An Investigation of the Shortcomings of the CONSORT 2010 Statement for the Reporting of Group Sequential Randomised Controlled Trials: A Methodological Systematic Review

Abigail Stevely[1], Munyaradzi Dimairo[2]*, Susan Todd[3], Steven A. Julious[2], Jonathan Nicholl[2], Daniel Hind[2], Cindy L. Cooper[2]

1 The Medical School, University of Sheffield, Sheffield, United Kingdom, 2 School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom, 3 Department of Mathematics and Statistics, University of Reading, Reading, United Kingdom

* m.dimairo@sheffield.ac.uk

Abstract

Background

It can be argued that adaptive designs are underused in clinical research. We have explored concerns related to inadequate reporting of such trials, which may influence their uptake. Through a careful examination of the literature, we evaluated the standards of reporting of group sequential (GS) randomised controlled trials, one form of a confirmatory adaptive design.

Methods

We undertook a systematic review, by searching Ovid MEDLINE from the 1st January 2001 to 23rd September 2014, supplemented with trials from an audit study. We included parallel group, confirmatory, GS trials that were prospectively designed using a Frequentist approach. Eligible trials were examined for compliance in their reporting against the CONSORT 2010 checklist. In addition, as part of our evaluation, we developed a supplementary checklist to explicitly capture group sequential specific reporting aspects, and investigated how these are currently being reported.

Results

Of the 284 screened trials, 68(24%) were eligible. Most trials were published in "high impact" peer-reviewed journals. Examination of trials established that 46(68%) were stopped early, predominantly either for futility or efficacy. Suboptimal reporting compliance was found in general items relating to: access to full trials protocols; methods to generate randomisation list(s); details of randomisation concealment, and its implementation. Benchmarking against the supplementary checklist, GS aspects were largely inadequately reported. Only 3(7%) trials which stopped early reported use of statistical bias correction.

Moreover, 52(76%) trials failed to disclose methods used to minimise the risk of operational bias, due to the knowledge or leakage of interim results. Occurrence of changes to trial methods and outcomes could not be determined in most trials, due to inaccessible protocols and amendments.

Discussion and Conclusions

There are issues with the reporting of GS trials, particularly those specific to the conduct of interim analyses. Suboptimal reporting of bias correction methods could potentially imply most GS trials stopping early are giving biased results of treatment effects. As a result, research consumers may question credibility of findings to change practice when trials are stopped early. These issues could be alleviated through a CONSORT extension. Assurance of scientific rigour through transparent adequate reporting is paramount to the credibility of findings from adaptive trials. Our systematic literature search was restricted to one database due to resource constraints.

Introduction

Appropriate use of adaptive designs has the potential to improve efficiency in the conduct of randomised clinical trials (RCTs). However, this can involve considerable extra work and effort during their planning, implementation, and reporting [1]. One form of adaptive design, namely a group sequential design, has been used in confirmatory RCTs for a number of years, which makes their review timely [2]. Regulators have described a group sequential design as well understood [3]. The design may offer ethical and economic benefits by allowing early stopping of RCTs, for instance, for futility or efficacy as soon as there is sufficient evidence to answer the research question(s). Nevertheless, adaptive designs are generally underused in routine practice, in relation to importance given to them in statistical literature [4,5]–although their use seems to be improving [1,6,7].

Various initiatives have been undertaken to facilitate discussions, and to address some of the barriers to the use of adaptive designs: predominantly from a pharmaceutical drug development perspective [1,4,8–14]. Recent publicly funded research, in the UK confirmatory setting, found some degree of conservatism towards the use of adaptive designs [15]. This appears to be influenced by, among others: concerns regarding robustness of adaptive designs in decision making, credibility of their findings to change medical practice, and fear of making wrong decisions when trials are stopped early; and some worry about potential introduction of operational bias during trial conduct [15]. It could be argued that one potential solution to alleviate these cited concerns, is assurance of the scientific rigour in the conduct of RCTs, through transparency and adequate reporting. Thus enabling consumers of research findings to make informed judgements regarding the quality of the research in front of them.

The CONSORT (CONsolidated Standards Of Reporting Trials) statement was first published in 1996, with the aim to enhance adequate reporting of RCTs, [16] and has since been revised in 2001 and 2010 [17,18]. There has been marked general improvement in the conduct and reporting of RCTs since the advent of the first CONSORT statement [19–21], although there are still some suboptimal areas requiring improvements [21,22]. Extensions to the CONSORT statement have since been made to accommodate other trial designs and hypotheses, such as: cluster RCTs, non-inferiority and equivalence trials, and pragmatic RCTs [23–25]. As of the 23rd September 2014, at least 30 reporting related guidance documents were being developed, to enhance transparency in the reporting and conduct of studies [26].

Although the CONSORT 2010 statement has some general items relating to "*interim analyses*", there is currently no CONSORT statement tailored for adaptive designs. Moreover, recent research suggested that the current reporting guidance framework, for adaptive designs, is inadequate for research consumers and policy makers to make informed judgements [15]. Some authors have recently suggested modifications to the CONSORT statement to accommodate various forms of adaptive designs [27,28]. However, although these proposals seem robust in capturing adaptive features, they were not informed by evidence on what is considered to be important by key research stakeholders. Hence, some cited concerns raised by researchers, decision makers and policymakers, and key aspects requiring improvement may have been overlooked [15]. For instance, the use of appropriate inference, to obtain unbiased or bias corrected trial results (point estimates, confidence intervals [CIs] and P-values) was previously neglected. This review therefore aims to:

a. Assess reporting compliance of group sequential designs, in confirmatory RCTs, against the CONSORT 2010 statement and some researcher-led proposed modifications (see items in S1 Table);

b. Investigate the shortcomings of the CONSORT 2010 statement in enhancing the reporting of group sequential RCTs;

c. Find exemplars of well reported group sequential RCTs, which could be used as references by researchers.

Methods

Eligibility criteria

This review was restricted to parallel group RCTs, conducted in humans, with confirmatory objectives. Eligible for inclusion were RCTs, which conducted prospectively planned interim analyses, within the class of group sequential designs using the Frequentist approach; regardless of the nature of: the primary endpoint(s), the number of intervention arms, and therapeutic area. Bayesian designed group sequential RCTs were excluded. In addition, only eligible RCTs with accessible, full-text, peer-reviewed reports, in the English language were included for the final examination of compliance in reporting.

Literature search

Inconsistencies in the indexing related to the reporting of RCTs that employ interim analyses, using group sequential methodology, made the systematic search more challenging. MD conducted a scoping exercise by searching MEDLINE with the assistance of an experienced systematic reviewer, in order to develop an efficient search strategy. This scoping exercise found one MeSH term "*Early Termination in Clinical Trials*" which could be used to index some group sequential RCTs. However, the drawback of this MeSH term is that it biases the findings, in favour of trials that were stopped early, which is also an outcome of interest. The review is focusing on reporting of trials regardless of their early stopping status, and the MeSH search term was also insensitive when used via Ovid MEDLINE. Hence, a free text search was employed of keywords often associated with group sequential methodology, such as: "*group sequential*", "*interim analys(i/e)s*", "*stopping rule(s) or boundar(y/ies)*", "*interim monitoring*", "*early stopping or termination*" and "*accumulating data or information*". Other more general terms such as: "*halted*", "*closed*", "*closure*", "*independent data monitoring committee*" and

"*data monitoring and safety board*" were excluded because they resulted in a very high number of irrelevant reports making the review impractical within time and resources constraints.

The search was used in combination with additional eligibility filters, namely publication type (*clinical trials, phase III*), check tags (*humans, full-text available, English language*) and publication year (*1st January 2001 to 23rd September 2014*). The final search combined independent searches with a Boolean operator "*OR*". MD implemented the final search on the 23rd September 2014 by searching Ovid MEDLINE. Systematic search was supplemented with some known group sequential RCTs, retrieved from an external audit study of adaptive designs on ClinicalTrials.gov [29] (see Fig 1). Duplicate records were checked and identified for exclusion based on the title, first author, and year of publication.

Eligibility screening and quality control

Two reviewers (*AS, MD*) working independently, screened trial reports for eligibility, extracted characteristics of eligible trials and examined compliance in their reporting. Compliance was assessed against the CONSORT 2010 checklist items, and some proposed modifications. These captured issues such as: the use of appropriate statistical methods for early stopping bias correction, mechanisms put in place to minimise operational bias (due to the leakage or knowledge of interim results), access to prior interim results, rationale for choosing a group sequential RCT (with any other add-on planned adaptations), and discussion of the lessons learned and value of using a group sequential design, to help the planning of future trials (see S1 Table). Accessible additional related reports, such as protocols and other prior publications, were also used to assess compliance in reporting. All discrepancies were reviewed and rectified in agreement between the two independent reviewers (*AS, MD*). Study investigators were contacted for clarification where possible and necessary.

