

REVIEW ARTICLES

An analysis of protocols and publications suggested that most discontinuations of clinical trials were not based on preplanned interim analyses or stopping rules

Mihaela Stegert^{a,b}, Benjamin Kasenda^{a,c}, Erik von Elm^d, John J. You^{e,f}, Anette Blümle^g, Yuki Tomonaga^h, Ramon Saccilotto^a, Alain Amstutz^a, Theresa Bengough^{d,i}, Matthias Briel^{a,e,*}, and the DISCO study group

^aBasel Institute for Clinical Epidemiology and Biostatistics, Department of Clinical Research, University Hospital Basel, Hebelstrasse 10, Basel 4031, Switzerland

^bClinic of Internal Medicine, University Hospital Basel, Petersgraben 4, Basel 4031, Switzerland

^cClinic of Medical Oncology, University Hospital Basel, Petersgraben 4, Basel 4031, Switzerland

^dCochrane Switzerland, Institute of Social and Preventive Medicine (IUMSP), Lausanne University Hospital, Route de la Corniche 10, Lausanne 1010, Switzerland

^eDepartment of Clinical Epidemiology and Biostatistics, McMaster University, 1280 Main Street West, HSC-2C8, Hamilton, Ontario, Canada L8S 4K1

^fDepartment of Medicine, McMaster University, 1280 Main Street West, HSC-2C8, Hamilton, Ontario, Canada L8S 4K1

^gGerman Cochrane Centre, Medical Center, University of Freiburg, Berliner Allee 29, Freiburg 79110, Germany

^hEpidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, Zurich 8001, Switzerland

ⁱDepartment of Health and Society, Austrian Federal Institute for Health Care, Stubenring 6, Vienna 1010, Austria

Accepted 19 May 2015; Published online 4 June 2015

Abstract

Objectives: To investigate the frequency of interim analyses, stopping rules, and data safety and monitoring boards (DSMBs) in protocols of randomized controlled trials (RCTs); to examine these features across different reasons for trial discontinuation; and to identify discrepancies in reporting between protocols and publications.

Study Design and Setting: We used data from a cohort of RCT protocols approved between 2000 and 2003 by six research ethics committees in Switzerland, Germany, and Canada.

Results: Of 894 RCT protocols, 289 prespecified interim analyses (32.3%), 153 stopping rules (17.1%), and 257 DSMBs (28.7%). Overall, 249 of 894 RCTs (27.9%) were prematurely discontinued; mostly due to reasons such as poor recruitment, administrative reasons, or unexpected harm. Forty-six of 249 RCTs (18.4%) were discontinued due to early benefit or futility; of those, 37 (80.4%) were stopped outside a formal interim analysis or stopping rule. Of 515 published RCTs, there were discrepancies between protocols and publications for interim analyses (21.1%), stopping rules (14.4%), and DSMBs (19.6%).

Conclusion: Two-thirds of RCT protocols did not consider interim analyses, stopping rules, or DSMBs. Most RCTs discontinued for early benefit or futility were stopped without a prespecified mechanism. When assessing trial manuscripts, journals should require access to the protocol. © 2016 Elsevier Inc. All rights reserved.

Keywords: Randomized controlled trials; Protocol; Early termination of clinical trials; Data monitoring committees; Interim analysis; Stopping Rules

1. Introduction

Randomized controlled trials (RCTs) are widely accepted as the principal research method for the assessment of health care interventions [1,2]. Stopping RCTs

prematurely for benefit, harm, futility, or other reasons (e.g., slow recruitment of participants) requires consideration of ethical, statistical, and logistical issues [3–6]. Preplanned interim analyses, stopping rules, and the presence of a data safety and monitoring board (DSMB) are means

Ethical approval: The participating Research Ethics Committees approved the study or explicitly stated that no ethical approval was necessary.

* Corresponding author: Tel.: +41-0-61-265-3100; fax: +41-0-61-265-3109.

E-mail address: matthias.briel@usb.ch (M. Briel).

What is new?

- Two-thirds of randomized controlled trial (RCT) protocols did not consider interim analyses, stopping rules, or data safety and monitoring boards (DSMBs).
- Discrepancies in the reporting of interim analyses, stopping rules, and DSMBs between protocols and publications were common.
- More than 80% (37 of 46) of RCTs discontinued for benefit or futility were stopped outside a matching interim analysis or stopping rule.
- Interim analysis plans, stopping guides, and DSMBs increase transparency and credibility when trials are stopped early. Their adequate reporting in both protocols (according to Standard Protocol Items Recommendations for Interventional Trial) and publications (according to Consolidated Standards of Reporting Trial) should be required.

to increase transparency and credibility of decision making during the course of RCTs, in particular with respect to premature trial discontinuation. The Consolidated Standards of Reporting Trials (CONSORT) statement recommends reporting of the timing of interim analyses, conditions required for initiating interim analyses, the number of analyses, whether they were planned or ad hoc, the presence of a priori stopping rules, and if an independent DSMB was in place [1]. Investigators planning interim analyses should document such plans in their trial protocols, including pre-specification of the statistical approach used to judge the possible need for premature discontinuation [4]. Since 2002, several regulatory guidelines about DSMBs have been published [7–11]. DSMBs should keep interim trial data confidential, and statistical criteria (stopping rules) should be treated as guidelines rather than strict rules for stopping early [12].

