Comparison of primary outcomes in protocols, public clinical-trial registries and publications: the example of oncology trials

A. Perlmutter[1], V-T Tran[2, 3], A Dechartres[2, 3, 4], P Ravaud[1, 2, 3, 4]

[1]Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA

[2] METHODS Team, Epidemiology and Statistics Sorbonne Paris Cité Research Center (INSERM 1153), Paris, France

[3]Center of Clinical Epidemiology, Hôtel-Dieu Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France

[4] Paris Descartes University, Paris, France

Corresponding author:

Prof. Viet-Thi Tran

Hôpital Hôtel Dieu, Centre d′Épidémiologie Clinique, Paris, France

1 place du Parvis Notre-Dame, Paris 75181, France

Tel: +33 1 42 34 89 87, fax: +33 1 42 34 87 90, email: thi.tran-viet@htd.aphp.fr

1

Abstract (298)

Background: Protocols are often unavailable to peer-reviewers and readers. To detect outcome reporting bias (ORB), readers usually have to resort to publicly available descriptions of study design such as clinical-trial registries. We compared primary outcomes in protocols, ClinicalTrials.gov and publications of oncology trials and evaluated the use of ClinicalTrials.gov as compared with protocols in detecting discrepancies between planned and published outcomes.

Method: We searched for phase III oncology trials registered in ClinicalTrials.gov and published in the *Journal of Clinical Oncology* and *New England Journal of Medicine* between January 2014 and June 2015. We extracted primary outcomes reported in the protocol, ClinicalTrials.gov and the publication. First, we assessed the quality of primary outcome descriptions by using a published framework. Second, we evaluated modifications of primary outcomes between each source. Finally, we evaluated the agreement, specificity and sensitivity of detecting modifications between planned and published outcomes by using protocols or ClinicalTrials.gov.

Results: We included 65 trials, with 81 primary outcomes common among the three sources. The proportion of primary outcomes reporting all items from the framework was 73%, 22% and 75% for protocols, ClinicalTrials.gov and publications, respectively. Eight (12 %) trials presented a discrepancy between primary outcomes reported in the protocol and in the publication. Twelve (18.5 %) trials presented a discrepancy between primary outcomes registered at ClinicalTrials.gov and in publications. We found a moderate agreement in detecting discrepant reporting of outcomes by using protocols or ClinicalTrials.gov (ƙ=0.53, 95% confidence interval [0.25-0.81]). Using ClinicalTrials.gov to detect discrepant reporting of outcomes showed high specificity (89.5 %) but lacked sensitivity (75 %) as compared with use of protocols.

Conclusion: In oncology trials, primary outcome descriptions in ClinicalTrials.gov are often of low quality and may not reflect what is in the protocol, thus limiting the detection of modifications between planned and published outcomes.

Key message: Identification of discrepancies between planned and published primary outcomes in oncology trials led to different results when using as reference ClinicalTrials.gov or protocols. These findings question the sole use of public clinical-trial registries to detect discrepancies between planned and published outcomes and underline the need for public access to protocols.

Introduction

Outcome reporting bias (ORB) refers to unacknowledged changes in trial outcomes from protocol to publication depending on the nature and direction of the results [1]. It involves a diverse group of practices that include under-reporting (not reporting planned outcomes), over-reporting (reporting unplanned outcomes) or misreporting (changing the definition and measures of outcomes)[2]. For approximately 40% to 62% of trials, at least one primary outcome is omitted, introduced or changed between what was planned in the protocol and what was published [3, 4]. Outcome reporting bias distorts the evidence available in the literature by favoring positive results [3].

Oncology trials are not safe from such practices. Although overall survival is the gold standard for demonstrating clinical benefit, many trials use different endpoints such as progression-free survival, tumor size, biologic markers, symptom control, quality of life or economic evaluations [5]. Studies have shown that 12% to 14% of clinical trials in oncology modified prespecified primary outcomes and that 38% reported an unplanned analysis[2, 6]. Such discrepant outcome reporting is important in oncology trials because such trials often assess new treatments that are both expensive and have a tight risk–benefit balance.

Detection of modifications between planned and published outcomes is complex. Protocols constitute the most comprehensive description of the study design before trial inception, but they often are confidential documents, unavailable to peer-reviewers and readers[7]. To overcome this problem and improve transparency in clinical research for patients, clinicians, researchers and policy makers, the International Committee of Medical Journal Editors requires access to key protocol information by registration of trials in public clinical-trial registries such as ClinicalTrials.gov before enrollment of the first participant [8]. However,

3

use of these public registries to identify discrepant outcome reporting is possible only if outcomes are fully and clearly registered before the beginning of the trial[9]. When examining ClinicalTrials.gov, only 63% of registered outcomes were precise enough for comparison with published findings [4]. This was also the case in oncology trials, for which only 37% of registry entries in ClinicalTrials.gov provided a sufficiently clear outcome description for comparison with publications [6].

To our knowledge, no study has compared the reporting of outcomes between protocols, public clinical-trial registries and publications. Studies usually compared publications with 1) protocols available from ethics committees [3, 10, 11], 2) protocols publicly available as supplemental material from journals [2], or 3) public clinical-trial registry entries [4, 6, 12]. One study compared reporting of outcomes between clinical study reports and publicly available materials in publications and ClinicalTrials.gov [13, 14]. The authors found that study reports were more complete than public clinical-trial registry entries and publications, but they did not describe in detail the quality of outcome reporting or the nature of outcome modifications.

In this study, we compared the primary outcomes reported in protocols, ClinicalTrials.gov registries and publications of oncology trials. Then we evaluated the use of ClinicalTrials.gov as compared with protocols in detecting modifications between planned and published outcomes.

Methods

We performed a methodological review of phase III oncology trials published in 2014-2015 in the *Journal of Clinical Oncology* and *New England Journal of Medicine* and compared the

description of primary outcomes reported in published articles, ClinicalTrials.gov and protocols. Then we evaluated the use of ClinicalTrials.gov as compared with protocols in detecting modifications between planned and published outcomes.

Study search

One investigator (AP) searched MEDLINE via PubMed for articles published between January 1, 2014 and June 29, 2015 by using the keywords *Cancer OR Oncol** and the Cochrane Highly Sensitive Search Strategy for identifying randomized trials in two journals, *Journal of Clinical Oncology* and *New England Journal of Medicine*. We chose these journals because they publish study protocols as supplementary material.

Selection of relevant studies

We included phase III randomized controlled trials in the field of oncology for which both an online protocol and a ClinicalTrials.gov registration were available. We excluded studies that involved a pediatric population (<18 years old) or hematologic malignancies, reported pooled data from two or more trials or were secondary reports of previously published trials. Two investigators (V-TT, AP) confirmed the eligibility of trials included in the selection.

Extraction of general characteristics

One investigator (AP) used a standardized extraction form to collect 1) publication details (journal name, year of publication), 2) disease site, 3) type of the intervention (chemotherapy, targeted therapy, radiation therapy, surgery, supportive care or screening and/or diagnosis), 3) trial design (superiority, non-inferiority, or equivalence), 4) number of study groups and 5) funding source (funding by industry or not as reported in ClinicalTrials.gov).

5

Extraction of primary outcomes from the three sources

For each trial, two investigators independently extracted the primary outcome(s) reported in the 1) the study protocol (including all amendments), 2) the entry in ClinicalTrials.gov at the time of publication, and 3) the published reports (including outcome modifications reported in methods as recommended by the CONSORT [15])

We considered as primary outcomes only those explicitly reported as such [4]. If no outcome was explicitly reported in publications or protocols, we used the outcome reported in sample size calculations. For each outcome extracted, we assessed results, which were considered positive if they significantly supported the superiority or non-inferiority of the intervention over the control.