Outcome measures

The primary outcome of the review is to establish compliance in reporting of the CONSORT 2010 checklist items, as well as the additional dimensions of interest specific to group sequential RCTs. This was subjectively examined and agreed upon by the two independent reviewers
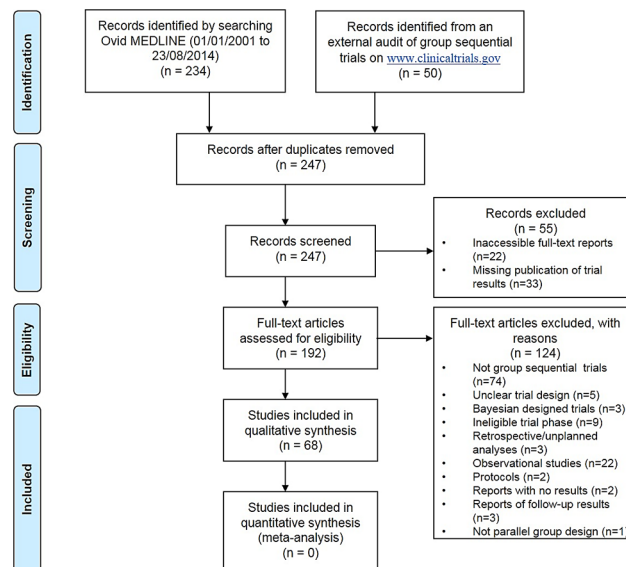

Fig 1. A modified PRISMA flowchart of the review process.

doi:10.1371/journal.pone.0141104.g001

(*AS*, *MD*), according to a predefined classification system of completeness: "*absent*", "*totally complete*", "*partially complete*", "*cannot access*" and "*not applicable*". This was then used to compute the number and proportion of RCTs meeting total and at least partial, compliance in reporting criteria for each checklist item. Furthermore, a global measure of the number and proportion of checklist items meeting total and at least partial compliance criteria was calculated.

Statistical analysis and reporting

The reporting of this review is in accordance with the PRISMA guidance [30,31]. Protocol registration correspondence is accessible (see S1 Appendix). Descriptive summary statistics including numbers (proportions) and median (IQR: Interquartile Range) for categorical and continuous data, respectively, were used to assess compliance in reporting. Clustered stacked bar charts and forest plots were used to aid visual interpretation. Two-sided 95% CIs around proportions were computed using the Wilson Score method [32]. Fisher's exact test was used to explore differences in proportions between subgroups of interest, and estimates are presented as risk ratio (RR), with associated 95% CIs.

A global measure of compliance in reporting, based on the number and proportion of checklist items meeting a certain completeness criterion, was also employed. In addition, bootstrap methods [33] were used with 10 000 replicates, to compute the median difference, or median ratio (95% CI), of the total number of checklist items meeting certain reporting compliance criteria between subgroups. This approach was used to explore whether the publication journal's CONSORT endorsement policy (yes or no), and publication period (pre- or post-publication of the CONSORT 2010 statement), was associated with improved compliance in reporting. The latter was used to explore the impact of the CONSORT 2010 statement in enhancing reporting. Comparability of compliance in reporting, between the standard CONSORT and researcher-led proposed items, was descriptive without any significance testing. Raw data are publicly accessible (see S1 Dataset).

Results

Eligibility screening

A total of 284 RCTs were screened for eligibility: of which, 234 were retrieved by searching the Ovid MEDLINE, and 50 were from an external audit study. Of these 284, 68(24%) peer-reviewed publications reporting main findings were eligible for examination of compliance in reporting. Reasons for exclusion and details of the screening process are shown on Fig 1.

Characteristics of reviewed group sequential RCTs

The majority of RCTs were published in "*high impact*" medical journals such as; *The New England Medical Journal*, *The Lancet Oncology*, *The American Society of Clinical Oncology*, and *The Journal of the American Medical Association*. The median (IQR) journal impact factor for the year 2013 to 2014 was 17.5 (6.6 to 30.4), and the maximum was 54.4. Eligible group sequential RCTs were predominantly in the disease area of oncology (76%). However, diverse therapeutic conditions under investigation included cardiovascular, respiratory and infectious diseases. The majority, 62(91%), of the RCTs investigated at least some form of pharmacological intervention, and 55(81%) were designed with two intervention arms, inclusive of the comparator arm. Forty-six (68%) of the publishing journals endorsed the CONSORT statement as part of their publication policy. Table 1 describes detailed characteristics of examined eligible

Table 1. Characteristics of eligible reviewed RCTs.

Variable	Scoring	Publication period		Total
		2001–2010 (n = 34)	2011–2014 (n = 34)	(n = 68)
Funder/sponsor	Private	16(47%)	19(56%)	35(51%)
	Public	8(24%)	11(32%)	19(28%)
	Private and Public	4(12%)	4(12%)	8(12%)
	None/independent	1(3%)	0(0%)	1(1%)
	Undisclosed	5(15%)	0(0%)	5(7%)
Nature of primary outcome(s)	Time-to-event	23(68%)	28(82%)	51(75%)
	Binary	6(18%)	3(9%)	9(13%)
	Continuous	3(9%)	3(9%)	6(9%)
	Binary and continuous	1(3%)	0(0%)	1(1%)
	Binary and time-to-event	1(3%)	0(0%)	1(1%)
Number of intervention arms	2	26(76%)	29(85%)	55(81%)
	3	6(18%)	3(9%)	9(13%)
	4	1(3%)	1(3%)	2(3%)
	5 or 6	1(3%)	1(3%)	2(3%)
Therapeutic area	Oncology	28(82%)	24(71%)	52(76%)
	HIV/AIDS	3(9%)	0(0%)	3(4%)
	Cardiac	0(0%)	2(6%)	2(3%)
	Musculoskeletal	1(3%)	1(3%)	2(3%)
	Optical	0(0%)	2(6%)	2(3%)
	Stroke	0(0%)	1(3%)	1(1%)
	Respiratory	1(3%)	0(0%)	1(1%)
	Diabetes	0(0%)	1(3%)	1(1%)
	Multiple Sclerosis	1(3%)	0(0%)	1(1%)
	Degenerative	0(0%)	1(3%)	1(1%)
	Epilepsy	0(0%)	1(3%)	1(1%)
	Kidney	0(0%)	1(3%)	1(1%)
Journal CONSORT endorsement status	No	13(38%)	9(26%)	22(32%)
	Yes	21(62%)	25(74%)	46(68%)
Publishing journal	The Lancet Oncology	3(9%)	9(26%)	12(18%)
	The New England Journal of Medicine	5(15%)	7(21%)	12(18%)
	American Society of Clinical Oncology	8(24%)	4(12%)	12(18%)
	Annals of Oncology	3(9%)	2(6%)	5(7%)
	The Journal of the American Medical Association	1(3%)	4(12%)	5(7%)
	Breast Cancer Research Treatment	2(6%)	1(3%)	3(4%)
	Journal of Clinical Oncology	2(6%)	1(3%)	3(4%)
	The Lancet	1(3%)	1(3%)	2(3%)
	The American Academy of Ophthalmology	0(0%)	2(6%)	2(3%)
	Arthritis and Rheumatology	0(0%)	1(3%)	1(1%)
	British Journal of Surgery	1(3%)	0(0%)	1(1%)
	Clinical Breast Cancer	0(0%)	1(3%)	1(1%)
	Clinical Cancer Research	1(3%)	0(0%)	1(1%)
	European Journal of Cancer	0(0%)	1(3%)	1(1%)
	HIV Clinical Trials	1(3%)	0(0%)	1(1%)
	Journal of the National Cancer Institute	1(3%)	0(0%)	1(1%)
	Journal of Urology	1(3%)	0(0%)	1(1%)

(Continued)

Table 1. (*Continued*)

