European Journal of Cancer 121 (2019) 19–28

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

Comparison of treatment effect from randomised controlled phase II trials and subsequent phase III trials using identical regimens in the same treatment setting

Fei Liang [a,*,1], Zhenyu Wu [b,1], Miao Mo [a], Changming Zhou [a], Jie Shen [a], Zezhou Wang [a], Ying Zheng [a,1]

[a] *Clinical Statistic Center, Fudan University Shanghai Cancer Center, Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, 200032, China*
[b] *Department of Biostatistics, School of Public Health, Key Laboratory of Public Health Safety and Collaborative Innovation Center of Social Risks Governance in Health, Fudan University, Shanghai, 200032, China*

Received 25 June 2019; received in revised form 7 August 2019; accepted 9 August 2019
Available online 14 September 2019

KEYWORDS
Treatment effect;
Phase II trials;
Phase III trials;
Randomised controlled trials

Abstract ***Purpose:*** We aimed to determine whether treatment effect size differed between randomised controlled phase II trials and subsequent phase III trials and examine potential predictor of positive phase III trials.
Methods: We searched MEDLINE for randomised controlled phase II studies published from January 2006 to December 2015. Matched phase III trials that investigated same intervention in the same setting of the same cancer were identified through Web of Science, ClinicalTrials.gov and conference proceedings. For each pair of phase II and phase III trials, we extracted hazard ratios (HRs) with 95% confidence intervals (CIs) for both overall survival (OS) and progression-free survival (PFS) and evaluated the differences by ratio of HRs (rHRs): the HR for phase II trial to that for phase III trial. A summary rHR was obtained through a random-effect meta-analysis. Univariable analyses were conducted to identify predictors of positive phase III trials.
Results: We identified 57 pairs of phase II and phase III trials. Compared with phase III trials, treatment effect sizes of PFS were, on average, 26% larger in phase II trials (rHR = 0.74, $P < 0.001$, 95% CI: 0.68–0.80). Treatment effect sizes of OS were 27% greater in phase II trials

* *Corresponding author*: Clinical Statistic Center, Fudan University Shanghai Cancer Center; Department of Oncology, Shanghai Medical College, Fudan University, 270 Dongan Road, 200032, Shanghai, China. Fax: +86 21 67820385.
E-mail address: liangfei0726@163.com (F. Liang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ejca.2019.08.006

than in phase III trials (rHR = 0.73, P < 0.001, 95% CI: 0.66–0.79). Fifteen (26.3%) phase III trials were positive, and the only predictor of positive phase III trials was positive phase II trials

Conclusion: Treatment effects in randomised controlled phase II trials were greater than those in matched phase III trials. Caution must be taken when interpreting promising results from randomised controlled phase II trials.

1. Introduction

Development progress from phase I to phase II to phase III studies is a well-established paradigm of developing new antitumour agents or regimens in cancer therapeutics. Phase II trials are undertaken to determine whether a novel drug (or combination) is promising enough to justify a definitive phase III study for efficacy, often referred to as making a go/no-go decision. Phase II trials in oncology have typically been single-arm trials with the response rate as a primary end-point. Single-arm trials, being simple and easily executed, with success based on comparison with historical controls, are generally inadequate to provide robust preliminary data for multiple reasons: considerable selection bias, variability in the historical control response rate established for comparison, patient heterogeneity and changes in patient outcomes based on supportive care, quality of diagnostic staging techniques, or other standards of care over time [1].

Simulation studies suggest that randomised phase II trials would have lower error rates and greater predictive power for phase III results [2,3]. It has been proposed that drug development could be made more efficient through greater use of randomised phase II trials [4]. Although theoretical superiority of randomised phase II trials has been postulated, a recent review of the perspectives on the use of randomisation in phase II oncology trials pointed out that no evidence showed that the use of randomised designs in phase II has improved outcomes in phase III [5]. In fact, several studies found that randomised phase II studies were not superior to single-arm phase II trials at predicting phase III study success [6,7], which introduced concerns about the value of randomised phase II trials which are more complex and requiring up to fourfold more patients as compared with a similar single-arm trials.

One possible explanation for the high failure rate in the phase II to phase III transition is a systematic overestimate of treatment effects in phase II trials. A previous study showed that response rates in most phase III oncology studies are lower than those in preceding single-arm phase II studies, with a mean difference of 12.9% [8]. Another study investigated the influence of trial sample size on treatment effect estimates of trials assessing therapeutic interventions with binary outcomes and found that compared with the 25% largest trials, treatment effects were, on average, 32% larger in trials in the 25% smallest trials [9]. But it is still unclear whether, or to what extent, treatment effect is overestimated in randomised phase II trials than in subsequent phase III trials.

Our study aimed to determine whether significant difference in treatment effect existed between randomised controlled phase II trials and subsequent matched phase III trials and examine which characteristics of randomised phase II trials might be the predictor of positive phase III trials.

2. Methods

2.1. Identification of randomised controlled phase II trials and matched phase III trials

We searched MEDLINE for all randomised controlled phase II studies published from January 2006 to December 2015. Detailed search strategy was provided in the appendix. Twelve major journals—where oncology randomised controlled trials (RCTs) are usually published—nine oncology journals (*The Lancet Oncology, the Journal of Clinical Oncology, JAMA Oncology, the Journal of the National Cancer Institute, the Annals of Oncology, Clinical Cancer Research, the European Journal of Cancer, the British Journal of Cancer and Cancer)* and three general medical journals *(The New England Journal of Medicine, Lancet and JAMA*) were also hand searched for any possible omission.

Randomised phase II trials that included a control group (placebo, best supportive care or active standard treatment) and designed to evaluate the efficacy and/or safety of a systemic cancer drug or combination therapy in solid tumours were eligible. Retrospective trials, phase I or phase III trials, non-randomised phase II trials, randomised phase II trials without a control group and randomised phase II trials of surgery, radiotherapy, life style intervention and supportive care were all excluded.

