all included studies. We evaluated the heterogeneity of RHR across different trials using the τ^2 and I^2 statistics. We prespecified and explored whether the direction or magnitude of detection bias would vary in the following subgroups: (1) trial phases, (2) cancer types and (3) drug classifications. We calculated the p values for interactions using a meta-regression model. We performed a meta-analysis of the ROR in the same manner. Note that unlike RHR, an ROR>1 indicates that the local investigators overestimate the effect compared with the findings of central reviewers, indicating detection bias. The 'metafor' package (V.3.8–1)²⁵ in R (V.4.0.3; R Foundation for Statistical Computing, Vienna, Austria) was used for meta-analysis. The collected data and codes used for our analysis are provided in the online supplemental materials.

Table 2 Estimated ratios of HRs between central and local adjudications in all the studies and in subgroups

	No of records	RHR (95% CI)	Tau ²	I ²	P value for heterogeneity	P value for interaction
Overall	92	0.95 (0.91 to 0.98)	0.00	0%	>0.99	
Trial phase						0.58
Phase III	68	0.94 (0.91 to 0.98)	0.00	0%	>0.99	
Phase II	24	0.97 (0.87 to 1.09)	0.00	0%	0.96	
Cancer type						0.86
Non-small cell lung cancer	23	0.91 (0.85 to 0.99)	0.00	0%	>0.99	
Breast cancer	20	0.98 (0.90 to 1.05)	0.00	0%	0.80	
Renal cell cancer	13	1.00 (0.92 to 1.09)	0.00	0%	0.95	
Melanoma	8	0.98 (0.85 to 1.13)	0.00	0%	0.86	
Ovarian cancer	8	0.91 (0.80 to 1.03)	0.00	0%	0.93	
Gastrointestinal stromal tumours	3	0.91 (0.58 to 1.42)	0.09	59%	0.08	
Head and neck cancer	3	0.97 (0.76 to 1.24)	0.00	0%	0.96	
Sarcoma	3	1.14 (0.70 to 1.85)	0.00	0%	0.60	
Pancreatic cancer	2	0.88 (0.70 to 1.12)	0.03	37.7%	0.21	
Endometrial carcinoma	1	0.74 (0.35 to 1.56)				
Gastric and gastro-oesophageal junction cancer	1	0.80 (0.62 to 1.02)				
Glioblastoma	1	0.83 (0.61 to 1.13)				
Hepatocellular cancer	1	0.76 (0.55 to 1.06)				
Liposarcoma or Leiomyosarcoma	1	1.05 (0.68 to 1.63)				
Malignant mesothelioma	1	0.98 (0.57 to 1.69)				
Neuroendocrine tumours	1	0.97 (0.70 to 1.34)				
Prostate cancer	1	0.73 (0.50 to 1.07)				
Squamous cell carcinoma of the lung	1	0.98 (0.78 to 1.22)				
Drug classification						0.45
Targeted therapy	62	0.95 (0.91 to 1.00)	0.00	0%	>0.99	
Immune checkpoint inhibitor	12	0.90 (0.83 to 0.98)	0.00	0%	0.73	
Chemotherapy	18	0.97 (0.89 to 1.05)	0.00	0%	0.71	

RHR, ratio of the HR.

Patient and public involvement

Patients and members of the public were not involved in this study because it was conducted to answer a methodological question that was not directly dependent on patient priorities, experiences or participant preferences.

Results

Study selection

Figure 1 illustrates the study selection process. We identified 6339 records, 1517 of which were eligible after screening titles and abstracts. From 1517 records, we retrieved 1197 eligible records after full-text screening and assessed the adjudicators of PFS and ORR. A total of 181 records (PFS: 157 records, ORR: 141 records) were adjudicated by both central reviewers and local investigators (online supplemental table 1), and the trend of the prevalence of outcome adjudicators did not change from 2010 to 2021 in both PFS (online supplemental figure 2) and ORR (online supplemental figure 3). Of the 181 records, we excluded 10 due to data duplication and included the remaining records that could extract the outcomes adjudicated by both central reviewers and local investigators. Finally, we included 114 records in this analysis, of which 92 were analysed for PFS and 74 for ORR. In other words, among the records that we judged were adjudicated by both central reviewers and local investigators for PFS, only 61.7% (92/149) reported results from both assessors and 54.4% (74/136) for ORR. Online supplemental materials list the exclusion criteria for records during the full-text screening stage (n=320), data extraction stage (n=57) and duplication records stage (n=10). We sent

121 emails to the corresponding authors to request details of the data (26 October 2022) and received 15 responses, of which only one provided sufficient information. The other 14 corresponding authors stated that they could not access the data.

Study characteristics

Table 1 presents the characteristics of the included records. The analysis included 114 records, 92 for PFS and 74 for ORR. The majority of the trials were phase III (n=74); the most common tumour types were non-small cell lung cancer (n=29), breast cancer (n=22) and renal cell cancer (n=14); and the most common type of drug was targeted therapy (n=70). PFS was the primary outcome in 82 records, whereas ORR was the primary outcome in 13. Most of the reports were from high-impact factor (>10) journals (n=102), with only 5 funded by the public sector.

Primary outcome: RHR and ROR

Table 2 and figure 2 present the RHR for PFS. Under the assumption of no dependency between central reviewers and local investigators in each trial ($\rho=0$), the RHR for PFS was 0.95 (95% CI 0.91 to 0.98), indicating that local investigators slightly overestimated the HR compared with the findings of central investigators. No heterogeneity was detected in the meta-analysis ($\tau^2=0$; $I^2=0\%$; $p>0.99$). Table 3 and figure 3 present ROR for ORR and, in contrast, the ROR for ORR was 1.00 (95% CI 0.91 to 1.09), and no heterogeneity was observed in meta-analyses ($\tau^2=0$; $I^2=0\%$; $p>0.99$). Sensitivity analysis showed the RHR for PFS and ROR for ORR under the assumption of dependency between

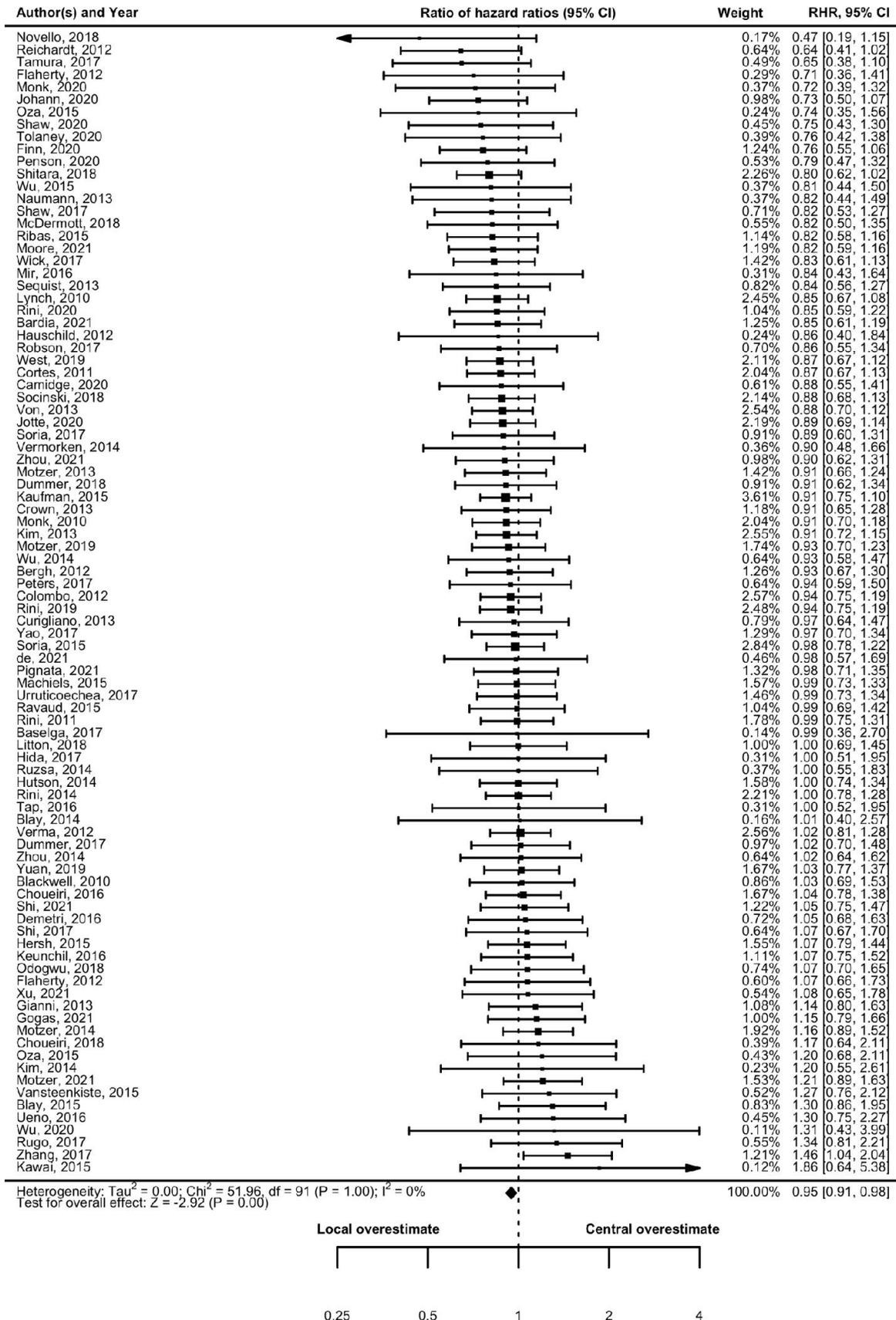


Figure 2 Comparison of treatment effect estimates (HR) between central reviewers and local investigators. RHR, ratio of HR.

Table 3 Estimated ratios of ORs between central and local adjudications in all the studies and in subgroups

	No of records	ROR (95% CI)	Tau ²	I ²	P value for heterogeneity	P value for interaction
Overall	74	1.00 (0.91 to 1.09)	0.00	0%	>0.99	
Trial phase						0.30
Phase III	45	0.98 (0.89 to 1.08)	0.00	0%	>0.99	
Phase II	29	1.14 (0.88 to 1.47)	0.00	0%	>0.99	
Cancer type						0.92
Non-small cell lung cancer	18	1.01 (0.85 to 1.20)	0.00	0%	0.99	
Breast cancer	14	0.94 (0.76 to 1.15)	0.00	0%	0.95	
Renal cell cancer	9	0.96 (0.81 to 1.14)	0.00	0%	0.95	
Head and neck cancer	5	1.36 (0.71 to 2.61)	0.00	0%	0.79	
Ovarian cancer	5	1.05 (0.77 to 1.42)	0.00	0%	>0.99	
Gastric and gastro-oesophageal junction cancer	4	0.93 (0.57 to 1.52)	0.00	0%	0.89	
Melanoma	4	0.91 (0.62 to 1.34)	0.00	0%	0.47	
Colorectal cancer	3	1.22 (0.42 to 3.50)	0.09	9.9%	0.29	
Hepatocellular cancer	2	1.94 (0.83 to 4.55)	0.00	0%	0.79	
Pancreatic cancer	2	1.22 (0.72 to 2.06)	0.00	0%	0.95	
Sarcoma	2	0.50 (0.13 to 1.91)	0.00	0%	0.59	
Endometrial carcinoma	1	8.64 (0.24 to 310.13)				
Gastrointestinal stromal tumours	1	1.10 (0.71 to 1.71)				
Malignant mesothelioma	1	0.87 (0.06 to 13.20)				
Pleural mesothelioma	1	0.39 (0.05 to 3.11)				
Small cell lung cancer	1	1.16 (0.26 to 5.24)				
Squamous cell carcinoma of the lung	1	1.40 (0.55 to 3.62)				
Drug classification						0.99
Targeted therapy	41	1.00 (0.87 to 1.14)	0.00	0%	>0.99	
Immune checkpoint inhibitor	10	1.00 (0.84 to 1.18)	0.00	0%	0.90	
Immunotherapy	3	0.89 (0.36 to 2.21)	0.00	0%	0.56	
Chemotherapy	20	1.01 (0.86 to 1.19)	0.00	0%	0.97	

ROR, ratios of OR.

the central reviewers and local investigators in each trial (online supplemental table 2 and online supplemental figures 4a–4d and 5a–5d). The RHR remained constant at 0.95, and the upper 95% CIs did not exceed 1.00, indicating that local investigators consistently overestimated the HR for PFS compared with the findings of central reviewers under any assumptions of dependency. Conversely, the ROR ranged from 1.00 to 1.03, with lower 95% CIs consistently below 1.00, indicating that local investigators did not overestimate the OR for ORR under any assumption of dependency. Heterogeneity increased for both PFS and ORR from $\rho=0$ (independent) to $\rho=0.95$ (nearly completely dependent). Tables 2 and 3 and online supplemental figures 6a,b and 8a,b show the prespecified subgroup analyses of trial phases, cancer types and drug classifications. No interactions were observed between these subgroups for either the PFS or ORR.