Assessment of quality of description of outcomes

Two investigators (V-TT, AP) assessed the quality of the description of each primary outcome reported in the three sources (excluding safety outcomes reported as primary outcomes), by using seven items inspired by the framework of Zarin et al. [16]. These 7 items are standard protocol items according to the SPIRIT guidelines [17] (**Appendix 1**):

1. *Domain*, defined as a clear description of what is being measured

2. *Specific measurement*, defined as a clear description of how it is being measured

3. *Specific metric*, defined as a description of how change was quantified (e.g., change from baseline, end value)

4. *Method of aggregation of data*, defined as a description of how data were managed (e.g., continuous value, proportion of patients achieving a given value)

5. *Time frame*, defined as a description of when the outcome was assessed

6

6. *Identity of outcome assessors*, defined as the presence of information on the identity and/or training of outcome assessors

7. *Blinding of outcome assessors*, defined as the presence of information on whether assessors were blinded to the intervention received, and how

We defined as an optimal outcome description the reporting of all seven items. We defined as an acceptable outcome description the reporting of all of the following five items: domain, specific measurement, specific metric, method of data aggregation and time frame.

Assessment of outcome modifications

Two investigators (V-TT, AP) independently looked for any modification to the primary outcomes between 1) protocols and published articles, 2) protocols and ClinicalTrials.gov, 3) ClinicalTrials.gov and published articles. Modifications could involve 1) a change from a primary outcome to a secondary outcome, 2) a change from a secondary outcome to a primary outcome, 3) introduction of a new primary outcome, 4) omission of a previously stated primary outcome, or 5) change in measurement method or time frame.

We considered as outcome modifications only flagrant discrepancies between the different sources. As a result, we did not consider the lack of precision in reporting outcomes as an outcome modification. For example, we considered that an outcome reported as *"Progression Free Survival"* in ClinicalTrials.gov and *"Progression-free Survival using RECIST criteria, measured every 8 weeks, as determined by blinded independent imaging review"* in a publication, contained no flagrant outcome modification.

Analysis

Data are presented as number (percentage) for qualitative data and median (interquartile range [IQR]) for continuous data.

7

First, we described the quality of primary outcome descriptions. We assessed the proportion of outcomes with an optimal description and an acceptable description in each source. We looked for the association between an acceptable outcome description and presence of modifications between the protocol and the published study by using Fisher's exact test. $P <$ 0.05 was considered statistically significant. Second, we described the modification of primary outcomes among each data source. Third, we evaluated the ability of ClinicalTrials.gov to detect modifications between planned and published outcomes as compared with protocols. We assessed the agreement in identifying discrepant reporting of outcomes by using the protocol or ClinicalTrials.gov with Cohen's Kappa (ƙ), then evaluated the sensitivity and specificity of using ClinicalTrials.gov to detect discrepant reporting of outcomes as compared with protocols.

All analyses involved use of R v3.2.2 (http://www.R-project.org), the R Foundation for Statistical Computing, Vienna, Austria).

Results

Our literature search yielded 651 references, from which 65 were included (**Figure 1**). Trials enrolled a median of 452 patients (IQR [253-704]). Approximately half of the trials were funded by industry (*n*=32, 49%) and half evaluated a targeted therapy (n=33, 51%) (**Table 1**). At the time of assessment, only 29 (44%) studies had results posted at ClinicalTrials.gov.

Quality of outcome descriptions

Accounting all outcome modifications (e.g., addition, omission and/or change from secondary to primary outcomes), we found a total of 81 primary outcomes common to the three sources (**Figure 1**). Approximately two thirds (66%) were overall survival or time-to-event outcomes

8

(e.g., progression-free survival, disease-free survival, etc.) (**Table 1**), and 48 (59%) were positive.

The proportion of primary outcomes with acceptable descriptions (i.e., reporting all elements of the Zarin et al. framework) was 59 (73%), 18 (22%), and 61 (75%) in protocols, ClinicalTrials.gov and publications, respectively (**Table 2** and **Figure 2**). Few outcome descriptions could be considered optimal, with 30%, 4%, and 26% of outcomes reporting all seven framework elements from in protocols, ClinicalTrials.gov and publications, respectively. Information about the blinding of outcome assessors was the least frequently reported information, with less than 45% of outcome descriptions reporting it in each source.

In our sample, less precise primary outcome descriptions in ClinicalTrials.gov was significantly associated with modifications of outcomes (*P*=0.03). Quality of outcome descriptions in protocols or publications was not associated with modification of outcomes nor with positive or negative results.

Outcome modifications

Comparison between protocol and publications

A total of 8 trials (12%) had at least one discrepancy between primary outcomes in the publication and the protocol (**Figure 3** and **Appendix 2**). Discrepancies involved the omission in the publication of one or several planned endpoints (*n*=2), the addition in the publication of one or several unplanned primary outcome (*n*=1), the change from one or several secondary outcomes in the protocol to primary outcomes in the publication (*n*=1), the change from a primary outcome in the protocol to a secondary outcome in the publication (*n*=2), and the modification of the measurement method of one or several outcomes (*n*=6). For example, in the published report of a trial evaluating early versus delayed initiation of palliative care, a

9

new primary outcome "Resource Use", absent from the protocol, was introduced in the publication[18].

Comparison between protocols and ClinicalTrials.gov

We found 12 (18%) studies with at least one discrepancy between primary outcomes in the protocol and in ClinicalTrials.gov (**Appendix 3**). In four cases, secondary outcomes in the protocol and publication were registered as primary outcomes.

Comparison between ClinicalTrials.gov and publications

We found 12 (18.5%) studies with at least one discrepancy between primary outcomes reported in ClinicalTrials.gov and the publication (**Figure 3 and Appendix 4**). Discrepancies involved the omission in the publication of one or several planned endpoints (n=5), the addition in the publication of one or several unplanned primary outcomes (n=3), a change from the primary outcome in ClinicalTrials.gov to a secondary outcome in the publication (n=4), modification of the measurement method of one or several outcomes (n=3) and an unclear entry in ClinicalTrials.gov preventing the assessment of outcome modification (n=1).

Comparison of identification of ORB by using protocols or ClinicalTrials.gov

We found moderate agreement in identifying studies with discrepant reporting of outcomes by using protocols or ClinicalTrials.gov, with ƙ=0.53 (95% CI [0.25-0.81]). This finding was due to both false-positive identifications of discrepant reporting of outcomes in ClinicalTrials.gov (n=9) (e.g., registration as primary outcomes of measurements reported as secondary outcomes in both the protocol and registration) and false-negative identification of discrepant reporting of outcomes (n=7) (e.g., modification of the measurement method between the protocol and the publication covered by an imprecise entry in the public clinical-trial registry [19]) (**Appendix 5).** Using ClinicalTrials.gov to detect discrepant reporting of outcomes

showed high specificity (89.5%) but lacked some sensitivity (75%) as compared with use of protocols.

Discussion

In the present study, we systematically compared primary outcomes reported in protocols, ClinicalTrials.gov and publications. We found evidence of distortion between protocols and published reports in 12% of trials. When using ClinicalTrials.gov to identify outcome reporting bias, 18.5% of trials had at least one discrepancy between what was presented in the registry and published report. Using ClinicalTrials.gov to detect modifications between planned and published outcomes resulted in both false-positive identification of discrepant reporting of outcomes (e.g., protocol amendments not reported in ClinicalTrials.gov) and false-negative identification of discrepant reporting of outcomes (e.g., discrepancy between protocols and publications covered by imprecise outcome descriptions in ClinicalTrials.gov).

In addition, we highlighted the low quality of primary outcome descriptions in ClinicalTrials.gov. Although guidance for registration in ClinicalTrials.gov insists on the importance of a clear description of the measurement method and time frame in registry entries [20, 21], only 62% and 59% of trial outcomes described in ClinicalTrials.gov contained a description of how and when the outcome would be measured. Some outcome modifications could have been covered by these imprecise descriptions. In our study, 25% of outcomes in publications were imprecisely reported and thus could suggest selective reporting based on results, not ascertainable by using protocols or ClinicalTrials.gov.

