Primary endpoint discrepancies were found in one in ten clinical drug trials. Results of an inception cohort study

van den Bogert, Cornelis A.; Souverein, Patrick C.; Brekelmans, Cecile T.M.; Janssen, Susan W.J.; Koëter, Gerard H.; Leufkens, Hubert G.M.; Bouter, Lex M.

published in
Journal of Clinical Epidemiology
2017

DOI (link to publisher)
10.1016/j.jclinepi.2017.05.012

document version
Publisher's PDF, also known as Version of record

Journal of Clinical Epidemiology 89 (2017) 199−208

Journal of Clinical Epidemiology

ORIGINAL ARTICLES

Primary endpoint discrepancies were found in one in ten clinical drug trials. Results of an inception cohort study

Cornelis A. van den Bogert[a,b,c], Patrick C. Souverein[a,*], Cecile T.M. Brekelmans[b], Susan W.J. Janssen[c], Gerard H. Koëter[b], Hubert G.M. Leufkens[a], Lex M. Bouter[d]

[a]*Division of Pharmacoepidemiology and Clinical Pharmacology, Department of Pharmaceutical Sciences, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Universiteitsweg 99, P.O. Box 80082, 3508 TB Utrecht, The Netherlands*

[b]*Central Committee on Research Involving Human Subjects (CCMO), Parnassusplein 5, P.O. Box 16302, 2500 BH The Hague, The Netherlands*

[c]*Department of Public Health and Health Services, National Institute for Public Health and the Environment (RIVM), Antonie van Leeuwenhoeklaan 9, P.O. Box 1, Bilthoven 3720 BA, The Netherlands*

[d]*Department of Epidemiology and Biostatistics, VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands*

Accepted 15 May 2017; Published online 20 May 2017

Abstract

Objective: To identify the occurrence and determinants of protocol-publication discrepancies in clinical drug trials.

Study Design and Setting: All published clinical drug trials reviewed by the Dutch institutional review boards in 2007 were analyzed. Discrepancies between trial protocols and publications were measured among key reporting aspects. We evaluated the association of trial characteristics with discrepancies in primary endpoints by calculating the risk ratio (RR) and 95% confidence interval (CI).

Results: Of the 334 published trials, 32 (9.6%) had a protocol/publication discrepancy in the primary endpoints. Among the subgroup of randomized controlled trials (RCTs; $N = 204$), 12 (5.9%) had a discrepancy in the primary endpoint. Investigator-initiated trials with and without industry (co-) funding were associated with having discrepancies in the primary endpoints compared with industry-sponsored trials (RR 3.7; 95% CI 1.4−9.9 and RR 4.4; 95% CI 2.0−9.5, respectively). Furthermore, other than phase 1−4 trials (vs. phase 1; RR 4.6; 95% CI 1.1−19.3), multicenter trials were also conducted outside the European Union (vs. single center; RR 0.2; 95% CI 0.1−0.6), not prospectively registered trials (RR 3.3; 95% CI 1.5−7.5), non-RCTs (vs. superiority RCT; RR 2.4; 95% CI 1.2−4.8) and, among the RCTs, crossover compared with a parallel group design (RR 3.7; 95% CI 1.1−12.3) were significantly associated with having discrepancies in the primary endpoints.

Conclusions: Improvement in completeness of reporting is still needed, especially among investigator-initiated trials and non-RCTs. To eliminate undisclosed discrepancies, trial protocols should be available in the public domain at the same time when the trial is published.

Keywords: Selective reporting; Discrepancies; Publication bias; Outcome reporting bias; Clinical trial; Primary endpoints; Clinical protocols

1. Introduction

Selective reporting is considered to be the most important cause of the poor reproducibility of biomedical research [1]. If mainly the positive results of a study are published, this may lead to overrepresentation of positive results and conclusions in the scientific literature [2]. Ignoring negative results can cause research waste, as futile experiments may be unnecessarily repeated [3]. Moreover, an inadequate description of the protocol of a study can frustrate replication of the study [4,5]. Transparency of the process from study protocols until publication remains therefore paramount in the responsible conduct of research. Complete and unbiased publication of clinical trials is an ethical and scientific obligation as recommendations and conclusions derived from clinical trials are often translated into clinical guidelines, and human participants were involved in obtaining the results [6].

One type of selective reporting is nonpublication [7]. Evidence across medical and geographical areas shows that approximately half of the clinical trials that are conducted

Conflict of interest: None.

Funding: This study was part of a research project funded by the Ministry of Health, Welfare and Sports of the Netherlands.

* Corresponding author. Tel.: +31-62-029-6516; fax: +31-30-253-9166.

E-mail address: p.c.souverein@uu.nl (P.C. Souverein).

http://dx.doi.org/10.1016/j.jclinepi.2017.05.012

What is new?

Key findings

- Ten percent of clinical drug trials (and 6% of the randomized clinical drug trials) had a discrepancy in the primary endpoint between the trial protocol and the subsequent publication.
- Investigator-initiated trials were more likely to have a protocol-publication discrepancy in the primary endpoint than industry-sponsored trials.

What this adds to what was known?

- The occurrence of protocol-publication discrepancies in the primary endpoints was substantially lower than what we expected based on the literature, especially among the randomized controlled trials.
- Numerous discrepancies were identified in the important reporting categories secondary endpoints, objectives, selection criteria, sample size, subgroup or other additional analyses, and methods used for data analysis.

What is the implication and what should change now?

- More transparency is needed in the trajectory from protocol until publication of clinical drug trials. A practical solution is that the original trial protocol is made available in the public domain at the same time with the trial publication.

are not being published in the scientific literature [8]. Another type is selective publication, meaning that at least some results are published, but with undisclosed discrepancies between the trial protocol and the publication. The first empirical study investigating the problem of selective publication found that 62% of the trials had at least one primary endpoint that was discrepant between the trial protocol and the trial publication [9]. As the main conclusions and recommendations of trials will be based on their primary outcome, this finding suggests that a substantial proportion of clinical evidence is biased due to selective publication. Other studies also showed an alarming amount of selective publication regarding subgroup analyses, sample size calculations, and sponsorship acknowledgment [10–13].

