ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

ESMO OPEN SCIENCE FOR OPTIMAL CANCER CARE

ORIGINAL RESEARCH

Accrual and statistical power failure in published adjuvant phase III oncology trials: a comprehensive analysis from 2013 to 2023

G. Villacampa[1,2], S. Dennett[1†], E. Mello[1†], J. Holton[1†], X. Lai[1], L. Kilburn[1], J. Bliss[1], J. Rekowski[1] & C. Yap[1*]

[1]Clinical Trials and Statistics Unit at The Institute of Cancer Research (ICR-CTSU), London, UK; [2]Statistics Unit, Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain

Available online xxx

Background: In a competitive landscape with many ongoing adjuvant randomised controlled trials (RCTs), the prevalence of trials that failed to recruit their targeted sample size and were inadequately powered is unclear. The aims of the study are (i) to determine the percentage of trials with accrual and statistical power failure and (ii) to evaluate their potential impact on the drug development process.

Materials and methods: A systematic review was carried out to identify adjuvant phase III oncology RCTs reported between 2013 and 2023 across all solid tumours. No restrictions were applied regarding the type of intervention or journal of publication. The percentage of trials with accrual failure and power failure was estimated as well as their association with the efficacy endpoints. Logistic regression models were used to estimate the odds ratio (OR) and its 95% confidence interval (CI).

Results: A total of 282 RCTs met the inclusion criteria with a median sample size of 661 patients and a median accrual period of 4.3 years. Most of these studies were superiority trials (83.0%). Accrual failure was observed in 22.0% of the studies, finishing recruitment without achieving the targeted sample size. Overall, 39.7% of the studies experienced power failure, having less power than specified in the protocol at the date of the read-out. Among superiority RCTs evaluating intermediate survival endpoints, only 31.1% presented statistically significant results. Trials with power failure were less likely to present statistically significant results (37.9% versus 21.9%, $P = 0.04$). The association was consistent across all cancer types. In the subset of non-inferiority trials, 35.0% formally demonstrated non-inferiority of the experimental arm.

Conclusions: Nearly 40% of adjuvant phase III RCTs experienced power failure, and the reduction in power significantly impacted the final study results. There is a need for procedural refinements in the design and implementation of future adjuvant RCTs to mitigate these fallacies.

Key words: accrual failure, power failure, oncology, adjuvant, phase III

INTRODUCTION

The success of clinical development in oncology over the past decade has generated a competitive landscape with many ongoing randomised controlled trials (RCTs). RCTs are the current gold standard to evaluate the efficacy of a new intervention. However, they usually require recruiting a large number of patients. In this context, the accrual capacity becomes one of the most relevant barriers for the completion and success of oncology trials.

It has been reported that around 8%-40% of trials failed to recruit the targeted sample size.[1,2] Sub-optimal accrual (called accrual failure hereafter) has several implications: (i) decrease in the statistical power of the study, (ii) negative effects on the reliability and interpretation of the results, (iii) waste of economical and human resources and (iv) delay of the drug development process.[3] Previous studies have recognised slow accrual and logistical challenges as the most common reasons for trials failing to complete.[3] Relaxing eligibility criteria is a straightforward strategy to facilitate more inclusive trials and to improve recruitment.[4,5] However, in some instances, this can negatively impact the clinical relevance of the study. Other aspects related to the study design can also play an important role for the completion of clinical trials.[6,7]

Additionally, even in studies that successfully recruited the targeted sample size, some trials with time-to-event endpoints may observe fewer events than required by

*Correspondence to: Prof. Christina Yap, Clinical Trials and Statistics Unit at The Institute of Cancer Research (ICR-CTSU), 15 Cotswold Rd, Sutton SM2 5NG, London, UK. Tel: 02087224022

E-mail: christina.yap@icr.ac.uk (C. Yap).

Twitter handles: @ChristinaBYap, @G_Villacampa

†These authors contributed equally.

the protocol, leading to insufficient statistical power for their primary objective. Using historical data to estimate the expected event rate could be the leading catalyst for this challenge. Efficacy outcomes in patients treated with the same drug tend to experience improvement over the years due to accumulated expertise in drug administration. This impacts the timing of the observation of events during the trial follow-up. Studies with inadequate follow-up time have a higher risk of not observing the required number of events, and consequently, to have power failure. This is particularly important in adjuvant cancer trials, a curative setting characterized by low event rates, and trials designed with large sample sizes and extended follow-up periods.

To date, the prevalence of accrual failure and power failure in contemporary trials and their impact on the clinical development process remain undetermined. Considering this background, here we present a systematic review of adjuvant phase III oncology RCTs reported over the previous 10 years (2013-2023). The aims of the study are (i) to determine the percentage of trials with accrual and power failure and (ii) to evaluate their impact on the trial results.

MATERIALS AND METHODS

Study design

A systematic review of the literature was carried out to identify reported phase III adjuvant RCTs from 2013 to 2023 across all solid tumours. The systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement.[8] The study was registered in the international prospective register of systematic reviews PROSPERO (registration no.: CRD42023451364).

Selection criteria

To be included in the systematic review, studies had to satisfy the following inclusion criteria: (i) phase III RCTs, (ii) in solid oncology tumours, (iii) conducted in the adjuvant setting, (iv) reported in English and (v) published between 01 January 2013 and 01 August 2023. To standardise the search, the study focused on adjuvant trials, given their curative nature, large sample sizes and extensive follow-up periods. To encompass a broader range of studies, no restrictions were applied regarding the type of intervention or journal of publication. The PubMed database was used to identify all potential eligible published studies as of 1 August 2023. Details of the systematic search can be found in the Supplementary Material, available at https://doi.org/10.1016/j.esmoop.2024.103603. Examination of potentially eligible papers was initially carried out by one author based on their titles and an initial screening of the full published papers. Additionally, three more authors re-evaluated the paper to confirm that the study met the inclusion criteria.