The proportion of discrepant outcomes found in our study concurs with the literature in oncology, in which authors found 12% and 14% modifications of primary outcomes by using

protocols [2] and ClinicalTrials.gov entries [6], respectively. Overall, modifications between planned and published outcomes seem less frequent in oncology trials than in other specialties [11], perhaps because of high standardization of methods to evaluate progression-free and overall survival. In our study, discrepant reporting of outcomes occurred mostly in studies involving patient-reported outcomes. Higher standardization of trial outcomes, as advocated by initiatives such as COMET may be a way to reduce the possibility of outcome modifications [22].

Our study is original because it is the first to compare primary outcome descriptions in protocols, public clinical-trial registries and publications and to show the limits of comparing published and registered outcomes to detect discrepant reporting of outcomes. Because trial protocols are often confidential documents not available to readers or peer reviewers[7, 23], our results question the ability of peer-reviewers or readers to identify deviations from the protocol as advocated by many journals [24]. For example, we found instances where modification of outcome measurements from the protocol to publication were covered by imprecise outcome descriptions in ClinicalTrials.gov.

Of note, discrepant outcome reporting is not always based on the result and may be due to a variety of reasons including loss of funding, poor quality of data or the non-analysis of secondary data because of no difference in the primary outcome [25]. However, in these cases, authors must identify any changes to the primary and secondary outcome measures after the trial started and explain the reasons for these changes. This important rule for the transparency of research was highlighted in the modification of the CONSORT reporting guidelines in 2010 [15]. In our sample, among 10 reports with discrepant reporting of outcomes between publications and protocols, only two gave the reasons for not reporting all primary outcomes [26, 27].

Our study has some limitations. First, because we studied recent trials, most did not yet have results posted at ClinicalTrials.gov. Information provided in public registries when results are posted is often more accurate than previous entries in ClinicalTrials.gov and/or publications [28, 29]. Thus, more research is needed to assess how outcomes are reported in this data source and how it may be used to investigate discrepant reporting of outcomes. Second, we considered only a limited number oncology trials published in two high-impact-factor journals providing open access to protocols. In addition, our sample comprised a small number of trials from each different sub-specialty of oncology. Therefore, our results and estimates for prevalence of outcome modifications may not be generalizable to other trials in oncology or other specialties and should be further investigated.

Because public clinical-trial registries may not precisely reflect protocols, and because peer editors and peer-reviewers often fail to detect discordance between planned and published outcomes in trials they assess, readers need to be allowed to evaluate the integrity of research themselves. Projects such as the COMPARE-trials initiative [30] require the public disclosure of all documents, including study protocols. Thus, the policy adopted by the *Journal of Clinical Oncology* or the *New England Journal of Medicine* to systematically append the study protocol to published reports [31] helps improve the identification of modifications between planned and published outcomes and should be considered by more journals.

Conclusion

Because protocols are confidential documents, public clinical-trial registries are the only option for readers and reviewers to compare primary outcomes reported in publications with a previous source. We have shown that outcome descriptions in public clinical-trial registries often lack precision and may not reflect what is in the protocol, thus limiting the ability to identify discrepancies between planned and published outcomes.

13

Funding

Competing interests

The authors have declared that no competing interests exist

References

1.   Kirkham JJ, Dwan KM, Altman DG et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ 2010; 340: c365.
2.   Raghav KP, Mahajan S, Yao JC et al. From Protocols to Publications: A Study in Selective Reporting of Outcomes in Randomized Trials in Oncology. J Clin Oncol 2015; 33: 3583-3590.
3.   Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. PLoS One 2013; 8: e66844.
4.   Mathieu S, Boutron I, Moher D et al. Comparison of registered and published primary outcomes in randomized controlled trials. JAMA 2009; 302: 977-984.
5.   Sylvester R, Van Glabbeke M, Collette L et al. Statistical methodology of phase III cancer clinical trials: advances and future perspectives. Eur J Cancer 2002; 38 Suppl 4: S162-168.
6.   You B, Gan HK, Pond G, Chen EX. Consistency in the analysis and reporting of primary end points in oncology randomized controlled trials from registration to publication: a systematic review. J Clin Oncol 2012; 30: 210-216.
7.   Chan AW. Bias, spin, and misreporting: time for full access to trial protocols and results. PLoS Med 2008; 5: e230.
8.   DeAngelis CD, Drazen JM, Frizelle FA et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. JAMA 2004; 292: 1363-1364.
9.   Zarin DA, Tse T. Trust but verify: trial registration and determining fidelity to the protocol. Ann Intern Med 2013; 159: 65-67.
10.   Chan AW, Hrobjartsson A, Haahr MT et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 2004; 291: 2457-2465.
11.   Redmond S, von Elm E, Blumle A et al. Cohort study of trials submitted to ethics committee identified discrepant reporting of outcomes in publications. J Clin Epidemiol 2013; 66: 1367-1375.
12.   Jones CW, Keil LG, Holland WC et al. Comparison of registered and published outcomes in randomized controlled trials: a systematic review. BMC Med 2015; 13: 282.

14

13.   Wieseler B, Wolfram N, McGauran N et al. Completeness of reporting of patient-relevant clinical trial outcomes: comparison of unpublished clinical study reports with publicly available data. PLoS Med 2013; 10: e1001526.

14.   Wieseler B, Kerekes MF, Vervoelgyi V et al. Impact of document type on reporting quality of clinical drug trials: a comparison of registry reports, clinical study reports, and journal publications. BMJ 2012; 344: d8141.

15.   Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Bmj 2010; 340: c332.

16.   Zarin DA, Tse T, Williams RJ et al. The ClinicalTrials.gov results database--update and key issues. N Engl J Med 2011; 364: 852-860.

17.   Chan AW, Tetzlaff JM, Gotzsche PC et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ 2013; 346: e7586.

18.   Bakitas MA, Tosteson TD, Li Z et al. Early Versus Delayed Initiation of Concurrent Palliative Oncology Care: Patient Outcomes in the ENABLE III Randomized Controlled Trial. J Clin Oncol 2015; 33: 1438-1445.

19.   Fernando HC, Landreneau RJ, Mandrekar SJ et al. Impact of brachytherapy on local recurrence rates after sublobar resection: results from ACOSOG Z4032 (Alliance), a phase III randomized trial for high-risk operable non-small-cell lung cancer. J Clin Oncol 2014; 32: 2456-2462.

20.   Tse T, Williams RJ, Zarin DA. Reporting "basic results" in ClinicalTrials.gov. Chest 2009; 136: 295-303.

21.   Zarin DA, Tse T, Ide NC. Trial Registration at ClinicalTrials.gov between May and October 2005. N Engl J Med 2005; 353: 2779-2787.

22.   Williamson P, Clarke M. The COMET (Core Outcome Measures in Effectiveness Trials) Initiative: Its Role in Improving Cochrane Reviews. Cochrane Database Syst Rev 2012; Ed000041.

23.   Chan AW, Upshur R, Singh JA et al. Research protocols: waiving confidentiality for the greater good. BMJ 2006; 332: 1086-1089.

24.   Chauvin A, Ravaud P, Baron G et al. The most important tasks for peer reviewers evaluating a randomized controlled trial are not congruent with the tasks most often requested by journal editors.

25.   Smyth RM, Kirkham JJ, Jacoby A et al. Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists. Bmj 2011; 342: c7153.

26.   Larkin J, Chiarion-Sileni V, Gonzalez R et al. Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. N Engl J Med 2015; 373: 23-34.

27.   Eijzenga W, Aaronson NK, Hahn DE et al. Effect of routine assessment of specific psychosocial problems on personalized communication, counselors' awareness, and distress levels in cancer genetic counseling practice: a randomized controlled trial. J Clin Oncol 2014; 32: 2998-3004.

28.   Hartung DM, Zarin DA, Guise JM et al. Reporting discrepancies between the ClinicalTrials.gov results database and peer-reviewed publications. Ann Intern Med 2014; 160: 477-483.