Little is known about the frequency of planned interim analyses and stopping rules in protocols or about the proportion of RCTs overseen by independent DSMBs. It also remains uncertain whether such plans have an impact on RCT discontinuation. In a cohort of RCT protocols approved by research ethics committees (RECs) between 2000 and 2003 (before most regulatory guidelines about DSMBs), we investigated (1) the prevalence of planned and reported interim analyses, stopping rules, and DSMBs in RCTs from various medical fields; (2) the association of these features with other trial characteristics; (3) these features across different reasons for trial discontinuation; and (4) discrepancies between planning of interim analyses, stopping rules, and DSMBs in RCT protocols and reporting in corresponding journal publications.

2. Methods*2.1. Study design*

This is a planned substudy of an international, multicenter cohort of RCT protocols approved by six RECs in Switzerland (Basel, Lucerne, Lausanne, and Zurich), Germany (Freiburg), and Canada (Hamilton, Ontario) between 2000 and 2003. The rationale and design of the retrospective cohort as well as other results have previously been published [13,14].

2.2. Eligibility criteria for protocols and corresponding publications

In the present study, we included RCT protocols regardless of publication status. We excluded protocols of RCTs that (1) compared different doses or routes of administration of the same drug (dose-finding studies), (2) enrolled only healthy volunteers, (3) were never started, or (4) were still ongoing as of April 2013. For the analysis of reporting, we included only full (peer-reviewed) journal publications corresponding to included RCT protocols. Any other formats such as research letters, letters to the editor, or conference abstracts were excluded.

2.3. Search for publications and data extraction

If the REC files provided no information about the publication status of a trial, we conducted comprehensive searches of electronic databases (MEDLINE, EMBASE, Google Scholar, Cochrane CENTRAL register of clinical trials, trial registries such as ClinicalTrials.gov and registries of sponsors, if publicly available) to find any associated publications [13]. If trial publication or completion status still remained unclear, the REC in charge conducted a survey of investigators by sending them a standardized questionnaire (see [Appendix](#) at www.jclinepi.com). The response rate was 80.3%; 240 of 299 investigators returned the questionnaire [14].

Twelve investigators trained in clinical research methodology independently extracted data from eligible trial protocols including study design, number of centers, clinical field, details about the intervention arms, trial funding and initiation, planned sample size, follow-up time, and planned statistical analyses [13]. The initial 30% of the protocol extractions were completed in duplicate. Twenty-two investigators trained in clinical research methodology, independently and in duplicate, extracted data from all corresponding publications. Any disagreements were resolved by consensus or third party adjudication. We used a Web-based password-protected database for all data abstraction (Squiekero, www.squiekero.org).

2.4. Definitions

We considered an RCT discontinued if the investigators indicated discontinuation with a reason in their correspondence with the REC, in a journal publication, or their response

to our survey. If we could not elucidate the reason for trial discontinuation or if poor participant recruitment was mentioned, we used a prespecified cutoff of less than 90% of achieved target sample size to determine discontinuation [13,14]. We considered a trial as discontinued based on a prespecified stopping rule if the reported reason for discontinuation matched the purpose of a planned interim analysis or stopping rule mentioned in the protocol. We defined an RCT protocol as having at least one interim analysis if it explicitly mentioned an interim analysis; and we defined an RCT protocol as having a stopping rule if it explicitly mentioned the use of a statistical rule or numerical threshold to check whether the trial should be stopped prematurely (e.g., O'Brien-Fleming, alpha spending function, conditional power, and so forth).

Reviewers classified RCT protocols as industry sponsored or investigator sponsored if the protocol clearly named the sponsor, displayed a company or institution logo prominently, mentioned industry affiliations for protocol authors, included statements about data ownership or publication rights, or statements about full funding by industry or public funding agencies [14,15]. Disagreements were resolved by consensus.

2.5. Statistical analysis

For binary data, we summarized results as frequencies and proportions and for continuous data as medians and interquartile ranges (IQRs). For comparison of proportions, we used the chi-squared test with continuity correction. We considered two data sets: (1) all RCT protocols involving patients ($n = 894$) and (2) the subgroup of RCT protocols involving patients with their corresponding publications ($n = 515$) to investigate discrepancies in the reporting of interim analyses, stopping rules, and DSMBs (Fig. 1).

We used three multivariable logistic regression models to investigate the association between RCT characteristics (independent variables) and the three outcomes: (1) interim

analyses (planned vs. not planned), (2) stopping rules (specified vs. not specified), and (3) DSMBs (appointed vs. not appointed) according to protocols. RCT characteristics included: length of follow-up (in years as continuous variable), sample size (in increments of 100), sponsorship (industry vs. investigator), number of centers (single center vs. multicentre), national vs. international, medical field (oncology, cardiovascular, infectious disease, and other), and type of control intervention (placebo/no intervention vs. active intervention). We calculated unadjusted and adjusted odds ratios (ORs) with 95% confidence intervals (CIs). We conducted a sensitivity analysis omitting length of follow-up as an independent variable in the regression models because of many missing values.