Variables

A standardised spreadsheet was created to extract data from the included studies. The following variables were collected for all studies if available: study identifier, name of the study, journal of publication, experimental arm(s), control arm(s), cancer type, superiority or non-inferiority, double-blinded or open-label, research funding (industry, non-industry or both), one country or international study, number of patients screened, number of patients included in the trial, randomisation ratio, number of sites, date recruitment started, date recruitment finished, accrual failure (yes/no), power failure (yes/no), primary endpoint, alpha error, planned statistical power, cross-over (yes/no), expected hazard ratio (HR) for the primary endpoint, reported HR for the primary endpoint with associated 95% confidence interval (CI), disease-free survival (DFS) or event-free survival (EFS) and overall survival HR with associated 95% CI. Regarding the primary endpoint (i) invasive DFS was considered as DFS and (ii) progression-free survival was pooled with EFS. In a more general analysis, we combined recurrence-free survival, DFS and EFS as intermediate (surrogate) survival endpoints.

Assessment strategy

All included studies were randomly allocated to at least one member of the study team (GV, SD, EM, JH and XL). Two roundtable discussions were held to standardise the data collection during the evaluation period. The initial roundtable took place to discuss the first 10 double-reviewed articles, where each study was independently reviewed by two authors. The second roundtable took place after 100 studies had been assessed to review, discuss and resolve potential discrepancies. At the end of the data extraction, another study member, distinct from the original reviewer, re-evaluated the co-primary outcomes of accrual and power failure status (yes/no) in all trials.

Endpoints

The co-primary endpoint for this study was (i) percentage of trials with accrual failure and (ii) percentage of trials with power failure. Accrual failure was defined as failing to achieve at least 90% of the targeted sample size. Trials that reported an amendment to the protocol to reduce the sample size by $>$10% due to low recruitment were also considered as accrual failures, but trials that amended the protocol due to new information were not considered as accrual failures. Trials that stopped early for efficacy or futility following pre-planned interim analysis, before reaching the targeted sample size, were not considered as accrual failures. Power failure was defined as observing $<$90% of the targeted number of events. Trials that amended the protocol to reduce the number of events due to low recruitment or low event rate were considered as power failures. Trials that stopped early for efficacy or futility reasons following pre-planned interim analysis and trials designed with a time-driven analysis rather than an event-driven analysis were not considered as power failures.

Secondary objectives were (i) the association between power failure and survival outcomes, (ii) factors associated with the accrual/power failure and (iii) factors associated with the duration of recruitment, defined as the time from the date of first patient randomised to the date of last patient randomised. The accrual rate was calculated as the ratio of the number of randomised patients to the number of months that the trial was open for recruitment.

Statistical methods

A descriptive analysis was carried out to summarise trial characteristics. The overall proportion of trials with accrual failure and power failure was reported with the corresponding 95% CI calculated using the Clopper—Pearson method. Logistic regression models were carried out to evaluate predictors of (i) accrual failure, (ii) power failure and (iii) trials with statistically significant results. The odds ratio (OR) and its 95% CI were reported. To select variables for the multivariable analysis, we carried out a least absolute shrinkage and selection operator regression with lambda 1-standard error using the R package glmnet to build a parsimonious multivariate model. Duration of recruitment in years was described as median time and first and third quartile (Q1-Q3). Funnel plots were used to detect publication bias. No data imputation was carried out and a significance level of 0.05 was set for two-sided tests. There was no adjustment for multiple testing. All analyses were undertaken using R statistical software (R Foundation for Statistical Computing, Vienna, Austria) version 4.1.2.

RESULTS

The literature search identified 846 records, out of which 282 RCTs met the inclusion criteria and were included in the analysis (Supplementary Figure S1, available at https://doi.org/10.1016/j.esmoop.2024.103603). Table 1 summarises trial characteristics. Briefly, 83.0% were superiority trials, 19.5% were double-blinded, 89.4% had two study groups, 59.2% were opened only in one country and 3.2% used an unequal randomisation ratio. The most prevalent cancer type was breast (38.7%, $n = 109$) followed by colorectal (13.1%), urinary (8.5%) and gastric (8.2%) cancers.

Sample size and accrual rate

The median sample size among the included studies was 661 patients (Q1-Q3 280-1299 patients), 653 patients (Q1-Q3 236-1201 patients) in superiority trials and 741 patients (Q1-Q3 383-1507 patients) in non-inferiority trials. The median accrual period was 4.3 years (Q1-Q3 2.8-5.8 years) and the median time from study initiation to the first publication of the results was 9.5 years (Q1-Q3 6.9-11.9

Table 1. Characteristics of included trials, overall and by study design (superiority and non-inferiority)