29.   Riveros C, Dechartres A, Perrodeau E et al. Timing and completeness of trial results posted at ClinicalTrials.gov and published in journals. PLoS Med 2013; 10: e1001566; discussion e1001566.

30.   Goldacre B, Drysdale H, Powell-Smith A. The COMPare Trials Project. In. 2016.

31.   Cannistra SA. New initiatives at Journal of Clinical Oncology. J Clin Oncol 2011; 29: 2609-2610.

15

List of tables

Table 1: Characteristics of randomized trials included in the study (n=65)

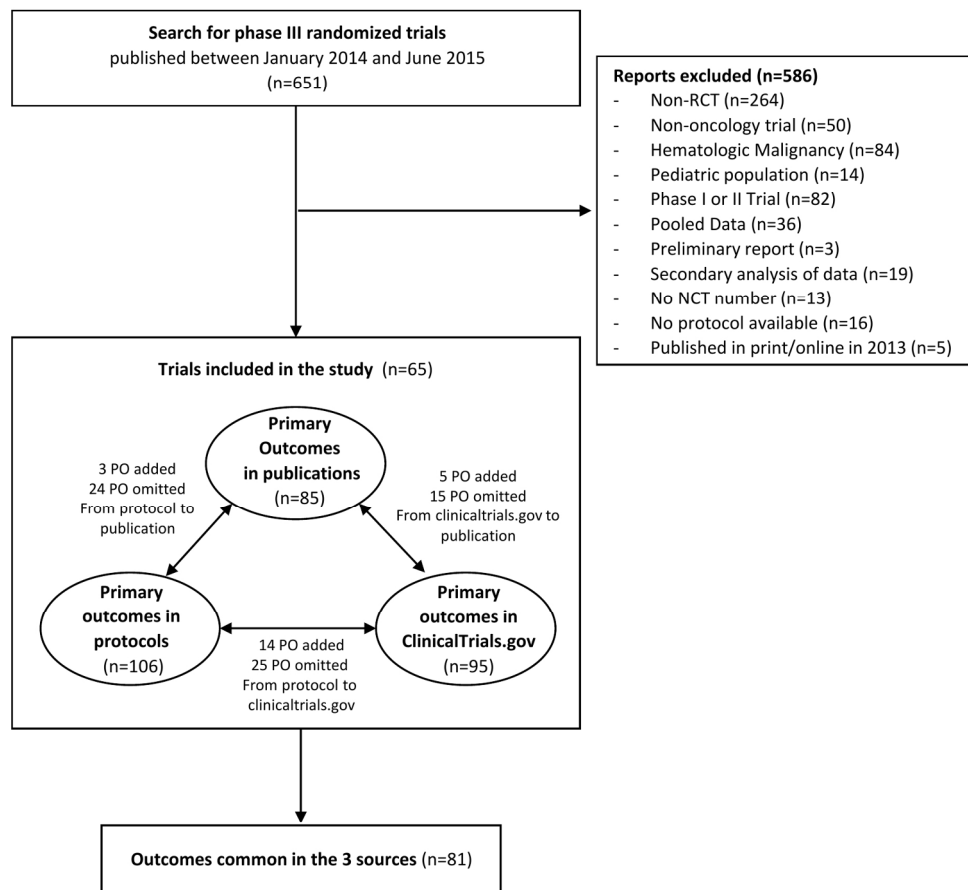Table 2: Quality of outcome descriptions in each data source (n=81) according to the Zarin et al. framework [11]

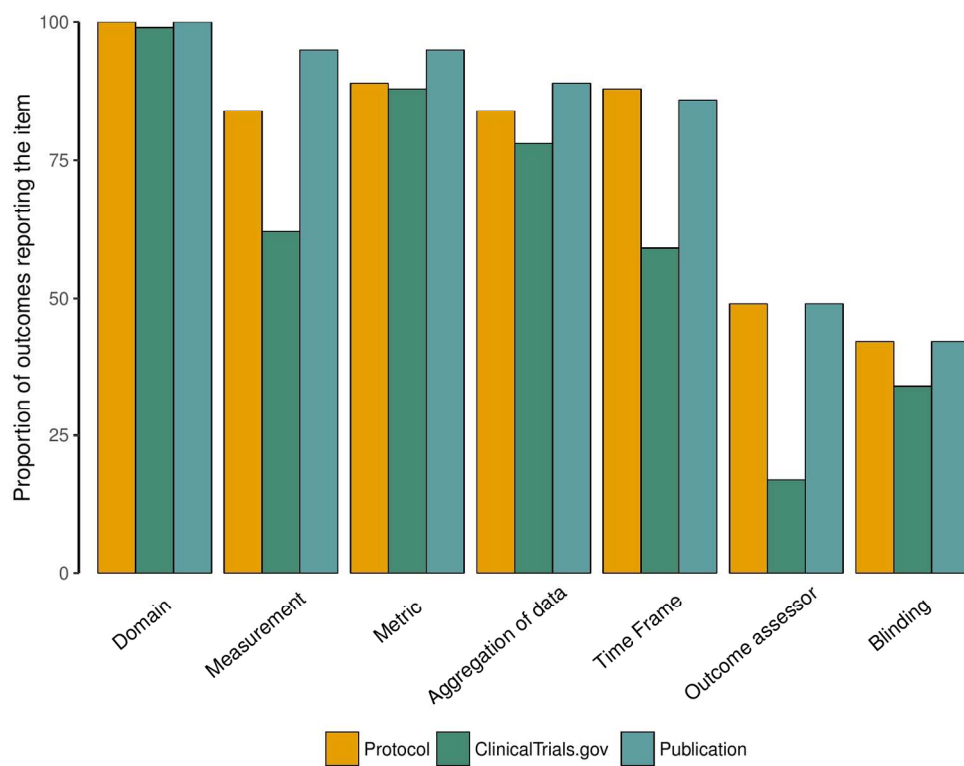**List of figures**

Figure 1: Flow chart of articles in the study. PO, primary outcome.

Figure 2: Outcome descriptions in each data source (n=81)

Figure 3: Discrepancies between primary outcomes in protocols and publications and in ClinicalTrials.gov and publications (n=65)

List of appendices

Supplementary table 1: Items assessed to describe the quality of outcome descriptions

Supplementary table 2: Modification of outcomes from protocol to publication

Supplementary table 3: Modification of outcomes from protocol to ClinicalTrials.gov

Supplementary table 4: Modification of outcomes from ClinicalTrials.gov to publication

Supplementary table 5: Reasons for discrepancies in identifying modifications between planned and published outcomes when using protocols or ClinicalTrials.gov (n=16). Comparison of protocols/publication is considered a reference for ORB identification.

16

17

Figure 1: Flow chart of articles in the study. PO, primary outcome.

180x170mm (270 x 270 DPI)

Search for phase III randomized trials
published between January 2014 and June 2015
(n=651)

Reports excluded (n=586)
- Non-RCT (n=264)
- Non-oncology trial (n=50)
- Hematologic Malignancy (n=84)
- Pediatric population (n=14)
- Phase I or II Trial (n=82)
- Pooled Data (n=36)
- Preliminary report (n=3)
- Secondary analysis of data (n=19)
- No NCT number (n=13)
- No protocol available (n=16)
- Published in print/online in 2013 (n=5)

Trials included in the study (n=65)

Primary Outcomes in publications (n=85)

3 PO added
24 PO omitted
From protocol to publication

5 PO added
15 PO omitted
From clinicaltrials.gov to publication

Primary outcomes in protocols (n=106)

Primary outcomes in ClinicalTrials.gov (n=95)

14 PO added
25 PO omitted
From protocol to clinicaltrials.gov

Outcomes common in the 3 sources (n=81)

Figure 2: Outcome descriptions in each data source (n=81)