Although the existence of selective publication has been convincingly established among clinical trials starting 15–20 years ago [9,10], its occurrence may have decreased due to subsequent countermeasures. Governments, journals, pharmaceutical companies, and research communities have implemented requirements for trial registration and data sharing [14–20]. However, more recent evidence suggests that only limited progress has been made [10]. Empirical evidence is very limited on whether other aspects of trials are transparently reported, such as the selection criteria, sample size, subgroup or other additional analyses, and the methods used for data analysis. Therefore, we studied the occurrence of protocol-publication discrepancies, determinants of discrepancies in primary endpoints, and the association between discrepancies and the direction of trial conclusions in a cohort of clinical drug trials.

2. Methods

The design of the study has been published before [7]. In short, we selected all the clinical drug trials that were reviewed by the Dutch-accredited institutional review boards (IRBs) in 2007, and we followed these trials until publication as peer-reviewed article in the scientific literature. The results of the study on nonpublication have been published and showed that of the 574 trials in the cohort, 240 (42%) remained unpublished [21]. For this follow-up study, we included the 334 trials in the cohort of which we found at least one publication by January 2016 in the scientific literature presenting results (Fig. 1). The data source was ToetsingOnline, a database containing all clinical drug trials submitted to accredited IRBs in the Netherlands, overseen by the competent authority (the Central Committee on Research Involving Human Subjects, CCMO). Hence, the cohort consists of all clinical drug trials that were IRB reviewed in the Netherlands in 2007. The data sources for the discrepancy assessment were the IRB files of the CCMO including the original trial protocols and substantial amendments (as required by law, these documents are submitted to both the IRB and the CCMO before start of the trial or before implementation of the amendment). We searched PubMed, Embase, and Google Scholar for publications of the trials in scientific journals. All publications containing results of the trials were downloaded as full text. The publication search was conducted in January and February 2016. Thus, the minimal follow-up between

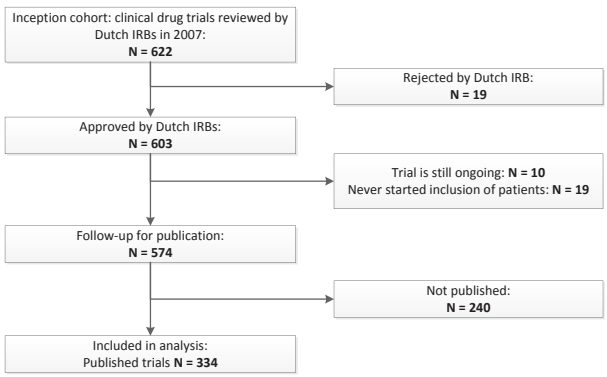

Fig. 1. Trials included in analysis. IRB, institutional review board.

IRB approval and publication was 8 years (December 2007 to January 2016).

We identified discrepancies between the most recently IRB-approved trial protocol (including IRB-approved substantial amendments) and the publications reporting results of the trial at issue. If multiple publications of the same trial were identified, all publications were included in the assessment. Discrepancies were scored by comparing the full text of the original protocol with the full texts of all identified publications of the trial. Five categories of protocol-publication discrepancies were measured: endpoints (the operationalization of events, symptoms, biomarkers, etc., that were measured in the trial); trial objectives (the general conceptual goals of the trial as stated in the introduction of the protocols and the publications); selection criteria; sample size; and sponsor acknowledgment. Two additional protocol-publication discrepancies were only scored among the randomized controlled trials (RCTs) in the cohort: discrepancies in additional or subgroup analyses, and discrepancies in the method used for the data analysis. The method used for data analysis was defined as how the trial arms were compared; for example, using the intention-to-treat, or the per protocol approach for the analysis (and whether the definition of the analysis population was similar in the protocol and in the publication). More details on the protocol-publication categories are provided in the Web-only Supplementary File, Table S1 at www.jclinepi.com.

Disclosure was the leading principle in scoring the discrepancies. Aspects were only scored as being discrepant if no reason for the discrepancy was provided in the publication, and we could not find another reason for the discrepancy that was disclosed to the IRB (such as approved substantial amendments). For example, if a primary endpoint was added in a publication compared with the trial protocol, and the publication also explained this addition, it was not considered discrepant. Similarly, a post hoc subgroup analysis was only considered discrepant if the post hoc nature of the subgroup analysis was not stated in the publication. In addition, if the trial was discontinued before the planned end of follow-up, and this was reported in the publication, we did not score the lower sample size as discrepancy. If multiple methods for data analysis were described in the publication and only one was specified in the protocol, we only scored this as being a discrepancy if the publication did not state which method was specified in the protocol. In case that multiple publications were found of one trial, omissions were only scored as discrepancy if the omitted item was not reported in any of the publications of the trial. Additions were scored as discrepancy if not labeled as post hoc in all publications of the trial. If we found in one of the publications an unexplained change compared with the protocol, it was scored as a discrepancy, regardless whether it was reported correctly in the other publications of the trial.

Trial characteristics were extracted from the ToetsingOnline database, from the form that all trial applicants filled out on a form at the time of submission of the trial application for IRB review. This form is mandatory and identical for all IRBs in the Netherlands. Other trial characteristics were prospective registration in the international registries of clinicaltrials.gov or ISRCTN, and whether the trial was completed as planned or discontinued. Other trial characteristics were the trial design (RCT superiority, RCT noninferiority, or non-RCT/exploratory pharmacology), and, only among the RCTs, the treatment arms (parallel group or crossover). Associations between these trial characteristics and protocol-publication discrepancies in the primary endpoints were evaluated.