		Overall ($n = 282$)	Superiority ($n = 234$)	Non-inferiority ($n = 48$)
Sample size, median (Q1-Q3)		661 (280-1299)	653 (236-1201)	741 (383-1507)
Number of sites, median (Q1-Q3)		57 (18-129)	58 (18-135)	51 (21-109)
Accrual time, years, median (Q1-Q3)		4.3 (2.8-5.8)	4.2 (2.7-5.8)	4.7 (3.0-6.0)
Cancer type, *n* (%)	Breast	109 (38.7)	85 (36.3)	24 (50.0)
	Colorectal	37 (13.1)	27 (11.5)	10 (20.8)
	Lung	18 (6.4)	18 (7.7)	0 (0.0)
	Gastric	23 (8.2)	21 (9.0)	2 (4.2)
	Gynaecological	10 (3.5)	10 (4.3)	0 (0.0)
	Prostate	9 (3.2)	7 (3.0)	2 (4.2)
	Skin	17 (6.0)	16 (6.8)	1 (2.1)
	Urinary	24 (8.5)	20 (8.5)	4 (8.3)
	Others	35 (12.4)	30 (12.8)	5 (10.4)
Study design, *n* (%)	Superiority	234 (83.0)	234 (100)	0 (0)
	Non-inferiority	48 (17.0)	0 (0)	48 (100)
Blinding, *n* (%)	Double blind	55 (19.5)	52 (22.2)	3 (6.2)
	Open label or single blind	227 (80.5)	182 (77.8)	45 (93.8)
Randomisation ratio, *n* (%)	1 : 1	273 (96.8)	226 (96.6)	47 (97.9)
	Unequal	9 (3.2)	8 (3.4)	1 (2.1)
Number of arms, *n* (%)	2	252 (89.4)	211 (90.2)	41 (85.4)
	>2	30 (10.6)	23 (9.8)	7 (14.6)
Country, *n* (%)	International	115 (40.8)	100 (42.7)	15 (31.2)
	One country	167 (59.2)	134 (57.3)	33 (68.8)
Funding, *n* (%)	Industry	81 (28.7)	68 (29.1)	13 (27.1)
	Non-industry	128 (45.4)	99 (42.3)	29 (60.4)
	Both	73 (25.9)	67 (28.6)	6 (12.5)
Alpha, *n* (%)	5% (two-sided)[a]	248 (87.9)	213 (91.1)	35 (72.9)
	≥5% (one-sided)	34 (12.1)	21 (8.9)	13 (27.1)
Power, *n* (%)	≤80%	175 (62.1)	145 (62.0)	30 (62.5)
	>80%	88 (31.2)	77 (33.0)	11 (22.9)
	Not reported	19 (6.7)	12 (5.1)	7 (14.6)
Primary endpoint, *n* (%)	DFS/RFS/EFS	183 (64.9)	151 (64.5)	32 (66.7)
	OS	33 (11.7)	31 (13.2)	2 (4.2)
	Toxicity	13 (4.6)	10 (4.3)	3 (6.2)
	Others	54 (19.1)	43 (18.4)	11 (22.9)

DFS, disease-free survival; EFS, event-free survival; OS, overall survival; Q1, first quartile; Q3, third quartile; RFS, relapse-free survival.
[a]Trials with 2.5% (one-sided) were also included in this category.

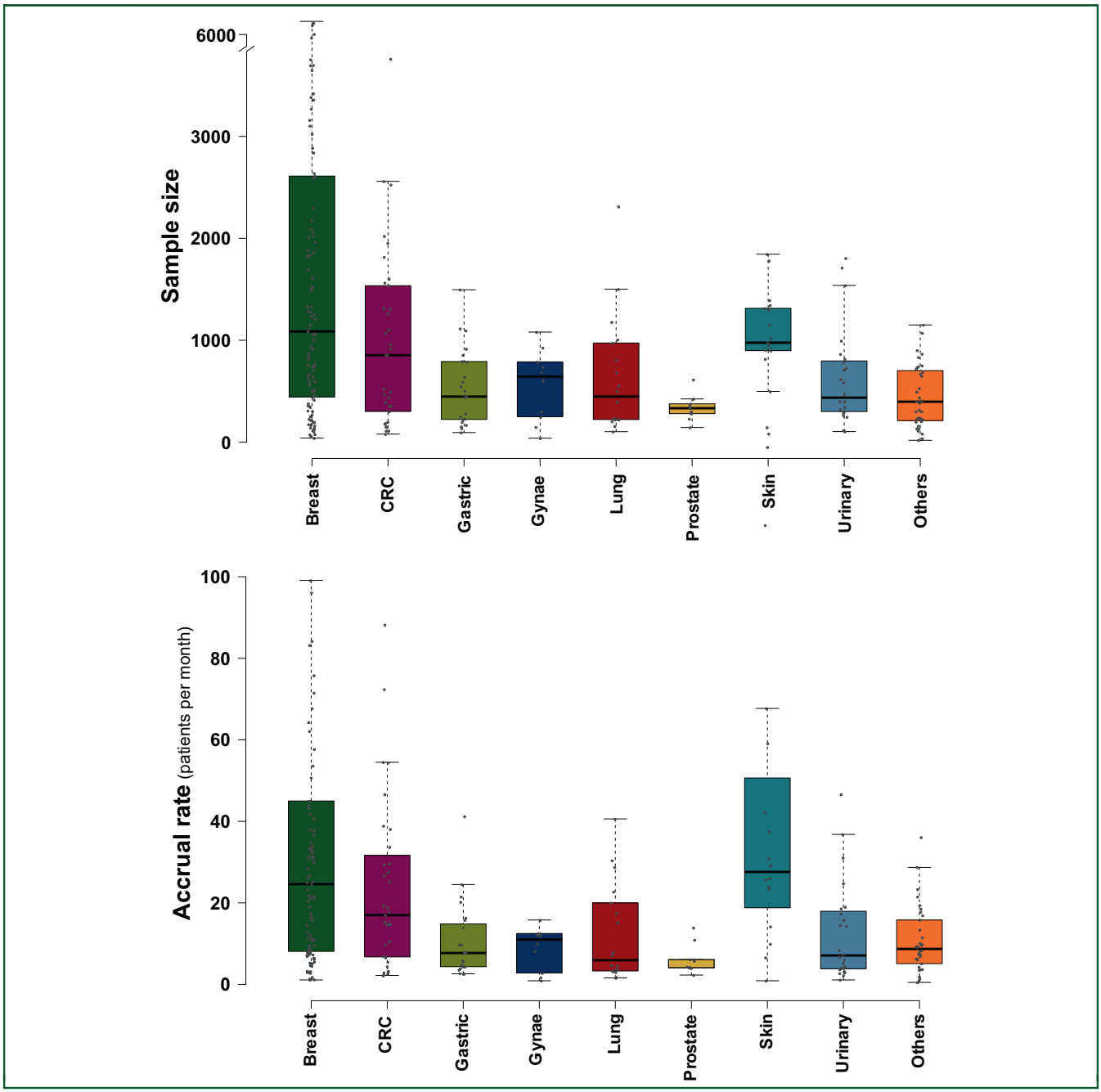

Figure 1. Sample size and the accrual rate of the included RCTs by cancer type.
CRC, colorectal; Gynae, gynaecological; RCTs, randomised controlled trials.

years). Breast and skin cancer trials presented numerically larger sample sizes and faster accrual rates (Figure 1).