176x139mm (277 x 277 DPI)
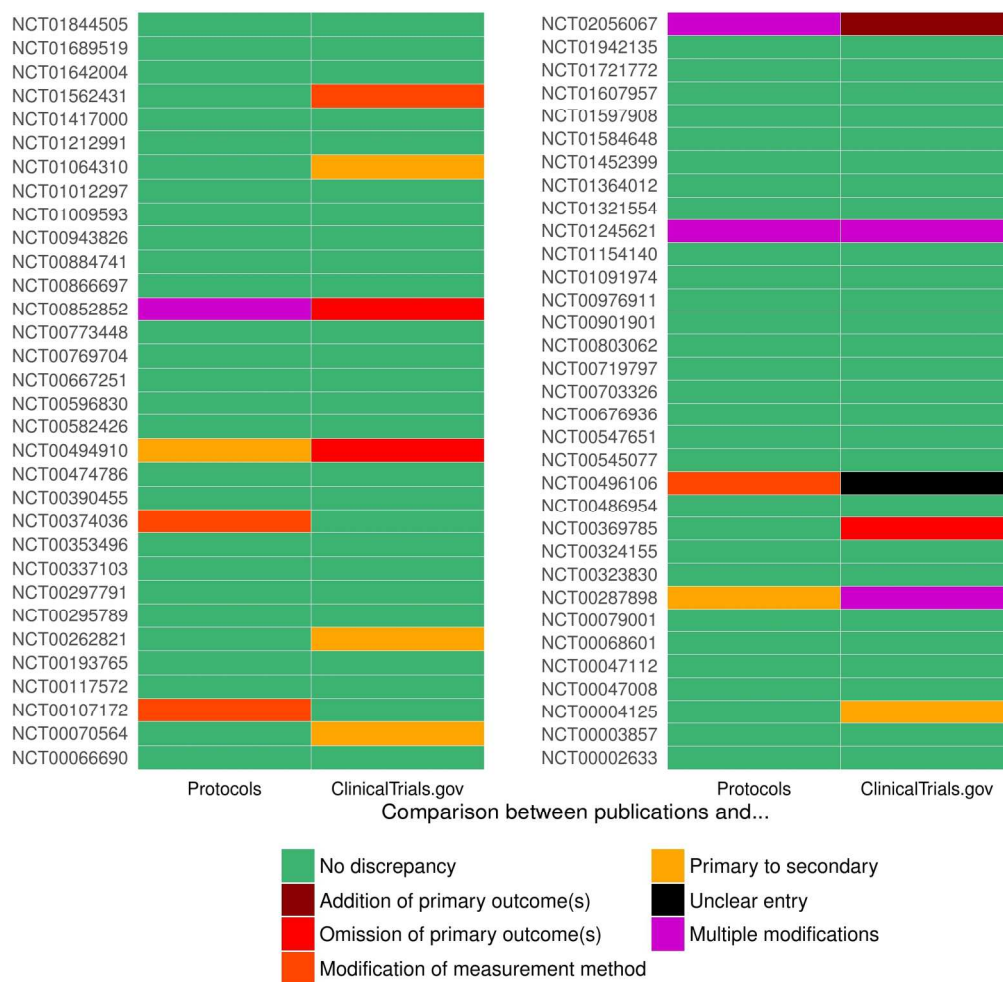
Figure 3: Discrepancies between primary outcomes in protocols and publications and in ClinicalTrials.gov and publications (n=65)

177x177mm (274 x 274 DPI)

Table 1. Characteristics of randomized trials included in the study (*n*=65)

Characteristic	Value
Journal - n (%)	
New England Journal of Medicine	22 (34%)
Journal of Clinical Oncology	43 (66%)
Type of tumor - n (%)	
Breast	12 (18%)
Colon/rectum	4 (6.1%)
Gastro intestinal (excluding colon/rectum cancer)	8 (12%)
Female reproductive tract	9 (14%)
Head & neck (including thyroid cancer)	7 (11%)
Kidney	2 (3.1%)
Lung	6 (9.2%)
Prostate	3 (4.6%)
Skin	7 (11%)
Any site	7 (11%)
Type of intervention - n (%)	
Chemotherapy	9 (14%)
Targeted therapy	33 (51%)
Radiation and chemotherapy	2 (3%)
Surgery and/or radiation therapy	8 (12%)
Supportive care	11 (17%)
Screening and/or diagnostic	2 (3%)
No. of study groups - n (%)	
2	60 (92%)
>2	5 (8%)
No. of patients included – median (IQR)	452 (253-704)
Funding source - n (%)	
Industry	32 (49%)
Non-industry	33 (51%)
Outcomes reported in the three sources - n (%)	
Total	81 (100%)
Overall survival	23 (28%)
Time to event*	30 (37%)
Response rate	2 (2%)
Patient-reported outcome	17 (21%)
Other	9 (11%)

*Time to event includes progression-free survival, disease-free survival, event-free survival, relapse-free survival, time to disease progression

Table 2. Quality of outcome descriptions in each data source (*n*=81) according to the Zarin et al. framework [11]

Item	Protocols	ClinicalTrials.gov	Publications
1) Domain	81 (100%)	80 (99%)	81 (100%)
2) Specific measurement	68 (84%)	50 (62%)	77 (95%)
3) Specific metric	72 (89%)	71 (88%)	77 (95%)
4) Method of aggregation of data	68 (84%)	63 (78%)	72 (89%)
5) Time frame	71 (88%)	48 (59%)	70 (86%)
6) Identity of outcome assessor	40 (49%)	14 (17%)	40 (49%)
7) Blinding of outcome assessor	34 (42%)	28 (34%)	34 (42%)
Minimal acceptable reporting of outcome[†]	59 (73%)	18 (22%)	61 (75%)
Optimal reporting of outcome[‡]	24 (30%)	3 (4%)	21 (26%)

[†]Minimal acceptable reporting of outcome involves the reporting of the five elements from the Zarin et al. framework [11]

[‡]Optimal reporting of outcome involves the reporting of the five elements from the Zarin et al. framework and information about the blinding and the identity of the outcome assessor.

2

Supplementary table 1. Items assessed to describe the quality of outcome descriptions

Item	Description	Example
Domain*	What is being measured?	Progression-free survival
Specific Measurement*	How was the outcome measured?	Radiographic evidence of disease progression
Specific Metric*	What metric is used to characterize each participant's results?	Time-to-event data
Method of data aggregation*	How were data aggregated in a single measure within each group?	Time-to-event data
Time frame*	Specifies the time point(s) at which the outcome measure was assessed and for which data are reported	Date of randomization to the date of disease progression (measured every 8 weeks) or death (whichever occurs first)
Assessor	Is the identity of the person who assessed the outcome reported?	Determined by blinded independent imaging review
Blinding	Is it specified if (or if not) outcome assessors blinded from intervention received	Determined by blinded independent imaging review

*Items from the framework of Zarin et al. [11]

Supplementary table 2. Modification of outcomes from protocol to publication. *We only present the necessary information to understand outcome modifications, and outcome descriptions may have been shortened.