Among the subgroup of RCTs, we categorized the direction of conclusion of the trials as positive or negative, as formulated in the publications. The direction of conclusion was positive if the trial results supported the trial objectives or hypotheses as stated in the protocol (eg, drug X is superior compared with placebo against disease Y). If the conclusion section of the publication stated that the results were negative, nonsignificant, or inconclusive, the direction of conclusion was classified as negative. If more than one publication was found of a trial, the first publication of the completed trial that reported primary endpoints was used to classify the direction of conclusion.

The protocol-publication discrepancies were described using univariate analysis, stratified for RCTs and non-RCTs. Randomized trials with exploratory objectives (eg, phase-1 trials investigating pharmacology, safety, and tolerability) were included in the non-RCT stratum. For the discrepancies in endpoints, objectives, and additional/subgroup analyses, we merged the outcome variable by calculating the sum total of primary endpoint discrepancies and primary objective discrepancies, and the sum total of discrepancies in additional or subgroup analyses.

We analyzed the association between the trial characteristics that were considered as being potential determinants of protocol-publication discrepancies. In addition to the sponsor type, we analyzed the trial characteristics that were significantly associated with nonpublication in the same cohort [21]: phase, centers involved, prospective registration, and completion. Furthermore, in line with previous studies [9,10], we also analyzed the association of the trial design and the treatment arms with discrepancies in the primary endpoints. The protocol of our study [7] prescribed the analysis of determinants for all protocol-publication discrepancies separately. In this study, we focus on determinants of the discrepancies in the primary endpoints, which are most likely to influence the direction of conclusions of the trials. Then, we analyzed the association between the protocol-publication discrepancies and the direction of conclusions of the trials. To estimate the overall associations, we used Pearson's chi-square test and indicate the associations of $P < 0.01$ and $P < 0.05$. Furthermore, risk ratios (RRs) and their 95% confidence interval (CIs) were

calculated to estimate the direction and precision of the associations. If zero outcomes (or zero reciprocal outcomes) were observed in categories with low numbers of trials, the RR was not calculated as a zero cell count will strongly bias the association toward statistical significance. The protocol of our study prescribed also multivariable logistic regression analysis. However, because of the relatively low number of discrepancies in the primary endpoints (32), the precision of the regression coefficients would have been low. Therefore, we decided to omit multivariable analysis.

One investigator (C.A.v.d.B.) performed the discrepancy scoring for all trials. A second investigator (P.C.S.) examined the reliability of the discrepancy scoring method. The protocol prescribed an additional double-check of 10% of the cohort and subsequently a randomly selected 20 trials. After comparing the seven discrepancy categories of the initial 35 trials selected for crosschecking (245 data entries in total), three data entries were changed after discussion. These included one discrepancy in the selection criteria, one discrepancy in the secondary endpoint, and one discrepancy in the subgroup or additional analyses. Thus, the interrater agreement was $(1 - (3/245)) \times 100 = 99\%$, with no disagreements about discrepancies in the primary endpoints. Based on the interrater agreement of 99%, we concluded there was sufficient proof of reliability of the scoring procedure and that the double-check could be restricted to the randomly selected 35 trials (10%). We included this protocol deviation in Table S2 (Web-Only Supplementary File at www.jclinepi.com).

3. Results

Of the 334 trials that were published by January 2016 (Fig. 1), we identified 506 articles. Of 91% of these trials, we found one or two articles (Table S3, Web-Only Supplementary File at www.jclinepi.com). The characteristics of the 334 trials are summarized in Table 1. The trials were mostly industry sponsored (62.3%), phase 3 (37.4%), and/or international multicenter (16.8% was also conducted in other European (EU) countries, and 40.7% was also conducted outside the EU). Oncology was the largest disease area (22.5%), and most trials (51.8%) were not prospectively registered at clinicaltrials.gov or ISRCTN. A small proportion (10.2%) was discontinued before the planned end of recruitment and/or follow-up. Sixty-one percent were RCTs (50.6% superiority and 10.5% noninferiority), and most trials had a parallel group design. Almost half of the trials (47.9%) planned to include less than 100 participants. Table 2 shows an overview of the protocol-publication discrepancies that were measured in all 334 trials. Omissions ($N = 17$; 5.1%) and changes ($N = 14$; 4.2%) of the primary endpoint were more common than additions ($N = 1$; 0.3%). The most common discrepancies were in secondary endpoints: 89 (43.6%) of the RCTs and 48

Table 1. Characteristics of the analyzed trials

Trial characteristics	Number of trials in analysis 334 (100%)	% (of 334)
Sponsor		
Pharmaceutical industry	208	62.3
Investigator [industry (co-)funded]	37	11.1
Investigator (no industry funding involved)	89	26.6
Phase		
Phase 1	41	12.3
Phase 2	78	23.4
Phase 3	125	37.4
Phase 4	32	9.6
Other than phases 1–4[a]	58	17.4
Centers		
Single center	113	33.8
Multicenter only in the Netherlands	29	8.7
Multicenter in the Netherlands and the EU	56	16.8
Multicenter in the Netherlands and outside the EU	136	40.7
Disease area		
Oncology	75	22.5
Endocrine diseases	40	12.0
Neurological diseases (including analgesia and anesthesia trials)	36	10.8
Infectious diseases (including vaccine trials)	32	9.6
Cardiovascular diseases	29	8.7
Respiratory diseases	25	7.5
Other disease areas	22	6.6
Musculoskeletal diseases	19	5.7
Mental and behavioral disorders	17	5.1
Hematological and immunological diseases	17	5.1
Digestive system diseases	12	3.6
Genitourinary system diseases	10	3.0
Prospective registration[b]		
Prospectively registered	161	48.2
Not (prospectively) registered	173	51.8
Completion		
Completed as planned	300	89.8
Terminated early	34	10.2
Design[c]		
RCT; superiority	169	50.6
RCT; noninferiority	35	10.5
Non-RCT and/or exploratory pharmacology trial	130	38.9
Treatment arms		
Parallel group	226	67.7
Crossover	39	11.7
Single arm	69	20.7
Number of participants planned in the trial protocol		
<100	160	47.9
≥100–500	89	26.6
≥500–1,000	50	15.0
≥1,000	35	10.5

Abbreviation: RCT, randomized controlled trial.