Accrual failure

A total of 62 trials (22.0%, 95% CI 17.4% to 27.4%) experienced accrual failure, completing recruitment without achieving their targeted sample sizes. Supplementary Table S1, available at https://doi.org/10.1016/j.esmoop.2024.103603, displays all these studies; the main reason for accrual failure was a low rate of accrual (72.6%, 45/62). Other reasons for accrual failure included safety concerns, and results from other studies. On average, trials with accrual failure included only 49.4% of the initially planned sample size. Trials with non-industry funding, open-label, designed with fewer than 1000 patients and trials with survival endpoints as primary endpoint presented a higher probability of accrual failure (Supplementary Figure S2, available at https://doi.org/10.1016/j.esmoop.2024.103603). In particular, the percentage of accrual failure was 26.4% in open-label, non-industry-funded trials and decreased to 5.6% in double-blinded, industry-funded trials.

Statistical power failure

At the date of their primary analysis, 112 trials (39.7%, 95% CI 34.0% to 45.7%) experienced power failure, observing fewer events than planned in the protocol. Among

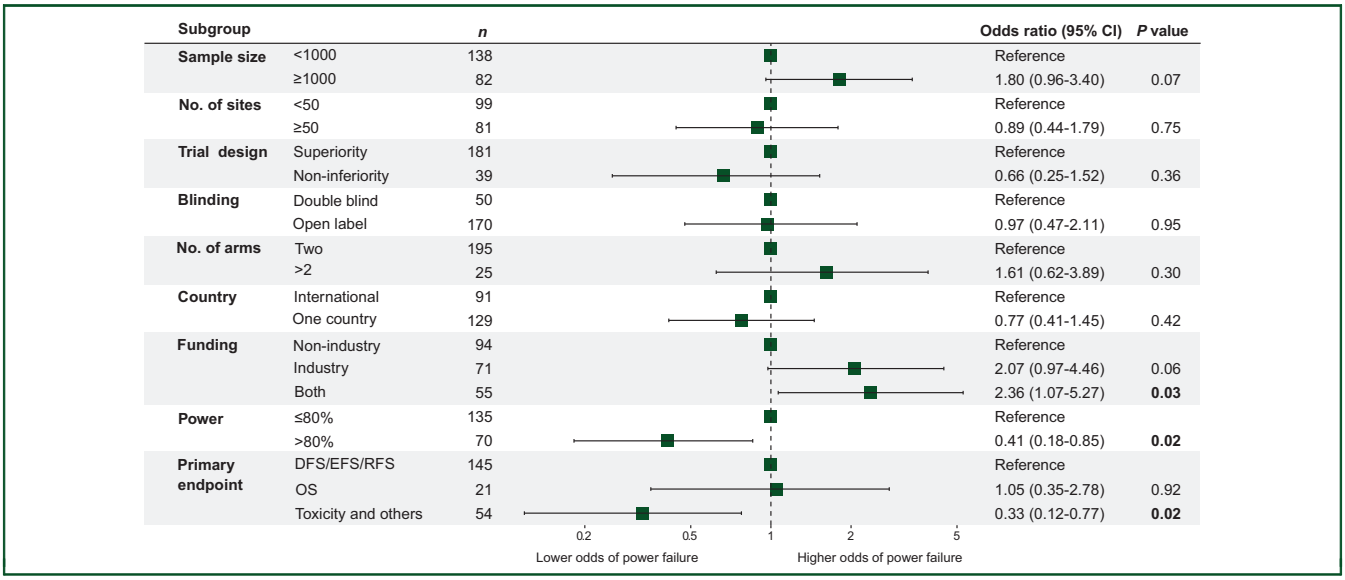

Figure 2. Univariable logistic regression models to evaluate the association between trial characteristics and statistical power failure, among trials that reached their targeted sample sizes ($n = 220$). *P*-values in bold represent the comparisons that reached statistical significance.
CI, confidence interval; DFS, disease-free survival; EFS, event-free survival; OS, overall survival; RFS, relapse-free survival.

superiority trials with intermediate survival endpoints, the power failure rate was 42.4% (64/151). Supplementary Table S2, available at https://doi.org/10.1016/j.esmoop.2024.103603, displays all studies with power failure. On one hand, 96.8% (60/62) of trials with accrual failure did not observe the pre-defined number of events. On the other hand, 23.6% (52/220) of the studies that recruited the expected sample size experienced power failure. Among 220 trials that reached their targeted sample sizes, trials evaluating survival endpoints and industry-funded trials were more likely to report results early in time without reaching the pre-specified number of events (28.2% industry-funded trials versus 15.9% in non-industry-funded trials, $P = 0.06$) (Figure 2). The main reason for power failure was a lower-than-expected event rate (71.2%, 37/52). Multivariable and subgroup analyses focusing on superiority and non-inferiority trials, separately, are presented in Supplementary Figures S3 and S4, available at https://doi.org/10.1016/j.esmoop.2024.103603.

Efficacy outcomes and impact of power failure

To evaluate the percentage of RCTs that met their efficacy objective, we analysed the subset of superiority trials with intermediate survival endpoint ($n = 151$). Among these trials, only 31.1% (47/151) presented statistically significant results. The percentage was 37.9% in trials without power failure but decreased to 21.9% in trials with power failure (OR 2.18, 95% CI 1.06-4.65, $P = 0.04$). The association between power failure and the lack of statistical significance was consistent across all cancer types (Figure 3). Overall, we observed an overestimation of the expected HR compared to the observed HR, with 74.2% of the studies presenting worse HRs than expected according to the protocol (Supplementary Figures S5 and S6, available at https://doi.org/10.1016/j.esmoop.2024.103603). Funnel plots showed no evidence of publication bias (Supplementary Figure S7, available at https://doi.org/10.1016/j.esmoop.2024.103603).