Author	Outcome reported in the protocol*	Outcome reported in publication*	Comment (protocol to publication)
Bakitas	• FACIT-Pal-Total Score, • *QUAL-E-Total Score,* and • Mood (CES-D) as measured at 6, 12,18, 24 and 36 weeks after randomization	• Functional Assessment of Chronic Illness Therapy–Palliative Care [FACIT-Pal]… • …and by Treatment Outcome Index (TOI). • Symptom impact (assessed by Quality of Life at End of Life [QUAL-E] symptom impact subscale) • Mood (assessed by Center for Epidemiologic Studies–Depression scale [CES-D] • One-year and overall survival. • Resource use and location of death. Patient-reported hospital and intensive care unit (ICU) days and emergency department(ED) visits	• Addition of new outcomes: TOI and one year survival are not mentioned in the protocol • Modification of an outcome : use of QUAL-E sub score and not total score as mentioned in the protocol • Possible change form secondary to primary outcome: Overall survival is not mentioned clearly as a primary outcome in the protocol • Possible addition of a new outcome: Resource use was a variable in the data collection forms but not specified as a study outcome in the protocol
Berry	• Frequencies of symptom/quality of life (QOL) issues (SQIs) initiated by patient or caregivers in the first post-treatment start date clinic visit. • Duration of conversation regarding patient/caregiver-initiated SQIs audio-recorded in the first post-treatment start date clinic visit. • Acceptance of and adherence to clinician-recommended therapies for SQIs at 6 weeks after treatment start date. • Frequencies of self-care strategies implemented for SQIs at 6 weeks after treatment start date. • Symptom distress scores (SDS) at 6 weeks after treatment start and 2-4 weeks after treatment end date.	• We used the original 13-item Symptom Distress Scale (SDS) plus two items (impact on sexual activity and interest, fever/chills) to form the SDS-15 at 2-4 weeks after treatment end.	• Modification of an outcome: use of a modified SDS scale not mentioned in the protocol • Omission of outcomes: other outcomes from the protocol are not mentioned in the publication. Of note, frequency of symptoms/QOL issues using audio recorded data was published in a different publication (Berry DL, BMC Cancer, 2014)
Breibart	• Spiritual wellbeing • QOL • Psychological distress (anxiety and depression) • End-of-life despair (hopelessness, desire for	• Spiritual well-being (Functional Assessment of Chronic Illness Therapy Spiritual Well-Being Scale) • Overall QOL (McGill Quality of Life Questionnaire).	• Change from primary outcome in protocol to secondary in publication for "Psychological distress" and "End-of-Life despair" Of note, primary outcomes are not clearly stated in the protocol.

	hastened death, and suicidal ideation)		
Eijzenga*	Outcomes are assessed at 4 weeks after randomization and 4 months after randomization. • Discussion of psychosocial problems. • Counselors' awareness of the psychosocial problems of the counselee • Management of psychosocial problems	Outcomes are assessed 4 weeks after randomization • Communication on psychosocial issues • Counselors' awareness of psychosocial problems of the counselee • Management of these psychosocial problems	• Omission of outcomes at 4 months after randomization. This discrepancy is acknowledged in the discussion: *"The supplemental telephone session was not a standard procedure within the genetic counseling process, and thus, the second phase of the trial cannot be viewed simply as a follow-up of the first phase. Therefore, the results of this second phase will be reported in a subsequent article."* Results at 4 months are reported in Eijzenga W, Clin Genet, 2015.
Fernando	• Local recurrence (LR) is indicated when a follow-up examination shows growth of primary tumor or abnormality in the resected lobe on CT scan. Because scarring may occur adjacent to the brachytherapy site, a CT scan will be obtained at 3 months. This will form the baseline study that local recurrence will be judged against. Increased parenchymal opacification (by 25% or more) adjacent to the staple line/mesh line will be considered suspicious for local recurrence. The repeated CT scans should allow any significant changes to be observed even if there is a slight image artifact from the metal seeds which may interfere with interpretation of the CT scan. Any suspicious areas should be confirmed by a needle biopsy.	• The primary end point was the time to Local Recurrence (LR). LR was defined as recurrence within the primary tumor lobe at the staple line (local progression), recurrence within the primary tumor lobe away from the staple line (involved lobe failure), or recurrence within hilar lymph nodes • Of note, in the publication: "Regional recurrence was defined as recurrence within another lobe on the same side as the resection or within ipsilateral mediastinal or subcarinal lymph nodes. Distant recurrence was defined as recurrence within contralateral, mediastinal, or hilar lymph nodes or distant metastatic disease."	• Possible modification of an outcome: Definition of LR in the protocol does not mention extent to ipsilateral hilar lymph nodes
Guimbaud	• The endpoint is the time to treatment failure (TTF), defined as the period between the date of randomization and the date of progression, or the date of premature discontinuation of treatment, or the date of relapse after response, or the death date.	• The primary end point was time-to-treatment failure (TTF) for the first-line therapy in the two treatment arms. TTF was defined as the time between random assignment and disease progression, treatment discontinuation, or death.	• Possible modification of an outcome: definition of TTF in the publication does not encompass relapse after response.
Irwin	• Arthralgia, BMD, QOL with a total of 20 different measurements. • Sample size is determined using the BPI score	• Arthralgia (BPI, DASH, WOMAC) • Grip strength using a bulb dynamometer	• Modification. In the protocol Arthralgia is measured using QuickDASH (11 items). In publication DASH is mentioned (30 items).

			• Omission of outcomes: Several measures stated in the protocol are not reported. Of note, some of the other outcomes are reported in a second publication (Arem H, J Cancer Surviv, 2016)
Larkin*	• In initial protocol (2013), primary endpoint is overall survival (OS) • In an amendment (2014), primary endpoint is modified to a co-primary outcomes of OS and progression free survival (PFS)	• Progression-free survival and overall survival were co-primary end points; results regarding progression-free survival are presented here	• Only one of the two co-primary outcome is reported in the publication. • Differential time of assessments for OS and PFS were planned in the protocol and ClinicalTrials.gov and explained in the publication
Schwartz	• Knowledge, • Satisfaction • Decisional conflict • Psychosocial distress • Quality of life. • Uptake of BRCA1/BRCA2 mutation testing	• *BRCA1/2* knowledge was measured at baseline and at the 2-week follow-up assessment with the breast Cancer Genetic Counseling Knowledge scale. • Decisional conflict regarding *BRCA1/2* testing was measured at all assessments with the 10-item version of the Decisional Conflict Scale (DCS). • Genetic counseling satisfaction was measured at 2 weeks with the Genetic Counseling Satisfaction Scale. • Cancer-specific distress was measured at all assessments with the Impact of Event Scale (IES) Perceived stress was measured at all assessments with the four-item version of the Perceived Stress Scale (PSS) • Quality of life was measured at baseline and at 3 months with the Short Form-12 (SF-12) Mental Component Summary (MCS) and Physical Component Summary (PCS). • Uptake of BRCA1/BRCA2 mutation testing has been switched to secondary outcome	• Change from primary to secondary: Uptake of BRCA1/BRCA2 mutation testing has been changed from primary to secondary outcome in the publication.
Wenzel	• Change of Quality of life (as measured by the FACT-Cx) between baseline (T1) and 3 months (T2)	• The primary outcome was change in FACT-Cx score from baseline to 4 months	• Modification of an outcome: change in time frame

*Trials not counted as containing discrepant outcome reporting because modification between planned and published outcomes are explained and discussed in the publication

Supplementary table 3. Modifications from protocol to ClinicalTrials.gov. *We only presented the necessary information to understand outcome modifications and outcome descriptions may have been shortened.

Author	Outcome reported in the protocol*	Outcome reported in ClinicalTrials.gov*	Comment (protocol to ClinicalTrials.gov)
Bakitas	The primary study endpoints are: • FACIT-Pal-Total Score, • QUAL-E-Total Score, and • Mood (CES-D) Overall survival is not mentioned clearly as a primary endpoint.	Quality of living assessments will include quality of life (QOL), mood, and symptom intensity measures using the following measures: • Functional Assessment of Chronic Illness Therapy-Palliative Care (FACIT-Pal): • Quality of Life at End of Life (QUAL-E). • Center for Epidemiological Study- Depression (CES-D). Quality of end of life care [Time Frame: chart review at time of death and caregiver proxy interview 2-3 months after patient death] • End of life (EOL) Care Data Collection • Quality of Dying and Death Measure (QODD). Estimate and compare the hazard ratios and median survival before and after 1 year from enrollment [Time Frame: From enrollment until patient death or end of study] Caregiver burden and QOL will be measured using: • QOL- Cancer- a self-report measure of QOL for caregivers of patients with cancer. • Montgomery Borgatta Caregiver Burden • Center for Epidemiological Study- Depression CESD is a measure of depressive symptoms. • Functional Assessment of Chronic Illness Therapy - Spiritual Module (FACIT-Sp) • Prigerson Inventory of Complicated Grief-Short form (ICG-SF) embedded in the Quality of Death and Dying (QODD).	• Change from secondary to primary: several measures reported as secondary outcomes in the protocol are registered as primary outcomes.

