Annals of Oncology Advance Access published January 9, 2013

original article

Annals of Oncology 00: 1–6, 2013
doi:10.1093/annonc/mds636

Bias in reporting of end points of efficacy and toxicity in randomized, clinical trials for women with breast cancer

F. E. Vera-Badillo, R. Shapiro, A. Ocana, E. Amir & I. F. Tannock*

Division of Medical Oncology & Hematology, Princess Margaret Hospital and University of Toronto, Toronto, Canada

Received 8 August 2012; revised 1 November 2012; accepted 5 November 2012

Background: Phase III randomized, clinical trials (RCTs) assess clinically important differences in end points that reflect benefit to patients. Here, we evaluate the quality of reporting of the primary end point (PE) and of toxicity in RCTs for breast cancer.

Methods: PUBMED was searched from 1995 to 2011 to identify RCTs for breast cancer. Bias in the reporting of the PE and of toxicity was assessed using pre-designed algorithms. Associations of bias with the Journal Impact Factor (JIF), changes in the PE compared with information in ClinicalTrials.gov and funding source were evaluated.

Results: Of 164 included trials, 33% showed bias in reporting of the PE and 67% in the reporting of toxicity. The PE was more likely to be reported in the concluding statement of the abstract when significant differences favoring the experimental arm were shown; 59% of 92 trials with a negative PE used secondary end points to suggest benefit of experimental therapy. Only 32% of articles indicated the frequency of grade 3 and 4 toxicities in the abstract. A positive PE was associated with under-reporting of toxicity.

Conclusion: Bias in reporting of outcome is common for studies with negative PEs. Reporting of toxicity is poor, especially for studies with positive PEs.

Key words: bias, breast cancer, clinical trials, methodology

introduction

Phase III randomized, clinical trials (RCTs) are designed to detect or exclude clinically important differences between experimental and control groups in end points that reflect benefit to patients [1]. Such trials provide the gold standard to evaluate the efficacy and toxicity of new drugs before approval by regulatory authorities [2, 3].

Appropriate design and objective reporting of RCTs in journals are essential to inform clinicians about the activity and safety of new medical interventions. It is good practice to design RCTs with no more than three outcomes for which hypothesis testing is planned [4]. Otherwise multiple significance testing may lead to apparently significant results that occur by chance. These outcomes should normally include at least one end point reflecting potential benefit and at least one reflecting potential harm (e.g. grade 3 and 4 adverse events). Reviews have shown that a substantial proportion of clinical trials have suboptimal reporting of harm—a large number of trials have shown deficiency in the report of toxicity, especially severe toxicity, graded as 3 and 4 [5]. Guidelines such as Consolidated Standards of Reporting Trials can improve the quality of reporting of clinical trials [6].

Bias in reporting of clinical trials and selective publication can create false perceptions of drug efficacy and safety. There is evidence for selective reporting of favorable results and suppression of unfavorable data from publication, leading to inappropriate conclusions [2, 7]. This may be influenced by publication bias—the association between positive results and acceptance of reports for publication [4, 8]. Selection bias can affect not only the interpretation of the trial itself but also the interpretation of subsequent systematic reviews or overviews, producing inaccurate summaries of research [2, 9] and misrepresentation of toxicity [10]. Reporting of harms may be viewed as discrediting the reporting of benefits.

Spin, a type of bias, is defined as use of reporting strategies to highlight that the experimental treatment is beneficial, despite a statistically non-significant difference in the primary outcome, or to distract the reader from statistically non-significant results [11]. It is important to recognize the presence of bias and spin in reports of clinical trials, and to evaluate their importance when placing a RCT in context and ascribing a level of credibility [12].

Here, we review the papers reporting RCTs for breast cancer to quantify the extent of biased reporting, and to guide readers in judging the credibility of their conclusions. Because busy

*Correspondence to: Prof. I. F. Tannock, Division of Medical Oncology & Hematology, Princess Margaret Hospital, 610 University Avenue, Toronto, ON, Canada M5G 2M9. Tel: +1-416-946-2245; Fax: +1-416-946-4563; E-mail: ian.tannock@uhn.ca

clinicians often read only the abstracts of publications [13], we have emphasized accurate reporting of the primary endpoint (PE) and toxicity in the abstract. We hypothesized that despite the availability of guidelines to minimize bias in reporting, this remains prevalent.

methods

literature search and study selection

We carried out an electronic search of MEDLINE (Host: PubMed) for publications from January 1995 to August 2011 using the following MeSH terms: randomized clinical trial, RCT, Phase III and breast neoplasms or breast cancer. The inclusion criteria were human studies published in English and including patients aged ≥18 years. We excluded trials with sample size less than 200 patients as they were unlikely to be definitive studies and more likely to have higher levels of bias. Furthermore, the focus of this study was to assess reporting of clinical trials that potentially change clinical practice. Other exclusion criteria included trials where the PE was not a time to event end point, commentaries, review articles, observational studies, meta-analyses, ongoing studies and articles for which only the abstract was available.