In RCTs that did not achieve statistical significance ($n = 104$), the median observed HR was similar in trials with and without power failure (0.92 and 0.94, respectively) but with a larger variability in power failure trials (Supplementary Figure S8, available at https://doi.org/10.1016/j.esmoop.2024.103603). In 10.0% of studies experiencing power failure, an HR lower than 0.8 was reported. The observation of a greater number of events in these studies could potentially have led to a statistically significant result. There was no evidence for an association between any study characteristics and obtaining statistically significant results (Supplementary Figure S9, available at https://doi.org/10.1016/j.esmoop.2024.103603).

Non-inferiority trials

A total of 17.0% (48/282) of included trials were designed to evaluate non-inferiority. Non-inferiority trials were more commonly funded by non-industry sources (60.4%), opened in only one country (68.8%) and using intermediate survival endpoints (66.7%) as the most frequent primary endpoint. A total of 20 RCTs used the CI of the HR to formally evaluate non-inferiority in intermediate survival endpoints.[9-28] The pre-specified margin to claim non-inferiority ranged from 1.13 to 3.01 (upper CI of the HR). These trials included a median sample size of 1098 patients with an accrual failure rate of 40.0% (8/20). Overall, 35.0% of the RCTs (7/20) demonstrated non-inferiority of the experimental arm, while the other studies could not demonstrate non-inferiority after crossing the pre-specified margin

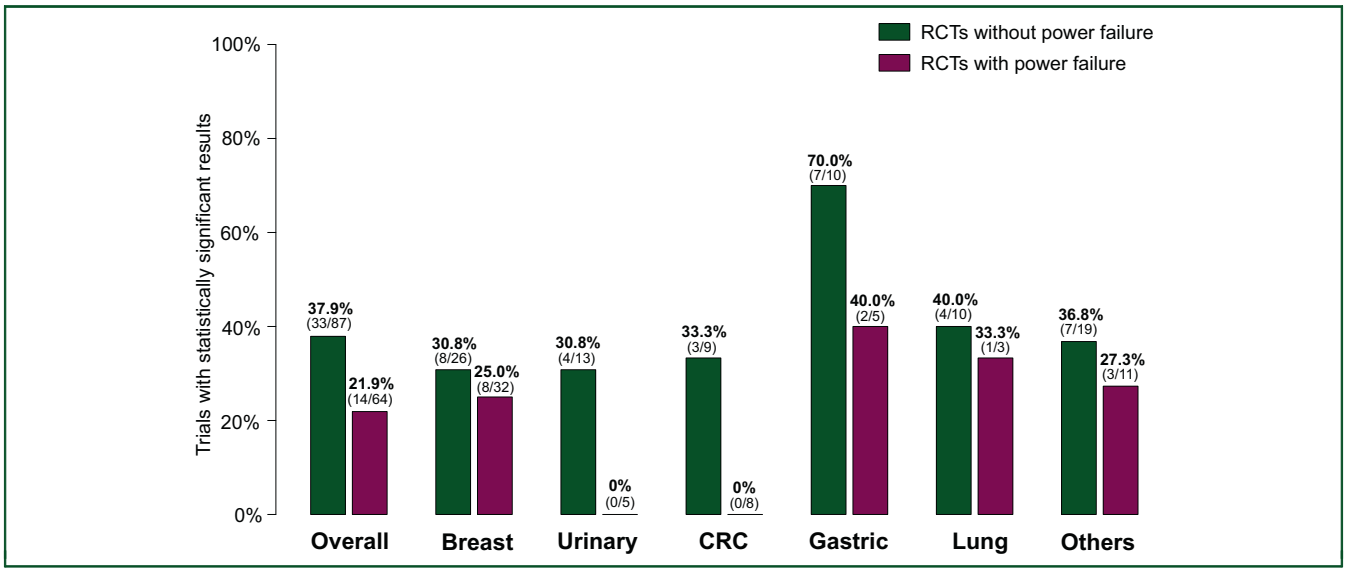

Figure 3. Association between power failure and the percentage of trials with statistically significant results among the subset of superiority studies with DFS/EFS/RFS as primary endpoint. Overall and by cancer type (only cancer types with at least 12 trials are represented).
CRC, colorectal; DFS, disease-free survival; EFS, event-free survival; RCTs, randomised controlled trials; RFS, relapse-free survival.

(Figure 4). Among trials that crossed the pre-specified margin, 53.8% (7/13) presented an HR <1.2.

DISCUSSION

Completion of accrual and having well-powered studies are essential for conducting high-quality oncology RCTs that can drive clinical decision making. While important throughout the entire drug development process, this is even more crucial in confirmatory RCTs where large sample sizes are typically required to assess the study objectives. In this study, we found that 22% of trials stopped recruitment without achieving the targeted sample size and around 40% experienced statistical power failure.

Almost 300 confirmatory adjuvant phase III RCTs have been reported during the past 10 years. Compared with RCTs in the metastatic setting, adjuvant studies tend to recruit more patients (661 versus 466), have longer accrual periods (4.3 versus 1.8 years) and less frequently use