Berry	Outcomes are assessed 6-8 weeks after treatment start date (T3). • Frequencies of symptom/QOL issues (SQIs) initiated by patient or caregivers • Duration of conversation regarding patient/caregiver-initiated SQIs • Acceptance of and adherence to clinician-recommended therapies for SQIs • Frequencies of self-care strategies implemented for SQIs • Symptom distress scores (SDS) at 6 weeks after treatment start and 2-4 weeks after treatment end date.	• Symptom burden… • …and QOL 2-4 weeks post treatment [Time Frame: 3 years]	• Possible omission of outcomes • Modification of outcomes. Time frame does not match with those in the protocol.
Breibart	• Spiritual wellbeing • Psychological distress (anxiety and depression) • QL • End-of-life despair (hopelessness, desire for hastened death, and suicidal ideation)	• The primary outcomes to be measured include measures of spiritual well-being (meaning)… • … and psychological distress (depression, hopelessness, optimism, QOL).	• Possible omission of outcomes
Budd	• The primary outcome is disease-free survival (DFS).	• Disease-free survival by medical history, physical exam, and mammograms every 6 months (annually for mammograms) for 5 years and then annually for 15 years or until death • Compare overall survival of patients among the 2 treatment arms by medical history and physical exam every 6 months for 5 years and then annually • Compare toxicity among the 2 treatment arms by medical history and physical exam every 6 months for 5 years and then annually	• Change from secondary to primary. Overall survival is a secondary outcome in the protocol. • Safety outcome registered as a primary outcome (counted as Primary outcome in CT.gov reported as secondary in Publication)
DiSilvestro	• Progression free survival will be the primary outcome	• Progression-free survival [Time Frame: Up to 5 years] • Frequency and severity of adverse events assessed by Common Terminology Criteria for Adverse Events version 3.0 [Time Frame: Up to 5 years]	• Safety outcome registered as a primary outcome (counted as Primary outcome in CT.gov reported as secondary in Publication)
Eijzenga	Outcomes are assessed at 4 weeks after randomization and 4 months after randomization. • Discussion of psychosocial problems.	• Communication on psychosocial issues at the counseling session [Time Frame: 3 weeks & 4 months after randomization].	• Modification of outcomes. Change in the time frame

	• Counselors' awareness of the psychosocial problems • Management of psychosocial problems	• Counselors' awareness of psychosocial problems of the counselee at the counseling session [Time Frame: 3 weeks & 4 months after randomization]. • Management of these psychosocial problems of the counselee during the counseling session [Time Frame: 3 weeks & 4 months after randomization]	
Escudier	• Patient preference (pazopanib vs. sunitinib) as assessed by patient preference Questionnaire	• Number of Participants With Preference for Pazopanib Versus Sunitinib as Assessed by the Patient Preference Questionnaire (PPQ) [Time Frame: End of treatment of both study drugs (maximum of 22 weeks)] • Number of Participants Answering "Yes," "no," or Not Applicable (N/A) to the Question of Whether the Indicated Factors Influenced Their Preference for Sunitinib or Pazopanib Treatment as Assessed by the Patient Preference Questionnaire [Time Frame: End of treatment of both study drugs (maximum of 22 weeks)]	• Change from secondary to primary. Reason for patient preference as assessed by patient preference questionnaire is noted in the protocol as a secondary outcome
Irwin	• Arthralgia, BMD, QOL with a total of 20 different measurements. • Sample size is determined using the BPI score	• Change from Baseline in Arthralgia Severity [Time Frame: 6 months] measured using the Brief Pain Inventory (BPI).	• Omission of outcomes. Several measurements reported in the protocol are not registered
Rapp	• Cognitive performance	• Fatigue... • …subjective confusion… • … and cognitive performance at 24 weeks	• Addition of an outcome. Imprecise registration.
Schwartz	• Knowledge, • Satisfaction • Decisional conflict • Psychosocial distress • QOL • Uptake of BRCA1/BRCA2 mutation testing	• Knowledge assessed by genetic testing knowledge measure at post-counseling and 3 months • Decisional Conflict Satisfaction at post-counseling and 3 months • QOL as assessed by SF-12 health survey at 3 and 6 months • Distress as assessed by Impact of Events Scale Brief Symptom Inventory MICRA at 3 and 6 months • Uptake of BRCA1/BRCA2 mutation testing as measured by genetic test results at 3 and 6 months	• Omission of an outcome. Satisfaction is not registered as a primary outcome.
Sparano	• The primary endpoint of this study is disease-free survival (DFS), defined to be time from	• Compare the disease-free survival… • …and overall survival in patients with node-	• Change from secondary to primary. Overall survival is a secondary outcome in the

	randomization to disease recurrence or death without recurrence	positive or high-risk node-negative operable stage II or IIIA breast cancer treated with docetaxel or paclitaxel after doxorubicin and cyclophosphamide.	protocol and registered as a primary outcome.
Wenzel	• QOL (as measured by the FACT-Cx)	• To evaluate the efficacy of a multicomponent biobehavioral psychosocial telephone counseling (PTC) intervention for cervical cancer survivors, compared to usual care [Time Frame: 10 years]	• Imprecise registration preventing assessment of outcome modification

Supplementary table 4. Modifications from ClinicalTrials.gov to publications

Author	Outcome reported in the ClinicalTrials.gov	Outcome reported in publication	Comment (ClinicalTrials.gov to publication)
Bakitas	Quality of living assessments will include quality of life (QOL), mood, and symptom intensity measures • Functional Assessment of Chronic Illness Therapy-Palliative Care (FACIT-Pal): • Quality of Life at End of Life (QUAL-E). • Center for Epidemiological Study- Depression (CES-D). Quality of end of life care [Time Frame: chart review at time of death and caregiver proxy interview 2-3 months after patient death] • End of life (EOL) Care Data Collection • Quality of Dying and Death Measure (QODD). Estimate and compare the hazard ratios and median survival before and after 1 year from enrollment Caregiver burden and QOL • QOL- Cancer- a self-reported measure of QOL for caregivers of patients with cancer. • Montgomery Borgatta Caregiver Burden • Center for Epidemiological Study- Depression CESD is a measure of depressive symptoms. • Functional Assessment of Chronic Illness Therapy - Spiritual Module (FACIT-Sp) • Prigerson Inventory of Complicated Grief-Short form (ICG-SF) embedded in the Quality of Death and Dying (QODD).	• QOL (assessed by 46-item Functional Assessment of Chronic Illness Therapy–Palliative Care [FACIT-Pal]… • …and by Treatment Outcome Index (TOI). • Symptom impact (assessed by four-item Quality of Life at End of Life [QUAL-E] symptom impact subscale) • Mood (assessed by 20-item Center for Epidemiologic Studies–Depression scale [CES-D] • One-year and overall survival. • Resource use and location of death. Patient-reported hospital and intensive care unit (ICU) days and emergency department(ED) visits	• Addition of a new outcome TOI not mentioned in the protocol • Addition of a new outcome (1-year survival) not mentioned in the protocol • Addition of a new outcome: resource use • Modification of an outcome: use of QUAL-E sub-score and not total score as mentioned in the protocol • Omission of outcomes related to quality of end of life • Omission of outcomes related to caregiver burden
Berry	• Symptom burden… • …and QOL 2-4 weeks post treatment [Time Frame: 3 years]	• We used the original 13-item Symptom Distress Scale (SDS) plus two items (impact on sexual activity and interest, fever/chills) to form the SDS-15.	• Omission of an outcome (QOL)
Breibart	• The primary outcomes to be measured include measures of spiritual well-being (meaning)…	• Spiritual well-being (Functional Assessment of Chronic Illness Therapy Spiritual Well-Being	• Omission of outcomes related to depression, hopelessness, optimism