To investigate the association between premature trial discontinuation for early benefit or futility and the presence of interim analyses or stopping rules and DSMBs adjusted for other trial characteristics, we used a multivariable analysis with early discontinuation for benefit or futility (yes vs. no) as the dependent variable and the following independent variables: interim analysis OR stopping rule (planned vs. not planned), presence of a DSMB (yes vs. no), length of follow-up (in years as continuous variable), sample size (in increments of 100), sponsorship (industry vs. investigator), number of centers (single center vs. multi-center), national vs. international. This analysis was carried out in a subsample including only RCTs discontinued for early benefit or futility and completed trials (sample $N = 303$). RCTs discontinued for other reasons than early benefit or futility were excluded because interim analyses or stopping rules for unexpected events are not plausible; in addition, we excluded RCTs with unclear completion status and RCTs planning an interim analysis only as part of an adaptive design, for sample size recalculation purposes only, or for other reasons without a stopping mechanism. Again, in a sensitivity analysis, we dropped length of follow-up from the model due to 286 missing values (increase in sample size from $N = 303$ to $N = 577$).

The goodness of fit of statistical models was evaluated using the Akaike information criterion [16]. The fit of the models did not improve by adding the participating REC as a random effect in a multilevel model; therefore, we omitted the random effect.

To investigate the agreement between protocols and publications with respect to the reporting of interim analyses, stopping rules, and DSMBs, we calculated (1) the crude agreement rate (in percent) and (2) the chance-corrected agreement (Cohen's kappa) [15,17]. All analyses were performed using R version 3.0.1 (www.r-project.org).

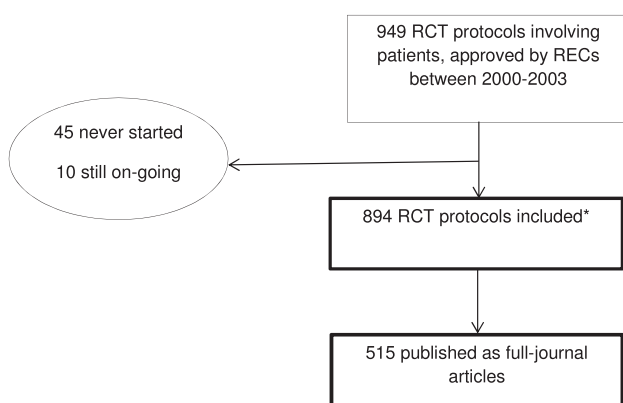


Fig. 1. Study flow. RCT, randomized controlled trial; REC, research ethics committee. *Number of protocols at each study center (collaborating REC): Basel (CH): 221; Freiburg (DE): 272; Lausanne (CH): 149; Lucerne (CH): 31; Hamilton (CA): 178; and Zürich (CH): 43. At Zürich, we only extracted RCT protocols from a subsidiary research ethics committee responsible for pediatric and surgical trials.

3. Results

Between 2000 and 2003, six participating RECs in Switzerland, Germany, and Canada approved 949 RCT

protocols enrolling patients (Fig. 1). Of these, 45 trials that were never started and 10 that were still ongoing at the time of analysis (April 2013) were excluded. We therefore included 894 RCT protocols of which 515 (57.6%) were published in one or more full journal articles. Most trials were multicenter, industry-sponsored, parallel group RCTs with a pharmacological intervention as experimental arm (Table 1). The median planned sample size was 260 participants (IQR, 100–610) and the median follow-up 0.5 years (IQR, 0.2–1.5).

3.1. Planning of interim analyses, stopping rules, and DSMBs

Of 894 RCT protocols, 289 specified interim analyses (32.3%), 153 stopping rules (17.1%), and 257 (28.7%) mentioned the presence of a DSMB (Table 2). More than one interim analysis was planned for 103 RCTs (11.5%). The most frequently reported purposes for planned interim analyses were early benefit (170 of 894; 19.0%) and harm

(133 of 894; 14.9%; Appendix Table 3 at www.jclinepi.com). Details about the planned DSMBs such as composition or blinding status were rarely provided. Almost all RCT protocols for which a stopping rule was planned [149 of 153 (99.6%)], also mentioned interim analyses (Appendix Table 2 at www.jclinepi.com). Both an interim analysis and a DSMB were planned in 164 of 894 RCTs (18.3%), and there were 125 of 894 RCT protocols (14.0%) that described an interim analysis but no DSMB. There were 100 of 894 RCT protocols (11.2%) mentioning all three features, whereas 509 of 894 (56.9%) mentioned none. Most of the 249 discontinued RCTs [139 (55.8%)] did not plan any form of early stopping mechanism (interim analysis, stopping rule, or DSMB).

In multivariable logistic regression, we investigated the association of interrelated trial characteristics with the planning of interim analyses, stopping rules, and DSMBs (Appendix Tables 4 and 5 at www.jclinepi.com). Results suggested that longer follow-up, larger planned sample size, and multicenter status in RCTs were independently