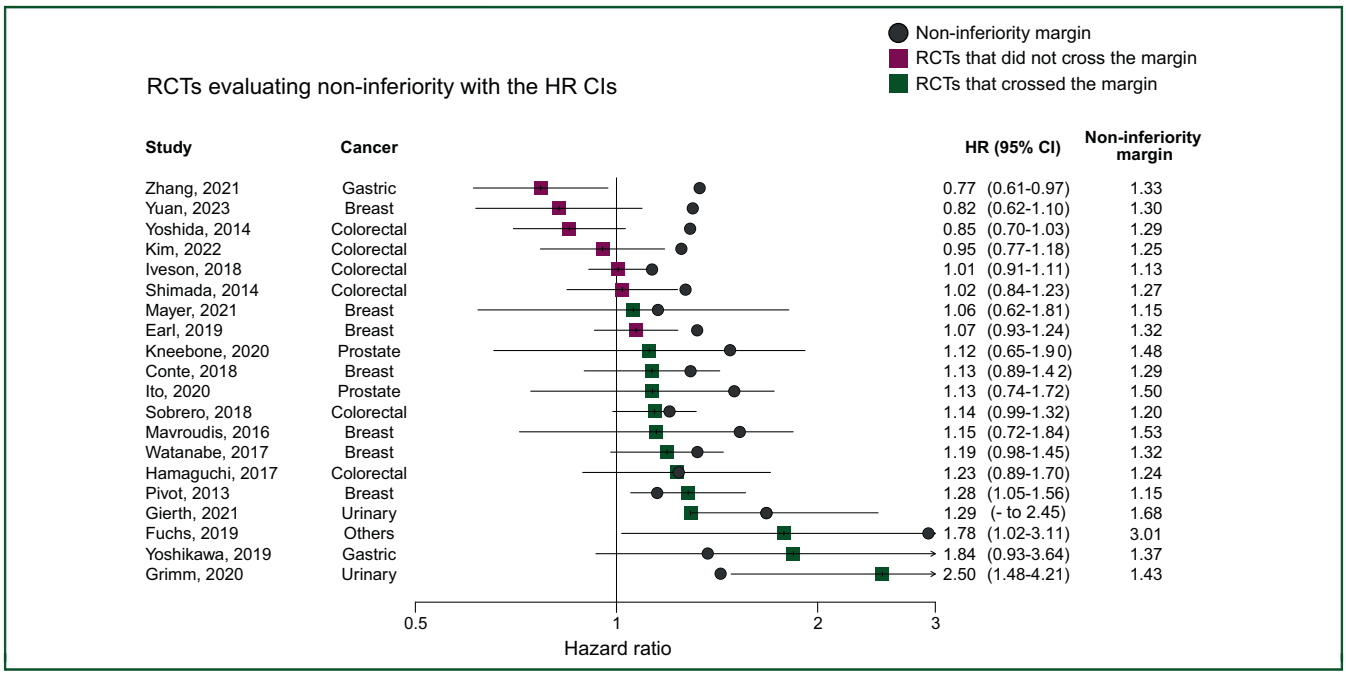

Figure 4. Forest plot with the reported HR of trials evaluating non-inferiority in DFS/EFS/RFS endpoints. Only trials that used the upper CI of the HR as non-inferiority margin were included ($n = 20$). The point estimation of the HR and its CI, as well as the non-inferiority margin, are presented per each study. The blue box represents studies that met the non-inferiority criteria (the CI did not cross the margin). The red box represents studies that did not meet the non-inferiority criteria (the CI did cross the margin).
CI, confidence interval; DFS, disease-free survival; EFS, event-free survival; HR, hazard ratio; RCTs, randomised controlled trials; RFS, relapse-free survival.

unequal randomisation ratios such as 2 : 1 or 3 : 1 (3.2% versus 29%).[29] Among adjuvant trials evaluating superiority, only 31.1% achieved statistical significance for the primary endpoint. These results highlight (i) the competitive landscape in drug development in the adjuvant setting, (ii) the need to design better phase I-II studies and (iii) the challenges in integrating new treatments into the therapeutic arsenal for early-stage disease.

Trials with power failure were less likely to report statistically significant results compared with trials that achieved the pre-defined statistical power (21.9% versus 37.9%). A reduction in statistical power increases the likelihood of false-negative decisions, making it less likely to detect true effects. Moreover, a less appreciated fact in the medical literature is that when a study finds a statistically significant result with low statistical power, two concerns may arise: (i) increased odds of a false-positive result (lower positive predictive value) and (ii) overestimation of the treatment effect.[30] This phenomenon can be even more problematic when future studies are designed under the inflated treatment effect, increasing the risk of obtaining future negative results.

The results presented in this study can also provide insights in improving the design of future RCTs. One of the most common reasons for power failure was the observation of a lower-than-expected event rate when evaluating survival outcomes. Some RCTs were designed assuming a similar event rate as those reported in previous studies, ignoring how treatment outcomes systematically experience improvement over the years. The incorporation of this knowledge in trial design could facilitate a more realistic estimation of the required sample size and the needed follow-up time. To mitigate the risk of underpowered trials, additional strategies include conducting regular interim monitoring of the event rate via an independent data monitoring committee. Such oversight can facilitate the early identification of trends, which may inform trial management, and potentially conserve resources if the trial appears unlikely to detect the intended effects. Implementing an adaptive design that features blinded sample size re-estimation is another effective strategy. It allows for pre-specified interim evaluations of the initial sample size assumptions, facilitating adjustments to enhance the likelihood of achieving the desired statistical power.[31] Additionally, synergizing efforts to incorporate patient involvement can help to design more efficient future oncology trials.[32]

A subgroup of special interest was the RCTs evaluating non-inferiority in efficacy outcomes. The classical strategy in the design of these studies was to pre-define a non-inferiority margin in terms of the upper CI of the HR. However, some limitations are associated with this approach: (i) the need for a large sample size, (ii) the arbitrary selection of the non-inferiority margin and (iii) the unclear clinical interpretability of the results. The non-inferiority RCTs evaluated in this study presented a larger sample size (median >1000) than their superiority counterparts. Interestingly, the pre-defined non-inferiority margins were considerably heterogeneous across studies, and only 35.0% of these studies could formally demonstrate non-inferiority. Some studies failed to demonstrate non-inferiority even when the de-escalation arm presented promising results.[12,15,16,21,25] These practical constraints, together with recent studies like the APT trial (a single-group breast cancer trial testing a de-escalation strategy that led to regulatory approval)[33] can open the door to design more non-inferiority single-group trials using a well-defined synthetic control group.[34-36]

The main limitation of this study is selection bias as we only focused on published phase III RCTs. Recent publications have noted a gradual improvement in result reporting within oncology trials in recent years, with a higher rate of reported studies in randomised phase III trials compared to non-randomised or phase I-II studies.[37] However, there still is a non-negligible number of unpublished phase III RCTs that were not included in this study, with the majority of these trials likely experiencing accrual and power failure. Therefore, the estimates of accrual and power failure presented in this study are conservative and would be underestimates if we were to incorporate the full spectrum of conducted adjuvant phase III trials. The impact of coronavirus disease 2019 was very small or negligible, considering that most trials (96.1%) had completed recruitment before 2020. Strengths of this systematic review include its wider scope in terms of including all published studies, the estimation of the accrual and power failure rate in contemporary trials, the quantification of its impacts in efficacy endpoints and the in-depth analysis of non-inferiority trials.