Results of each identified phase II trial were determined to be either ‘positive’ or ‘negative’. A study was

considered to be positive if the stated primary end-points were met. For positive phase II trials or negative phase II trials that authors clearly stated that further phase III studies were needed, matched phase III trials were searched.

To ensure the comparable of treatment effect between phase II trials and phase III trials, we used the PICO (population, intervention, comparator and outcome) principle to define matched phase III trials. Eligible matched phase III trial was defined as randomised phase III trials that investigated the same intervention using the same control in the same population (same treatment setting of the same cancer type) and reported at least one efficacy end-point (primary or non-primary) that also reported in preceding phase II trials. Dose modification of the experimental drug in phase III trials was allowed. For trials investigating the efficacy of one experimental drug plus standard chemotherapy, replacing one cytotoxic drug with similar cytotoxic drug (for example, carboplatin replaced cisplatin) was also allowed.

For each identified randomised controlled phase II trials, matched phase III trials were searched in the following order: (1) Screening all articles in the database of Web of Science that cited this phase II trials. (2) Clincaltrials.gov was searched using drug name and cancer type as a keyword with filters of 'With Results' and 'Phase 3'. (3) American Society of Clinical Oncology(ASCO) and European Society for Medical Oncology (ESMO) meeting abstracts were searched using drug name, cancer type and phase III as keywords. A cut-off date of December 31, 2018, was used for retrieving matched phase III publications.

2.2. Data extraction

For both phase II and phase III trials, hazard ratio (HR) and the associated 95% confidence intervals (CIs) of overall survival (OS) and other time-to-event outcomes (progression-free survival [PFS] for trials in the metastatic setting and disease-free survival [DFS] in the adjuvant setting) were extracted. When these were unavailable, we used log-rank p-values and the number of reported events to calculate the ln(HR) following established methods [10]. If necessary, HRs were also computed based on the published survival curves [11]. The following study characteristics were also recorded for each phase II trial: sample size, year of publication, cancer type, treatment regimen, study results (positive vs negative), primary end-point, treatment setting, funding source, trial design (comparative vs non-comparative), patient selection (all-comer vs biomarker-selected) and basis for recommending phase III trials (primary analyses vs subgroup analyses). Results, sample size and primary outcome of phase III trials were also extracted. Two authors (F.L. and Z.W.) screened trials independently for eligibility and extracted data from each included trial. Any discrepancies were resolved by consensus.

2.3. Statistical analysis

We evaluated the differences in treatment effect size between phase II trials and matched phase III trials by using ratio of HR (rHR), which has been used in previous studies [12,13]. rHR referred to the ratio of HR in the phase II trial to that in the phase III trial. All HRs were standardised so that an HR < 1.0 indicates benefit of intervention; therefore, an rHR <1.0 indicated that the treatment effect size was greater in phase II trials than in matched phase III trials. CIs (95%) of rHR were estimated by the methods used by Tan et al. [12]. For each pair of matched phase II and phase III trials, rHR of OS (HR−OS−phase II trial/HR−OS−phase III trial) and rHR of PFS or rHR of DFS were calculated. We then estimated a combined rHR and 95% CI across matched pair of trials by combining the individual rHR using a random-effect meta-analysis model weighted by variance of rHR. Heterogeneity across rHRs was assessed using the I^2 statistics, the Cochran Q χ^2 test and the between-pair variance τ.

We planned two sensitivity analyses to test the robustness of the estimation of the combined rHR: combining the rHRs using a fixed-effect meta-analysis model, excluding matched pairs with modification of drug dose or background chemotherapy.

The following subgroup analyses according to characteristics of phase II trials were performed: treatment mode (combination therapy vs monotherapy), study results (positive vs negative), double-blind design (yes vs no), patient selection (all-comer vs biomarker-selected), first-line setting (yes vs no) and basis for recommending phase III (primary analyses vs subgroup analyses).

Association of positive phase III trials and characteristics of phase II trials was estimated using χ^2 test or Fisher exact test for univariable analyses. Multivariable analyses included variables associated with positive phase III trials in the univariable analysis with $P < 0.10$.

3. Results

3.1. Characteristics of the included phase II and phase III randomised controlled trials

We identified 3262 unique articles published from January 2005 to December 2015 through search in PubMed. Of these, 367 randomised controlled phase II trials met the study inclusion criteria. Eligible phase III trials were identified for 57 trials (Fig. 1). Full list of included 57 pairs of phase II and phase III trials were provided in appendix Table 1.

The characteristics of the included phase II and phase III trials are detailed in Table 1. The most studied cancer