Variable	Scoring	Publication period		Total
		2001–2010 (n = 34)	**2011–2014 (n = 34)**	**(n = 68)**
	Journal of the National Cancer Institute	1(3%)	0(0%)	1(1%)
	Nutrition	1(3%)	0(0%)	1(1%)
	Radiotherapy and Oncology	1(3%)	0(0%)	1(1%)
	The Journal of Infectious Diseases	1(3%)	0(0%)	1(1%)
Type of intervention	Drug	29(85%)	30(88%)	59(87%)
	Dietary	1(3%)	1(3%)	2(3%)
	Device	0(0%)	1(3%)	1(1%)
	Physiological	1(3%)	0(0%)	1(1%)
	Radiotherapy	1(3%)	0(0%)	1(1%)
	Drug and radiotherapy	0(0%)	1(3%)	1(1%)
	Drug and dietary	1(3%)	0(0%)	1(1%)
	Surgical	1(3%)	0(0%)	1(1%)
	Vaccine	0(0%)	1(3%)	1(1%)
Class of intervention	Pharmacological	30(88%)	32(94%)	62(91%)
	Non-pharmacological	4(12%)	2(6%)	6(9%)
Stage of reporting	Interim analysis	25(74%)	22(65%)	47(69%)
	Final analysis	7(21%)	6(18%)	13(19%)
	Unplanned interim analysis	2(6%)	6(18%)	8(12%)
Number of planned interims	1	16(47%)	12(35%)	28(41%)
	2	9(26%)	14(41%)	23(34%)
	3	3(9%)	2(6%)	5(7%)
	4	0(0%)	4(12%)	4(6%)
	5 or 7	3(9%)	0(0%)	3(4%)
	Undisclosed	3(9%)	2(6%)	5(7%)
Trials stopped early	No	11(32%)	9(26%)	20(29%)
	Yes	22(65%)	24(71%)	46(68%)
	No, but interim arm discontinued at interim	1(3%)	1(3%)	2(3%)
Reasons for early stopping (N = 46)	Futility	12(55%)	10(42%)	22(48%)
	Efficacy	5(23%)	5(21%)	10(22%)
	Safety	1(5%)	1(4%)	2(4%)
	Futility and safety	0(0%)	5(21%)	5(11%)
	Poor recruitment and/or financial	3(14%)	3(13%)	6(13%)
	Futility and external information	1(5%)	0(0%)	1(2%)
Planned stopping criteria	Undisclosed	16(47%)	6(18%)	22(32%)
	Futility or efficacy	8(24%)	12(35%)	20(29%)
	Futility	3(9%)	6(18%)	9(13%)
	Efficacy	0(0%)	6(18%)	6(9%)
	Efficacy or safety	3(9%)	1(3%)	4(6%)
	Futility or efficacy or safety	1(3%)	3(9%)	4(6%)
	Non-inferiority	2(6%)	0(0%)	2(3%)
	Safety	1(3%)	0(0%)	1(1%)
Planned total sample size	Min to Max	160–8028	100–15000	100–15000
	Median(IQR)	604(350–1071)	784(428–1200)	724(357–1155)

group sequential RCTs, stratified by publication period (pre- and post-publication of the CONSORT 2010 statement).

Reporting of the universal CONSORT 2010 checklist items

Fig 2 shows a clustered bar chart of compliance in reporting against the CONSORT 2010 checklist items. Additional summary data are provided (see S2 Table). The median proportions (IQR) of RCTs meeting complete and at least partial compliance in reporting criteria of checklist items was 81% (53% to 91%) and 93% (78% to 97%), and a minimum of 12% and 22%, respectively. Figs 3 and 4 are forest plots showing the proportions of RCTs meeting total and at least partial completeness compliance criteria, with associated 95% CIs, respectively.

Suboptimal reporting was observed among checklist items relating to: the disclosure and access to full trial protocols 53(53%), methods used to generate the randomisation list(s) 32
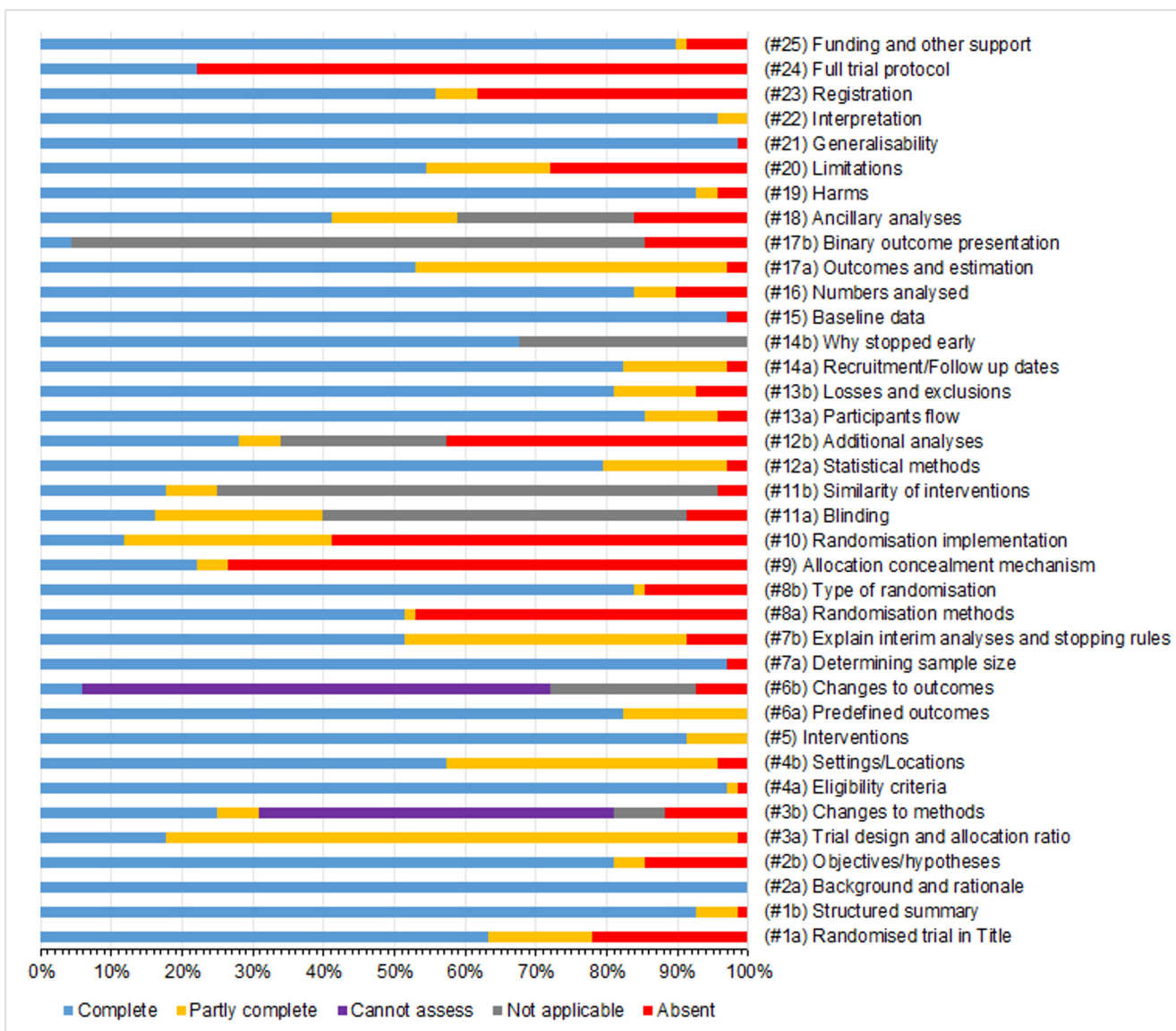

Fig 2. Clustered stacked bar charts of compliance in the reporting of general CONSORT 2010 checklist items.
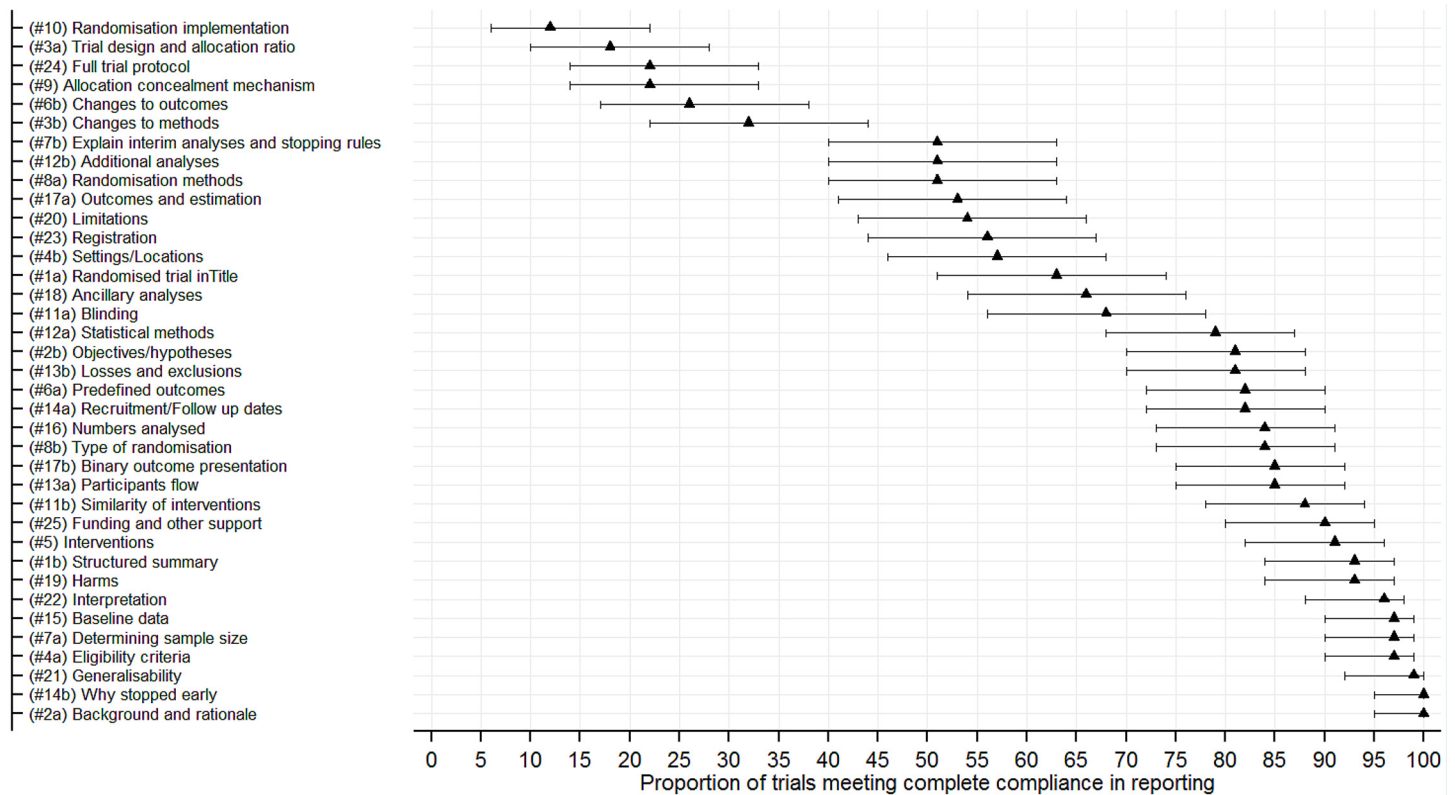
doi:10.1371/journal.pone.0141104.g002

Fig 3. Forest plot of the proportion of trials meeting total completeness in the reporting of general CONSORT 2010 checklist items.