Discussion

Principal findings

This meta-epidemiological study found that local investigators tended to slightly overestimate the HR for PFS compared with the findings of central reviewers in oncological open-label trials. In contrast, there was no evidence of overestimation in ORR. These results remained consistent in the sensitivity analysis which accounted for various assumptions, and the subgroup analysis did not identify any factors that might influence the findings. However, it is important to note that these results were based on a small subset of oncological open-label trials, and only approximately half of the reports that claimed the outcomes were

adjudicated by both central reviewers and local investigators reported both results.

Strengths and limitations of the study

This study has several strengths. First, because it excluded any terms related to outcome adjudicators in its search strategy, this study represents the most comprehensive meta-analysis to date, including the largest number of studies of any existing reports. Second, we were able to provide a breakdown of adjudicated outcomes in recent oncological open-label trials. This enabled us to detect how many studies adjudicated PFS and ORR by central reviewers and local investigators, as well as the corresponding number of reported outcomes. Third, the results remained unchanged after sensitivity analysis, highlighting their robustness. Fourth, as our study compared the same outcomes between central and local investigators within the same trial, the risk of confounding was low.

Although this study has several strengths, it also has some limitations. First, the records in which both central reviewers and local investigators adjudicated the outcomes were only a small proportion of the total number of oncological open-label trials. Furthermore, only half of all reports were available for analysis, with the rest not reporting results from either assessor. Additionally, our preliminary study showed that outcome adjudicators might differ from those prespecified in the protocol or trial registry,¹⁷ suggesting the possibility of selective outcome reporting bias in our results. In other words, among studies that claimed that both central reviewers and local investigators

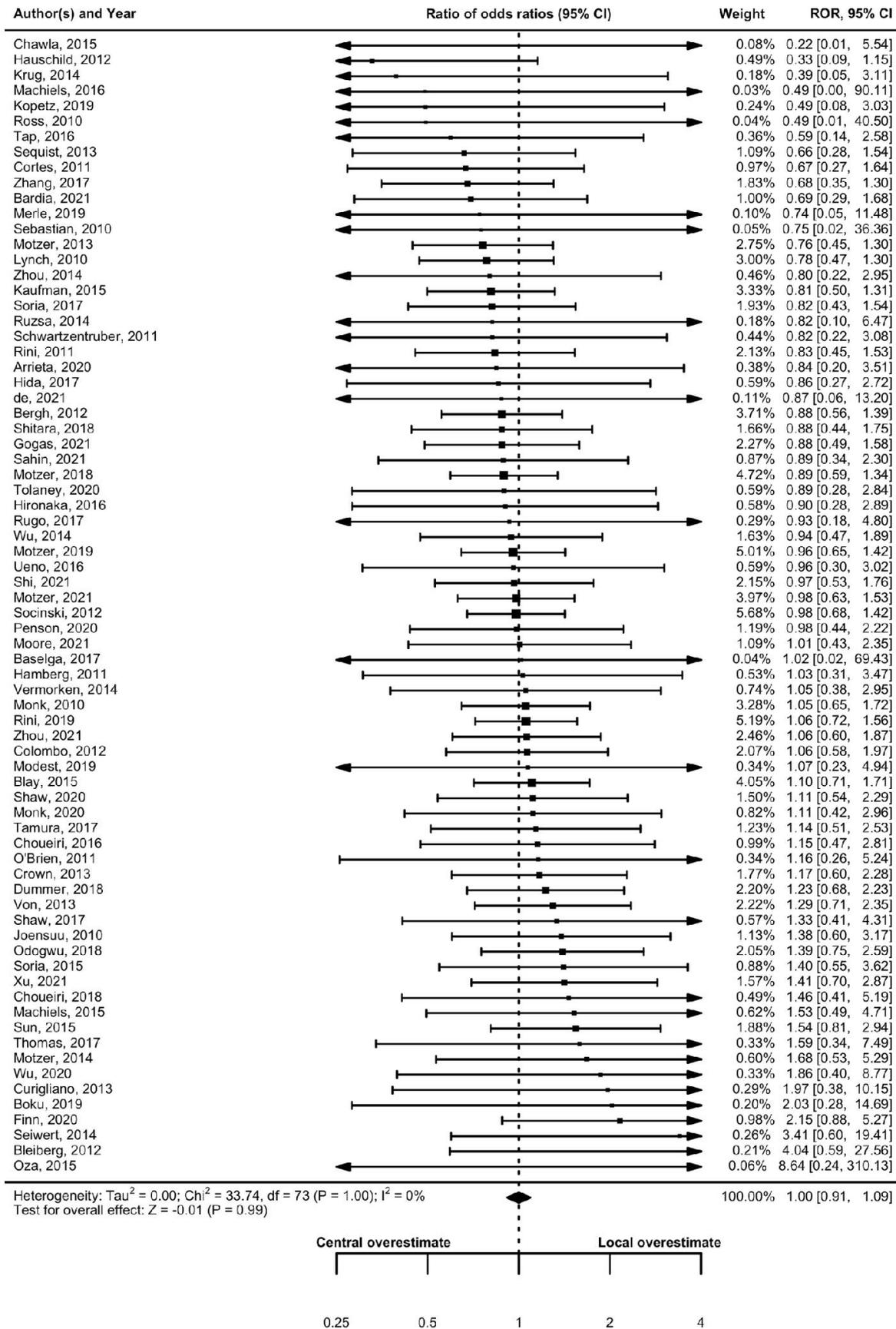


Figure 3 Comparison of treatment effect estimates (OR) between central reviewers and local investigators. ROR, ratio of OR.

adjudicated the outcome, we expect studies that reported results by both assessors (ie, studies included in our analysis) to show less significant differences between the two assessors. Moreover, when we contacted the investigators of those trials that reported only central or local adjudicators' outcomes, very few investigators provided data regarding local adjudicator outcomes. Because trial sponsors are unlikely to publish the results of trials in which the central and local adjudications are inconsistent, the low response rate may underestimate the extent of detection bias by local investigators due to selective outcome reporting bias. However, it should be noted that although the number of studies included in the analysis was limited, they showed a slight overestimation in the HR of PFS by local investigators. Actual detection bias could be greater in real-world settings. Second, we assumed that the dependence (ρ) between central and local adjudicators in all studies was 0 in the primary analysis, and between 0.25 and 0.95 in the sensitivity analysis. However, the presence of no dependence ($\rho=0$) between central and local adjudications is implausible, and the dependence may vary from trial to trial; therefore, this assumption may not be entirely accurate. A more rigorous synthesis could be achieved by calculating the true dependence in each study using individual patient data. Nevertheless, as our sensitivity analysis indicated that changing the values of ρ from 0 to 0.95 did not affect the final results, we assume that the variation in ρ observed among the studies would not affect the final RHR or ROR. Third, the period of the search was limited to studies published after 2010. Although this meta-analysis includes more studies than previously published reviews, this search strategy was not entirely systematic.

Comparison with other studies

Regarding general medicine, the MetaBLIND study, which evaluated the effect of blinding outcome assessors to the intervention, did not find an apparent overestimation by non-blinded assessments.⁶ However, as this meta-epidemiological study compared the outcomes of different trials, there was a risk of confounding. Contrarily, when examining studies that compared blinded and non-blinded outcome adjudicators within the same trials, several previous meta-epidemiological studies found an overestimation of non-blinded adjudication by 27%–36%.^{2–4} Nevertheless, as these reviews did not focus on the same medical field and outcomes and included both double-blinded and open-label trials, these findings may not be generalisable to oncology.

Narrowing the focus to oncology and the outcomes to PFS and ORR, several meta-epidemiological studies have been published to date, with inconsistent findings. Two studies that evaluated PFS found no apparent differences between central and local investigators.^{12 13} However, these studies did not focus exclusively on open-label trials, and included 36 and 21 open-label trials, respectively. Of the three studies that evaluated ORR, one study found an overestimation by local investigators of 17.5% in 33 trials. However, this was limited to phase II trials and did not distinguish between double-blind and open-label trials.¹⁴ Two studies, which included 22 and 23 open-label trials, found no evidence of overestimation.^{11 12} The most recent published review included 38 and 33 open-label trials in the meta-analysis of PFS and ORR, respectively, and found that local investigators overestimated PFS but not ORR, which is consistent with the findings of our study.¹⁵ In comparison to these meta-epidemiological studies, we specifically focused on open-label trials and included a large number of reports using the updated search strategy. Our results show that local investigators slightly overestimated PFS by 5%, but not did not overestimate ORR. While there were some discrepancies in the

results between the studies; however, the impact of the discrepancies between central and local adjudications is not substantial.

Mechanisms and implications

There are many possible reasons for discrepancies between central reviewers and local investigators in the adjudication of tumour outcomes.²⁶ Local investigators often lack formal training in radiology, and adjudication may be influenced by knowledge of a patient's clinical status. Central reviewers help ensure consistency in data collection and adjudication across sites, reducing the potential for measurement error and bias. In this study, we found detection bias in the estimation of the HR of PFS but not in the estimation of the OR of ORR. This could be because PFS is a more subjective outcome than ORR. ORR may be less prone to bias, as it is usually defined according to imaging evaluation using manuals such as RECIST, making it more objective. Although there was bias in HR of PFS, the bias was not marked and may not have significantly impacted the estimation and interpretation of the effect. This suggests that central reviewers are not necessary in oncological open-label trials.

However, it is important to note that methodological biases may have distorted the true values and affected these results. In other words, the results represent three possibilities: (1) low risk of detection bias, (2) underestimation of detection bias due to selective outcome reporting bias and (3) underestimation of detection bias due to the use of central reviewers. (1) If there is no methodological bias in this study, these results are true values and there is low risk of detection bias in oncological open-label trials. (2) If selective outcome reporting is present in this study (ie, the analysed studies that reported results from both assessors tended to show smaller differences between the two reviewers than studies that reported results from only one assessor), these results might underestimate the detection bias. (3) If the central adjudicators' judgement was influenced by the local investigators,²⁷ or the influence of informative censoring is higher than that assumed,²⁸ the difference between the central and local adjudications would be reduced, resulting in an underestimation of the detection bias. Informative censoring occurs when patients whose progression is judged by local investigators but not by central reviewers are treated as censored. Previous studies, including this study, have only analysed openly available data, so it is difficult to verify these possibilities. To address this challenge, all studies with central adjudication should also report the results of the local investigators, and a meta-analysis of the ratios should be performed. We expect that further epidemiological studies will accumulate over time, enabling our findings to be investigated.

Conclusions

This meta-epidemiological study found that compared with the findings of central reviewers, local investigators may slightly overestimate the PFS, but not the ORR, in oncological open-label oncological trials. These findings suggest that detection bias of local investigators may not be substantial in the field of oncology. However, this analysis did not extract data from all identified trials and thus may not reflect true detection bias in oncological trials. Further studies that overcome this limitation are necessary before conclusions can be drawn.