Table 1. Characteristics of included RCT protocols by completion status

Characteristics	Discontinued (N = 249) (%)	Completed (N = 575) (%)	Unclear (N = 70) (%)	All (N = 894) (%)
Study design				
Parallel	235 (94.4)	534 (92.9)	67 (95.7)	836 (93.5)
Crossover	12 (4.8)	27 (4.7)	2 (2.9)	41 (4.6)
Factorial	2 (0.8)	13 (2.3)	0	15 (1.7)
Unclear	0	1 (0.3)	1 (1.4)	2 (0.2)
Planned centers				
Multicenter	195 (78.3)	496 (86.3)	50 (71.4)	741 (82.9)
Single center	54 (21.7)	75 (13)	20 (28.6)	149 (16.7)
Unclear	0	4 (0.7)	0	4 (0.4)
Trial setting				
National	117 (47)	155 (26.9)	37 (52.8)	309 (34.5)
International	127 (51)	401 (69.7)	27 (3.8)	555 (62.1)
Unclear	5 (2.0)	19 (3.3)	6 (8.6)	30 (3.4)
Type of control intervention ^a				
No active treatment/standard care	65 (26.1)	122 (21.2)	18 (25.7)	205 (22.9)
Active drug/other treatment	113 (45.4)	282 (49)	39 (55.7)	434 (48.5)
Placebo/sham procedure	88 (35.3)	240 (41.7)	18 (25.7)	346 (38.7)
Trial sponsorship				
Industry	119 (47.8)	394 (68.5)	38 (54.3)	551 (61.6)
Investigator	130 (52.2)	181 (31.5)	32 (45.7)	343 (38.4)
Clinical area				
Oncology	47 (18.9)	92 (16)	8 (11.4)	147 (16.4)
Cardiovascular	23 (9.2)	78 (13.6)	6 (8.6)	107 (12.0)
Infectious disease	19 (7.6)	45 (7.8)	12 (17.1)	76 (8.5)
Other ^b	160 (64.3)	360 (62.6)	44 (62.9)	564 (63.1)
Planned sample size				
Median (IQR)	200 (90–530)	330 (120–735)	150 (78–300)	260 (100–610) ^c
Planned follow-up time (yr)				
Median (IQR)	0.52 (0.25–2)	0.5 (0.23–1.5)	0.3 (0.08–1)	0.5 (0.23–1.5)
Not reported	133 (53.4)	279 (48.5)	47 (67.1)	459 (51.3) ^d

Abbreviations: RCT, randomized controlled trial; IQR, interquartile range.

Values are frequencies (column percentages) unless otherwise specified.

^a Categories were not mutually exclusive; it was possible to select several options for one RCT.

^b Includes pediatrics, surgery, and 20 different medical specialties each representing <7% of the total. A complete list is provided in Appendix Table 1 at www.jclinepi.com.

^c Missing data for target sample size in 12 protocols.

^d 18 of 459 RCTs without reported follow-up time had a time to event analysis.

Table 2. Information on interim analyses, stopping rules, and DSMBs in RCT protocols by completion status

Characteristics	Discontinued (N = 249) (%)	Completed (N = 575) (%)	Unclear (N = 70) (%)	All (N = 894) (%)
Interim analysis planned	88 (35.3)	180 (31.3)	21 (30)	289 (32.3)
Interim period definition				
After a certain number of recruited patients	46 (18.5)	86 (15.0)	15 (21.4)	147 (16.4)
After a specific interval	13 (5.2)	34 (5.9)	4 (5.7)	51 (5.7)
After reaching a specific number of events	10 (4)	26 (4.5)	0	36 (4)
At least two of the above three	5 (2)	18 (3.1)	1 (1.4)	24 (2.7)
After every event or patient	0	2 (0.3)	0	2 (0.2)
Other	0	2 (0.3)	1 (1.4)	3 (0.3)
Not reported	14 (5.6)	12 (2.1)	0	26 (2.9)
Number of planned interim analyses				
1	53 (21.3)	100 (17.4)	10 (14.3)	163 (18.2)
2	18 (7.2)	35 (6.0)	2 (2.9)	55 (6.2)
3	4 (1.6)	17 (3.0)	4 (5.7)	25 (2.8)
≥4	4 (1.6)	15 (2.6)	4 (5.7)	23 (2.6)
Not reported	9 (3.6)	13 (2.3)	1 (1.4)	23 (2.6)
Stopping rule defined	51 (20.5)	96 (16.7)	6 (8.6)	153 (17.1)
DSMB planned	63 (25.3)	181 (31.5)	13 (18.6)	257 (28.7)
Blinding status				
Blind	8 (3.2)	32 (5.6)	3 (4.3)	43 (4.8)
Definitely not blind	14 (5.6)	47 (8.2)	5 (7.1)	66 (7.4)
Unclear	41 (16.5)	102 (17.7)	5 (7.1)	148 (16.6)
Information on composition of DSMB				
Member expertise provided	10 (4)	47 (8.2)	3 (4.3)	60 (6.7)
Member names provided	5 (2.0)	11 (1.9)	1 (1.4)	17 (1.9)
Member affiliation provided	2 (0.8)	0	1 (1.4)	3 (0.3)
Two of above three	6 (2.4)	17 (2.9)	1 (1.4)	24 (2.7)
All of above three	6 (2.4)	10 (1.7)	0	16 (1.8)
No information provided	32 (12.9)	81 (14)	6 (8.6)	119 (13.3)
Sponsor representative was member	4 (1.6)	24 (4.2)	2 (2.8)	30 (3.4)

Abbreviations: DSMB, data safety and monitoring board; RCT, randomized controlled trial.

associated with the planning of interim analyses, stopping rules, and DSMBs. Trials conducted in oncology compared with other clinical areas and investigator-sponsored trials (vs. industry-sponsored) more often planned interim analyses and stopping rules (analysis adjusted for other examined trial characteristics). We found a similar pattern for discontinued RCTs only, that is, discontinued RCTs without a preplanned stopping mechanism were on average smaller, mostly investigator-sponsored and single-center trials (Appendix Table 6 at www.jclinepi.com).