(47%), details of randomisation concealment 50(74%) and implementation of randomisation 40(59%), details of additional analysis 29(43%) and disclosure of trial registration information 26(38%). Furthermore, changes to methods and outcomes could not be assessed in 34(50%) and 45(66%) RCTs, respectively, due to inaccessible protocols and related amendments for most RCTs. Aspects relating to the trial design were partially reported in 55(81%) RCTs. However, most other checklist items were well reported.

Of the 37 CONSORT checklist items, the median number (proportion) [IQR] that were completely reported was 26(70%) [24(65%) to 28(76%)], and a minimum of 15(41%). However, this distribution was 30(81%) [29(78%) to 32(87%)], and a minimum of 24(65%) for checklist items that met partial compliance criterion. The median number (proportion) [95% CI] of items that met complete compliance increased by 2 (5%) [1(1%) to 4(10%); P = 0.009] post-publication of the CONSORT 2010 statement. The median difference (proportion) [95% CI] in items that met complete compliance in favour of journals that endorse the CONSORT statement as part of their publication policy was 1.5(4.1%) [-0.3(-0.9%) to 3.3(9.0%); P = 0.112].

Reporting of group sequential specific checklist items and proposed modifications

Most items relating to group sequential aspects were poorly reported (see Fig 5). Additional summary data are provided (see S1 Table). Only 3(4%) RCTs were identifiable by the term "*group sequential*" in the Title or Abstract. An additional 39(57%) were identifiable by the terms "*interim analyses*" or "*interim analysis*". The rationale for choosing a group sequential design (with any other add-on forms of trial adaptation) was only explained in 11(16%) RCTs.
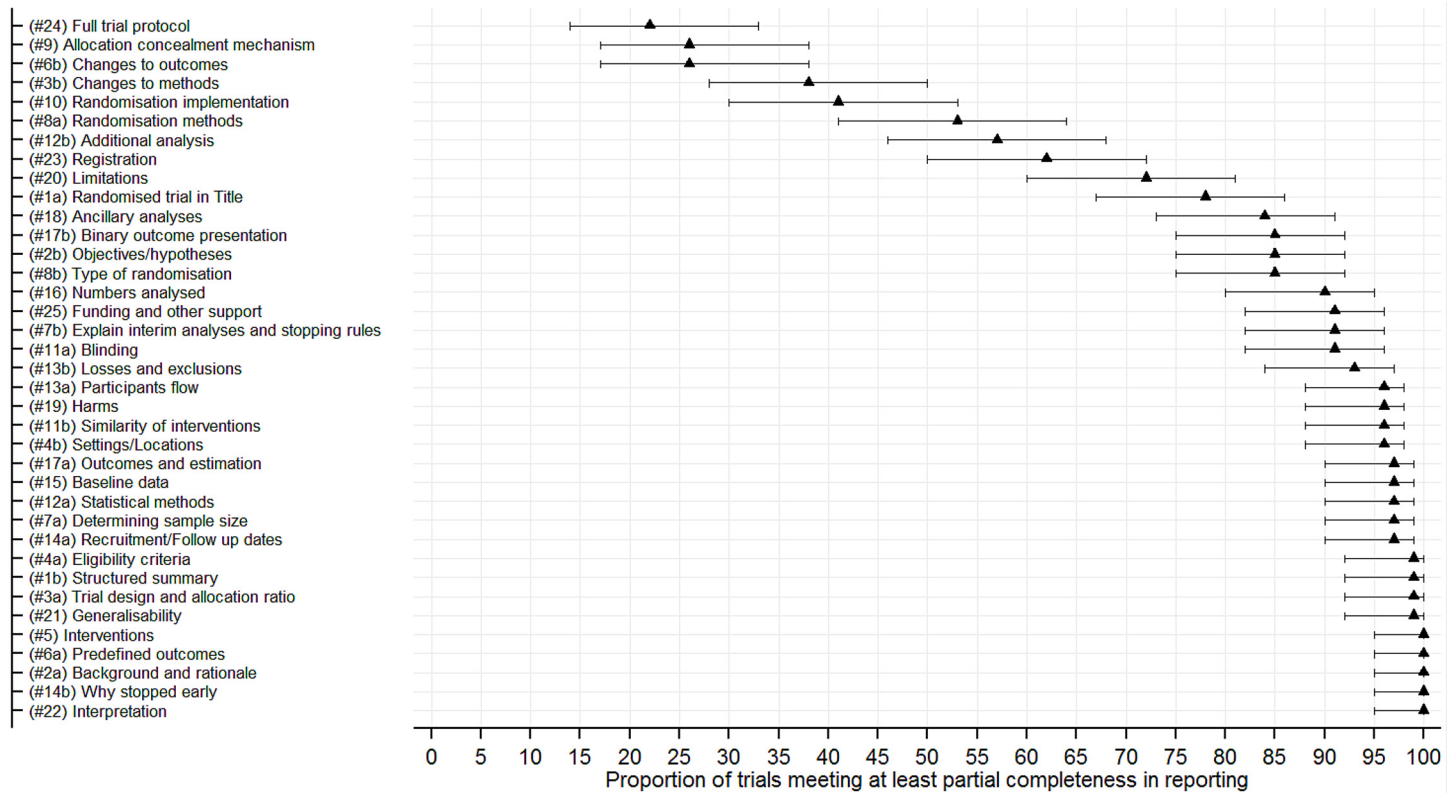
Fig 4. Forest plot of the proportion of trials meeting at least partial completeness in the reporting of general CONSORT 2010 checklist items.

Just 11(16%) RCTs adequately reported the mechanism used to minimise operational bias due to the knowledge or leakage of the interim results; 7 of these cited relevant prior publications. Of the 33 RCTs that were reporting interim results after the first interim, 9(27%) reported or disclosed prior interim results. Only 3 RCTs reported unplanned deviations from planned group sequential design and its potential implications on the findings. However, unplanned deviations could not be assessed in 41(60%) RCTs due to inaccessibility of protocols and associated amendments. Only 2 RCTs described the lessons learned from using the group sequential design, and its value in helping the planning of future group sequential RCTs.

The following aspects of a group sequential design were deemed to be adequately reported: description of the total (and per group) sample size, the planned number of interims and associated interim sample sizes (or number of events), and clarification on whether the RCT was stopped early with reasons where applicable. In addition, planned stopping criteria and rules or boundaries were deemed fairly reported. These particular interim analyses related items are well covered in the CONSORT 2010 checklist.