[a] Trials carried out using medicinal products in connection with objectives other than those referred to in the phase definitions 1–4. Such trials are not intended primarily to provide information about the product itself, but a medicinal product is needed to address the objective of the trial.

[b] Prospective registration was defined as registration of the trial at www.clinicaltrials.gov or www.isrctn.com, at latest 1 month after institutional review board approval.

[c] Exploratory pharmacology trials that involved randomization but no formal hypothesis testing (which is common in, eg, phase-1 randomized dose-escalation trials) were also excluded from the RCT subgroup.

Table 2. Occurrence of protocol-publication discrepancies stratified for trial design

Discrepancy categories	All trials, *N* (%) 334 (100)	RCTs, *N* (%) 204 (100)	Non-RCTs[b], *N* (%) 130 (100)
Discrepancies in endpoints			
Primary endpoint added in publication	1 (0.3)	0 (0.0)	1 (0.8)
Primary endpoint omitted in publication	17 (5.1)	3 (1.5)	14 (10.8)
Primary endpoints changed in publication	14 (4.2)	9 (4.4)	5 (3.8)
Sum total of discrepancies in primary endpoint[a]	32 (9.6)	12 (5.9)	20 (15.4)
No discrepancy in primary endpoint, but discrepancy in secondary endpoint	137 (41.0)	89 (43.6)	48 (36.9)
No discrepancies in endpoints	163 (48.8)	102 (50.0)	61 (46.9)
No information in protocol and/or publication on endpoints	2 (0.6)	1 (0.5)	1 (0.8)
Discrepancies in objectives			
Primary objective added in publication	2 (0.6)	2 (1.0)	0 (0.0)
Primary objective omitted in publication	12 (3.6)	2 (1.0)	10 (7.7)
Primary objective changed in publication	11 (3.3)	6 (2.9)	5 (3.8)
Sum total of discrepancies in primary objective[a]	25 (7.5)	10 (4.9)	15 (11.5)
No discrepancy in primary objective, but discrepancy in secondary objective	64 (19.2)	47 (23.0)	17 (13.1)
No discrepancies in objectives	241 (72.2)	144 (70.6)	97 (74.6)
No information in protocol and/or publication on objectives	4 (1.2)	3 (1.5)	1 (0.8)
Discrepancies in selection criteria			
Changed in publication	37 (11.1)	21 (10.3)	16 (12.3)
No discrepancies in selection criteria	295 (88.3)	183 (89.7)	112 (86.2)
No information in protocol and/or publication on selection criteria	2 (0.6)	0 (0.0)	2 (1.5)
Discrepancies in sample size			
<80% of sample size as calculated in protocol included	32 (9.6)	10 (4.9)	22 (16.9)
>120% of sample size as calculated in protocol included	6 (1.8)	3 (1.5)	3 (2.3)
No discrepancies in selection criteria	296 (88.6)	191 (93.6)	105 (80.8)
No information in protocol and/or publication on sample size	0 (0.0)	0 (0.0)	0 (0.0)

Abbreviation: RCT, randomized controlled trial.
[a] Sum total of all discrepancies in primary endpoints or primary objectives.
[b] Randomized exploratory pharmacology trials were also included in the non-RCT stratum.

(36.9%) of the non-RCTs had no discrepancy in the primary endpoints but a discrepancy in the secondary endpoints. Discrepancies in the sample size were mainly due to inclusion of <80% of the sample as calculated in the protocol, which occurred in 10 (4.9%) of the RCTs and in 22 (16.9%) of the non-RCTs. Three (1.5%) of the RCTs and three (2.3%) of the non-RCTs included more than 120% of the sample size as calculated in the protocol.

Table 3. Occurrence of protocol-publication discrepancies that were only scored among the subgroup of randomized controlled trials (RCTs)

Discrepancy categories	*N* (%)
Number of trials assessed	204 (100)
Discrepancies in additional or subgroup analyses	
Additional/subgroup analysis added in publication	25 (12.3)
Additional/subgroup analysis omitted in publication	62 (30.4)
Additional/subgroup analysis changed in publication	4 (2.0)
Sum total of discrepancies in additional or subgroup analysis[a]	91 (44.6)
No discrepancies in additional or subgroup analysis	111 (54.4)
No information in protocol and/or publication on subgroup analysis	2 (1.0)
Discrepancies in method used for data analysis	
Method used for data analysis changed in the publication	21 (10.3)
No discrepancies in method used for data analysis	147 (72.1)
No information in protocol and/or publication on data analysis	36 (17.6)

[a] Sum total of all discrepancies in additional or subgroup analysis.

Table 3 summarizes the discrepancy categories that were only measured among the subgroup of 204 RCTs. Among the 204 RCTs, 91 (44.6%) had a discrepancy in the subgroup analysis. Furthermore, 21 (10.3%) of the RCTs had a discrepancy in the methods used for data analysis. None of the trials had a discrepancy in sponsorship acknowledgments. Seventy-eight trials (23.3%) had no discrepancy at all (Table S4, Web-Only Supplementary File at www.jclinepi.com). In 36 (17.6%) of the RCT protocols and/or publications, the methods used for data analysis were not specified (Table 3). In 21 of the 204 RCTs (10.3%), it was only missing in the protocol, in nine RCTs (4.4%) only in the publications, and in six RCTs (2.9%) both in the protocol and in the publications. The information on the methods used for data analysis was missing in 30 (55.6%) of the protocols and/or publications of the 54 investigator-initiated RCTs, and in 6 (4.0%) of the 150 industry-sponsored RCTs. The other discrepancy categories were missing in the protocols and/or publications of zero to four trials (0–1.5%).