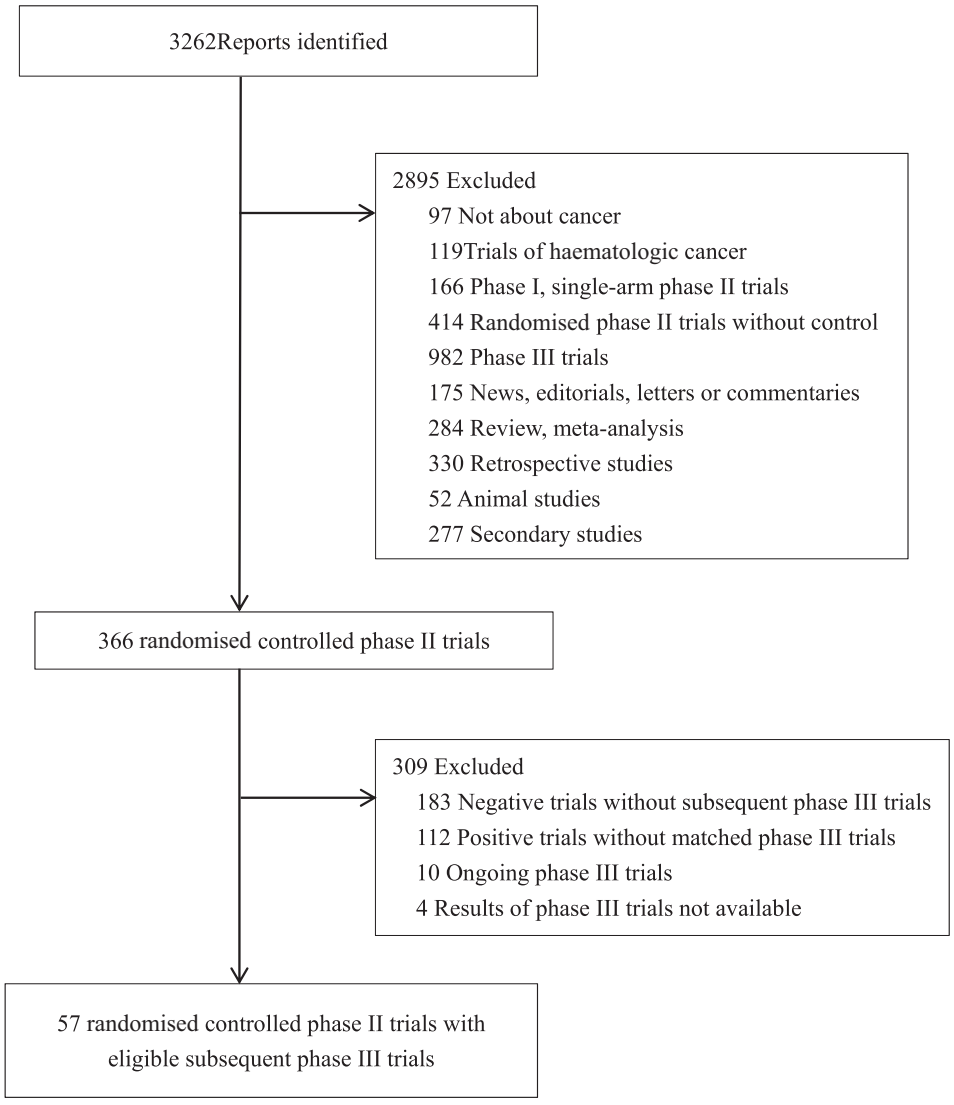

Fig. 1. Study flow chart.

type was lung cancer (21 trials), followed by breast cancer (6 trials). The median sample size of phase II trials was 141 (range, 38–595). Fifty-one (89.5%) phase II trials were at least partially funded by the industry. Among phase II trials, 44 (75.9%) were comparative trial and 44 (75.9%) reported positive results. Forty-six (80.7%) of phase III trials were reported as full articles, 4 (7.0%) trials posted these results on the ClinicalTrials.gov without full publications and 7 (12.3%) trials were reported as meeting abstracts. Only 15 (26.3%) of the phase III trials were positive (Table 2).

3.2. Treatment comparison between phase II trials and subsequent phase III trials

HRs of PFS were reported or estimable for 45 pairs of phase II and phase III trials. For 41 (91.1%) of matched trials, treatment effect size of PFS was larger in phase II trial than in phase III trial and only 4 (8.9%) phase III trials reported larger treatment effect size of PFS than that in preceding phase II trials (Fig. 2A). Compared with phase III trials, treatment effects of PFS were, on average, 26% larger in phase II trials (rHR = 0.74, $P < 0.001$, 95% CI: 0.68–0.80, Fig. 3A). Heterogeneity across trials were moderate ($I^2 = 16\%$, $P = 0.18$, $\tau^2 = 0.01$).

HRs of OS were reported or estimable for 42 pairs of phase II and phase III trials. For 37 (88.1%) of matched trials, treatment effect size of OS was larger in phase II trial than in phase III trial and only 4 (9.5%) phase III trials reported larger treatment effect size of OS than that in preceding phase II trials (Fig. 2B). Treatment effects of OS were 27% greater in phase II trials than in phase III trials (rHR = 0.73, $P < 0.001$, 95% CI: 0.66–0.79, Fig. 3B). Heterogeneity across trials were moderate ($I^2 = 17\%$, $P = 0.16$, $\tau^2 = 0.01$).

Table 1
Characteristics of included phase II and phase III trials (N = 57).

Characteristics	No. Of trials (%)
Phase II trials	
Cancer type	
Lung	21 (36.8)
Breast	6 (10.5)
Pancreatic	4 (7.0)
Prostate	5 (8.8)
Ovarian	4 (7.0)
Melanoma	3 (5.3)
Gastric	3 (5.3)
Colorectal	2 (3.5)
Other	9 (15.8)
Year of publication	
2006–2010	23 (40.4)
2011–2015	34 (59.6)
Treatment type	
Chemotherapy	10 (17.5)
Targeted therapy	40 (70.2)
Hormone therapy	1 (1.8)
Immunotherapy	6 (10.5)
Study result	
Positive	43 (75.4)
Negative	14 (24.6)
Sample size, median (range)	141 (38–595)
Treatment mode	
Monotherapy	17 (29.8)
Combination therapy	40 (70.2)
Primary end-point	
OS	9 (15.8)
PFS or TTP	31 (54.4)
DFS	1 (1.8)
Response rate	13 (22.8)
Other	3 (5.3)
Funding source	
Industry	51 (89.5)
Non-industry	3 (5.3)
Not reported	3 (5.3)
Treatment setting	
(Neo)Adjuvant	2 (3.5)
Non-curative	55 (96.5)
Patient selection	
All-comer	9 (15.8)
Biomarker-selected	48 (84.2)
Double-blinded	
Yes	31 (54.4)
No	26 (45.6)
Comparative trials	
Yes	44 (75.9)
No	13 (24.1)
Basis for recommending phase III trials	
Primary analyses	49 (86.0)
Subgroup analyses	8 (14.0)
Phase III trials	
Form of publication	
Full articles	46 (80.7)
Meeting abstracts	7 (12.3)
Results posted on registry	4 (7.0)
Sample size, median (range)	609 (107–2312)
Results	
Positive	15 (26.3)
Negative	42 (73.7)
Primary end-points	
OS	34 (59.6)
PFS	21 (36.8)
DFS	1 (1.8)
PCR	1 (1.8)