Early stopping of trials or treatment arms. Of the 68 RCTs, it was found that 46(68%) were stopped early; predominantly for futility, 61% (28/46), and efficacy, 22% (10/46) (see Table 1). The proportion of RCTs stopped early for any reasons before and after 2010 appeared to be similar; 22(65%) versus 24(71%), respectively: RR (95% CI, P-value); 0.92(0.66 to 1.27, P = 0.796). Of the 22 RCTs which were not stopped early, 6(27%) had multiple intervention arms. Of these 6 RCTs, 2 had discontinued one intervention arm at previous interim analyses. In 46 RCTs that were stopped early, the median (IQR) of the distribution of the proportion of
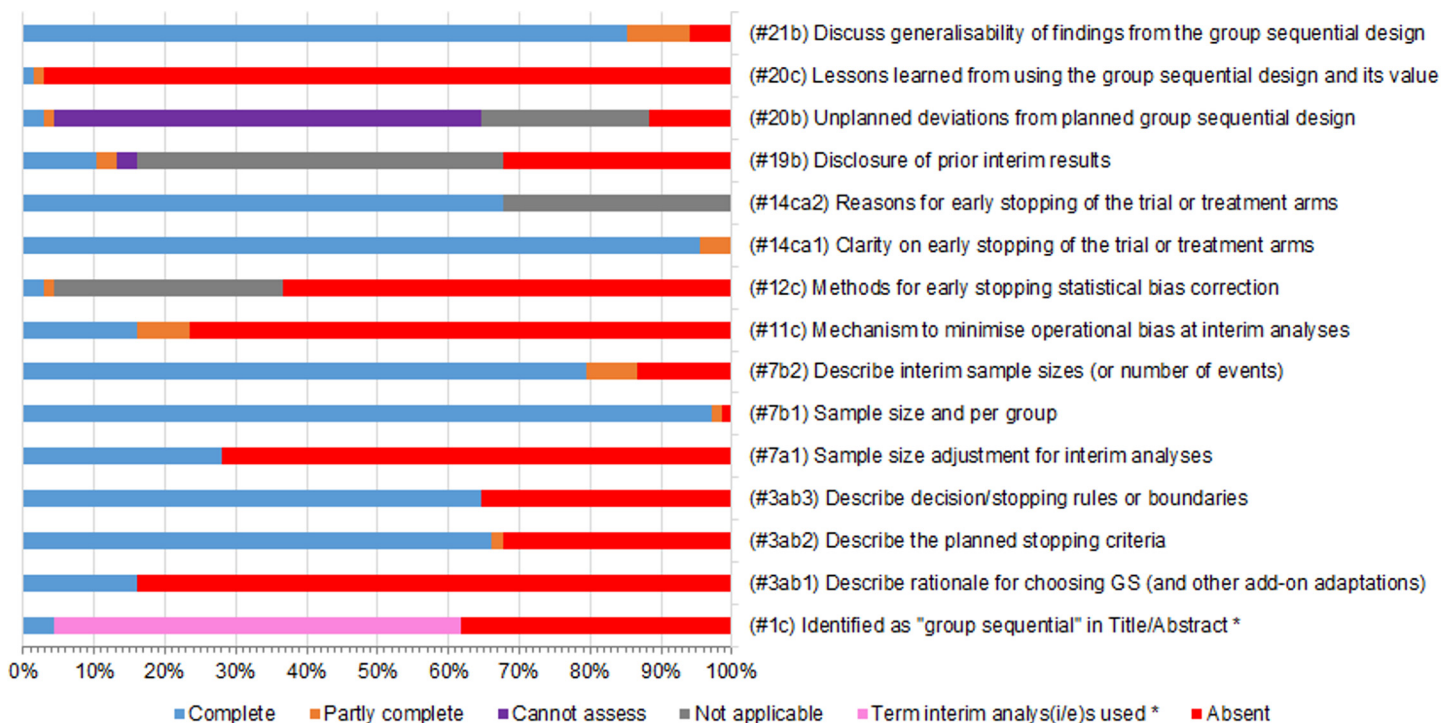
Fig 5. Clustered stacked bar charts of compliance in the reporting of group sequential-specific items. Only items marked (#3ab2), (#3ab3), (#7b1), (#7b2), (#14ca1), and (#14ca2) are partly or fully covered in the current CONSORT checklist. GS: Group Sequential.

doi:10.1371/journal.pone.0141104.g005

interim sample size (or observed interim events) at the time of trial stopping relative to the planned was 65% (50% to 85%), and a minimum of 19%.

Type of planned stopping criteria. Stopping criteria and rules or boundaries planned at the design stage, were unreported in 22(32%) and 24(35%) of RCTs, respectively. Of the RCTs that reported planned stopping criteria and/or stopping boundaries, 11(16%) cited additional relevant information accessible in the form of prior publications or protocols. Thirty-three (49%) RCTs were planned with at least some form of futility early stopping criteria; 9(13%) for futility only, 20(29%) for either futility or efficacy, and 4(6%) for futility, efficacy or safety.

Type of planned stopping rules or boundaries. Twenty-four (35%) RCTs did not disclose the stopping rules or boundaries used. Of the 44(65%) RCTs that reported stopping rules or boundaries, the most frequently used stopping boundaries were: 15(34%) Lan-DeMets (LD) [34] error spending function mimicking O'Brien-Fleming (OBF) type properties [35], and 12 (27%) OBF. Other stopping boundaries or rules that were rarely used were: Pocock [36]; Hay-bittle-Peto [37,38]; Pampallona and Tsiatis [39], in combination with LD error spending function of the OBF type; Pampallona and Tsiatis [39]; Gamma family ($\gamma = -8$) [40]; Rho family ($\rho = 3$) [41], in combination with OBF; Wang and Tsiatis (shape parameter of 0) [42], in combination with OBF; Fleming [43], in combination with LD error spending function of the OBF type; Whitehead's double triangular test [44]; conditional power based; and reported in terms of number of events or hazard ratios.

Number of planned interim analyses and stage of reporting. The majority (75%) of RCTs were planned with either one or two interim analyses. There were very few RCTs planned with large numbers of interim analyses (see Table 1). Only 5(7%) RCTs did not report the number of planned interim analyses. 55(81%) RCTs were reporting interim results; of

which, 47(69%) were as intended. Poor recruitment and/or financially related issues were the main reasons for reporting unplanned interim results in the remaining 8(12%) RCTs.

Early stopping statistical bias correction. Of the 46 RCTs that were stopped early, only 3 (7%) reported the use of appropriate statistical methods for bias correction of point estimates of the intervention effect, and associated CIs and P-values; 2 of these were stopped early for futility and/or safety. Only 1 of the 10 RCTs that were stopped early for efficacy reported the use of bias corrected statistical methods to conduct inference.

Exemplars to enhance reporting of group sequential RCTs

Although there were no publications that met complete compliance on all checklist items, we found exemplars of group sequential RCTs that reported most items adequately [45–47]. The PRIMO trial is an exemplar that provided a comprehensive rationale of choosing, and detailed description of a group sequential design incorporating sample size re-estimation used [48,49]. Some publications described aspects of the randomisation process, allocation concealment and its implementation, which were found to be problematically reported in most group sequential RCTs, better than others [47,50]. A further exemplar reported detailed description of protocol changes [51]. Another useful exemplar gave a description, and graphical representation, of prior interim trends of the intervention effect, and explored the trend using regression methods [52].

Roger et al [45] gave a clear description of an exact statistical method, used to obtain unbiased inference following early stopping of group sequential RCTs:

> "*The study design is based on a group-sequential test procedure with pre-planned analyses after 220, 320 and 428 patients meeting one of the off-study criteria. An alpha-spending approach as suggested by Lan and DeMets [34] with an O'Brien/Fleming-like alpha spending function was used to define the test boundaries of the group-sequential procedure. The primary analysis regarding OS uses a Cox Proportional Hazard Model with treatment and prognosis groups as predictor variables to calculate the Z score needed for the group- sequential procedure. Stagewise ordering was used to compute the unbiased median estimate and confidence limits for the prognosis-group-adjusted hazard rates [53].*"

One exemplar explained deviations from the planned interim analyses and the implications on their findings [54]. Finally, two RCTs discussed lessons learned from, value of, and implications of using a group sequential approach [55,56].

Discussion

Main findings

We found a significant number of confirmatory RCTs that employ interim analyses using group sequential methods, particularly in oncology, although the therapeutic areas of application appear to be diverse. Moreover, the majority of the group sequential RCTs were published in "*high impact*" peer-reviewed medical journals, and were often stopped early, predominantly for futility or efficacy.

We found inadequate reporting of CONSORT checklist items relating to the disclosure and access to trial protocols, methods used to generate the randomisation list(s), details of randomisation concealment and its implementation, and trial design aspects. Most concerning, is the lack of access to full trial protocols, with related amendments, in the public domain, for most of the examined group sequential RCTs. Hence, we could not ascertain the completeness in

reporting of other key checklist items, such as changes to methods and outcomes. Despite this, compliance in reporting of most CONSORT 2010 checklist items appeared to be very high.

In general, additional important features of group sequential RCTs are poorly reported. These encompass; rationale for choosing a group sequential design with any other add-on planned adaptations, mechanisms put in place to minimise operational bias due to the knowledge of interim results, lessons learned from using a group sequential design and its value, and clarification on whether sample size was adjusted for interim analyses. Most importantly, we found suboptimal reporting of the use of appropriate statistical methods for early stopping bias correction of point estimates, with associated CIs and P-values.

Reporting of planned stopping criteria, and stopping rules or boundaries employed, is still unsatisfactory, despite the fact that aspects are covered in the CONSORT 2010 statement. Nonetheless, some aspects of interim analyses, such as clarification of early stopping with reasons where applicable, description of; interim sample sizes (or number of events), number of planned interim analyses and timing, were adequately reported.

Interpretation of the findings

Our review findings are predominantly based upon group sequential RCTs, published in "*high impact*" medical journals, particularly in oncology. Hence, the general quality of compliance in reporting, may exaggerate what might be observed based on reports in other therapeutic areas, or lower impact journals for some checklist items [57–59]. For instance, suboptimal compliance to most checklist items has been reported in previous reviews in other therapeutic areas [58,59]. Regardless of the publishing journal's impact factor, trial design and therapeutic area, suboptimal reporting of randomisation methods, and details of randomisation concealment and its implementation has been widely reported and is consistent with our findings [57–61]. Similar findings on these checklist items were also found in oncology, and moreover, inadequate reporting was associated with exaggerated, biased intervention effects [61].