3.2. Association of premature discontinuation with interim analyses, stopping rules, and DSMBs

Of 249 discontinued RCTs, poor recruitment was the most frequent reason for discontinuation; 70 RCTs were primarily discontinued for harm, early benefit, or futility (Table 3). For 15 of 70 RCTs (21.4%), the protocol mentioned the purpose of the interim analysis that matched the actual reason for trial discontinuation (6 of 24 for harm, 5 of 9 for early benefit, and 4 of 37 for futility). When we leave aside the RCTs discontinued for harm (because stopping in case of unexpected harm is exactly what responsible investigators and expert DSMBs should do), we still note that 46 RCTs were discontinued for early benefit or futility, of which 37 (80.4%) were stopped outside a formal interim

analysis. A multivariable regression analysis that included only RCTs discontinued for early benefit or futility and completed trials showed that RCTs discontinued for early benefit or futility more often planned any interim analysis or stopping rule in their protocol than completed RCTs (adjusted OR, 3.74; 95% CI: 1.1, 12.7; Appendix Table 7 at www.jclinepi.com). However, the purpose of the interim analysis or stopping rule in the protocol did most of the time not match the actual reason for discontinuation.

3.3. Agreement between protocols and publications

Of 515 published RCTs, 132 (25.6%) reported interim analyses, 66 (12.8%) a stopping rule, and 147 (28.5%) a DSMB. Of 132 publications reporting an interim analysis, 33 (25.0%) did not specify its purpose. In protocols, the proportion of planned interim analysis with no reported purpose tended to be smaller [32 of 289 (11.1%); Appendix Table 3 at www.jclinepi.com]. Both interim analysis and DSMB were reported in 92 of 515 publications (17.9%). All three (i.e., interim analyses, stopping rule, and DSMB) were reported in 47 (9.1%).

The crude agreement between planning in protocols and reporting in publications was for interim analyses 78.9%, for stopping rules 85.6%, and for DSMBs 80.4% (Appendix Table 8 at www.jclinepi.com). Planning an

Table 3. Planned DSMBs, interim analyses, or stopping rules in RCT protocols by completion status and reasons of discontinuation

Trial status	All (N = 894) (%)	Planned DSMB ^a (%)	Planned interim analysis or stopping rule ^a (%)
Completed	575 (64.3)	181 (31.5)	186 (32.3)
Unclear	70 (7.8)	13 (18.6)	21 (30)
Discontinued	249 (27.9)	63 (25.3)	88 (35.3)
Poor recruitment ^b	100 (11.2)	18 (18.0)	27 (27.0)
Futility ^c	37 (4.1)	14 (37.8)	26 (70.2)
Administrative reasons ^d	36 (4.0)	8 (22.2)	8 (22.2)
Harm	24 (2.7)	11 (45.8)	11 (45.8)
Unknown reason ^e	24 (2.7)	3 (12.5)	5 (20.8)
Benefit	9 (1.0)	2 (22.2)	7 (77.8)
External evidence	8 (0.9)	4 (50)	2 (25)
Lack of funding	5 (0.6)	1 (20)	1 (20)
Other	6 (0.7)	1 (16.6)	0

Abbreviations: DSMB, data safety and monitoring board.

^a Columns include numbers and row percentages.

^b Some trials had an additional reason for discontinuation: benefit ($n = 1$), futility ($n = 2$), and other reasons ($n = 3$).

^c Includes randomized trials with adaptive designs that have been stopped after the first ($n = 5$) or second stage ($n = 1$).

^d Includes strategic decisions from companies, consequence of new requirements from regulatory bodies, and change of workplace of principal investigators.

^e Reason for not achieving 90% of target sample size remained unclear.

interim analysis in the protocol and not reporting it in the publication occurred in 80 (15.5%) of 515 RCTs and was more common than reporting an interim analysis in the publication without corresponding documentation in the protocol, which happened in 29 (5.6%) of 515 RCTs. The chance-corrected agreement between protocols and publications was moderate with Cohen's kappa values of 0.51 (95% CI: 0.43, 0.59) for the reporting of interim analyses, 0.47 (95% CI: 0.37, 0.57) for stopping rules, and 0.55 (95% CI: 0.47, 0.63) for DSMBs.

Of 103 RCTs reporting an interim analysis in both protocol and publication, 32 (31.1%) reported discrepant numbers of interim analyses, 55 (53.4%) reported consistent numbers, and 16 (15.5%) did not specify the number of interim analyses. In 70 RCTs (68.0%), the purpose of interim analyses was reported consistently in protocols and publications; there were discrepancies in the remaining 33 RCTs (32.0%).

All nine RCTs discontinued for early benefit and 18 of the 37 RCTs discontinued for futility were published as full journal articles. Of the nine publications of RCTs discontinued for early benefit, six mentioned a prespecified stopping mechanism, and three did not. Furthermore, for two RCTs discontinued for early benefit, a DSMB was specified in the protocol, whereas for three RCTs, a DSMB was mentioned in the publication. Of the 18 publications of RCTs discontinued for futility, 16 mentioned that the RCT was prematurely discontinued and 9 explicitly stated that the trial discontinuation was based on the recommendation of a DSMB.