OS, overall survival; PFS, progression-free survival; TTP, time-to-progression; DFS, disease-free survival; PCR, pathologic complete response; ITT, intention-to-treat.

Only 1 pair of phase II [14] and phase III trial [15] reported results of DFS, with HR of 0.76 (95% CI: 0.48–1.21) and 1.02 (95% CI: 0.89–1.18), respectively.

3.3. Subgroup analyses and sensitivity analyses

In the subgroup analysis, we did not find any significant differences of rHR for both PFS and OS by characteristics of phase II trials except basis for recommending phase III (primary analyses vs subgroup analyses). The difference in treatment effects of PFS between phase II trials and matched phase III trials was significantly greater for those phase III trials that were conducted based on the subgroup analyses of phase II trials (rHR of PFS, 0.53; 95% CI: 0.43–0.66) than for those pairs that phase III trials were conducted based on the primary analyses of phase II trials (rHR of PFS, 0.78; 95% CI: 0.72–0.84), with an interaction P value of 0.001. Subgroup analyses of rHR for OS revealed the same results (0.51 [0.38–0.68] vs 0.76 [0.70–0.83], interaction P = 0.009)(Table 3).

Sensitivity analyses using fixed-effect model or excluding trials with modification of drug dose or background chemotherapy were also consistent with our primary analysis.

3.4. Predictor of positive phase III trials

In the univariable analyses, only positive phase II trials were associated with positive phase III trials. None of the 14 phase III trials based on negative phase II trials met the primary outcome (Table 2). No multivariable analyses were conducted.

4. Discussion

Despite the theoretical advantage of randomised controlled phase II trials over single-arm phase II trials, our study showed that promising results from randomised phase II trials are seldom reproduced in subsequent phase III trials, which echoed the conclusion of the recent review by Grayling *et al.* [5]. The treatment effects in phase II trials were, on average, nearly 30% larger than those in subsequent phase III trials investigating the same intervention in the same treatment setting. The only possible predictor for the success of phase III trials was the positivity of phase II trials.

Table 2
Univariable analysis of phase II study characteristics as predictors of positive phase III results.

Characteristics of phase II trials	Positive phase III trials (%)	Negative phase III trials (%)	P value
Treatment mode			0.316
Monotherapy	6 (35.3)	11 (64.7)	
Combination therapy	9 (22.5)	31 (77.5)	
Funding source			0.918
Industry	13 (25.5)	38 (74.5)	
Non-industry	1 (33.3)	2 (66.7)	
Not reported	1 (33.3)	2 (66.7)	
Patient selection			0.178
All-comer	4 (44.4)	5 (55.6)	
Biomarker-selected	11 (22.9)	37 (77.1)	
Double-blinded			0.924
Yes	8 (25.8)	23 (74.2)	
No	7 (26.9)	19 (73.1)	
Setting			0.674
Adjuvant	0	2 (100.0)	
First-line	7 (25.9)	20 (74.1)	
Second or more	8 (28.6)	20 (71.4)	
Comparative trials			0.308
Yes	13 (29.5)	31 (70.5)	
No	2 (15.4)	11 (84.6)	
Basis for recommending phase III trials			0.339
Primary analyses	14 (28.6)	35 (71.4)	
Subgroup analyses	1 (12.5)	7 (87.5)	
Primary end-point			0.259
Overall survival	1 (12.5)	7 (87.5)	
Others	14 (29.2)	34 (70.8)	
Results			0.010
Positive	15 (34.9)	28 (65.1)	
Negative	0	14 (100.0)	

The relationship between phase II trials and phase III trials has been extensively studied. Zia et al. [8] compared the response rate in single-arm phase II studies and subsequent randomised control studies using identical chemotherapeutic regimens and found that response rates in most phase III studies are lower than those in preceding phase II studies. El-Maraghi and Eisenhauer [16] retrieved reports of single-agent phase II trials in six solid tumours for 19 targeted drugs and found that higher overall response rates were predictive of regulatory approval. Both Chan et al. [17] and Monzon [6] found that randomised phase II studies were not superior to single-arm phase II trials at predicting phase III study success. These previous studies either focused on the relationships between single-arm phase II trials and phase III trials or compared the efficiency of randomised phase II trials and single-arm phase II trials in predicting positive phase III trials. To the best of our knowledge, our study is the first to study the relationship between randomised controlled phase II trials and matched phase III trials and quantified the difference in treatment effects from randomised controlled phase II trials and subsequent phase III trials. Furthermore, owing to the widely known publication bias of RCTs that positive phase III trials are more likely to be published than negative ones [18], previous articles that started searching for published phase III studies and then looked for matching phase II may miss some negative phase III trials that never be published. In contrast, we first looked for phase II trials and then searched for matching phase III trials; thus, we could

Table 3
Subgroup analysis of differences in treatment effect size between phase II and matched phase III trials.