Most importantly, our findings of very poor reporting and use, of statistical methods for bias correction for early stopping, are consistent with previous findings of a systematic review, focusing on trials which stopped early for benefit [62]. This dimension has been overlooked in some proposed modifications to the CONSORT 2010 statement [27].

We believe that our findings on areas requiring improvement provide a conservative picture, of the scale of the problem regarding compliance in reporting of group sequential specific aspects, which are vital for research consumers to make informed judgements about the quality of findings.

Implications to practice

Our review uncovered group sequential RCTs published predominantly in "*high impact*" medical journals. To some extent, this may provide assurance to sceptical researchers, who may have concerns pertaining to poor receptiveness by journal editors and reviewers towards adaptive designs, when RCTs are stopped early [15]. In contrast, suboptimal reporting of appropriate statistical methods for early stopping bias correction, may influence some research consumers, who are aware of the phenomenon of exaggerated intervention effects when a naïve statistical approach is used, to consider findings from group sequential RCTs with a degree of scepticism. Research consumers, such as clinicians and regulators, may be reluctant to accept these findings in order to change medical practice, when trials are stopped early coupled with failure to implement bias correction, and poor communication and reporting of the corrective actions taken [15,62].

The phenomenon of exaggeration of the intervention effects in group sequential RCTs, following early stopping when a naïve statistical approach is used, has been widely debated and highlighted [62–69]. Although much attention has been paid to group sequential RCTs that are stopping early for benefit [62,63,66,68], the consequences could be similar when trials are stopped early for futility, since the evidence can be used to withdraw intervention(s) already in the care pathway. More so, it could be argued that the consequences on future evidence synthesis, through meta-analysis, should be treated similarly regardless of the reasons for early stopping.

Although there does not exist a unique solution to early stopping correction of statistical bias, various statistical procedures have been proposed [67,70–76], and it appears that these methods are rarely implemented by statisticians in routine practice. What is unclear, is the extent of the impact of statistical bias correction, on the results and decision making of these group sequential RCTs, particularly those that are stopped early. Some examples of RCTs, where interpretation of findings changed after bias correction, have been reported [62]. In contrast, one case study reported consistent interpretation of findings from using a naïve statistical approach, and bias correction under various methods [67]. The lack of knowledge of the impact that inaccurate analyses has on decision making among statisticians, lack of awareness of bias adjustment methods and unfamiliarity with mainstream statistical software(s) offering options to implement these procedures, could be contributing to their poor uptake. The scale of this problem in routine practice has been honestly articulated [67]:

> "One purpose of this article is to right a wrong I committed about a decade ago as part of the group reporting the results of The Randomized Aldactone Evaluation Study (RALES). . . . The trial stopped early after crossing an O'Brien–Fleming-like boundary (calculated through a Lan–DeMets spending function), and we reported the results as we saw them; we corrected neither the p-value nor the effect size for having stopped early. I have consoled myself for this lapse because I knew that other trials that had stopped early had also failed to correct the observed results for having stopped early."

The increasing use of add-on trial adaptations (such as sample size re-estimation and treatment selection in multiple arm studies), within the group sequential framework, adds more complexities requiring more transparency in the design, conduct and reporting of these trials, beyond the checklist items covered by the CONSORT 2010 statements [27].

Our findings support initiatives for mandatory publication of, not only full trial protocols, but also all related protocol amendments. Assessing the quality of reporting of some key aspects of RCTs, proved challenging without access to these important trial documents, most imperative for complex adaptive designs. We absolutely concur with previous findings, that assessment of methodological quality should be based on evaluation of both protocols and publications [61].

Interim analyses heighten anxiety among some research consumers, due to the potential for introducing operational bias to the conduct of RCTs, thereby undermining the scientific integrity, credibility and validity of the findings. The potential introduction of operational bias, due to the leaking or knowledge of interim results in adaptive RCTs, has been well described [13,77]. However, the extent and impact of operational bias on the findings and decision making in routine practice is less well-understood [27]. Therefore, it is imperative to report mechanisms put in place, to minimise and/or control for operational bias such as; who conducted the interim analyses, how data was transferred and results were communicated, who were the stakeholders in interim decision-making process, and how decisions were made. Although it is

difficult to prevent indirect inference of interim results, due to decisions made following an interim analysis; careful planning, implementation, and optimal reporting of the mechanisms put in place, may go a long way in alleviating research consumers' worries about operational bias.

Our findings of poor reporting or inaccessibility of, prior interim results at the interim reporting, or point of early stopping, could hinder the ability of consumers of research findings, to assess trends of the direction of the treatment effects and the potential effect of population drift. Although it is challenging to distinguish between natural and population drift induced by operation bias, access to prior interim results may help research consumers to make their own informed judgements, and alleviate some of the cited concerns.

Recommendations to practice

Based upon our findings and desire to improve: the reproducibility and planning of future trials, the acceptability of findings from group sequential RCTs and to reduce waste in trials research [78]; we recommend urgent cross-disciplinary extension to the CONSORT 2010 statement, specifically aspects tailored for group sequential designs, most of which are consistent with recent suggestions [27]. We believe this would improve the reporting of group sequential RCTs and any other add-on adaptations considered. These aspects may encompass, but are not limited to:

- The use of the term "*group sequential*" or at least "*interim analys(i/e)s*" in the Title or Abstract for easy identification of group sequential RCTs for future systematic reviews and meta-analyses;

- Provision for the rationale of choosing a group sequential design and any add-on adaptations used;

- Detailed description of the mechanisms put in place to minimise (and control for) operational bias due to the leaking or knowledge of the interim results;

- Clarification on the use of, and description of, appropriate statistical methods to obtain unbiased results (point estimates, CIs, and P-values) when trials are stopped early. This should also encompass description of appropriate statistical inferential methods used to account for add-on trial adaptations where appropriate;

- Disclosure of prior interim results on the primary endpoint(s) and related decisions made. Summary tables or figures showing trends in the results or citation of previous related publications may suffice;

- Discussion of the lessons learned by using a group sequential design, and any other add-on adaptations where appropriate, and its value, with implications to aid the planning of future trials;

- Discussion of the deviations from the planned group sequential design and any other add-on adaptations, and their implications on generalisability of the findings.

We also encourage better reporting of group sequential aspects, which are already covered briefly in the CONSORT 2010 statement, such as planned stopping criteria and rules or boundaries used. Finally, more attention by journal editors and reviewers, and researchers, should be given to complete reporting of trial design, including any add-on adaptations; methods to generate randomisation list(s); randomisation concealment, and its implementation; and disclosure of full trial protocols (with related amendments).

Strengths and limitations

The need for this review has been based upon concerns raised by researchers, decision makers and policymakers regarding robustness and credibility of adaptive designs in decision making, to change practice when trials are stopped early [15]. We provided a comprehensive examination of compliance in the reporting of group sequential RCTs, utilising all accessible publication related reports using an improved classification system. Furthermore, the systematic search was supplemented with known group sequential RCTs from another source, and provided exemplars which could be used to enhance adequate reporting. However, one of the major limitations is that the literature search was restricted to Ovid MEDLINE due to resources and time limitations. Moreover, group sequential RCTs are not systematically indexed in the titles and abstracts and their key characteristics are poorly reported. Hence, our searching of Ovid MEDLINE could have missed a significant number of eligible trials. Inaccessibility of trials protocols and associated amendments hampered the assessment of some key CONSORT checklist items, such as changes to methods and outcomes. Finally, exploration of factors associated with suboptimal reporting was out of scope of this research.

Conclusions

There are issues with suboptimal reporting of key group sequential specific characteristics, such as disclosure of the use of appropriate adjusted inferential methods following early stopping. Suboptimal reporting of bias correction methods could potentially imply most group sequential trials stopping early are giving biased results of treatment effects. These issues may partly explain the cited concerns about robustness and acceptability of the methodology to change practice, when trials are stopped early. These concerns could be alleviated by modifications to the CONSORT 2010 statement. Assurance of scientific rigour through transparent and adequate reporting of RCTs is paramount to the acceptability of findings from adaptive designs in general. There is an urgent need for CONSORT statement(s) tailored for adaptive design(s). Improvements in the reporting of general CONSORT checklist items relating to access to trial protocols, trial design aspects, methods to generate randomisation list(s), concealment of the randomisation and its implementation is required.

We hope our findings will enlighten researchers on potential ways to address some of the concerns we highlighted. We hope our results will also inform the CONSORT working group regarding the need for modifications to enhance the conduct and reporting of group sequential RCTs and adaptive trials in general.