Author affiliations

¹Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine and Faculty of Medicine / School of Public Health, Kyoto, Japan

²Department of Preventive Medicine and Public Health, Keio University School of Medicine, Tokyo, Japan

³Department of Internal Medicine, Kyoto Min-iren Asukai Hospital, Kyoto, Japan

⁴Section of Clinical Epidemiology, Department of Community Medicine, Kyoto University Graduate School of Medicine and Faculty of Medicine, Kyoto, Japan

⁵Department of Healthcare Epidemiology, Kyoto University Graduate School of Medicine and Faculty of Medicine / School of Public Health, Kyoto, Japan

⁶Scientific Research Works Peer Support Group (SRWS-PSG), Osaka, Japan

⁷Department of Surgery, Kyoto University Graduate School of Medicine and Faculty of Medicine, Kyoto, Japan

⁸Department of Surgery, Osaka Red Cross Hospital, Osaka, Japan

⁹Department of Surgery, Okinawa Chubu Hospital, Okinawa, Japan

¹⁰Human Health Science, Kyoto University Graduate School of Medicine and Faculty of Medicine, Kyoto, Japan

¹¹Department of Neurology, Emergency Medicine, Nagoya City University East Medical Center, Nagoya, Japan

¹²Department of Preventive Services, Kyoto University Graduate School of Medicine and Faculty of Medicine / School of Public Health, Kyoto, Japan

¹³Department of Gastroenterology and Gastrointestinal Oncology, National Cancer Center-Hospital East, Kashiwa, Japan

¹⁴Translational Research Support Section, National Cancer Center Hospital East, Kashiwa, Japan

¹⁵Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine and Faculty of Medicine, Kyoto, Japan

¹⁶Department of Industrial Engineering and Economics, Tokyo Institute of Technology, Tokyo, Japan

Twitter Satoshi Funada @funada_satoshi, Yan Luo @yluo06 and Yuki Kataoka @YukiKataoka3

Acknowledgements We thank Yasushi Tsujimoto for his valuable comments on this protocol and Makoto Ueno for responding to our query. We acknowledge the National Institute of Public Health for the support of the language editing/ article publishing charge.

Contributors Conceptualisation, data curation, formal analysis, investigation, methodology, project administration, resources, software, visualisation, writing—original Draft: SF. Conceptualisation, investigation, methodology, resources, writing—review and editing: YL, YK. Investigation, resources, writing—review and editing: TY, YF, SY, MK, MT, NN. Conceptualisation, writing—review and editing, supervision: YN, RU. Conceptualisation, writing—review and editing: KU. Conceptualisation, methodology, writing—review and editing, supervision: TAF. SF is the guarantor and accepts full responsibility for the work and the conduct of the study, had access to the data and controlled the decision to publish.

Funding This study was funded by the Pfizer Health Research Foundation, and a Kyoto University School of Public Health Crossover Award.

Competing interests SF received a research grant from Japan Society for the Promotion of Science KAKENHI (grant number JP 20K18964) and the Pfizer Health Research Foundation. YL received a Grant-in-Aid for JSPS Fellows (grant number 21J15050) for research. YK received a research grant from the Systematic Review Workshop Peer Support Group, the Japan Osteoporosis Foundation, and the Yasuda Memorial Medical Foundation for other research purposes. TY received a research

grant from JSPS KAKENHI (grant number JP 21K17228) and the National Cancer Center Research Grant (grant number 2022-A-25) outside this work. NN was supported by a grant-in-aid for multicentre clinical research from Japanese Association for Acute Medicine. YN reported research funding from Taiho Pharmaceutical, Chugai Pharmaceutical, Guardant Health, Genomedia, Daiichi Sankyo and Seagen. KU reported personal fees from Sumitomo Pharma for other than the submitted work. RU reported personal fees from Eisai, Sawai Pharmaceutical and EPS Corporation for other than the submitted work, and a research grant from JSPS KAKENHI (grant number JP 20H04147 and 21KK0205). TAF reported grants and personal fees from Mitsubishi-Tanabe, personal fees from MSD, grants and personal fees from Shionogi, outside the submitted work; TAF has a patent pending (2020-548587) concerning a smartphone CBT app, and intellectual properties for Kokoro-app licensed to Mitsubishi-Tanabe.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. The collected data and codes used for our analysis are provided in online supplemental materials (https://github.com/SatoshiFunada/2023_detection_bias).

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iDs

Satoshi Funada <http://orcid.org/0000-0002-8925-2348>

Yan Luo <http://orcid.org/0000-0002-5271-5126>

Yuki Kataoka <http://orcid.org/0000-0001-7982-5213>

Toshi A Furukawa <http://orcid.org/0000-0003-2159-3776>

References

- Higgins JPT, Altman DG, Gotzsche PC, *et al*. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
- Hróbjartsson A, Thomsen ASS, Emanuelsson F, *et al*. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ* 2012;344:e1119.

- 3 Hróbjartsson A, Thomsen ASS, Emanuelsson F, *et al.* Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. *CMAJ* 2013;185:E201–11.
- 4 Hróbjartsson A, Thomsen ASS, Emanuelsson F, *et al.* Observer bias in randomized clinical trials with time-to-event outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *Int J Epidemiol* 2014;43:937–48.
- 5 Ndounga Diakou LA, Trinquart L, Hróbjartsson A, *et al.* Comparison of central adjudication of outcomes and onsite outcome assessment on treatment effect estimates. *Cochrane Database Syst Rev* 2016;3:MR000043.
- 6 Moustgaard H, Clayton GL, Jones HE, *et al.* Impact of blinding on estimated treatment effects in randomised clinical trials: meta-epidemiological study. *BMJ* 2020;368:l6802.
- 7 Hirsch BR, Califf RM, Cheng SK, *et al.* Characteristics of oncology clinical trials: insights from a systematic analysis of Clinicaltrials.Gov. *JAMA Intern Med* 2013;173:972–9.
- 8 Chen EY, Haslam A, Prasad V. FDA acceptance of surrogate end points for cancer drug approval: 1992–2019. *JAMA Intern Med* 2020;180:912–4.
- 9 U.S. Food and Drug Administration. Clinical trial endpoints for the approval of cancer drugs and Biologics. 2018. Available: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-endpoints-approval-cancer-drugs-and-biologics>
- 10 European Medicines Agency. Appendix 1 to the guideline on the evaluation of anticancer medicinal products in man. Methodological consideration for using progression-free survival (PFS) or disease-free survival (DFS) in Confirmatory trials. 2012. Available: https://www.ema.europa.eu/en/documents/scientific-guideline/appendix-1-guideline-evaluation-anticancer-medicinal-products-man-methodological-consideration-using_en.pdf
- 11 Zhang J, Zhang Y, Tang S, *et al.* 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;5:481.
- 12 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.
- 13 Dello Russo C, Cappoli N, Navarra P. A comparison between the assessments of progression-free survival by local investigators versus blinded independent central reviews in phase III oncology trials. *Eur J Clin Pharmacol* 2020;76:1083–92.
- 14 Dello Russo C, Cappoli N, Pilunni D, *et al.* Local investigators significantly overestimate overall response rates compared to blinded independent central reviews in phase 2 oncology trials. *J Clin Pharmacol* 2021;61:810–9.
- 15 Lian Q, Fredrickson J, Boudier K, *et al.* Meta-analysis of 49 Roche oncology trials comparing blinded independent central review (BICR) and local evaluation to assess the value of BICR. *Oncologist* 2023;oyad012.
- 16 Funada S, Luo Y, Kataoka Y, *et al.* Detection bias in open-label trials of cancer drug: a meta-epidemiological study. 2021.
- 17 Funada S, Luo Y, Kataoka Y, *et al.* Inadequate reporting of adjudicators in open-label trials of anticancer drugs between 2017 and 2021: a methodological review. *J Clin Epidemiol* 2022;150:80–9.
- 18 Page MJ, McKenzie JE, Bossuyt PM, *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev* 2021;10:89.
- 19 Nussbaumer-Streit B, Klerings I, Dobrescu AI, *et al.* Excluding non-English publications from evidence-syntheses did not change conclusions: a meta-epidemiological study. *J Clin Epidemiol* 2020;118:42–54.
- 20 Therasse P, Arbus SG, Eisenhauer EA, *et al.* New guidelines to evaluate the response to treatment in solid tumors. *J Natl Cancer Inst* 2000;92:205–16.
- 21 Eisenhauer EA, Therasse P, Bogaerts J, *et al.* New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009;45:228–47.
- 22 Miller AB, Hoogstraten B, Staquet M, *et al.* Reporting results of cancer treatment. *Cancer* 1981;47:207–14.
- 23 Altman DG, Bland JM. How to obtain the confidence interval from a P value. *BMJ* 2011;343:d2090.
- 24 Sterne JAC, Jüni P, Schulz KF, *et al.* Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-Epidemiological' research. *Stat Med* 2002;21:1513–24.
- 25 Viechtbauer W. 'Package Metafor' 2022. Available: <https://www.metafor-project.org/doku.php/metafor>
- 26 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.
- 27 Raunig D, Goldmacher G, Conklin J. Local evaluation and blinded central review comparison: a victim of meta-analysis shortcomings. *Ther Innov Regul Sci* 2013;47:1–2.
- 28 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.

Supplementary Appendix

We used the search terms to search the MEDLINE via PubMed (1 Jan 2010 to 30 June 2021) as follows:

#1 "neoplasms"[MeSH Terms] OR "cancer*"[Title/Abstract] OR "carcinoma*"[Title/Abstract] OR "oncol*"[Title/Abstract] OR "tumor*"[Title/Abstract] OR "tumour*"[Title/Abstract] OR "adenocarcinoma*"[Title/Abstract] NOT Leukemia[MeSH Terms] NOT Lymphoma[MeSH Terms] NOT "Multiple Myeloma" [MeSH Terms]

#2 "random*"[Title/Abstract]

#3 "Antineoplastic Agents"[MeSH Terms] OR "immunotherapy"[All Fields] OR "Immune Checkpoint Inhibitors"[All Fields] OR "target therapy"[All Fields] OR "hormone"[All Fields] OR "chemotherapy"[All Fields]

#4 "humans"[MeSH Terms]

#5 "english"[Language]

#6 randomizedcontrolledtrial[Filter]

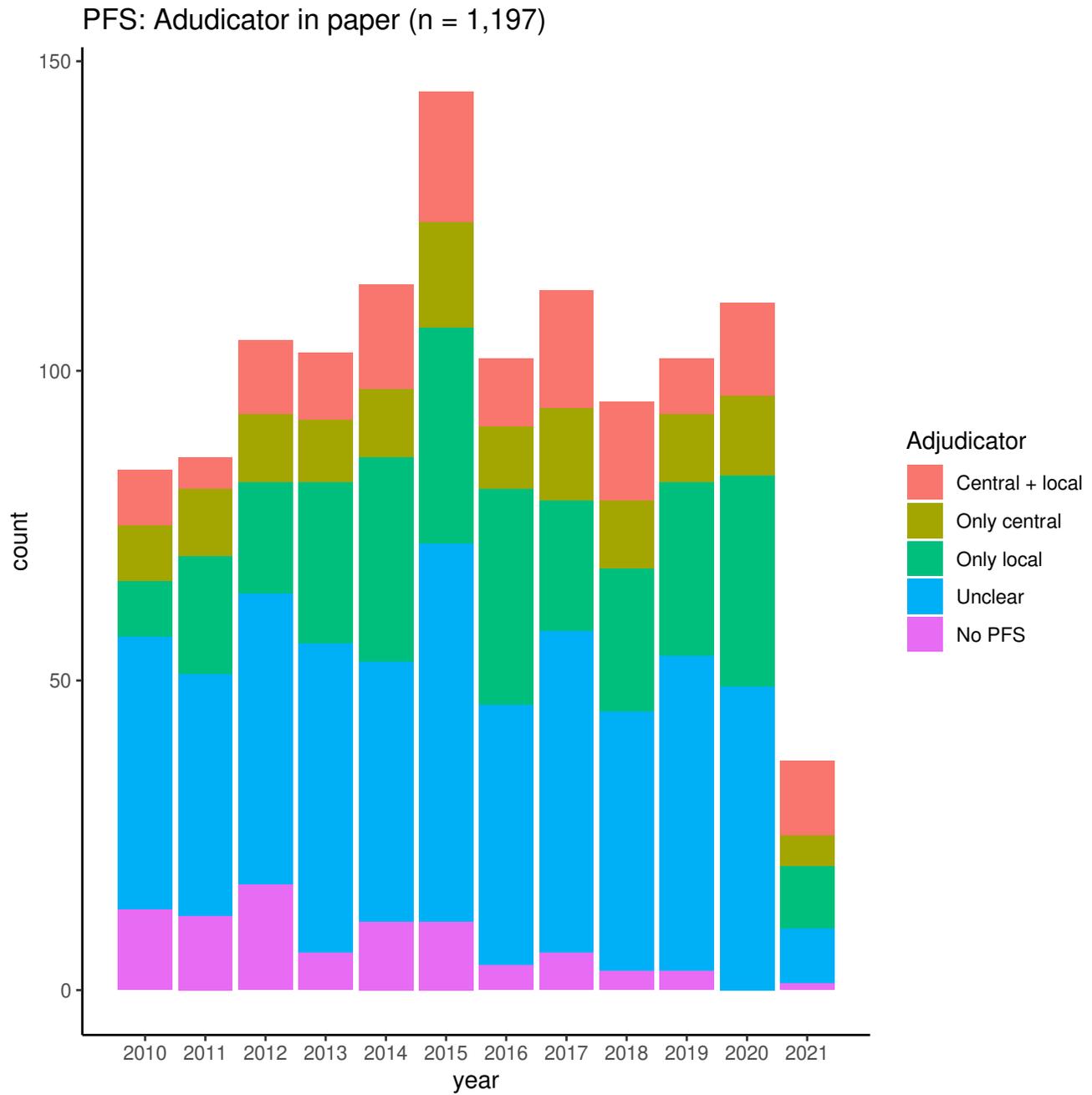
#7 2010/01/01:2021/06/30[Date - Publication]