Group	PFS		OS	
	rHR (95% CI)	Interaction P	rHR (95% CI)	Interaction P
Basis for recommending phase III trials		0.001		0.009
Primary analyses	0.78 [0.72, 0.84]		0.76 [0.70, 0.83]	
Subgroup analyses	0.53 [0.43, 0.66]		0.51 [0.38, 0.68]	
Treatment mode		0.74		0.72
Combination therapy	0.73 [0.67, 0.80]		0.73 [0.65, 0.82]	
Monotherapy	0.75 [0.64, 0.88]		0.70 [0.61, 0.81]	
Comparative trials		0.24		0.82
Yes	0.73 [0.66, 0.79]		0.73 [0.66, 0.81]	
No	0.84 [0.67, 1.05]		0.71 [0.58, 0.86]	
Double-blind		0.98		0.17
Yes	0.74 [0.66, 0.82]		0.69 [0.60, 0.78]	
No	0.74 [0.65, 0.84]		0.78 [0.69, 0.89]	
Biomarker-enriched trial		0.52		0.33
Biomarker-selected	0.70 [0.59, 0.82]		0.79 [0.65, 0.96]	
All-comer	0.75 [0.68, 0.82]		0.71 [0.64, 0.79]	
Study result of phase II trials		0.42		0.96
Positive	0.75 [0.68, 0.83]		0.72 [0.64, 0.82]	
Negative	0.70 [0.61, 0.81]		0.73 [0.64, 0.83]	
Line of treatment		0.50		0.21
First-line	0.76 [0.68, 0.85]		0.78 [0.67, 0.90]	
Second or later line	0.72 [0.64, 0.81]		0.69 [0.62, 0.77]	

PFS, progression-free survival; OS, overall survival; rHR, ratio of hazard ratio.

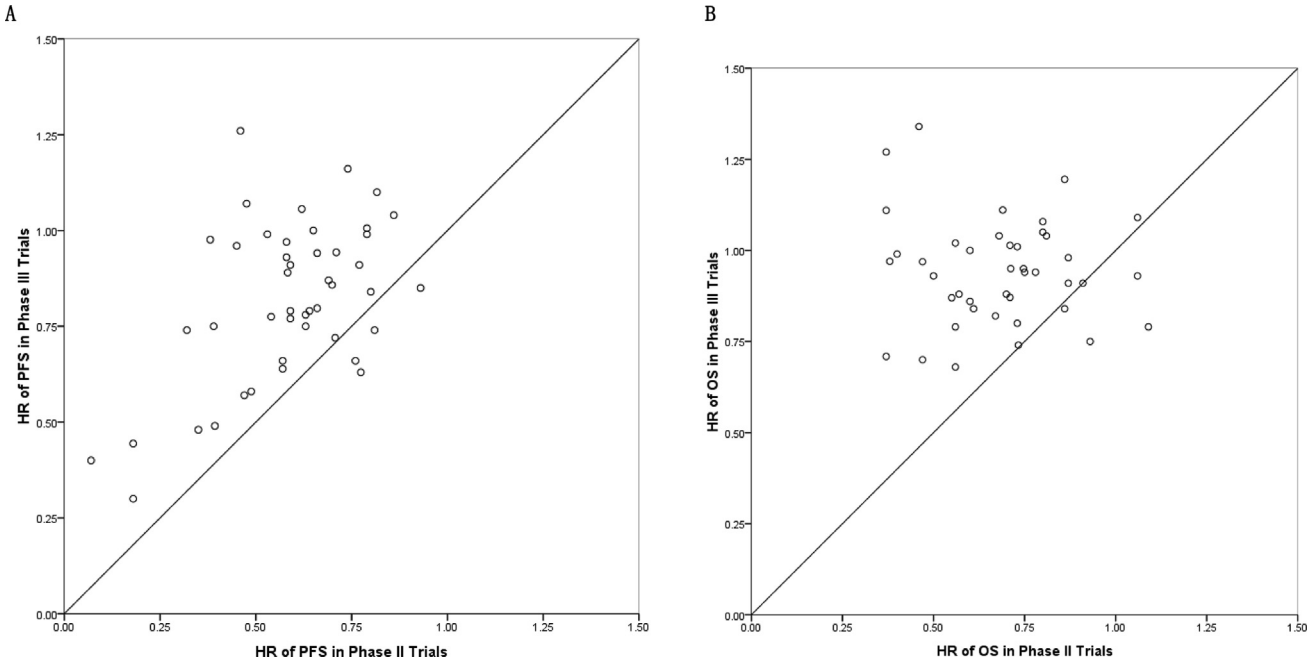

Fig. 2. Comparison of HR of (A) PFS and (B) OS reported in phase II trials and phase III trials. HR, hazard ratio; PFS, progression-free survival; OS, overall survival.

identify some negative phase III trials that have not been published in medical journals. In fact, we identified 11 phase III RCTs (4 trials posted results in ClinicalTrial.gov and 7 published as meeting abstracts) that have not been published as full articles, and 10 of these trials were negative. Our study provided a benchmark to properly evaluate the results from randomised controlled phase II studies to select experimental drugs or combinations that can bring clinically meaningful benefit to patients to be tested in phase III trials. Gan et al. [19] found that investigators consistently make overly optimistic assumptions regarding treatment benefits when designing RCTs. Our study can help investigators to set the appropriate expected treatment effect in sample size calculation in phase III trials to prevent the trials being underpowered.

Several mechanisms could help explain the significant difference in treatment effects between randomised controlled phase II trials and phase III trials. For one particular drug, many randomised phase II trials could be conducted, investigating different doses, schedules or combinations in different types of cancers. By chance, some trials may overestimate the true treatment effect, while some may underestimate the true effect. However, trials that overestimated the true effect may be more likely to have subsequent phase III trials, while no further phase III trials will be initiated for those that underestimated the true treatment effect. In subsequent phase III trials, the true treatment effect was approached, making the treatment effect in phase III trials smaller than those in the preceding randomised phase II trials. Moreover, the greater heterogeneity in selecting participants or implementing interventions in large phase III trials may also contribute to the dilution of treatment effect [20].