4. Discussion

4.1. Main findings

We found that 32% of RCT protocols planned one or more interim analyses, 17% specified stopping rules, and

29% planned a DSMB. These design features were independently more common in RCT protocols with longer follow-up, multicenter status, and larger sample size and in oncology rather than other specialties. Of all started patient RCTs, 28% were discontinued prematurely; mostly due to unexpected reasons (poor recruitment, administrative reasons, and unexpected harm). Only 5% (46 of 894) of RCTs were discontinued for early benefit or futility. More than 80% of those (37 of 46) were stopped outside a formal interim analysis or stopping rule. Approximately 20% of the 515 published RCTs had discrepant reporting between protocols and publications for the presence of interim analyses, stopping rules, or DSMBs; chance-corrected agreement was moderate.

4.2. Strengths and limitations

The present analyses were part of a larger project investigating the prevalence and publication of discontinued RCTs in an international, multicenter cohort of RCT protocols across all medical fields [13,14]. We worked in close collaboration with RECs and had unrestricted access to their archived RCT protocols and corresponding files. We involved only trained methodologists in data abstraction and considered only a limited number of variables in the statistical models to minimize spurious associations. We allowed for a long follow-up period (10–13 years) between protocol approval and our last search for corresponding publications (April 2013) to maximize the number of published RCTs.

On the other hand, the long follow-up time may represent a limitation of our study. The included protocols from 2000 to 2003 might no longer reflect current practices of planning interim analyses, stopping rules, and DSMBs because in the meantime considerable efforts have been

made to improve the monitoring of RCTs through several regulatory guidelines [7–11,18]. A further limitation of our study was that we extracted only 30% of the protocol data in duplicate due to limited resources, that is, we used single data extraction for 70% of the protocol data potentially increasing extraction errors [14]. However, we used piloted extraction forms with detailed written instructions, conducted formal calibration exercises with all data extractors, and checked extractions from a random sample of protocols at several points during the process. Agreement was good, with no more than two discrepancies among 30 extracted variables [14]. A second investigator verified all data regarding discontinuation and publication of RCTs. The suboptimal quality of information found in protocols, in particular those of investigator-sponsored trials, limited our ability to extract details regarding interim analyses, stopping rules, and DSMBs or other trial characteristics such as trial phase. For many included RCTs, lack of clarity regarding the length of follow-up substantially limited the number of protocols without missing data for our regression analyses. All regression analyses were post hoc, and although sensitivity analyses omitting length of follow-up yielded similar results, findings should be interpreted cautiously.

4.3. Comparison with other studies

Most previous research addressing the use of interim analyses, stopping rules, and DSMBs has been based on data from RCT publications without taking into account whether trials were discontinued or completed [6,7,19–21]. One exception, a study by Chan et al. [22] comparing 70 protocols approved by a REC in Denmark with corresponding publications, reported that 13 (18%) mentioned interim analyses in protocols, but only 5 (7%) were mentioned in corresponding publications [22]. We found results similar to Chan et al., including a lower reporting rate of interim analyses in publications (26%) than in protocols (35%).

In a cross-sectional study including RCTs published in 2000 in 24 high-impact general medical and specialist journals, Sydes et al. found that interim analyses were reported in 107 (16%) and DSMBs in 120 (18%) of 662 trials. In the same study, 58 to 150 RCTs (39%) reporting a DSMB and/or interim analyses also described a statistical stopping rule [19]. Another study included 1,772 RCTs published in eight major general medical and specialist journals between 2000 and 2005; 586 (33%) reported an interim analysis and 470 (27%) a DSMB [20]. This study found that reporting of both interim analyses and DSMBs increased from 2000 to 2005. These may also explain the higher prevalence of interim analyses (26%) and DSMBs (29%) in our study, in which most of the 515 RCTs were published after 2005; Sydes et al. [19] examined RCTs published in 2000, and Chan et al. [22] included RCT protocols that were approved during 1994–95.

Similar to Sydes et al. [19], we found a positive association of DSMB presence with multicenter study design, larger sample size, and trials including a placebo arm. In addition, we showed that these trial characteristics were also associated with the use of interim analyses.

4.4. Implications

To decide if results from a particular RCT are trustworthy, clinicians and policy makers need to know whether interim analyses, stopping rules, or DSMBs were planned. Our study suggests that among RCTs discontinued for early benefit or futility, few mentioned a matching interim analysis or stopping rule in the protocol. DSMBs can help to increase transparency and credibility of the decision making during the course of RCTs, in particular with respect to early discontinuation. However, we found that only about a third of RCTs discontinued for early benefit or futility mentioned a DSMB in their protocol. This raises the possibility that many RCTs were discontinued after review of interim results, thus potentially introducing substantial bias. Most trialists did not inform readers about whether and how often they had planned to review outcome data or actually did so.

The concept of stopping for harm refers to both formal inferiority of a tested intervention compared with a reference intervention and unexpected safety concerns. In our cohort, of the 24 RCTs stopped early for harm, we found six corresponding protocols with a matching interim analysis or stopping rule, but we noted that of the 12 published RCTs discontinued for harm, 11 were actually stopped due to unexpected safety concerns. As Pocock [23] argues “it is generally difficult to define formal statistical guidelines when it comes to the plethora of potential safety problems that a new treatment might give rise to.” We should not draw any inferences from the small number of matching prespecified interim analyses and stopping rules in RCTs discontinued for harm because picking up unexpected safety problems when looking at all the available information is how an expert DSMB should function. However, with RCTs discontinued for harm, it is of concern that less than half of protocols (46%) mention a DSMB in the first place and only half (50%) are fully published RCTs.