	• … and psychological distress (depression, hopelessness, optimism QOL).	Scale) • Overall QOL (McGill Quality of Life Questionnaire).	
Budd	• Disease-free survival by medical history, physical exam, and mammograms every 6 months (annually for mammograms) for 5 years and then annually for 15 years or until death • Compare overall survival of patients among the 2 treatment arms by medical history and physical exam every 6 months for 5 years and then annually • Compare toxicity among the 2 treatment arms by medical history and physical exam every 6 months for 5 years and then annually	• The primary outcome was disease-free survival (DFS) defined as time from registration (random assignment) to first instance of disease recurrence (local, regional, or distant), new breast primary tumor, or death as a result of any cause.	• Change form primary to secondary. Overall survival is registered as a primary outcome but is a secondary outcome in publication (and protocol) • Safety outcome registered as primary outcome (counted as primary outcome in CT.gov published as a secondary outcome)
DiSilvestro	• Progression-free survival [Time Frame: Up to 5 years] • Frequency and severity of adverse events assessed by Common Terminology Criteria for Adverse Events version 3.0	• Primary endpoint was progression-free survival (PFS).	• Safety outcome registered as primary outcome (counted as primary outcome in CT.gov published as a secondary outcome)
Eijzenga	• Communication on psychosocial issues at the counseling session [Time Frame: 3 weeks & 4 months after randomization] • Counselors' awareness of psychosocial problems of the counselee at the counseling session [Time Frame: 3 weeks & 4 months after randomization] • Management of these psychosocial problems of the counselee during the counseling session [Time Frame: 3 weeks & 4 months after randomization]	Assessed 4 weeks after randomization • Discussion of psychosocial problems. • Counselors' awareness. • Management of psychosocial problems.	• Modification of outcomes: change in the time frame • Omission of outcomes at 4 months is explained in the discussion.
Escudier	• Number of Participants With Preference for Pazopanib Versus Sunitinib as Assessed by the Patient Preference Questionnaire (PPQ) [Time Frame: End of treatment of both study drugs (maximum of 22 weeks)]	• The primary end point was patient preference. Patient preference was assessed by questionnaire at study end, before unblinding and before patients were informed of the final disease assessment. Patients were asked whether they preferred to continue treatment with the drug administered	• Change from primary to secondary. number of participants Answering "Yes," "no," or not applicable (N/A) to the question of whether the indicated factors influenced their preference for sunitinib or pazopanib treatment is a secondary

	• Number of Participants Answering "Yes," "no," or Not Applicable (N/A) to the Question of Whether the Indicated Factors Influenced Their Preference for Sunitinib or Pazopanib Treatment as Assessed by the Patient Preference Questionnaire [Time Frame: End of treatment of both study drugs (maximum of 22 weeks)]	during period 1 or period 2 or whether they had no preference, assuming both treatments were equally effective.	outcome in the publication
Irwin	• Change from Baseline in Arthralgia Severity [Time Frame: 6 months] measured using the Brief Pain Inventory (BPI).	• Arthralgia (BPI, DASH, WOMAC) • Grip strength using a bulb dynamometer	• Addition of primary outcomes in the publication
Larkin	• Overall Survival (OS) • Progress Free Survival (PFS)	• Progression-free survival and overall survival were co-primary end points; results regarding progression-free survival are presented here	• Only one of the 2 primary outcomes are reported • Differential time of assessments for OS and PFS were planned in the protocol and ClinicalTrials.gov and explained in the publication
Rapp	• Fatigue... • ...subjective confusion… • … and cognitive performance at 24 weeks	• The primary objective of this randomized trial was to assess the effect of donepezil on overall cognitive performance after 24 weeks of therapy. The summary cognitive composite score was computed by standardizing (z score) eight individual test scores representing the major cognitive domains (HVLT-R total recall, HVLT-R DR, mROCF delayed figural recall, DST total, Controlled Oral Word Association Test, TMT-A, TMT-B, and GP-D)	• Omission of outcome related to fatigue • Omission of outcome related to subjective confusion
Schwartz	• Knowledge assessed by genetic testing knowledge measure at post-counseling and 3 months • Decisional Conflict Satisfaction at post-counseling and 3 months • QOL as assessed by SF-12 health survey at 3 and 6 months • Distress as assessed by Impact of Events Scale Brief Symptom Inventory MICRA at 3 and 6 months • Uptake of BRCA1/BRCA2 mutation testing as	• *BRCA1/2* knowledge was measured at baseline and at the 2-week follow-up assessment with the Breast Cancer Genetic Counseling Knowledge scale. • Decisional conflict regarding*BRCA1/2* testing was measured at all assessments with the 10-item version of the Decisional Conflict Scale (DCS). • Genetic counseling satisfaction was measured at 2 weeks with the Genetic Counseling Satisfaction Scale. • Cancer-specific distress was measured at all assessments with the Impact of Event Scale (IES)	• Change from primary to secondary: uptake of BRCA1/BRCA2 mutation testing is a secondary outcome in publication • Addition of an outcome in publication: genetic counseling satisfaction

	measured by genetic test results at 3 and 6 months	Perceived stress was measured at all assessments with the four-item version of the Perceived Stress Scale (PSS) • QOL was measured at baseline and at 3 months with the Short Form-12 (SF-12) Mental Component Summary (MCS) and Physical Component Summary (PCS).	
Sparano	• Compare the disease-free survival… • …and overall survival in patients with node-positive or high-risk node-negative operable stage II or IIIA breast cancer treated with docetaxel or paclitaxel after doxorubicin and cyclophosphamide.	• The primary endpoint was DFS	• Change from primary to secondary. overall survival is a secondary outcome in publication (and protocol)
Wenzel	• To evaluate the efficacy of a multicomponent biobehavioral psychosocial telephone counseling (PTC) intervention for cervical cancer survivors, compared to usual care [Time Frame: 10 years]	• The primary outcome was change in FACT-Cx score from baseline to 4 months	• Imprecise registration preventing assessment of outcome modifications

Supplementary table 5. Reasons for discrepancies in identifying modifications between planned and published outcomes when using protocols or ClinicalTrials.gov (n=16). Comparison of protocols/publication is considered a reference for ORB identification.

Reason	Example	N (%)
False negative ORB: imprecise registration covers a possible modification from protocol to publication	*In the protocol*, time to therapeutic failure (TTF) is defined as the period between the date of randomization and the date of progression, or the date of premature discontinuation of treatment, or the date of relapse after response, or the death date. I*n the publication*, TTF is defined as the time between random assignment and disease progression, treatment discontinuation, or death. *In ClinicalTrials.gov*, primary outcome is simply "Time to therapeutic failure"	6 (37%)
False negative ORB: a secondary outcome in the protocol is registered and reported as a primary outcome in publication	In a trial of early versus delayed initiation of concurrent palliative oncology care, overall survival is not reported as a primary outcome in the protocol but is registered as such in ClinicalTrials.gov and reported as such in the publication.	1 (6 %)
False positive ORB because a secondary outcome in protocol and publication is registered as a primary outcome	In a trial assessing donepezil in treatment for patients who have undergone radiation therapy for brain tumors, fatigue and confusion are registered as primary outcomes but appear as secondary in both the publication and protocol	1 (6%)
False positive ORB because a primary outcome specified in the protocol was not registered but reported	In a study assessing telephone versus in-person genetic counseling for hereditary breast and ovarian cancer, the outcome "satisfaction" was a primary outcome in the protocol and publication but was not registered at all.	1 (6%)
False positive ORB because a secondary outcome in the protocol (and publication) is registered as a primary outcome in ClinicalTrials.gov	In a trial assessing Cisplatin and Radiation Therapy with or without Tirapazamine in treating patients with cervical cancer, frequency and severity of adverse events is registered as a primary outcome whereas it is not in protocol and publication	5 (31 %)
Unclear, misleading outcome in ClinicalTrials.gov	Outcome registered is "To evaluate the efficacy of a bio behavioral psychosocial telephone counseling intervention for cervical cancer survivors, compared to usual care"	2 (12%)