In conclusion, the results of this study show that at least 22% of the studies did not achieve their targeted sample sizes, 40% suffered from power failure and only 31% achieved statistically significant results. The reduction of power has a significant negative impact on the final study results. Future phase III RCTs should be designed to minimise the risk of not achieving the target sample size and to incorporate close monitoring of recruitment progress and event rates. This is crucial for reliable detection of clinically meaningful effects, and for generating robust evidence that will inform regulatory decisions, public health policy and medical practice.

FUNDING

This work was supported by the National Institute for Health Research (NIHR) Senior Investigator Award [grant number NF-SI-0616-10107]. ICR-CTSU also receives programme grant funding from Cancer Research UK [grant number C1491/A25351]. The findings, interpretations and conclusions expressed in this paper are entirely those of the authors. The funding source did not have a role in the writing or decision to submit for publication. All authors have full access to the full data in the study and accept responsibility to submit for publication.

DISCLOSURE

GV has received a speaker's fee from Merck Sharp & Dohme, Pfizer, GlaxoSmithKline and Pierre Fabre, and received consultant fees from Reveal Genomics. JB has received research funding from AstraZeneca, Merck Sharp & Dohme, Puma Biotechnology, Pfizer, Roche, GlaxoSmithKline/Novartis, Lilly, Janssen-Cilag, Clovis Oncology and received travel/accommodations/expenses from Pfizer. All other authors have declared no conflicts of interest.