Initiatives to improve the quality of RCT protocols such as the Standard Protocol Items Recommendations for Interventional Trials are urgently needed and should be endorsed by both RECs and funding agencies [11]. Similarly, the CONSORT statement for RCT reports, first issued in 1996 and updated last in 2010, could be used more effectively by academic institutions and journals to inform the research community about important design features of trials in a transparent and standardized manner [1]. Our results are based on protocols from 2000 to 2003, and practices might have changed over time; future research will need to examine whether updated or new reporting guidelines actually improved the planning and reporting of interim analyses, stopping rules, and DSMBs over the

last decade. As a simple measure, we suggest that investigators routinely submit their RCT protocol together with the manuscript reporting results to a journal (unless the protocol has already been published) and that editors and peer-reviewers consider both documents together.

5. Conclusions

This empirical study found that of 894 RCT protocols approved by one of six RECs between 2000 and 2003, two-thirds did not consider interim analyses, stopping rules, or DSMBs. Five percent of RCTs were discontinued for early benefit or futility, and most of these stopped without a formal mechanism specified in the protocol. Discrepancies in reporting between protocols and publications were common. When assessing trial manuscripts, journals should require the trial protocol, and adherence to reporting guidelines should be ensured.

Acknowledgments

The authors thank the presidents and staff of participating Research Ethics Committees from Switzerland (Basel, Lausanne, Zurich, and Lucerne), Germany (Freiburg), and Canada (Hamilton) for their continuous support and cooperation. DISCO study group: Mihaela Stegert^{1,24}, Benjamin Kasenda^{1,19}, Erik von Elm², John J. You^{3,5}, Anette Blümle⁴, Yuki Tomonaga⁶, Ramon Saccilotto¹, Alain Amstutz¹, Theresa Bengough^{2,28}, Joerg J. Meerpohl⁴, Kari A. O. Tikkinen^{3,7}, Ignacio Neumann^{3,29}, Alonso Carrasco-Labra^{3,25}, Markus Faulhaber³, Sohail Mulla³, Dominik Mertz^{3,20,21}, Elie A. Akl^{3,8}, Dirk Bassler⁹, Jason W. Busse^{3,26,27}, Ignacio Ferreira-González¹⁰, Francois Lamontagne¹¹, Alain Nordmann¹, Viktoria Gloy^{1,18}, Kelechi Kalu Olu¹, Heike Raatz¹, Lorenzo Moja¹², Rachel Rosenthal¹³, Shanil Ebrahim³, Stefan Schandelmaier^{1,14}, Xin Sun¹⁵, Per O. Vandvik¹⁶, Bradley C. Johnston^{17,22,23}, Martin A. Walter¹⁸, Bernard Burnand², Matthias Schwenkglenks⁶, Lars G. Hemkens¹, Heiner C. Bucher¹, Gordon H. Guyatt³, and Matthias Briel^{1,3}. ¹Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital of Basel, Hebelstrasse 10, 4031 Basel, Switzerland; ²Cochrane Switzerland, Institute of Social and Preventive Medicine (IUMSP), Lausanne University Hospital, Lausanne, Switzerland; ³Department of Clinical Epidemiology and Biostatistics, McMaster University, 1280 Main Street West, HSC-2C8, Hamilton, Ontario, Canada L8S 4K1; ⁴German Cochrane Center, Medical Center - University of Freiburg, Berliner Allee 29, 79110 Freiburg, Germany; ⁵Department of Medicine, McMaster University, 1280 Main Street West, HSC-2C8, Hamilton, Ontario, Canada L8S 4K1; ⁶Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland; ⁷Departments of Urology and Public Health, Helsinki