#8 adjuvant[Title] OR neoadjuvant[Title]

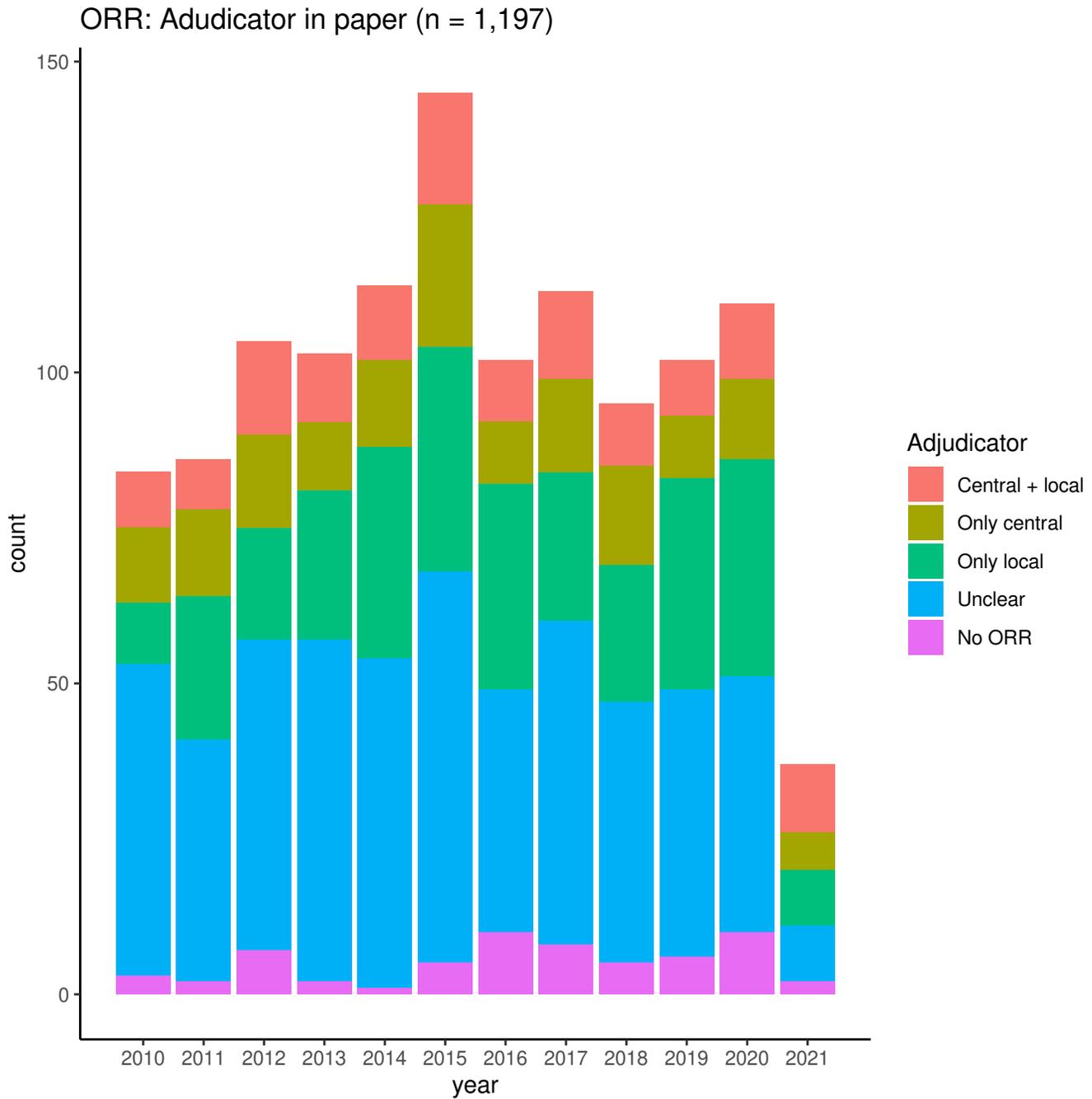
#9 "placebo"[Title] OR "blind"[Title]

#1 AND #2 AND #3 AND #4 AND #5 AND #6 AND #7 NOT #8 NOT #9

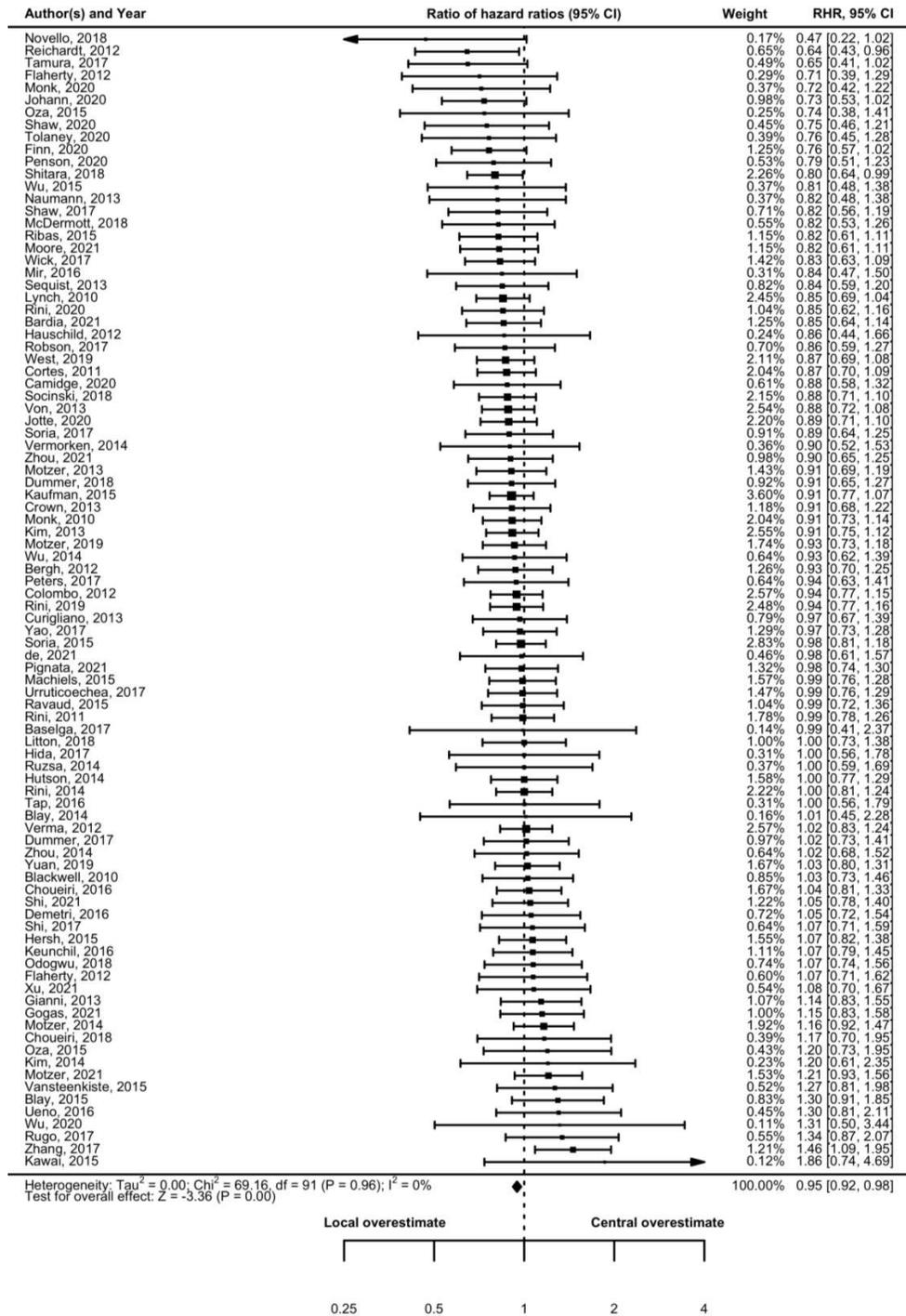
Supplementary Figure 2. The trend in PFS adjudicators from 2010 to 2021



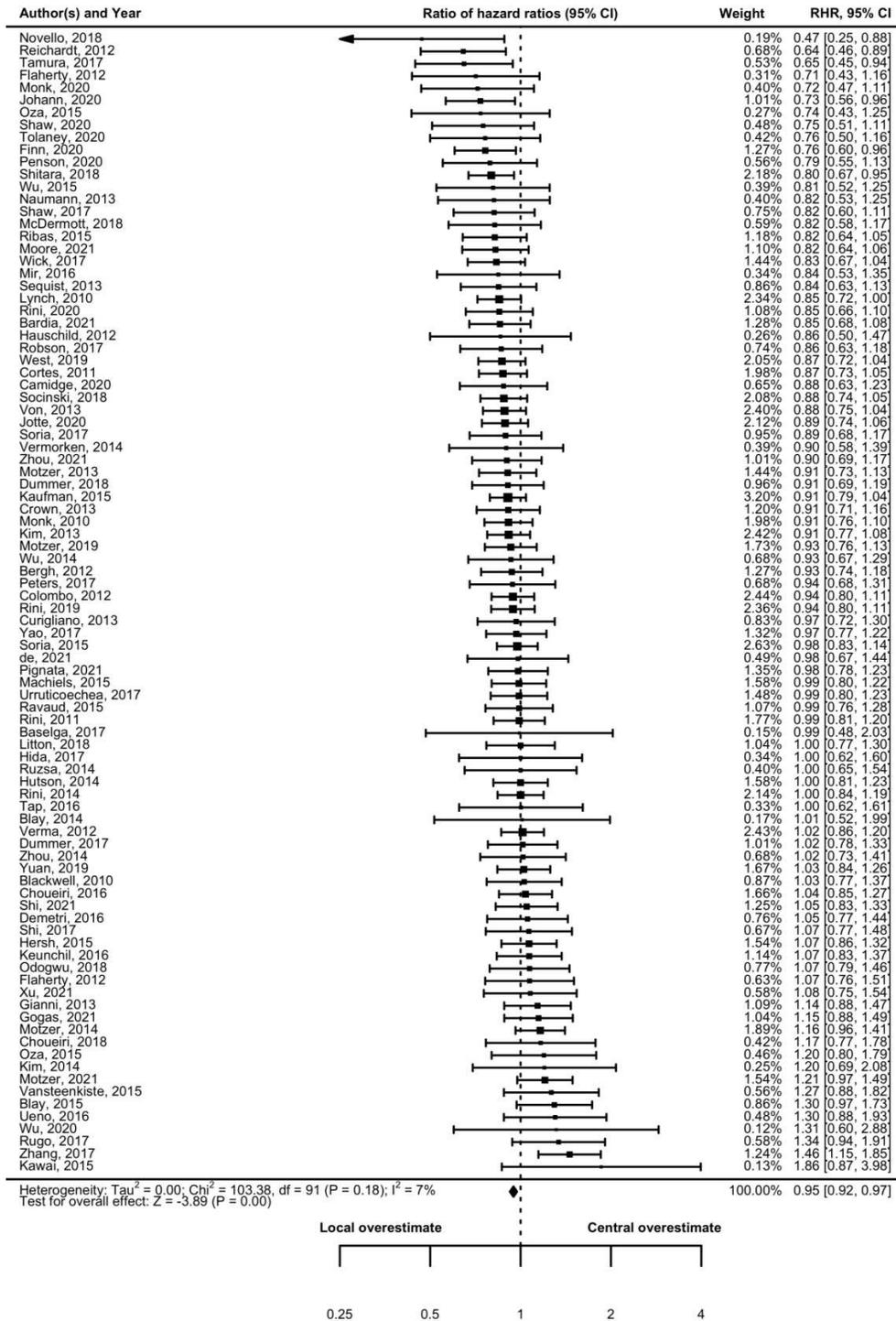
Supplementary Figure 3. The trend in ORR adjudicators from 2010 to 2021