Supporting Information

S1 Appendix. PROSPERO protocol registration correspondence.
(PDF)

S1 Dataset. Raw data on completeness in reporting of reviewed RCTs.
(XLSX)

S1 Table. Summary data of compliance in the reporting of group sequential specific and researcher-proposed modified checklist items.
(DOCX)

S2 Table. Summary data of compliance in the reporting of general CONSORT 2010 checklist items.
(DOCX)

Acknowledgments

Disclaimer

The views expressed are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research, the Department of Health or organisations affiliated to or funding them.

Author Contributions

Conceived and designed the experiments: MD. Performed the experiments: MD AS. Analyzed the data: MD. Wrote the paper: MD AS ST SAJ CLC JN DH. Contributed to the development of the study protocol: MD AS ST SAJ CLC JN DH. Wrote the analysis plan and conducted the analysis: MD. Led the writeup of the manuscript: AS MD. Supervised MD as part of the NIHR Doctoral Research Fellowship: ST JN SAJ. Advised MD as part of the NIHR Doctoral Research Fellowship: DH CLC.

References

1. Quinlan J, Gaydos B, Maca J, Krams M (2010) Barriers and opportunities for implementation of adaptive designs in pharmaceutical product development. Clinical trials (London, England) 7: 167–173. doi: 10.1177/1740774510361542

2. Todd S (2007) A 25-year review of sequential methodology in clinical studies. Statistics in medicine 26: 237–252. doi: 10.1002/sim.2763 PMID: 17139691

3. FDA (2010) Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics.

4. Coffey CS, Levin B, Clark C, Timmerman C, Wittes J, et al. (2012) Overview, hurdles, and future work in adaptive designs: perspectives from a National Institutes of Health-funded workshop. Clinical trials (London, England) 9: 671–680. doi: 10.1177/1740774512461859

5. Kairalla J a, Coffey CS, Thomann M a, Muller KE (2012) Adaptive trial designs: a review of barriers and opportunities. Trials 13: 145. doi: 10.1186/1745-6215-13-145 PMID: 22917111

6. Elsäßer A, Regnstrom J, Vetter T, Koenig F, Hemmings RJ, et al. (2014) Adaptive clinical trial designs for European marketing authorization: a survey of scientific advice letters from the European Medicines Agency. Trials 15: 383. doi: 10.1186/1745-6215-15-383 PMID: 25278265

7. Morgan CC, Huyck S, Jenkins M, Chen L, Bedding a., et al. (2014) Adaptive Design: Results of 2012 Survey on Perception and Use. Therapeutic Innovation & Regulatory Science 48: 473–481.

8. Chang M, Chow S- C, Pong A (2006) Adaptive design in clinical research: issues, opportunities, and recommendations. Journal of biopharmaceutical statistics 16: 299–309; discussion 311–312. doi: 10.1080/10543400600609718 PMID: 16724486

9. Chuang-Stein C, Beltangady M (2010) FDA draft guidance on adaptive design clinical trials: Pfizer's perspective. Journal of biopharmaceutical statistics 20: 1143–1149. doi: 10.1080/10543406.2010.514456 PMID: 21058110

10. Gaydos B, Anderson KM, Berry D, Burnham N, Chuang-Stein C, et al. (2009) Good Practices for Adaptive Clinical Trials in Pharmaceutical Product Development. Therapeutic Innovation & Regulatory Science 43: 539–556.

11. Liu Q, Chi GYH (2010) Understanding the FDA guidance on adaptive designs: historical, legal, and statistical perspectives. Journal of biopharmaceutical statistics 20: 1178–1219. doi: 10.1080/10543406.2010.514462 PMID: 21058114

12. Chow S-C (2014) Adaptive clinical trial design. Annual review of medicine 65: 405–415. doi: 10.1146/annurev-med-092012-112310 PMID: 24422576

13. Chow S-C, Chang M (2008) Adaptive design methods in clinical trials—a review. Orphanet journal of rare diseases 3: 11. doi: 10.1186/1750-1172-3-11 PMID: 18454853

14. Quinlan J, Krams M (2006) Implementing adaptive designs: logistical and operational considerations. Drug Information Journal 40: 437–444.

15. Dimairo M, Boote J, Julious SA, Nicholl JP, Todd S (2015) Missing steps in a staircase: a qualitative study of the perspectives of key stakeholders on the use of adaptive designs in confirmatory trials. Trials 16: 430. doi: 10.1186/s13063-015-0958-9 PMID: 26416387

16. Begg C, Cho M, Eastwood S, Horton R, Moher D, et al. (1996) Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 276: 637–639. PMID: 8773637

17. Schulz KF, Altman DG, Moher D (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. Annals of internal medicine 152: 726–732. doi: 10.7326/0003-4819-152-11-201006010-00232 PMID: 20335313

18. Moher D, Schulz KF, Altman DG (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 357: 1191–1194. PMID: 11323066

19. Egger M, Jüni P, Bartlett C (2001) Value of flow diagrams in reports of randomized controlled trials. JAMA 285: 1996–1999. PMID: 11308437

20. Moher D, Jones A, Lepage L (2001) Use of the CONSORT Statement and Quality of Reports of Randomized Trials. JAMA 285: 1992. doi: 10.1001/jama.285.15.1992 PMID: 11308436

21. Turner L, Shamseer L, Altman DG, Schulz KF, Moher D (2012) Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. Systematic reviews 1: 60. doi: 10.1186/2046-4053-1-60 PMID: 23194585

22. Altman DG, Moher D, Schulz KF (2012) Improving the reporting of randomised trials: the CONSORT Statement and beyond. Statistics in medicine 31: 2985–2997. doi: 10.1002/sim.5402 PMID: 22903776

23. Campbell MK, Piaggio G, Elbourne DR, Altman DG (2012) Consort 2010 statement: extension to cluster randomised trials. BMJ (Clinical research ed) 345: e5661. doi: 10.1136/bmj.e5661

24. Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG (2012) Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA 308: 2594–2604. doi: 10.1001/jama.2012.87802 PMID: 23268518

25. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, et al. (2008) Improving the reporting of pragmatic trials: an extension of the CONSORT statement. BMJ (Clinical research ed) 337: a2390. doi: 10.1136/bmj.a2390

26. The EQUATOR Network (n.d.) Enhancing the QUAlity and Transparency Of health Research: Reporting guidelines under development. Available: http://www.equator-network.org/library/reporting-guidelines-under-development/. Accessed 23 September 2014.

27. Detry M, Lewis R, Broglio K, Connor J (2012) Standards for the Design, Conduct, and Evaluation of Adaptive Randomized Clinical Trials.

28. SAACTD Workshop Committee (2009) Connecting Non-Profits to Adaptive Clinical Trial Designs: Themes and Recommendations from the Scientific Advances in Adaptive Clinical Trial Designs Workshop. Available: https://custom.cvent.com/536726184EFD40129EF286585E55929F/files/2627e73646ce4733a2c03692fab26fff.pdf. Accessed 3 October 2014.

29. U.S. National Institutes of Health (n.d.) ClinicalTrials.gov. Available: https://clinicaltrials.gov/. Accessed 1 June 2014.

30. Stovold E, Beecher D, Foxlee R, Noel-Storr A (2014) Study flow diagrams in Cochrane systematic review updates: an adapted PRISMA flow diagram. Systematic reviews 3: 54. doi: 10.1186/2046-4053-3-54 PMID: 24886533

31. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS medicine 6: e1000097. doi: 10.1371/journal.pmed.1000097 PMID: 19621072

32. Newcombe RG (1998) Two-sided confidence intervals for the single proportion: Comparison of seven methods. Statistics in Medicine 17: 857–872. doi: 10.1002/(SICI)1097-0258(19980430)17:8<857::AID-SIM777>3.0.CO;2-E PMID: 9595616

33. Efron B (1979) Bootstrap Methods: Another Look at the Jackknife. The Annals of Statistics 7: 1–26.

34. Lan K, DeMets D (1983) Discrete sequential boundaries for clinical trials. Biometrika 70: 659–663.

35. O'Brien P, Fleming T (1979) A multiple testing procedure for clinical trials. Biometrics 35: 549–556. PMID: 497341

36. Pocock SJ (1977) Group Sequential Methods in the Design and Analysis of Clinical Trials. Biometrika 64: 191. doi: 10.2307/2335684

37. Haybittle J (1971) Repeated assessment of results in clinical trials of cancer treatment. The British journal of radiology 44: 793–797. PMID: 4940475

38. Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, et al. (1977) Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. British journal of cancer 35: 1–39. PMID: 831755

39. Pampallona S, Tsiatis A a. (1994) Group sequential designs for one-sided and two-sided hypothesis testing with provision for early stopping in favor of the null hypothesis. Journal of Statistical Planning and Inference 42: 19–35. doi: 10.1016/0378-3758(94)90187-2

40. Hwang IK, Shih WJ, De Cani JS (1990) Group sequential designs using a family of type I error probability spending functions. Statistics in medicine 9: 1439–1445. PMID: 2281231

41. Jennison C, Turnbull BW (2000) Flexible Monitoring: The Error Spending Approach. Group Sequential Methods with Applications to Clinical Trials. Florida, USA: CHAPMAN & HALL/CRC. pp. 148–149.