Associations between trial characteristics and discrepancies in the primary endpoints are shown in Table 4. Investigator-initiated trials were associated with a higher likelihood to have a discrepancy in the primary endpoints compared with industry-sponsored trials, whether the industry was involved as (one of) the funder(s) of the trial (RR 3.7, 95% CI 1.4–9.9 and RR 4.4, 95% CI 2.0–9.5, respectively). Furthermore, trials not being classified as phases 1–4 had a

Table 4. Association between trial characteristics and protocol-publication discrepancies in primary endpoints

Trial characteristics	Trials with primary endpoint discrepancy, *N* (%)[c]	Chi-square	RR (95% CI)
Sponsor		17.82[a]	
Pharmaceutical industry (*N* = 208)	9 (4.3)		Ref
Investigator [industry (co-)funded; *N* = 37]	6 (16.2)		3.7 (1.4–9.9)
Investigator (no industry funding involved; *N* = 89)	17 (19.1)		4.4 (2.0–9.5)
Phase		17.04[a]	
Phase 1 (*N* = 41)	2 (4.9)		Ref
Phase 2 (*N* = 78)	6 (7.7)		1.6 (0.3–7.5)
Phase 3 (*N* = 125)	6 (4.8)		1.0 (0.2–4.7)
Phase 4 (*N* = 32)	5 (15.6)		3.2 (0.7–15.4)
Other than phases 1–4[d] (*N* = 58)	13 (22.4)		4.6 (1.1–19.3)
Centers		15.42[a]	
Single center (*N* = 113)	19 (16.8)		Ref
Multicenter only in the Netherlands (*N* = 29)	5 (17.2)		1.0 (0.4–2.5)
Multicenter in the Netherlands and the EU (*N* = 56)	3 (5.4)		0.3 (0.1–1.0)
Multicenter in the Netherlands and outside the EU (*N* = 136)	5 (3.7)		0.2 (0.1–0.6)
Prospective registration[e]		9.83[a]	
Prospectively registered (*N* = 161)	7 (4.3)		Ref
Not (prospectively) registered (*N* = 173)	25 (14.5)		3.3 (1.5–7.5)
Completion		0.03	
Completed as planned (*N* = 300)	29 (9.7)		Ref
Discontinued before planned end (*N* = 34)	3 (8.8)		0.9 (0.3–2.8)
Design[f]		8.72[b]	
RCT; superiority (*N* = 169)	11 (6.5)		Ref
RCT; noninferiority (*N* = 35)	1 (2.9)		0.4 (0.1–3.3)
Non-RCT and/or exploratory pharmacology trial (*N* = 130)	20 (15.4)		2.4 (1.2–4.8)
Subgroup of RCTs	204 (100)		
Discrepancies in primary endpoints[c]	12 (5.9)		
Treatment arms		4.64[b]	
Parallel group (*N* = 187)	9 (4.8)		Ref
Crossover (*N* = 17)	3 (17.6)		3.7 (1.1–12.3)

Abbreviations: RR, risk ratio; CI, confidence interval; RCT, randomized controlled trial.

[a] *P*-value < 0.01 (based on Pearson's chi-square).

[b] *P*-value between 0.01 and 0.05 (based on Pearson's chi-square).

[c] The sum total of discrepancies in primary endpoints (*N* = 32; see Table 2), or for the subgroup of RCTs (*N* = 12).

[d] Trials carried out using medicinal products in connection with objectives other than those referred to in the phase definitions 1–4. Such trials are not intended primarily to provide information about the product itself, but a medicinal product is needed to address the objective of the trial [31].

[e] Prospective registration was defined as registration of the trial at the international public registers www.clinicaltrials.gov or www.isrctn.com, at latest 1 month after institutional review board approval.

[f] Exploratory pharmacology trials that involved randomization but no formal hypothesis testing (which is common in, eg, phase-1 randomized dose-escalation trials) were also excluded from the RCT subgroup.

higher likelihood of discrepancies in the primary endpoints compared with phase-1 trials (RR 4.6, 95% CI 1.1–19.3). Multicenter trials also conducted outside the EU had a lower likelihood of having a discrepancy in the primary endpoints compared with single-center trials (RR 0.2, 95% CI 0.1–0.6). Trials that were not prospectively registered in clinicaltrials.gov or the ISRCTN registry were more likely to have a discrepancy in the primary endpoints compared with trials that were prospectively registered (RR 3.3, 95% CI 1.5–7.5). Compared with superiority RCTs, non-RCTs had a higher likelihood to have a discrepancy in the primary endpoint (RR 2.4, 95% CI 1.2–4.8). This association was not observed when comparing superiority RCTs with noninferiority RCTs (RR 0.4, 95% CI 0.1–3.3). Finally, crossover RCTs had a higher likelihood of discrepancies in the primary endpoint compared with parallel group RCTs (RR 3.7, 95% CI 1.1–12.3).

Table 5 shows the association between the protocol-publication discrepancies and a positive direction of the trial conclusions. In none of the discrepancy categories, having a protocol-publication discrepancy in that category was associated with a positive direction of trial conclusions.

4. Discussion

We found that 9.6% of all clinical drug trials, and 5.9% of the RCTs, in our study had a protocol-publication discrepancy in the primary endpoints. This is a substantially lower proportion than reported by the two previous studies investigating this issue. Chan et al. found discrepancies in primary endpoints among 62% of RCTs [9]. Berendt et al. conducted a study among academic (investigator-initiated) trials and found discrepancies in primary endpoints in 38% and 43% in non-RCTs and

Table 5. Association between protocol-publication discrepancies of the subgroup of randomized controlled trials (RCTs) and the direction of conclusion