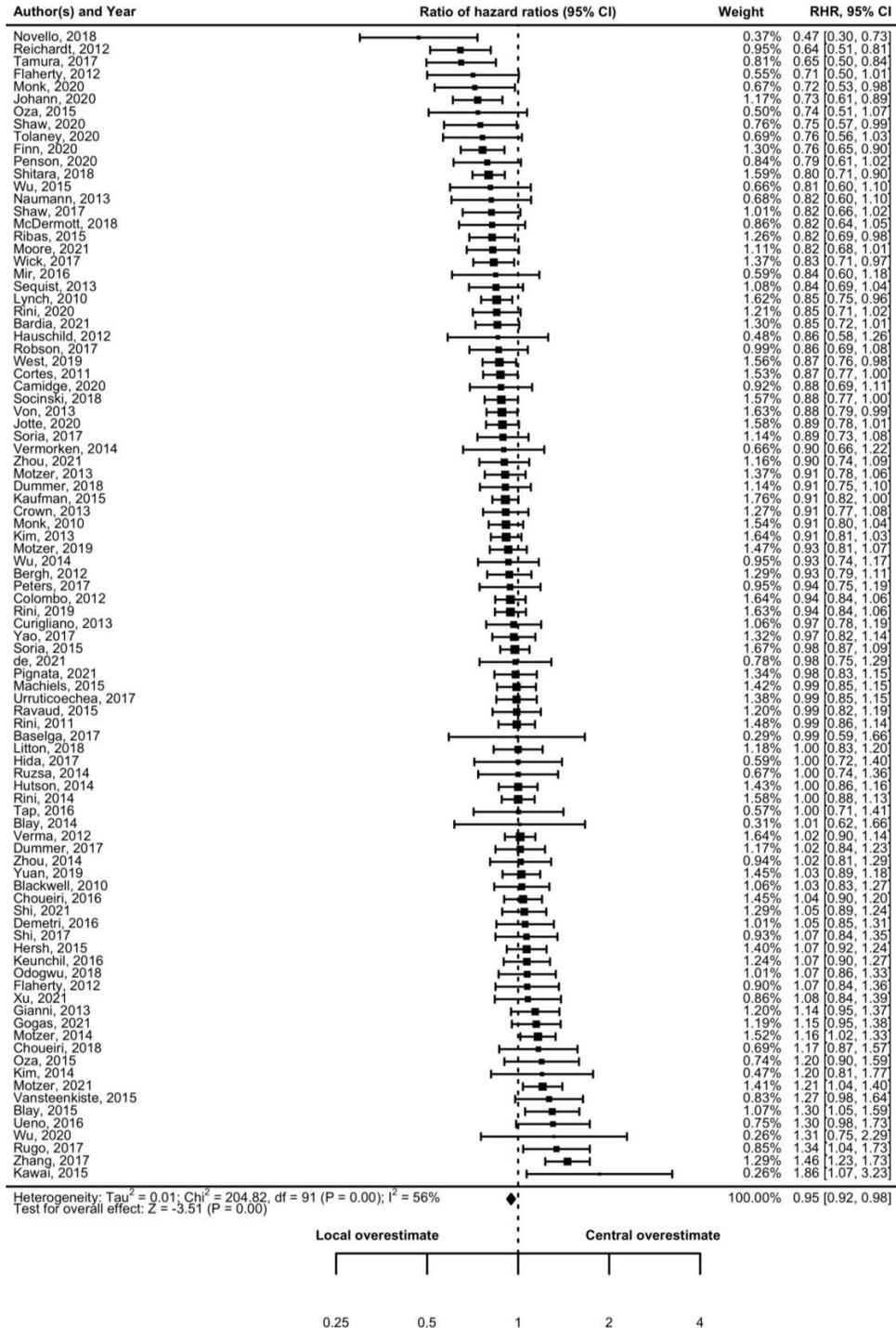
Supplementary Figure 4a. Comparison of treatment effect estimates (Hazard Ratio) between central reviewers and local investigators. Sensitivity analysis ($\rho = 0.25$).



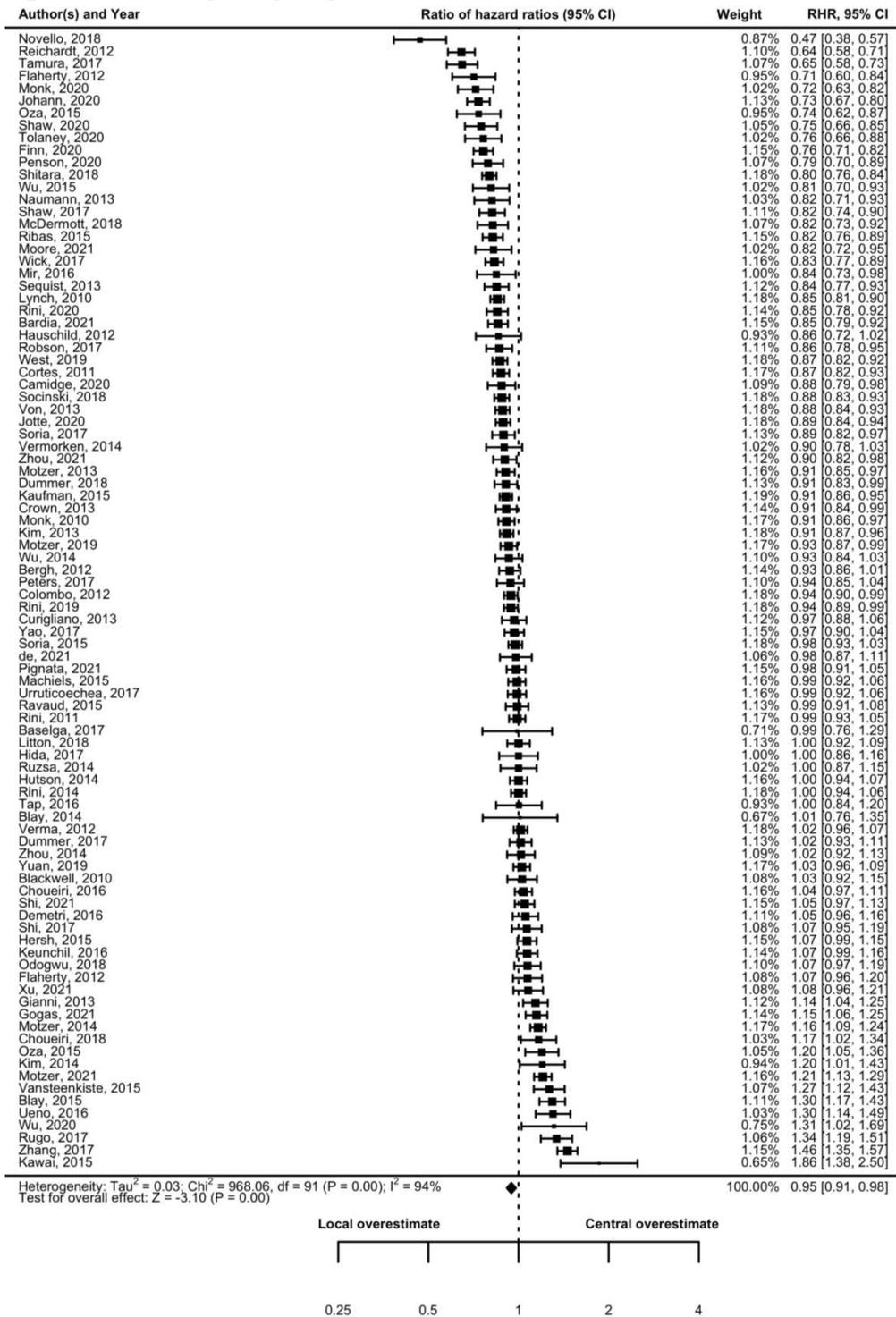
Supplementary Figure 4b. Comparison of treatment effect estimates (Hazard Ratio) between central reviewers and local investigators. Sensitivity analysis ($\rho = 0.50$).



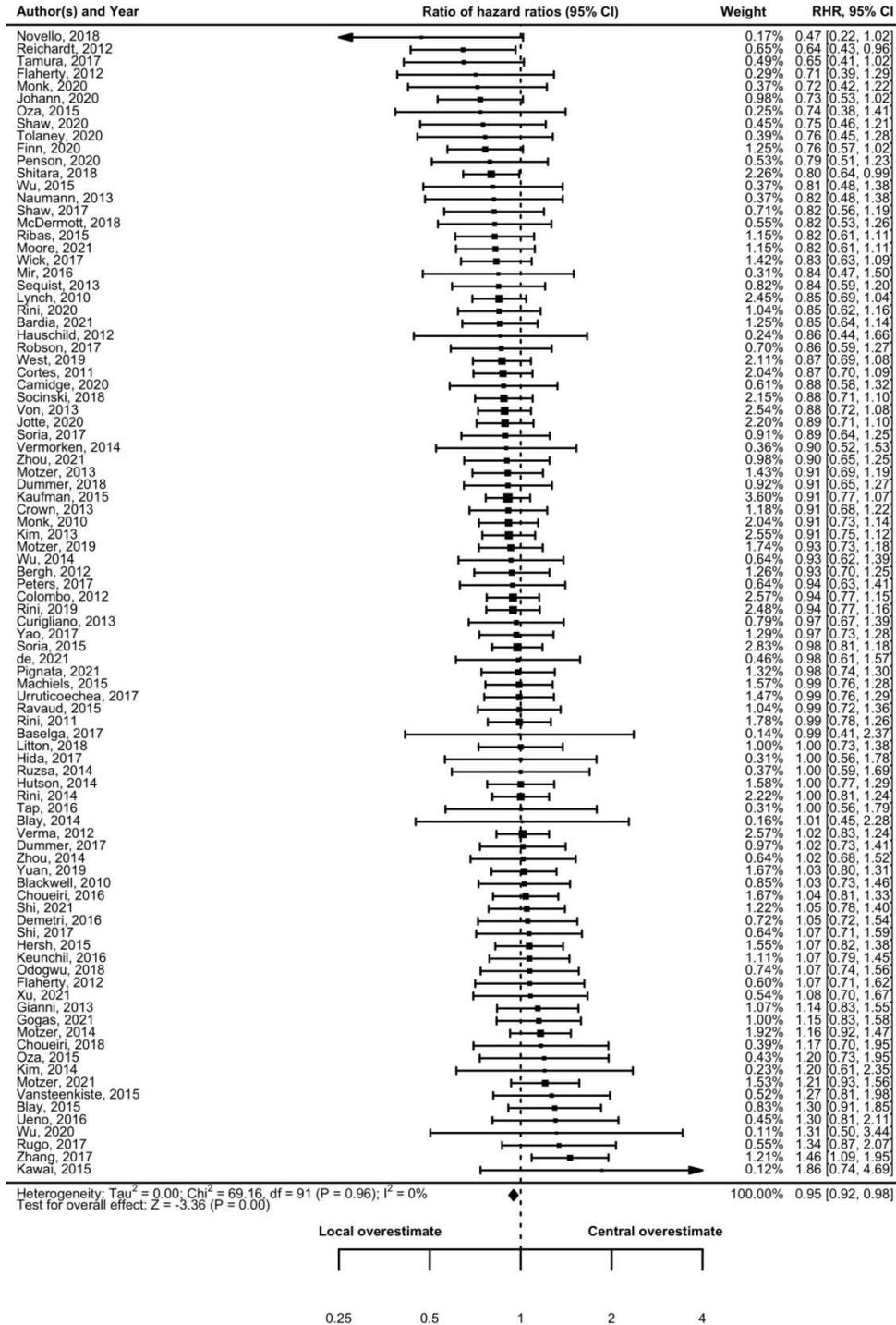
Supplementary Figure 4c. Comparison of treatment effect estimates (Hazard Ratio) between central reviewers and local investigators. Sensitivity analysis ($\rho = 0.75$).