One major advantage of randomised controlled phase II trials over single-arm trials is the ability to conduct subgroup analysis to identify predictive biomarkers in a more rigorous fashion [1]. Without the randomised control group, there is no ability to determine whether a biomarker is prognostic and/or predictive [21]. However, our results showed that the differences in treatment effect between phase II trials and matched phase III trials was significantly greater for those pairs that phase III trials were initiated based on the subgroup analyses of phase II trials than for those based on the primary analyses of phase II trials. Treatment effect observed in the subgroup analysis of phase II trials were nearly reduced by half in subsequent confirmatory phase III trials. Our previous study showed that even in phase III trials, most claimed subgroup effects were not based on prespecified subgroup analysis and without using interaction test and addressing multiplicity, which increased possibility of false-positive subgroup results [22]. Wallach et al [23], in their study assessing the credibility of subgroup claims in RCTs, reported that none of the initially observed subgroup effect in RCTs was reproduced in subsequent corroboration attempt by a meta-analysis or an RCT. In our study, only one out of eight phase III trials that was initiated based on subgroup analysis of phase II trials was positive, with a 40% reduction in treatment

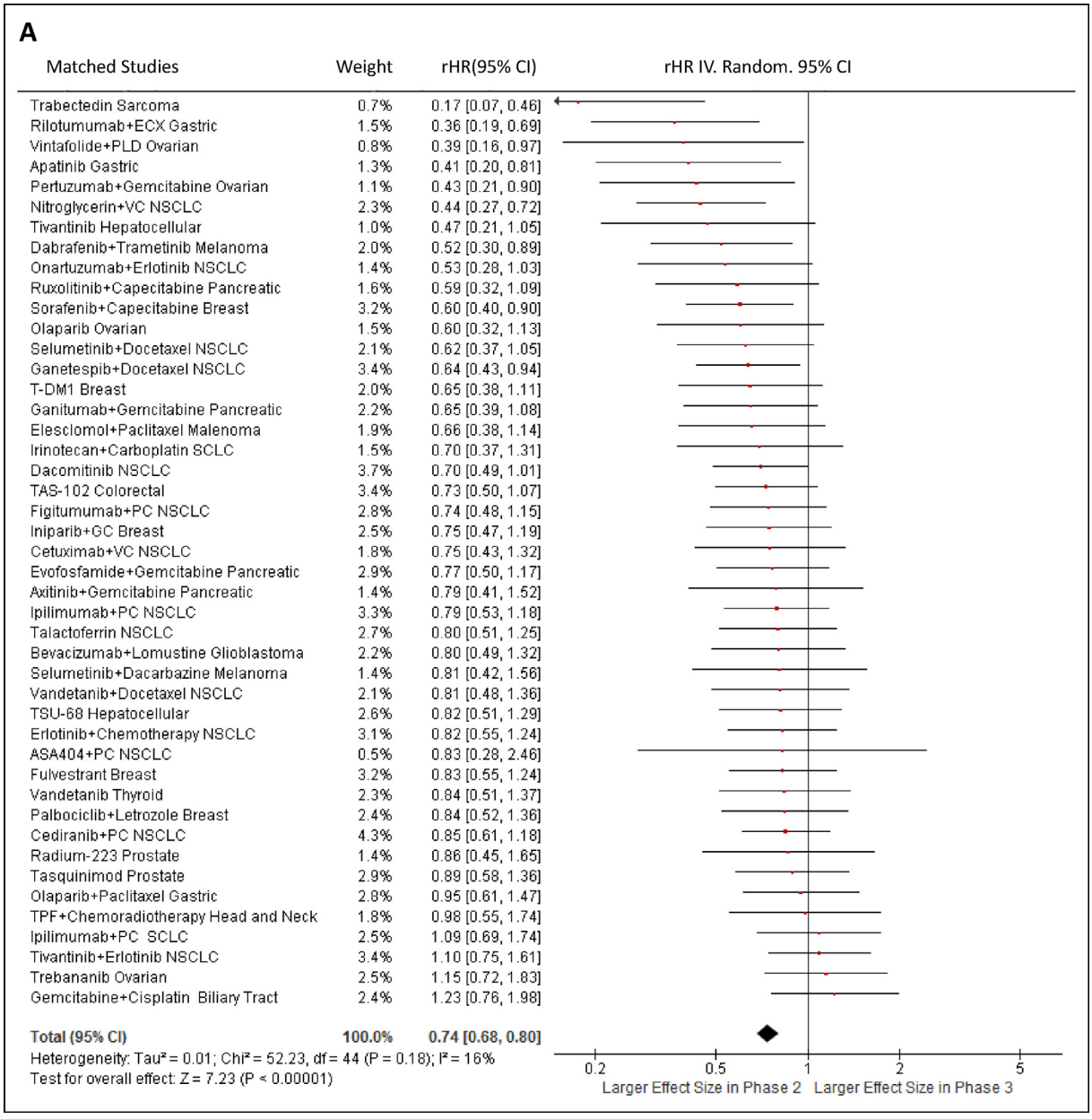

Fig. 3. rHR of (A) progression-free survival or (B) overall survival between phase II trials and phase 3 trials. An rHR <1.0 indicates larger treatment effect size for phase II trials than that for phase III trials. Matched studies were denoted by experimental drugs and cancer type. RHR, ratio of hazard ratio; CI, confidence interval; ECX, epirubicin, cisplatin and capecitabine; PLD, Pegylated liposomal doxorubicin; NSCLC, non–small-cell lung cancer; VC, vinorelbine and cisplatin; T-DM1, trastuzumab emtansine; SCLC, small-cell lung cancer; GC, gemcitabine and carboplatin; PC, paclitaxel and carboplatin.

effect (HR = 0.30 for PFS in phase III trials compared with 0.18 in phase II trials) [24–26]. Initiating large phase III trials based on positive findings from a subgroup analysis in phase II RCTs should be the exception rather than the rule for evaluating novel therapies in oncology.