Discrepancy categories	RCTs with positive direction of conclusion, *N* (%)	Chi-square[a]	RR (95% CI)
Endpoints		4.29	
No discrepancies (*N* = 102)	75 (73.5)		Ref
Primary endpoint added in publication (*N* = 0)	—		—
Primary endpoint omitted in publication (*N* = 3)	2 (66.7)		0.9 (0.4–2.0)
Primary endpoint changed in publication (*N* = 9)	4 (44.4)		0.6 (0.3–1.3)
Secondary endpoint omitted/added/changed in publication (*N* = 89)	59 (66.3)		0.9 (0.7–1.1)
No information in protocol (*N* = 1)	1 (100)		—[b]
Objectives		8.37	
No discrepancies (*N* = 144)	105 (72.9)		Ref
Primary objective added in publication (*N* = 2)	2 (100)		—[b]
Primary objective omitted in publication (*N* = 2)	0 (0)		—[b]
Primary objective changed in publication (*N* = 6)	4 (66.7)		0.9 (0.5–1.6)
Secondary objective omitted/added/changed in publication (*N* = 47)	28 (59.6)		0.8 (0.6–1.1)
No information in protocol (*N* = 3)	2 (66.7)		0.9 (0.4–2.0)
Selection criteria		0.06	
No discrepancies (*N* = 183)	126 (68.9)		Ref
Changed in publication (*N* = 21)	15 (71.4)		1.0 (0.8–1.4)
No information in protocol (*N* = 0)	—		—
Sample size		3.26	
No discrepancies (*N* = 191)	133 (69.6)		Ref
<80% of sample size as calculated in protocol included (*N* = 10)	6 (60.0)		0.9 (0.5–1.4)
>120% of sample size as calculated in protocol included (*N* = 3)	2 (66.7)		1.0 (0.4–2.1)
No information in protocol (*N* = 0)	—		—
Additional/subgroup analyses		3.06	
No discrepancies (*N* = 111)	76 (68.5)		Ref
Additional/subgroup analysis added in publication (*N* = 25)	18 (72.0)		1.1 (0.8–1.4)
Additional/subgroup analysis omitted in publication (*N* = 62)	41 (66.1)		1.0 (0.8–1.2)
Additional/subgroup analysis changed in publication (*N* = 4)	4 (100)		—[b]
No information in protocol (*N* = 2)	2 (100)		—[b]
Methods used for data analysis		1.85	
No discrepancies (*N* = 147)	101 (68.7)		Ref
Changed in publication (*N* = 21)	17 (81.0)		1.2 (0.9–1.5)
No information in protocol (*N* = 36)	23 (63.9)		0.9 (0.7–1.2)

Abbreviations: RR, risk ratio; CI, confidence interval; RCT, randomized controlled trial.

[a] None of the Pearson's chi-square tests indicated a statistically significant association ($P < 0.05$) between protocol-publication discrepancies and a positive direction of trial conclusions.

[b] In case 100%, or 0%, of the RCTs within a category with a low number of trials had a positive direction of conclusion, the risk ratio was not calculated because a zero cell count would bias the estimation of the 95% CI toward statistical significance.

RCTs, respectively [10]. In the subgroup of the 126 academic trials in our cohort, discrepancies in primary endpoints were found in 10 of 54 RCTs (19%), and in 13 of 72 non-RCTs (18%). This finding suggests that also the reporting of academic trials has been improved. Furthermore, both in RCTs and in non-RCTs, protocol-publication discrepancies were substantial in secondary endpoints: 89 (44%) and 48 (37%), respectively. These proportions were also considerably lower than those of the recent COMP are initiative [22], which reported discrepancies in endpoints among 87% of the trials (not differentiating between primary and secondary endpoints). Discrepancies in primary objectives were found in 7% (*N* = 25) of the trials, and in 5% (*N* = 10) of the RCT subgroup. This is also lower than the previous study that investigated this discrepancy [10]. In line with secondary endpoints, discrepancies in secondary objectives occurred in 19% (*N* = 64) of the trials and in 23% (*N* = 47) of the RCTs. We found a discrepancy in selection criteria in 11% (*N* = 37) of the trials, and in 10% (*N* = 21) of the RCTs. To our knowledge, our study is the first to investigate this. Discrepancies in the sample size were also found in 11% (*N* = 38) of the trials, and in 6% (*N* = 13) of the RCTs. This finding is considerable lower than reported in a previous study, which found discrepancies in sample size calculations in 53% of the RCTs [23].

Among 91 (45%) of the 204 RCTs, we also found discrepancies in subgroup or other additional analyses. In this discrepancy category, omissions of subgroup or additional analysis that were planned in the protocol were most common. Omissions of planned subgroup analyses are, to our knowledge, not investigated in previous studies. The 25 RCTs (12%) that added an unplanned subgroup analysis and did not label it as being post hoc are lower than a

previous study that found 35% [11]. Finally, 21 (10%) of the RCTs changed the method used for data analysis. This is also considerably lower compared with findings by a previous study [23].

The reason for the lower occurrence of discrepancies compared with the three previous cohort studies [9−11] may be that more clinical investigators and journals are aware of the importance of complete and accurate reporting of all protocol aspects. In 2007, the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) Statement initiative was launched. Furthermore, the CONsolidated Standards of Reporting Trials (CONSORT) statement was updated in 2010. The broad attention to and implementation of these initiatives by the major medical journals might have contributed to better reporting [19,24−29]. Another reason could be related to the information that was available to explain the observed differences between protocols and publications. If a potential discrepancy was identified, we exhaustively searched substantial amendments and follow-up publications to explain the difference. If these explanations were available, for example, in online supplementary files or in publications other than the main results publication, we did not consider it as being a discrepancy. Hence, it might have made a difference whether the research team was determined to find as many discrepancies as possible, or find as many explanations for potential discrepancies as possible. Another reason for the difference with two of the previous cohorts is that we included all available publications in the assessment, whereas these studies only included one publication for each trial protocol [9,11,23]. The third previous cohort study [10] had only a follow-up from IRB approval until publication of 5 years, which is likely too short to identify all relevant (additional) publications [30].