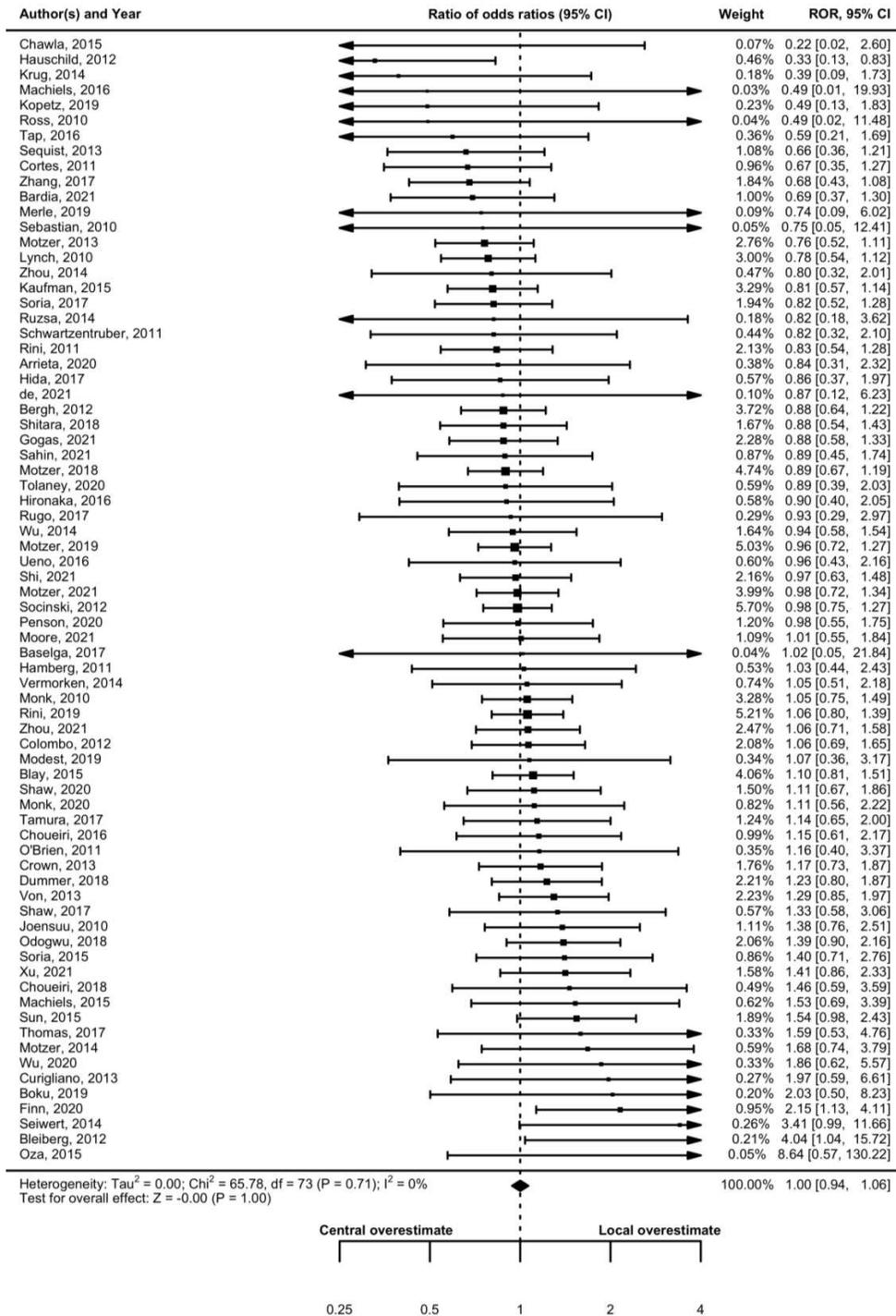
Supplementary Figure 4d. Comparison of treatment effect estimates (Hazard Ratio) between central reviewers and local investigators. Sensitivity analysis ($\rho = 0.95$).



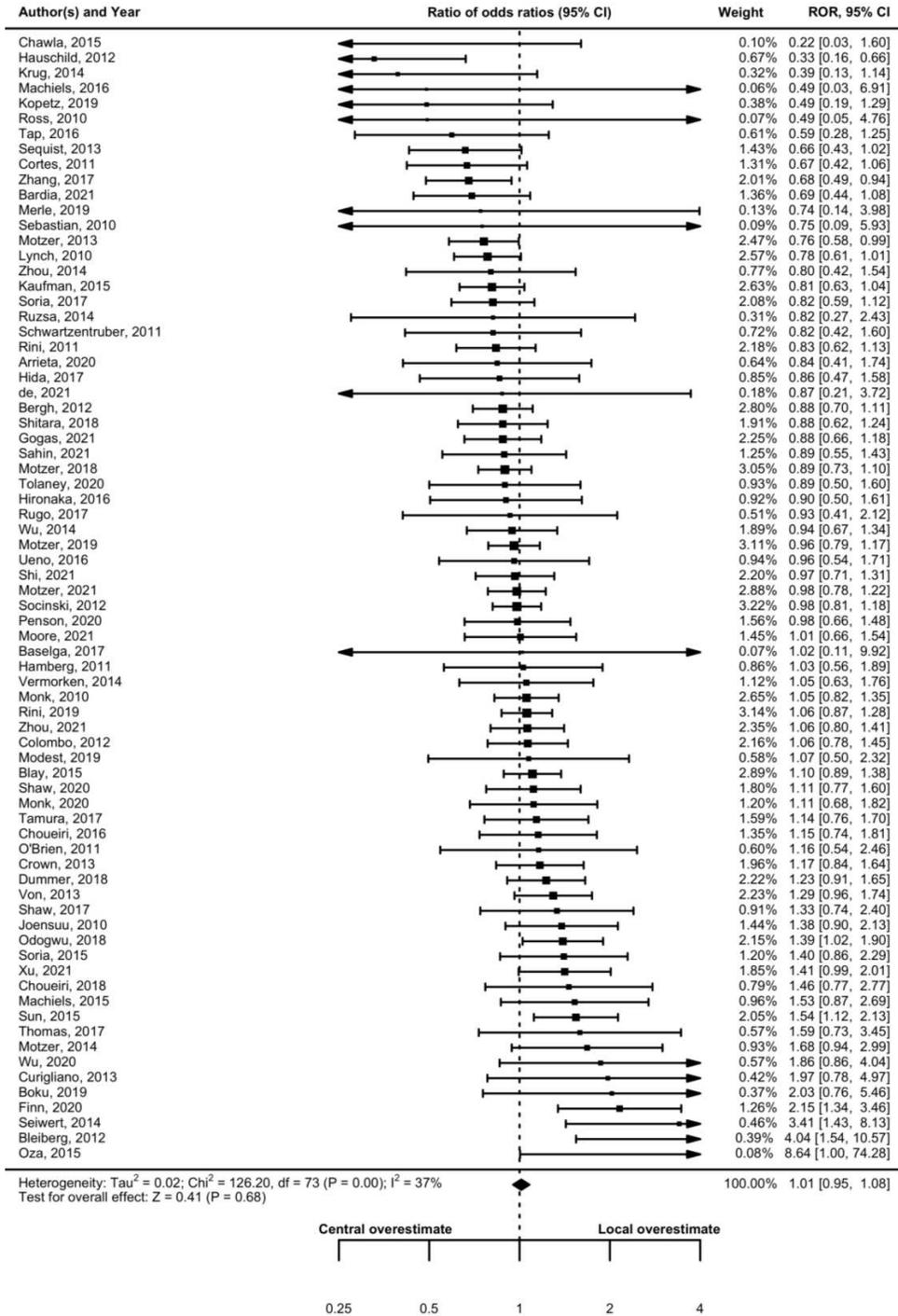
Supplementary Figure 5a. Comparison of treatment effect estimates (Odds Ratio) between central reviewers and local investigators. Sensitivity analysis ($\rho = 0.25$).



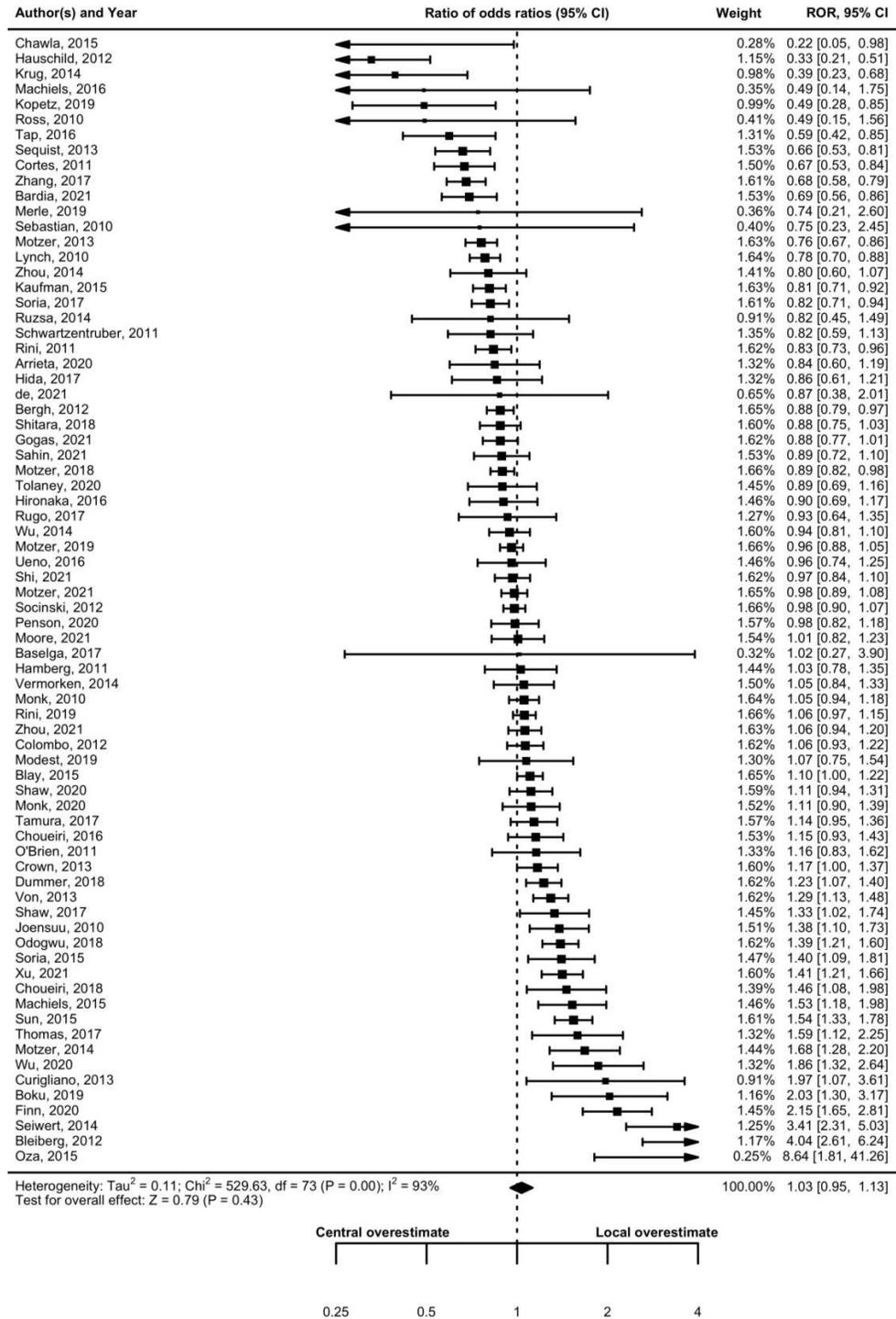
Supplementary Figure 5b. Comparison of treatment effect estimates (Odds Ratio) between central reviewers and local investigators. Sensitivity analysis ($\rho = 0.50$).