The only statistically significant predictor of positive phase III outcome was a positive randomised controlled phase II outcome, which is consistent with a previous study [6]. All phase III trials that were initiated based on positive secondary end-point or subgroup analysis of negative phase II trials failed to meet the primary end-point. But even for phase III trials based on positive randomised controlled phase II trials, the overall success rate is only 34.9%. Investigators should be cautious in interpreting the positive findings from randomised phase II trials, and the bar should also be raised when making decisions to proceed from randomised controlled phase II to phase III trials. Nevertheless, our results should not be used to discourage the conduct of randomised controlled phase II trials; instead, they should be used to properly interpret the results of randomised controlled

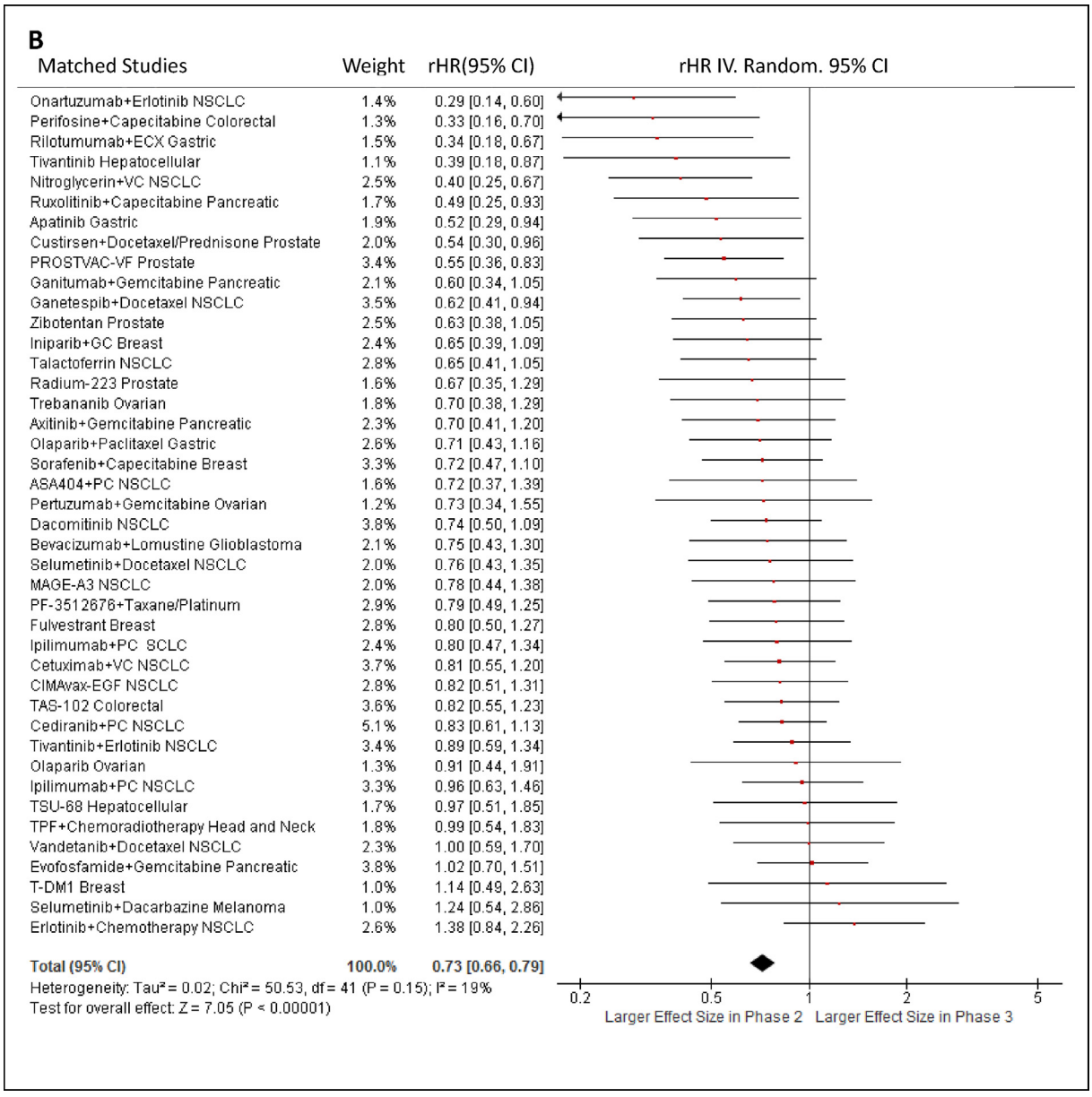

Fig. 3. (*continued*).

phase II trials and better inform the decision to proceed to phase III evaluation.

Our study has several limitations. Although we conducted a comprehensive research for subsequent phase III trials, the results of some phase III trials may not be reported as full article, meeting abstract or press release. But these disappeared phase III trials are more likely to be negative; thus, inclusion of these trials may have strengthened our results. Only 57 paired trials were included in our analysis, which may thus limit the generalisability of our results.

In conclusion, our study demonstrated that success rate of phase III trials based on randomised controlled phase II trials was still low. Treatment effect sizes in randomised controlled phase II trials were significantly overestimated by almost 30%. These results suggest that one should be more critical and cautious in interpreting the results of phase II trials and proceeding into phase III trials.