University Central Hospital and University of Helsinki, Helsinki, Finland; ⁸Department of Internal Medicine, American University of Beirut, Department of Internal Medicine, P.O. Box: 11-0236 Riad-El-Solh, Beirut 1107 2020, Lebanon; ⁹Department of Neonatology, University Hospital Zurich, Zurich, Switzerland; ¹⁰Epidemiology Unit, Department of Cardiology, Vall d'Hebron Hospital and CIBER de Epidemiología y Salud Publica (CIBER-ESP), Passeig Vall d'Hebron 119-129, 08005 Barcelona, Spain; ¹¹Center de Recherche Clinique du Center Hospitalier Universitaire de Sherbrooke, Université de Sherbrooke, CHUS, 3001 12e avenue nord, Sherbrooke, PQ, Canada J1H5N4; ¹²IRCCS Orthopedic Institute Galeazzi, Via Riccardo Galeazzi, 4 20161 Milano, Italy; ¹³Department of Surgery, University Hospital Basel, Spitalstrasse 26, 4031 Basel, Switzerland; ¹⁴Academy of Swiss Insurance Medicine, University Hospital of Basel, Schanzenstrasse 55, 4031 Basel, Switzerland; ¹⁵Center for Health Research, Kaiser Permanente Northwest, Portland, OR, USA; ¹⁶Norwegian Knowledge Center for the Health Services, Oslo, Norway; ¹⁷Department of Anesthesia & Pain Medicine, Hospital for Sick Children, 555 University Ave, Toronto, Ontario, Canada; ¹⁸Institute of Nuclear Medicine, University Hospital Bern, Freiburgstrasse 4, 3010 Bern, Switzerland; ¹⁹Clinic of Medical Oncology, University Hospital Basel, Basel Switzerland; ²⁰Department of Medicine, McMaster University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada; ²¹Michael G. DeGroot Institute for Infectious Diseases Research, McMaster University, 1280 Main Street West, Hamilton, ON, L8S 4K1, Canada; ²²Institute of Health Policy, Management and Evaluation, University of Toronto, 150 College St, Toronto, Ontario, Canada; ²³The Hospital for Sick Children Research Institute, 686 Bay Street, Toronto, Ontario, Canada, M5G-0A4; ²⁴University Hospital Basel, Clinic of Internal Medicine, Basel, Switzerland; ²⁵Evidence-Based Dentistry Unit, Faculty of Dentistry, Universidad de Chile, Santiago, Chile; ²⁶The Michael G. DeGroot Institute for Pain Research and Care, McMaster University, Hamilton, Canada; ²⁷Department of Anesthesia, McMaster University, Hamilton, Canada; ²⁸Austrian Federal Institute for Health Care, Department of Health and Society, Stubenring 6, 1010 Vienna, Austria; ²⁹Department of Internal Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile.

Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jclinepi.2015.05.023>.

References

- [1] Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010;63:e1–37.

- [2] Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663–94.
- [3] Bassler D, Montori VM, Briel M, Glasziou P, Guyatt G. Early stopping of randomized clinical trials for overt efficacy is problematic. *J Clin Epidemiol* 2008;61:241–6.
- [4] Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. *Lancet* 2005;365:1657–61.
- [5] Pocock SJ. When to stop a clinical trial. *BMJ* 1992;305:235–40.
- [6] Mills E, Cooper C, Wu P, Rachlis B, Singh S, Guyatt GH. Randomized trials stopped early for harm in HIV/AIDS: a systematic survey. *HIV Clin Trials* 2006;7:24–33.
- [7] Grant AM, Altman DG, Babiker AB, Campbell MK, Clemens FJ, Darbyshire JH, et al. Issues in data monitoring and interim analysis of trials. *Health Technol Assess* 2005;9:1–238, iii-iv.
- [8] Food and Drug Administration. Guidance for Clinical Trial Sponsors: Establishment and Operation of Clinical Trial Data Monitoring Committees. Food and Drug Administration 2006. Available at: <http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm127073.pdf>. Accessed September 2, 2015.
- [9] Committee for Medicinal Products for Human Use (CHMP). Guideline on Data Monitoring Committees. European Medicines Agency 2005. Available at: <http://osp.od.nih.gov/sites/default/files/resources/WC500003635.pdf>. Accessed September 2, 2015.
- [10] Council for International Organizations of Medical Sciences (CIOMS) in collaboration with the World Health Organization (WHO): International ethical guidelines for biomedical research involving human subjects. Geneva: Council for International Organizations of Medical Sciences (CIOMS); 2002.
- [11] Chan AW, Tetzlaff JM, Altman DG, Dickersin K, Moher D. SPIRIT 2013: new guidance for content of clinical trial protocols. *Lancet* 2013;381:91–2.
- [12] A proposed charter for clinical trial data monitoring committees: helping them to do their job well. *Lancet* 2005;365:711–22.
- [13] Kasenda B, von Elm EB, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Learning from failure—rationale and design for a study about discontinuation of randomized trials (DISCO study). *BMC Med Res Methodol* 2012;12:131.
- [14] Kasenda B, von Elm E, You J, Blumle A, Tomonaga Y, Saccilotto R, et al. Prevalence, characteristics, and publication of discontinued randomized trials. *JAMA* 2014;311:1045–51.
- [15] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- [16] Bozdogan H. Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika* 1987;52:345–70.
- [17] Grouven U, Bender R, Ziegler A, Lange S. [The kappa coefficient]. *Dtsch Med Wochenschr* 2007;132(Suppl 1):e65–8.
- [18] Food and Drug Administration. Code of Federal Regulations Title 21. Food and Drug Administration 2014. Available at: <http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?CFRPart=312&showFR=1>. Accessed September 2, 2015.
- [19] Sydes MR, Altman DG, Babiker AB, Parmar MK, Spiegelhalter DJ. Reported use of data monitoring committees in the main published reports of randomized controlled trials: a cross-sectional study. *Clin Trials* 2004;1:48–59.
- [20] Tharmanathan P, Calvert M, Hampton J, Freemantle N. The use of interim data and Data Monitoring Committee recommendations in randomized controlled trial reports: frequency, implications and potential sources of bias. *BMC Med Res Methodol* 2008;8:12.
- [21] Booth CM, Ohorodnyk P, Zhu L, Tu D, Meyer RM. Randomised controlled trials in oncology closed early for benefit: trends in methodology, results, and interpretation. *Eur J Cancer* 2011;47:854–63.
- [22] Chan AW, Hrobjartsson A, Jorgensen KJ, Gotzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ* 2008;337:a2299.
- [23] Pocock SJ. Current controversies in data monitoring for clinical trials. *Clin Trials* 2006;3:513–21.