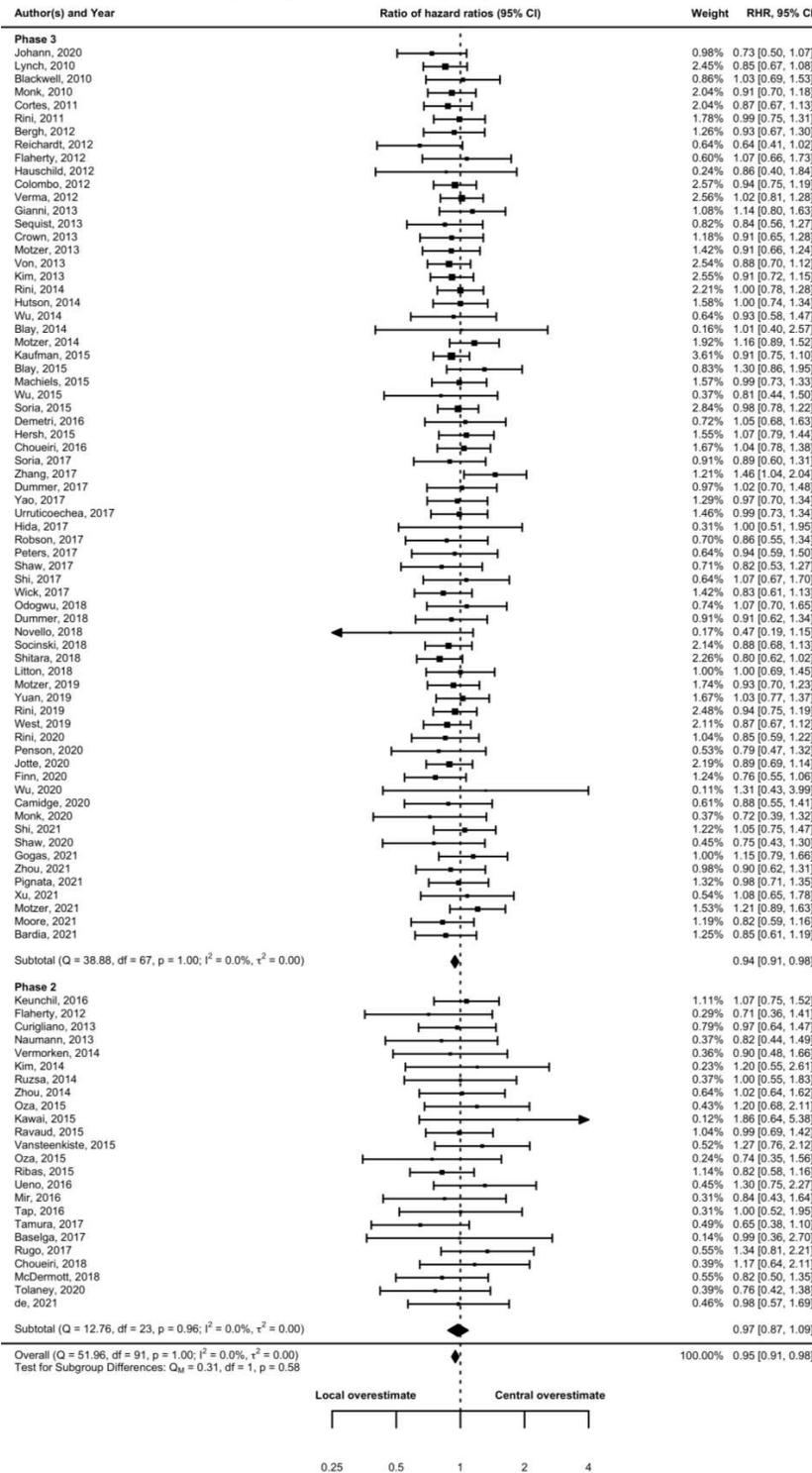
Supplementary Figure 5c. Comparison of treatment effect estimates (Odds Ratio) between central reviewers and local investigators. Sensitivity analysis ($\rho = 0.75$).



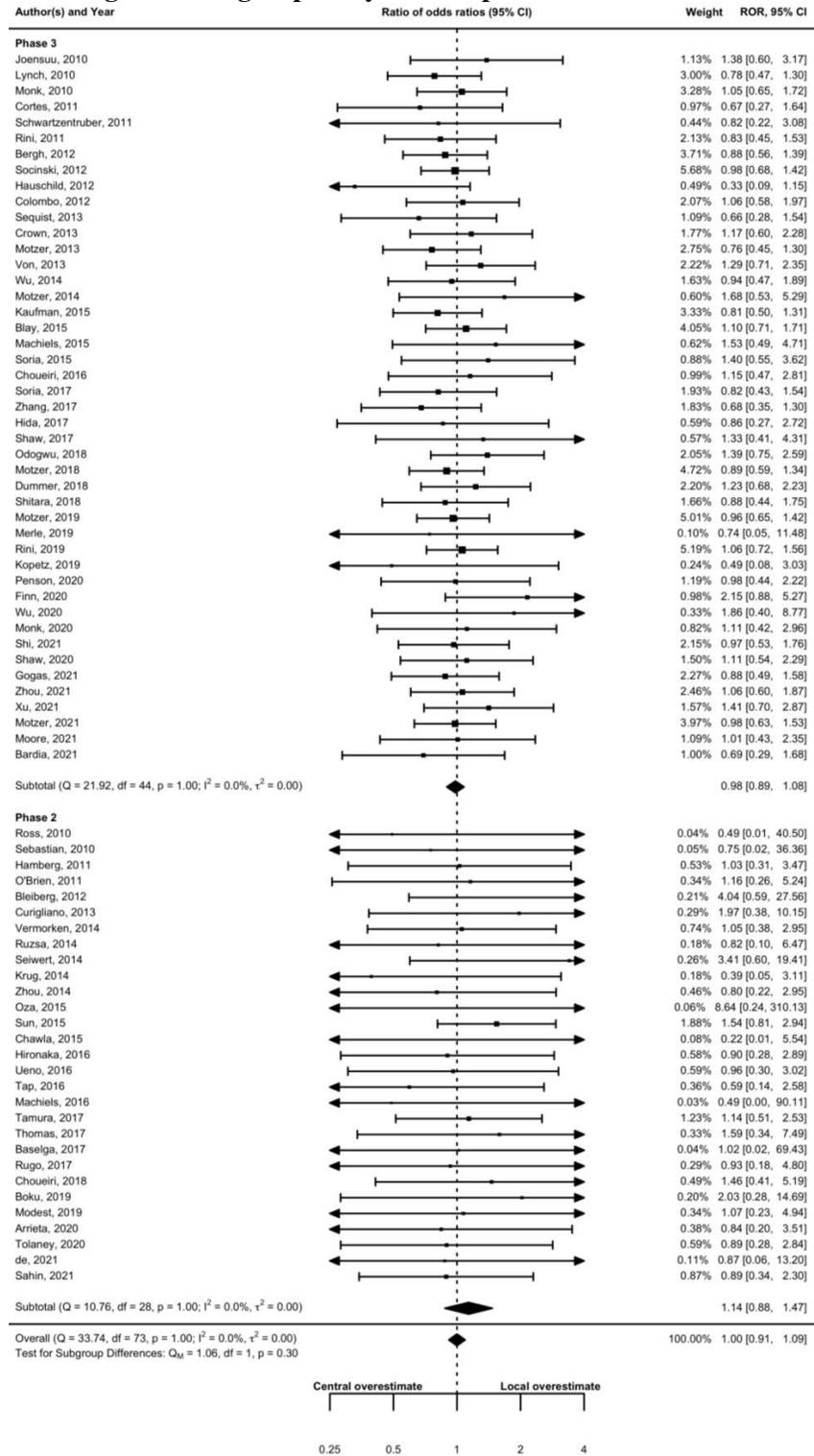
Supplementary Figure 5d. Comparison of treatment effect estimates (Odds Ratio) between central reviewers and local investigators. Sensitivity analysis ($\rho = 0.95$).



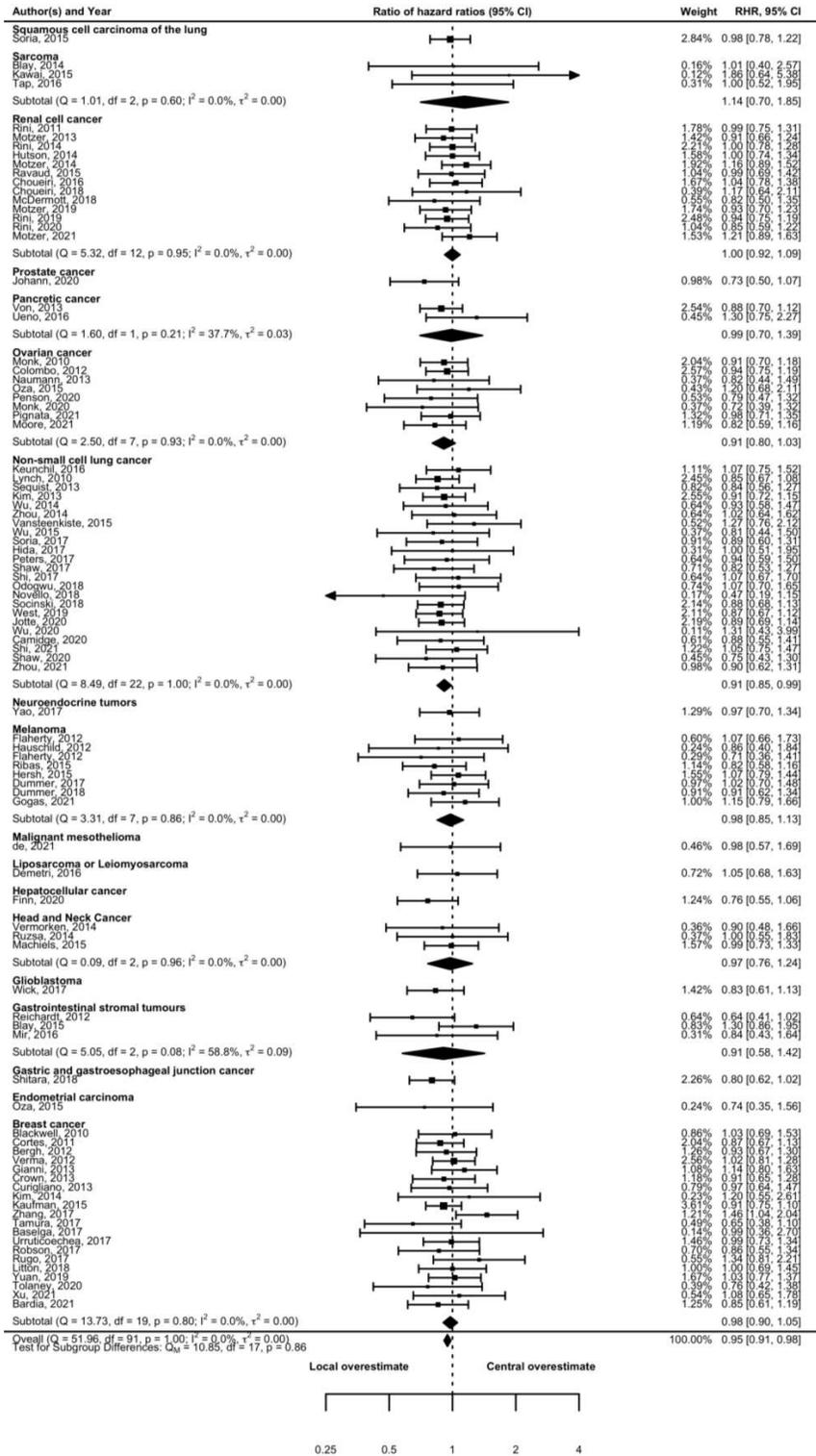
Supplementary Figure 6a. Comparison of treatment effect estimates (Hazard Ratio) between central reviewers and local investigators. Subgroup analysis: trial phase.