42. Wang S, Tsiatis A (1987) Approximately optimal one-parameter boundaries for group sequential trials. Biometrics 43: 193–199. PMID: 3567304

43. Fleming TR (1982) One-sample multiple testing procedure for phase II clinical trials. Biometrics 38: 143–151. PMID: 7082756

44. Whitehead J A unified theory for sequential clinical trials. Statistics in medicine 18: 2271–2286. PMID: 10474138

45. Tröger W, Galun D, Reif M, Schumann A, Stanković N, et al. (2013) Viscum album [L.] extract therapy in patients with locally advanced or metastatic pancreatic cancer: a randomised clinical trial on overall survival. European journal of cancer (Oxford, England: 1990) 49: 3788–3797. doi: 10.1016/j.ejca.2013.06.043

46. Butts C, Socinski M a, Mitchell PL, Thatcher N, Havel L, et al. (2014) Tecemotide (L-BLP25) versus placebo after chemoradiotherapy for stage III non-small-cell lung cancer (START): a randomised, double-blind, phase 3 trial. The Lancet Oncology 15: 59–68. doi: 10.1016/S1470-2045(13)70510-2 PMID: 24331154

47. Middleton G, Silcocks P, Cox T, Valle J, Wadsley J, et al. (2014) Gemcitabine and capecitabine with or without telomerase peptide vaccine GV1001 in patients with locally advanced or metastatic pancreatic cancer (TeloVac): an open-label, randomised, phase 3 trial. The Lancet Oncology 15: 829–840. doi: 10.1016/S1470-2045(14)70236-0 PMID: 24954781

48. Pritchett Y, Jemiai Y, Chang Y, Bhan I, Agarwal R, et al. (2011) The use of group sequential, information-based sample size re-estimation in the design of the PRIMO study of chronic kidney disease. Clinical trials (London, England) 8: 165–174. doi: 10.1177/1740774511399128

49. Thadhani R, Wenger J, Tamez H, Cannata J, Thompson BT, et al. (2012) Vitamin D Therapy and Cardiac Structure and Function in Patients With Chronic Kidney Disease. 02114.

50. Wolff AC, Lazar AA, Bondarenko I, Garin AM, Brincat S, et al. (2013) Randomized phase III placebo-controlled trial of letrozole plus oral temsirolimus as first-line endocrine therapy in postmenopausal women with locally advanced or metastatic breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 31: 195–202. doi: 10.1200/JCO.2011.38.3331

51. Mehta RS, Barlow WE, Albain KS, Vandenberg T a, Dakhil SR, et al. (2012) Combination Anastrozole and Fulvestrant in Metastatic Breast Cancer. The New England journal of medicine 367: 435–444. doi: 10.1056/NEJMoa1201622 PMID: 22853014

52. Chew EY, Clemons TE, Bressler SB, Elman MJ, Danis RP, et al. (2014) Randomized trial of a home monitoring system for early detection of choroidal neovascularization home monitoring of the Eye (HOME) study. Ophthalmology 121: 535–544. doi: 10.1016/j.ophtha.2013.10.027 PMID: 24211172

53. Emerson S, Fleming T (1990) Parameter estimation following group sequential hypothesis testing. Biometrika 77: 875–892.

54. Mascia L, Pasero D, Slutsky AS, Arguis MJ, Berardino M, et al. (2010) Effect of a lung protective strategy for organ donors on eligibility and availability of lungs for transplantation: a randomized controlled trial. JAMA 304: 2620–2627. doi: 10.1001/jama.2010.1796 PMID: 21156950

55. Moore MJ, Hamm J, Dancey J, Eisenberg PD, Dagenais M, et al. (2003) Comparison of gemcitabine versus the matrix metalloproteinase inhibitor BAY 12–9566 in patients with advanced or metastatic adenocarcinoma of the pancreas: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. Journal of clinical oncology: official journal of the American Society of Clinical Oncology 21: 3296–3302. doi: 10.1200/JCO.2003.02.098

56. Markman M, Liu PY, Wilczynski S, Monk B, Copeland LJ, et al. (2003) Phase III randomized trial of 12 versus 3 months of maintenance paclitaxel in patients with advanced ovarian cancer after complete response to platinum and paclitaxel-based chemotherapy: a Southwest Oncology Group and Gynecologic Oncology Group trial. Journal of clinical oncology: official journal of the American Society of Clinical Oncology 21: 2460–2465. doi: 10.1200/JCO.2003.07.013

57. Yao AC, Khajuria A, Camm CF, Edison E, Agha R (2014) The reporting quality of parallel randomised controlled trials in ophthalmic surgery in 2011: a systematic review. Eye (London, England) 28: 1341–1349. doi: 10.1038/eye.2014.206

58. Hurst D (2011) Quality of reporting randomised controlled trials in major dental journals suboptimal. Evidence-based dentistry 12: 52–53. doi: 10.1038/sj.ebd.6400796 PMID: 21701551

59. Sjögren P, Halling A (2002) Quality of reporting randomised clinical trials in dental and medical research. British dental journal 192: 100–103. doi: 10.1038/sj.bdj.4801304a PMID: 11838001

60. Camm CF, Chen Y, Sunderland N, Nagendran M, Maruthappu M, et al. (2013) An assessment of the reporting quality of randomised controlled trials relating to anti-arrhythmic agents (2002–2011). International journal of cardiology 168: 1393–1396. doi: 10.1016/j.ijcard.2012.12.020 PMID: 23298560

61. Mhaskar R, Djulbegovic B, Magazin A, Soares HP, Kumar A (2012) Published methodological quality of randomized controlled trials does not reflect the actual quality assessed in protocols. Journal of clinical epidemiology 65: 602–609. doi: 10.1016/j.jclinepi.2011.10.016 PMID: 22424985

62. Montori VM, Devereaux PJ, Adhikari NKJ, Burns KEA, Eggert CH, et al. (2005) Randomized trials stopped early for benefit: a systematic review. JAMA 294: 2203–2209. doi: 10.1001/jama.294.17.2203 PMID: 16264162

63. Bassler D, Montori VM, Briel M, Glasziou P, Guyatt G (2008) Early stopping of randomized clinical trials for overt efficacy is problematic. Journal of clinical epidemiology 61: 241–246. doi: 10.1016/j.jclinepi.2007.07.016 PMID: 18226746

64. Bassler D, Briel M, Montori VM, Lane M, Glasziou P, et al. (2010) Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. JAMA 303: 1180–1187. doi: 10.1001/jama.2010.310 PMID: 20332404

65. Wears RL (2015) Are We There Yet? Early Stopping in Clinical Trials. Annals of Emergency Medicine 65: 214–215. doi: 10.1016/j.annemergmed.2014.12.020 PMID: 25601251

66. Zhang JJ, Blumenthal GM, He K, Tang S, Cortazar P, et al. (2012) Overestimation of the effect size in group sequential trials. Clinical cancer research: an official journal of the American Association for Cancer Research 18: 4872–4876. doi: 10.1158/1078-0432.CCR-11-3118

67. Wittes J (2012) Stopping a trial early—and then what? Clinical trials (London, England) 9: 714–720. doi: 10.1177/1740774512454600

68. Zannad F, Stough WG, McMurray JJ V, Remme WJ, Pitt B, et al. (2012) When to stop a clinical trial early for benefit: Lessons learned and future approaches. Circulation: Heart Failure 5: 294–302. doi: 10.1161/CIRCHEARTFAILURE.111.965707 PMID: 22438522

69. Freidlin B, Korn EL (2009) Stopping clinical trials early for benefit: impact on estimation. Clinical trials (London, England) 6: 119–125. doi: 10.1177/1740774509102310

70. Kim K (1989) Point estimation following group sequential tests. Biometrics 45: 613–617. PMID: 2765641

71. Todd S, Whitehead J, Facey K (1996) Point and interval estimation following a sequential clinical trial. Biometrika 83: 453–461.

72. Liu A, Hall W (1999) Unbiased estimation following a group sequential test. Biometrika 86: 71–78.

73. Whitehead J (1986) On the bias of maximum likelihood estimation following a sequential test. Biometrika 73: 573–581.

74. Emerson SS (1993) Computation of the uniform minimum variance unbiased estimator of a normal mean following a group sequential trial. Computers and biomedical research, an international journal 26: 68–73. PMID: 8444028

75. Emerson S, Kittelson J (1997) A computationally simpler algorithm for the UMVUE of a normal mean following a group sequential trial. Biometrics 53: 365–369. PMID: 9147601

76. Milanzi E, Molenberghs G, Alonso A, Kenward MG, Tsiatis AA, et al. (2014) Estimation After a Group Sequential Trial. Statistics in Biosciences.

77. Chow S- C, Corey R (2011) Benefits, challenges and obstacles of adaptive clinical trial designs. Orphanet journal of rare diseases 6: 79. doi: 10.1186/1750-1172-6-79 PMID: 22129361

78. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, et al. (2014) Reducing waste from incomplete or unusable reports of biomedical research. Lancet 383: 267–276. doi: 10.1016/S0140-6736(13)62228-X PMID: 24411647