REFERENCES

1. Schroen AT, Petroni GR, Wang H, et al. Achieving sufficient accrual to address the primary endpoint in phase III clinical trials from U.S. Cooperative Oncology Groups. *Clin Cancer Res*. 2012;18(1):256-262.
2. Khunger M, Rakshit S, Hernandez AV, et al. Premature clinical trial discontinuation in the era of immune checkpoint inhibitors. *Oncologist*. 2018;23(12):1494-1499.
3. Stensland KD, McBride RB, Latif A, et al. Adult cancer clinical trials that fail to complete: an epidemic? *J Natl Cancer Inst*. 2014;106(9):1-6.
4. Liu R, Rizzo S, Whipple S, et al. Evaluating eligibility criteria of oncology trials using real-world data and AI. *Nature*. 2021;592(7855):629-633.
5. Fowler RA, Van Spall HGC. Eligibility criteria of randomized controlled trials - reply. *J Am Med Assoc*. 2007;298(1):39-40.
6. Bennette CS, Ramsey SD, McDermott CL, Carlson JJ, Basu A, Veenstra DL. Predicting low accrual in the National Cancer Institute's Cooperative Group Clinical Trials. *J Natl Cancer Inst*. 2016;108(2):1-7.
7. Tate WR, Cranmer LD. Development and validation of a clinical trial accrual predictive regression model at a single NCI-designated comprehensive cancer center. *J Natl Compr Cancer Netw*. 2016;14(5):561-569.
8. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *Br Med J*. 2021;372:n160.
9. Yuan P, Kang Y, Ma F, et al. Effect of epirubicin plus paclitaxel vs epirubicin and cyclophosphamide followed by paclitaxel on disease-free survival among patients with operable ERBB2-negative and lymph node-positive breast cancer: a randomized clinical trial. *JAMA Netw Open*. 2023;6(2):E230122.
10. Kim ST, Kim SY, Lee J, et al. Oxaliplatin (3 months v 6 months) with 6 months of fluoropyrimidine as adjuvant therapy in patients with stage II/III colon cancer: KCSG CO09-07. *J Clin Oncol*. 2022;40(33):3868-3877.
11. Zhang X, Liang H, Li Z, et al. Perioperative or postoperative adjuvant oxaliplatin with S-1 versus adjuvant oxaliplatin with capecitabine in patients with locally advanced gastric or gastro-oesophageal junction adenocarcinoma undergoing D2 gastrectomy (RESOLVE): an open-label, superiority and non-inferiority, phase 3 randomised controlled trial. *Lancet Oncol*. 2021;22(8):1081-1092.
12. Mayer IA, Zhao F, Arteaga CL, et al. Randomized phase III postoperative trial of platinum-based chemotherapy versus capecitabine in patients with residual triple-negative breast cancer following neoadjuvant chemotherapy: ECOG-ACRIN EA1131. *J Clin Oncol*. 2021;39(23):2539-2551.
13. Gierth M, Breyer J, Zeman F, et al. The HELENA study: Hexvix®-TURB vs. white-light TURB followed by intravesical adjuvant chemotherapy—a prospective randomized controlled open-label multicenter non-inferiority study. *World J Urol*. 2021;39(10):3799-3805.
14. Sobrero A, Lonardi S, Rosati G, et al. FOLFOX or CAPOX in stage II to III colon cancer: efficacy results of the Italian three or six colon adjuvant trial. *J Clin Oncol*. 2018;36(15):1478-1485.
15. Kneebone A, Fraser-Browne C, Duchesne GM, et al. Adjuvant radiotherapy versus early salvage radiotherapy following radical prostatectomy (TROG 08.03/ANZUP RAVES): a randomised, controlled, phase 3, non-inferiority trial. *Lancet Oncol*. 2020;21(10):1331-1340.
16. Ito K, Kobayashi M, Komiyama M, et al. Oncological outcomes for patients with locally advanced prostate cancer treated with neoadjuvant endocrine and external-beam radiation therapy followed by adjuvant continuous/intermittent endocrine therapy in an open-label, randomized, phase 3 trial. *Cancer*. 2020;126(17):3961-3971.
17. Grimm MO, van der Heijden AG, Colombel M, et al. Treatment of high-grade non—muscle-invasive bladder carcinoma by Standard Number and Dose of BCG Instillations Versus Reduced Number and Standard Dose of BCG Instillations: results of the European Association of Urology Research Foundation randomised phase III clinical trial "NIMBUS". *Eur Urol*. 2020;78(5):690-698.
18. Fuchs M, Goergen H, Kobe C, et al. Positron emission tomography-guided treatment in early-stage favorable Hodgkin lymphoma: final results of the international, randomized phase III HD16 trial by the German Hodgkin Study Group. *J Clin Oncol*. 2019;37(31):2835-2845.
19. Earl HM, Hiller L, Vallier AL, et al. 6 Versus 12 months of adjuvant trastuzumab for HER2-positive early breast cancer (PERSEPHONE): 4-year disease-free survival results of a randomised phase 3 non-inferiority trial. *Lancet*. 2019;393(10191):2599-2612.
20. Yoshikawa T, Terashima M, Mizusawa J, et al. Four courses versus eight courses of adjuvant S-1 for patients with stage II gastric cancer (JCOG1104 [OPAS-1]): an open-label, phase 3, non-inferiority, randomised trial. *Lancet Gastroenterol Hepatol*. 2019;4(3):208-216.
21. Conte P, Frassoldati A, Bisagni G, et al. Nine weeks versus 1 year adjuvant trastuzumab in combination with chemotherapy: final results of the phase III randomized Short-HER study. *Ann Oncol*. 2018;29(12):2328-2333.
22. Iveson TJ, Kerr RS, Saunders MP, et al. 3 Versus 6 months of adjuvant oxaliplatin-fluoropyrimidine combination therapy for colorectal cancer (SCOT): an international, randomised, phase 3, non-inferiority trial. *Lancet Oncol*. 2018;19(4):562-578.
23. Hamaguchi T, Shimada Y, Mizusawa J, et al. Capecitabine versus S-1 as adjuvant chemotherapy for patients with stage III colorectal cancer (JCOG0910): an open-label, non-inferiority, randomised, phase 3, multicentre trial. *Lancet Gastroenterol Hepatol*. 2018;3(1):47-56.
24. Watanabe T, Kuranami M, Inoue K, et al. Comparison of an AC-taxane versus AC-free regimen and paclitaxel versus docetaxel in patients with lymph node-positive breast cancer: final results of the National Surgical Adjuvant Study of Breast Cancer 02 trial, a randomized comparative phase 3 study. *Cancer*. 2017;123(5):759-768.
25. Mavroudis D, Matikas A, Malamos N, et al. Dose-dense FEC followed by docetaxel versus docetaxel plus cyclophosphamide as adjuvant chemotherapy in women with HER2-negative, axillary lymph node-positive early breast cancer: a multicenter randomized study by the Hellenic Oncology Research Group (HORG). *Ann Oncol*. 2016;27(10):1873-1878.
26. Shimada Y, Hamaguchi T, Mizusawa J, et al. Randomised phase III trial of adjuvant chemotherapy with oral uracil and tegafur plus leucovorin versus intravenous fluorouracil and levofolinate in patients with stage III colorectal cancer who have undergone Japanese D2/D3 lymph node dissection: final results of JCOG0205. *Eur J Cancer*. 2014;50(13):2231-2240.
27. Yoshida M, Ishiguro M, Ikejiri K, et al. S-1 as adjuvant chemotherapy for stage III colon cancer: a randomized phase III study (ACTS-CC trial). *Ann Oncol*. 2014;25(9):1743-1749.
28. Pivot X, Romieu G, Debled M, et al. 6 months versus 12 months of adjuvant trastuzumab for patients with HER2-positive early breast cancer (PHARE): a randomised phase 3 trial. *Lancet Oncol*. 2013;14(8):741-748.
29. Jenei K, Haslam A, Olivier T, Miljkovíc M, Prasad V. What drives cancer clinical trial accrual? An empirical analysis of studies leading to FDA authorisation (2015-2020). *BMJ Open*. 2022;12(10):1-8.
30. Button KS, Ioannidis JPA, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013;14(5):365-376.

31. Pallmann P, Bedding A, Choodari-Oskooei B, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Med*. 2018;16:29.
32. Crocker JC, Ricci-Cabello I, Parker A, et al. Impact of patient and public involvement on enrolment and retention in clinical trials: systematic review and meta-analysis. *Br Med J*. 2018;363:k4738.
33. Tolaney SM, Barry WT, Dang CT, et al. Adjuvant paclitaxel and trastuzumab for node-negative, HER2-positive breast cancer. *N Engl J Med*. 2015;372(2):134-141.
34. Amiri-Kordestani L, Xie D, Tolaney SM, et al. Food and Drug Administration analysis of survival outcomes comparing the Adjuvant Paclitaxel and Trastuzumab trial with an external control from historical clinical trials. *Ann Oncol*. 2020;31(12):1704-1708.
35. Mishra-Kalyani PS, Amiri Kordestani L, Rivera DR, et al. External control arms in oncology: current use and future directions. *Ann Oncol*. 2022;33(4):376-383.
36. Serrano C, Rothschild S, Villacampa G, et al. Rethinking placebos: embracing synthetic control arms in clinical trials for rare tumors. *Nat Med*. 2023;29:2689-2692.
37. Liu X, Zhang Y, Li W, et al. Evaluation of oncology trial results reporting over a 10-year period. *JAMA Netw Open*. 2021;4(5):e2110438.