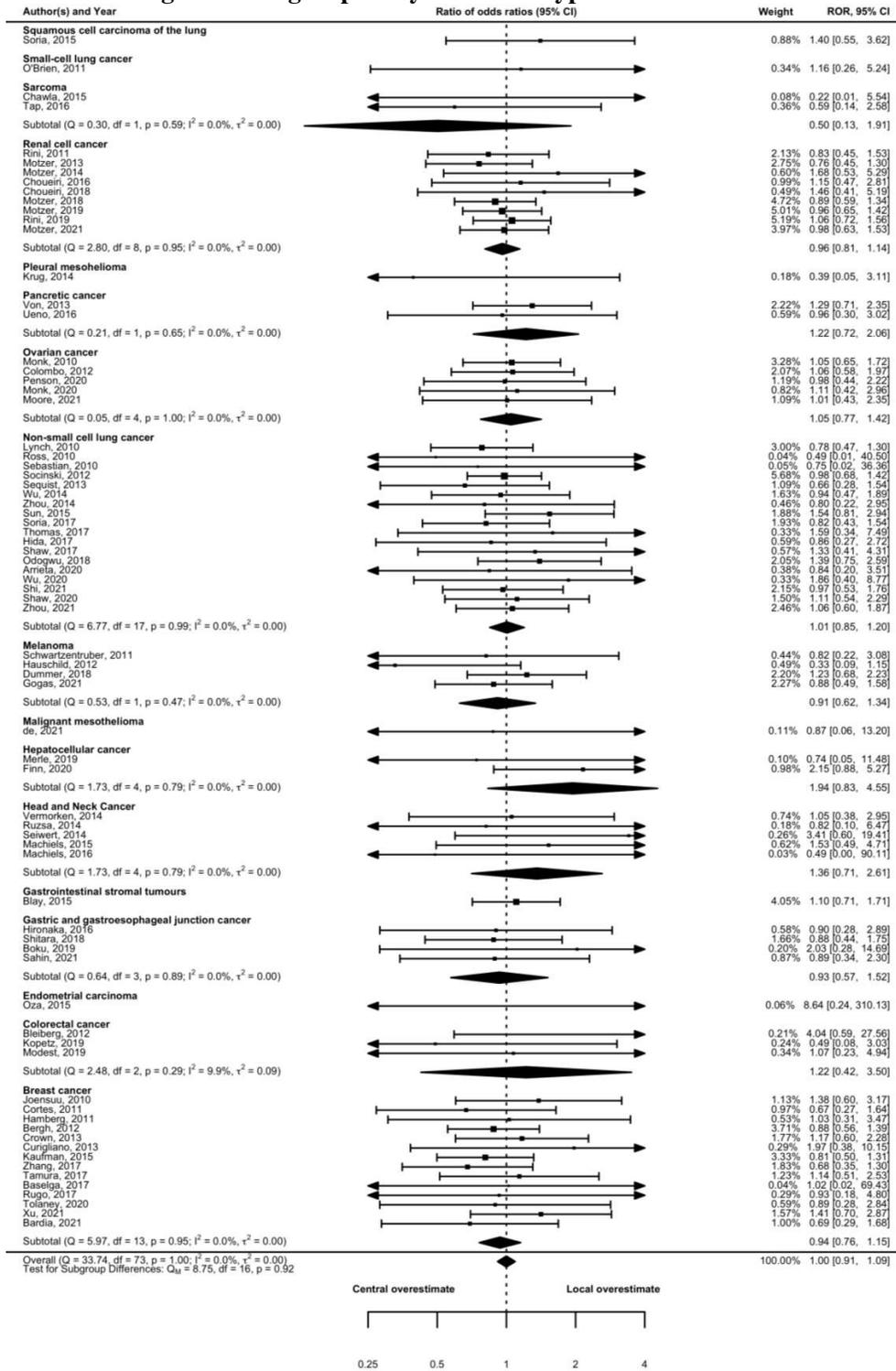
Supplementary Figure 6b. Comparison of treatment effect estimates (Odds Ratio) between central reviewers and local investigators. Subgroup analysis: trial phase.



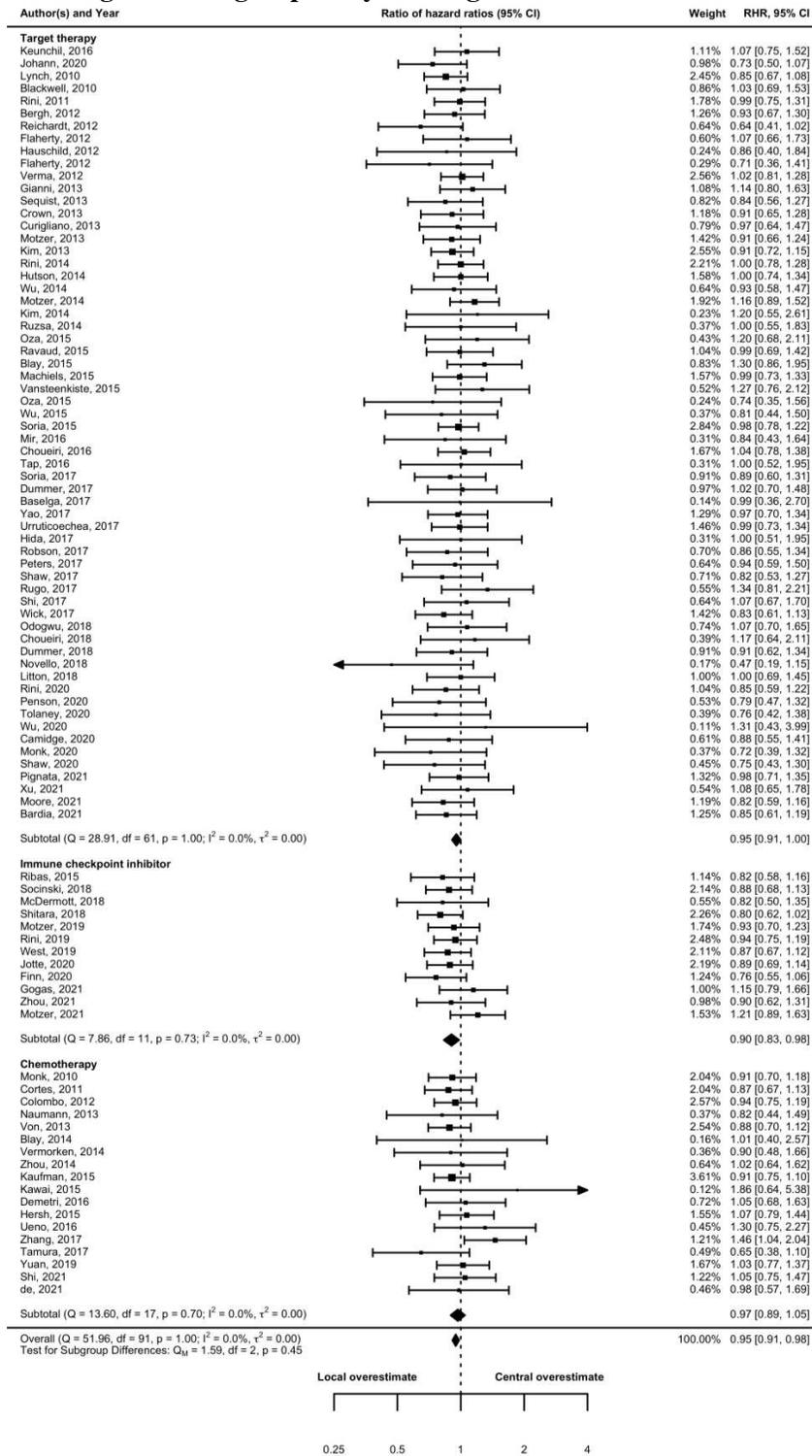
Supplementary Figure 7a. Comparison of treatment effect estimates (Hazard Ratio) between central reviewers and local investigators. Subgroup analysis: cancer type.



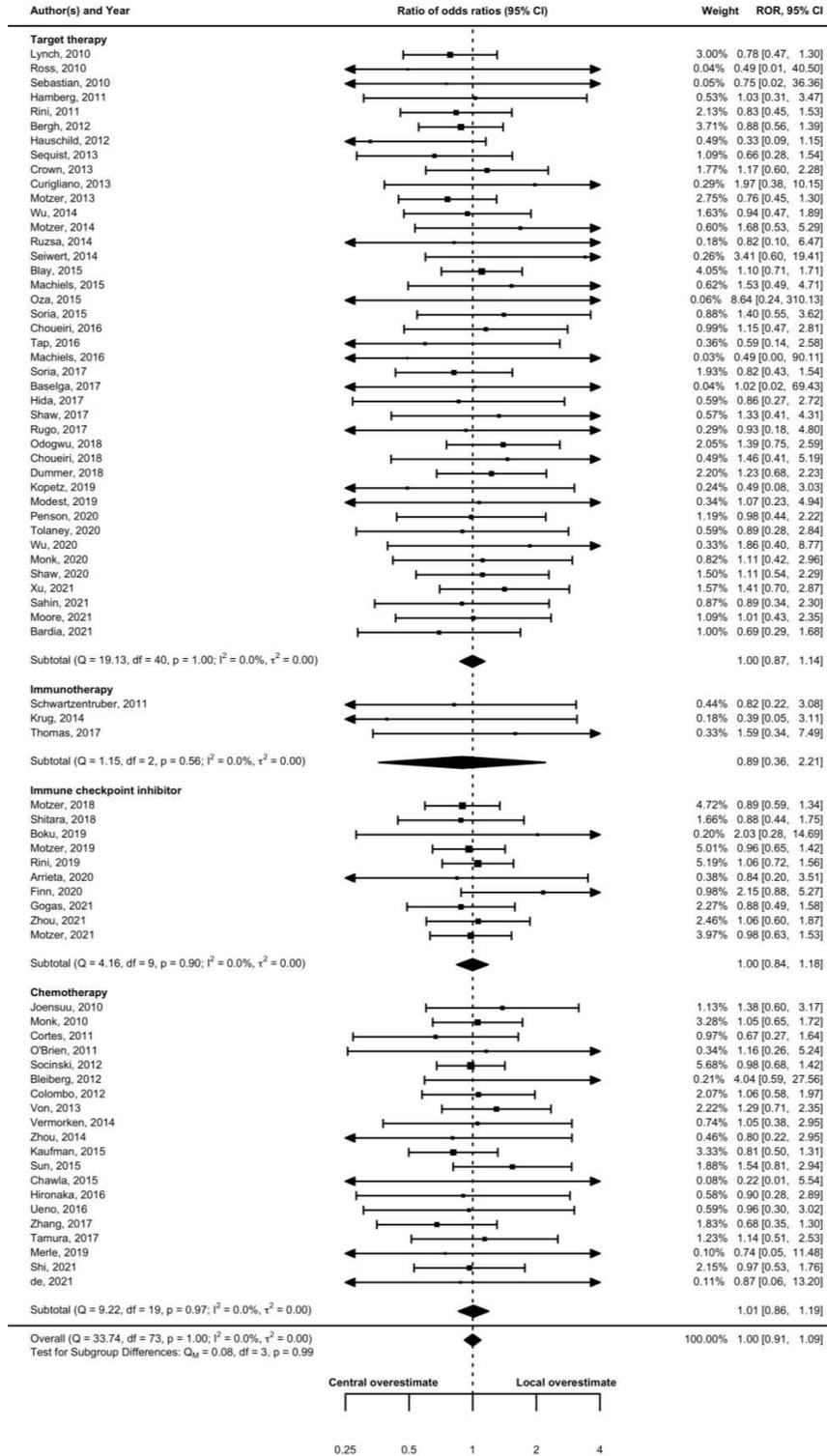
Supplementary Figure 7b. Comparison of treatment effect estimates (Odds Ratio) between central reviewers and local investigators. Subgroup analysis: cancer type.



Supplementary Figure 8a. Comparison of treatment effect estimates (Hazard Ratio) between central reviewers and local investigators. Subgroup analysis: drug classification.



Supplementary Figure 8b. Comparison of treatment effect estimates (Odds Ratio) between central reviewers and local investigators. Subgroup analysis: drug classification.



Supplementary Table 1. Adjudicators of open-label trials of anticancer drugs (n = 1,197)

PFS	
Central + local	157 (13.1)
Only central	134 (11.2)
Only local	291 (24.3)
Unclear	528 (44.1)
No PFS	87 (7.3)
ORR	
Central + local	141 (11.8)
Only central	158 (13.2)
Only local	301 (25.2)
Unclear	536 (44.8)
No ORR	61 (5.1)

Abbreviations: ORR, objective response rate; PFS, progression-free survival.

Supplementary Table 2. Effect of correlation between central and local adjudications (sensitivity analysis).

PFS					
ρ	RHR (95%CI)	Tau ²	I ²	P value for heterogeneity	
0	0.95 (0.91 to 0.98)	0.00	0%	>0.99	
0.25	0.95 (0.92 to 0.98)	0.00	0%	0.96	
0.50	0.95 (0.92 to 0.97)	0.00	7.0%	0.18	
0.75	0.95 (0.92 to 0.98)	0.01	56.0%	< .0001	
0.95	0.95 (0.91 to 0.98)	0.03	94.0%	< .0001	
ORR					
ρ	ROR (95%CI)	Tau ²	I ²	P value for heterogeneity	
0	1.00 (0.91 to 1.09)	0.00	0%	>0.99	
0.25	1.00 (0.93 to 1.08)	0.00	0%	>0.99	
0.50	1.00 (0.94 to 1.06)	0.00	0%	0.71	
0.75	1.01 (0.95 to 1.08)	0.02	37.0%	< .0001	
0.95	1.03 (0.95 to 1.13)	0.11	93.0%	< .0001	

Abbreviations: ORR, objective response rate; PFS, progression-free survival; RHR, ratios of hazard ratio; ROR, ratios of odds ratio.

A ρ is correlation coefficients between central and local adjudications in each trial.

A ρ of 0 indicates no dependency and a ρ of 0.95 indicates almost complete dependency.