We replicated the previous findings that non-RCTs/exploratory trials have more discrepancies than RCTs [10], and that, among the RCTs, trials with a crossover treatment arm have more discrepancies compared with parallel group treatment arms [9]. Furthermore, the analysis of trial characteristics suggests that discrepancies in primary endpoints mainly occur in small, local, investigator-initiated trials not in the context of drug product development (other than phases 1−4 [31]), that also were less likely to be prospectively registered. These characteristics often coincide. A reason for this may be that such trials are more flexible in the choice of endpoints, as the protocol and subsequent publications will not be reviewed by the drug marketing authorization authorities (who request standardized endpoints for a given disease area [32]). These analyses elucidate trial types that have remained outside the reach of the past initiatives and countermeasures against selective reporting.

No discrepancy categories were significantly associated with a positive direction of conclusion of the RCT subgroup. This suggests that the observed discrepancies have not been introduced in the publications to change the overall direction of conclusion of the trials. However, this was only assessed using a binary classification of direction of conclusion, leaving no room for nuances. In particular, the discrepancies in primary and secondary endpoints may still have led to a ‘‘more positive,’’ or ‘‘less negative’’ conclusion, thereby introducing reporting bias [22,33]. If investigators measure several endpoints and can decide afterward which to report, the likelihood is high that those endpoints are reported that fit expectations or desires [34]. If objectives are discrepant, the publication may fail to provide an accurate description of the original rationale and research question of the trial. This can be relevant for the interpretation of results. Discrepancies in the sample size can bias the interpretation of results, as the likelihood of erroneous chance findings is high if the sample size is too small [35]. Discrepancies in the methods used for data analysis can also be a way to spin the interpretation of results toward the preferred conclusion, for example, by excluding or including outliers or cases with partially missing data entries. Not reporting planned subgroup analyses occur likely because of the absence of effect, and unplanned subgroup analyses were probably added post hoc because there was an (unexpected) effect. Although the latter can serendipitously lead to important discoveries, their exploratory nature should be clearly acknowledged when they are reported [2,36]. Furthermore, protocol-publication discrepancies in the selection criteria can affect the ethical justification to include certain participants in the trial. For example, a protocol in the cohort prescribed inclusion of only patients with a given tumor characterized as grade 4 (the tumor grade indicating the most severe grade of illness). The publication reported inclusion of patients with tumor grades 3−4. However, as the protocol stated only the inclusion of grade-4 tumors, the IRB had approved the trial to be conducted specifically in the population with grade 4. The IRB had not considered grade-3 tumors in their evaluation, and investigators were therefore not permitted to include these patients. In addition, including more participants than needed according to the protocol can also be unethical. The research question could then have been answered at the cost of a smaller number of participants being exposed to risks and burdens of the trial [37].

In our cohort, the magnitude of nonpublication likely exceeded the magnitude of selective publication in terms of causing research waste and publication bias. Of the 574 drug trials initially followed until publication, 42% remained unpublished [21], and of the published trials, 32 (10%) had discrepancies in the primary endpoints. Nevertheless, the identified discrepancies could still have introduced spin and bias in the trial publications [38]. Trial publications should therefore become more transparent and provide a clear track record of the process of a clinical trial, from the initial research protocol until the publications presenting the results [39]. Some journals published the trial protocol as well as protocol amendments as online supplement, but this was rather an exception than common practice. To further facilitate independent interpretation of

protocol deviations, simple checklists can indicate which part of the protocol changed, when and why, and to which extent this may have influenced the conclusions [40]. This discrepancy checklist could then also be included in the assessment of bias (such as outcome reporting bias), which should be done when the trials are included in systematic reviews [4,41,42]. The finding that 20% of the RCTs missed information in the protocol and/or the publication about the methods used for data analysis attends IRBs and journals to always request this important information [20,24]. This attention is especially needed for the investigator-initiated RCTs.

A strength of our study is that the selection of protocols was not limited to those that are publicly available, thereby avoiding selection bias. We had access to all clinical drug trials that were submitted to an IRB in 2007. Furthermore, we did not limit our discrepancy assessment to the endpoints but assessed seven essential trial aspects that should be consistent between trial protocol and subsequent publications. As we included trials across all medical specialties, and 57% of the trials were multicenter international trials, our findings can be considered as being generalizable across geographical and medical areas. Although a higher sample size would have enabled a more precise conclusion, the number of 334 trials included in our cohort is higher than most previous studies, with the exception of the study by Kasenda et al. [9−11,23]. A limitation of our study was that we might have missed some documents explaining the discrepancies that were not included in the CCMO-archive. However, this missing information would then be incidental and therefore unlikely systematic or differential. And, if substantial amendments were missing in the archives, a record of these as well as nonsubstantial amendments should nevertheless have been provided in the publication, thus discrepancies scored as a result from these missing documents can be justified. Another limitation is that we might have missed some publications that were published after the follow-up period or were missed in the publication search. Finally, a limitation was the low number of cases in some trial characteristic categories (eg, investigator-initiated trials), which limited the precision of the risk estimates. We could, therefore, not perform multivariable analysis. Conclusions regarding the associations between trial characteristics and discrepancies should be interpreted with caution.

To conclude, protocol-publication discrepancies in clinical drug trials were not unusual in primary endpoints, but common in secondary endpoints, secondary objectives, and subgroup or other additional analyses. Despite the improvement compared with previous studies, the occurrence of discrepancies was still substantial, indicating that selective publication remains a problem in clinical research. Investigator-initiated, not prospectively registered and nonrandomized trials were determinants of discrepancies in the primary endpoint. Full transparency of the process of clinical trial protocols to publications can eliminate these opportunities for selective reporting. Practically, this could mean that the original trial protocol and substantial amendments are made publically available at the moment of publication of the results of the trial. That is likely the essential way forward in pursuing ethically sound and scientifically valid clinical research.

Acknowledgments

The authors thank our colleagues from the Central Committee on Research Involving Human Subjects (CCMO) and the accredited Institutional Review Boards (METCs) in the Netherlands for their collaboration and for providing access to their data.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2017.05.012.