Ethics approval and consent to participate

Ethics approval was not required for this study because it was based on publicly available data and involved no individual patient data collection or analysis.

Consent for publication

Consent for publication is not needed as no individual patient data or images are involved in this research.

Availability of data and materials

All relevant data have been provided in the text and online supplement. Data extracted from the published manuscript are available from the senior author at liangfei0726@163.com.

Conflict of interest statement

All authors report no conflict of interest.

Funding

None.

Authors' contributions

Conceptual design of the study was performed by F.L. All authors contributed to the drafting of the article; statistical analysis; acquisition of data and final approval of the article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejca.2019.08.006.

References

[1] Gan HK, Grothey A, Pond GR, Moore MJ, Siu LL, Sargent D. Randomized phase II trials: inevitable or inadvisable? J Clin Oncol 2010;28(15):2641–7.
[2] Tang H, Foster NR, Grothey A, Ansell SM, Goldberg RM, Sargent DJ. Comparison of error rates in single-arm versus randomized phase II cancer clinical trials. J Clin Oncol 2010;28(11):1936–41.
[3] Sharma MR, Karrison TG, Jin Y, et al. Resampling phase III data to assess phase II trial designs and endpoints. Clin Cancer Res 2012;18(8):2309–15.
[4] Sharma MR, Stadler WM, Ratain MJ. Randomized phase II trials: a long-term investment with promising returns. J Natl Cancer Inst 2011;103(14):1093–100.
[5] Grayling MJ, Dimairo M, Mander AP, Jaki TF. A review of perspectives on the use of randomization in phase II oncology trials. J Natl Cancer Inst 2019;111(12). pii: djz126 [Epub ahead of print].
[6] Monzon JG, Hay AE, McDonald GT, et al. Correlation of single arm versus randomised phase 2 oncology trial characteristics with phase 3 outcome. Eur J Cancer 2015;51(17):2501–7.
[7] Seruga B, Ocana A, Amir E, Tannock IF. Failures in phase III: causes and consequences. Clin Cancer Res 2015;21(20):4552–60.
[8] Zia MI, Siu LL, Pond GR, Chen EX. Comparison of outcomes of phase II studies and subsequent randomized control studies using identical chemotherapeutic regimens. J Clin Oncol 2005;23(28):6982–91.
[9] Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. BMJ 2013;346:f2304.
[10] Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Stat Med 1998;17(24):2815–34.
[11] Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol 2012;12:9.
[12] Tan A, Porcher R, Crequit P, Ravaud P, Dechartres A. Differences in treatment effect size between overall survival and progression-free survival in immunotherapy trials: a meta-epidemiologic study of trials with results posted at ClinicalTrials.gov. J Clin Oncol 2017;35(15):1686–94.
[13] Wayant C, Vassar M. A comparison of matched interim analysis publications and final analysis publications in oncology clinical trials. Ann Oncol 2018;29(12):2384–90.
[14] Vansteenkiste J, Zielinski M, Linder A, et al. Adjuvant MAGE-A3 immunotherapy in resected non-small-cell lung cancer: phase II randomized study results. J Clin Oncol 2013;31(19):2396–403.
[15] Vansteenkiste JF, Cho BC, Vanakesa T, et al. Efficacy of the MAGE-A3 cancer immunotherapeutic as adjuvant therapy in patients with resected MAGE-A3-positive non-small-cell lung cancer (MAGRIT): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Oncol 2016;17(6):822–35.
[16] El-Maraghi RH, Eisenhauer EA. Review of phase II trial designs used in studies of molecular targeted agents: outcomes and predictors of success in phase III. J Clin Oncol 2008;26(8):1346–54.
[17] Chan JK, Ueda SM, Sugiyama VE, et al. Analysis of phase II studies on targeted agents and subsequent phase III trials: what are the predictors for success? J Clin Oncol 2008;26(9):1511–8.
[18] Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA 1998;279(4):281–6.
[19] Gan HK, You B, Pond GR, Chen EX. Assumptions of expected benefits in randomized phase III trials evaluating systemic treatments for cancer. J Natl Cancer Inst 2012;104(8):590–8.
[20] Yusuf S, Held P, Teo KK, Toretsky ER. Selection of patients for randomized controlled trials: implications of wide or narrow eligibility criteria. Stat Med 1990;9(1–2):73–83. discussion 83-76.
[21] Ballman KV. Biomarker: predictive or prognostic? J Clin Oncol 2015;33(33):3968–71.
[22] Zhang S, Liang F, Li W, Hu X. Subgroup Analyses in reporting of phase III clinical trials in solid tumors. J Clin Oncol 2015;33(15):1697–702.
[23] Wallach JD, Sullivan PG, Trepanowski JF, Sainani KL, Steyerberg EW, Ioannidis JP. Evaluation of evidence of statistical support and corroboration of subgroup claims in randomized clinical trials. JAMA Intern Med 2017;177(4):554–60.
[24] Ledermann J, Harter P, Gourley C, et al. Olaparib maintenance therapy in patients with platinum-sensitive relapsed serous ovarian cancer: a preplanned retrospective analysis of outcomes by BRCA status in a randomised phase 2 trial. Lancet Oncol 2014;15(8):852–61.
[25] Ledermann J, Harter P, Gourley C, et al. Olaparib maintenance therapy in platinum-sensitive relapsed ovarian cancer. N Engl J Med 2012;366(15):1382–92.
[26] Pujade-Lauraine E, Ledermann JA, Selle F, et al. Olaparib tablets as maintenance therapy in patients with platinum-sensitive, relapsed ovarian cancer and a BRCA1/2 mutation (SOLO2/-ENGOT-Ov21): a double-blind, randomised, placebo-controlled, phase 3 trial. Lancet Oncol 2017;18(9):1274–84.