References

[1] Baker M. 1,500 scientists lift the lid on reproducibility. Nature 2016; 533:452−4.
[2] Ioannidis JP. Why most published research findings are false. PLoS Med 2005;2(8):e124.
[3] Dickersin K, Chalmers I. Recognizing, investigating and dealing with incomplete and biased reporting of clinical research: from Francis Bacon to the WHO. J R Soc Med 2011;104(12):532−8.
[4] Bouter LM. Open data are not enough to realize full transparency. J Clin Epidemiol 2015.
[5] Knottnerus JA, Tugwell P. Selection-related bias, an ongoing concern in doing and publishing research. J Clin Epidemiol 2014;67:1057−8.
[6] Strech D. Normative arguments and new solutions for the unbiased registration and publication of clinical trials. J Clin Epidemiol 2012;65:276−81.
[7] Van den Bogert CA, Souverein PC, Brekelmans CT, Janssen SW, van Hunnik M, Koeter GH, et al. Occurrence and determinants of selective reporting of clinical drug trials: design of an inception cohort study. BMJ Open 2015;5(7):e007827.
[8] Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gotzsche PC, et al. Increasing value and reducing waste: addressing inaccessible research. Lancet 2014;383:257−66.
[9] Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 2004;291:2457−65.
[10] Berendt L, Callreus T, Petersen LG, Bach KF, Poulsen HE, Dalhoff K. From protocol to published report: a study of consistency in the reporting of academic drug trials. Trials 2016;17:100.
[11] Kasenda B, Schandelmaier S, Sun X, von Elm E, You J, Blumle A, et al. Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications. BMJ 2014;349:g4539.
[12] Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. JAMA 2009;302:977−84.
[13] Raghav KP, Mahajan S, Yao JC, Hobbs BP, Berry DA, Pentz RD, et al. From protocols to publications: a study in selective reporting of outcomes in randomized trials in oncology. J Clin Oncol 2015; 33:3583−90.
[14] Merck Guidelines for Publication of Clinical Trials in the Scientific Literature. Available at https://www.merck.com/research/discovery-and-

development/clinical-development/Merck-Guidelines-for-Publication-of-Clinical-Trials-in-the-Scientific-Literature.pdf. Accessed February 16, 2016.

[15] Food and Drug Administration Amendments Act of 2007, Title VIII - Clinical trial databases. Silver Spring, Maryland, USA: Public Law; 2007:110−85.

[16] Amendment of the Medical Research Involving Human Subjects Act with regard to the evaluation of the act and recovery of incomplete implementation of guideline no. 2001/20/EG. Dossier 31452. Tweede Kamer der Staten Generaal (Dutch national Parliament), The Hague, the Netherlands. Enacted as per 1 July 2012 (full text in Dutch only).

[17] Trans-NIH BioMedical Informatics Coordinating Committee (BMIC). NIH Data Sharing Repositories. Available at https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html. Accessed October 18, 2016.

[18] European Federation of Pharmaceutical Industries and Associations (EFPIA) Position - Transparency of Information on Clinical Trials included in the Proposed EU Database (Article 78 of the Proposal for a Regulation on Clinical Trials). Available at http://www.efpia.eu/uploads/Modules/Documents/efpia_key_messages_transparency_final_10.12.2012-updated_0.pdf. Accessed November 2, 2016.

[19] ICMJE. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals 2015. Available at http://www.icmje.org/icmje-recommendations.pdf Accessed November 2, 2016.

[20] Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332.

[21] van den Bogert CA, Souverein PC, Brekelmans CT, Janssen SW, Koeter GH, Leufkens HG, et al. Non-publication is common among phase 1, single-center, not prospectively registered, or early terminated clinical drug trials. PLoS One 2016;11:e0167709.

[22] Goldacre B, Drysdale H, Powell-Smith A, Dale A, Milosevic I, Slade E, et al. The COMPare Trials Project. Available at www.COMPare-trials.org; 2016. Accessed October 18, 2016.

[23] 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.

[24] Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gotzsche PC, Krleza-Jeric K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med 2013;158:200−7.

[25] 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.

[26] Deangelis CD, Fontanarosa PB. Strengthening the credibility of clinical research. Lancet 2010;376:234.

[27] Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016;355:i5239.

[28] Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA 2012;308:2594−604.

[29] Schulz KF, Moher D, Altman DG. CONSORT 2010 comments. Lancet 2010;376:1222−3.

[30] Sune P, Sune JM, Montoro JB. Positive outcomes influence the rate and time to publication, but not the impact factor of publications of clinical trial results. PLoS One 2013;8:e54583.

[31] Central Committee on Research Involving Human Subjects. Explanation of the General Assessment and Registration (ABR) Form (version 15 December 2015). Available at http://www.ccmo.nl/attachments/files/b1-abr-toelichting-15-dec-2015-eng.pdf. Accessed October 30, 2016.

[32] European Medicines Agency. Clinical efficacy and safety guidelines. Available at http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000085.jsp. Accessed October 18, 2016.

[33] Hewitt CE, Mitchell N, Torgerson DJ. Listen to the data when results are not significant. BMJ 2008;336:23−5.

[34] Munafo MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Du Sert NP, et al. A manifesto for reproducible science. Nat Hum Behav 2017;1:0021.

[35] Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, et al. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 2013;14(5):365−76.

[36] van der Valk FM, van Wijk DF, Stroes ES. Serendipity of post-hoc surrogate marker research. Eur Heart J 2012;33:2897−8.

[37] World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013;310:2191−4.

[38] Chan AW. Bias, spin, and misreporting: time for full access to trial protocols and results. Plos Med 2008;5(11):e230.

[39] Santori G. Research papers: journals should drive data reproducibility. Nature 2016;535:355.

[40] National Institutes of Health. Principles and guidelines for reporting preclinical research. Available at https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research. Accessed October 19, 2016.

[41] Knottnerus JA, Tugwell P, Tricco AC. Individual patients are the primary source and the target of clinical research. J Clin Epidemiol 2016;76:1−3.

[42] Koletsi D, Valla K, Fleming PS, Chaimani A, Pandis N. Assessment of publication bias required improvement in oral health systematic reviews. J Clin Epidemiol 2016;76:118−24.