data extraction and analysis

The following data were extracted independently from each RCT by two authors (FV and RS): setting of treatment (adjuvant versus metastatic), sponsorship (industry versus non-industry or not stated), year of publication, impact factor of the journal where the trial was published (as a continuous variable), the primary and secondary end points [overall survival (OS), progression-free survival (PFS), disease-free survival (DFS), response rate, toxicity or quality of life], whether the PE was defined in the abstract and/or in the paper and whether the secondary end points were reported in the abstract and in the paper. If data on PFS or DFS were not reported, similar end points such as time to progression, time-to-treatment failure or event-free survival were extracted instead. For papers with more than two arms, the assessment of efficacy was not considered if at least one arm was positive (see Figure 1); if all were negative, it was included for analysis. For toxicity, all trials were included regardless of the number of study arms. The journal impact factor (JIF) was extracted from the Gerstein Science Information Center of the University of Toronto through the Web of Science (Host:BIOSIS) and was retrieved from the Journal Citation reports up to 2012.

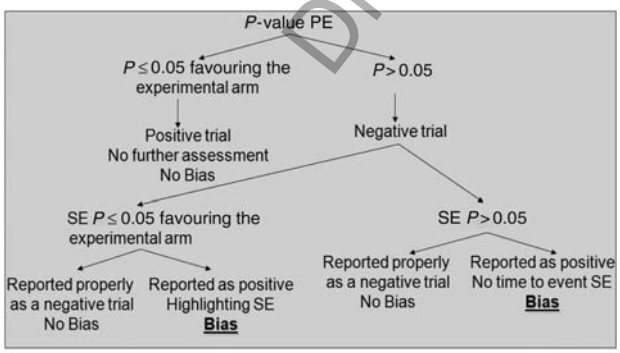

Figure 1. Decision tree for assessment of reporting of the primary end point in the concluding statement of the abstract. PE, primary end point; SE, secondary end point.

For recent trials, we also explored whether the PE listed in the trial registry ClinicalTrials.gov was the same as that reported in abstracts or papers reporting the same trial. We initially searched articles for any reference to trial registration. For those not reporting such data, we manually searched the ClinicalTrials.gov database for trial-related information.

end points

The primary analysis included the assessment of the prevalence of bias and spin in reporting the PE of the study, and in reporting toxicity; the secondary analysis evaluated predictors of bias and spin.

Bias was defined as inappropriate reporting of the PE and toxicity, with emphasis on reporting of these outcomes in the abstract. A decision tree was used to assess whether the PE was reported with bias, and whether a secondary end point was used to imply benefit of the experimental arm (Figure 1). Studies where multiple PEs were reported and where at least one end point was positive were not considered for assessment of bias.

Bias in reporting of toxicity was assessed using a hierarchy scale from 1 (excellent) to 7 (very poor) to indicate whether reporting of grade 3 and 4 toxicities occurred in the concluding statement of the abstract, elsewhere in the abstract, in the results section of the paper, only in a table or not at all, with lower scores if they were also included in the discussion section of the paper (Figure 2). We defined reporting of grade 3 and 4 toxicities as poor if they were not mentioned in the abstract (scale of 5–7 in our hierarchy), and good (scale 1–2) if they were mentioned in the concluding statement of the abstract. When there were no statistically significant differences in toxicity, a general statement in the abstract was deemed to be sufficient; when statistically significant differences were seen, it was expected that they would be reported in the abstract.

Spin was defined as the use of words in the concluding statement of the abstract to suggest that a trial with a negative PE was positive based on some apparent benefits shown in one or more secondary end points.

Predictors of bias included the impact of source of funding on reporting of the PE, the relationship between the quality of reporting of the PE and of toxicity and the frequency of a change in the PE from the original protocol (when this could be obtained from ClinicalTrials.gov) to a published paper. Also whether the JIF, setting of the trial (adjuvant versus metastatic), definitive versus surrogate end point (e.g. OS compared with PFS or DFS) and modification of the PE influenced the prevalence of bias or spin.

statistical analysis

Data were presented descriptively as means or medians. Predictors of bias were assessed by the chi-squared test and by univariable logistic regression (categorical variables) or univariable linear regression (continuous variables). Correlations between variables were tested using Spearman's correlation and the magnitude of association was assessed as described by Burnand et al. [14]. All statistical analyses were conducted using SPSS statistical software version 17 (IBM Corp, Arkmon, New York). All significance tests were two-sided using an alpha level of 0.05. No correction was applied for multiple statistical testing.

results

A total of 568 articles were identified initially and 164 RCTs (148 for systemic therapy, 11 for radiation therapy and 5 for surgical therapy) were eligible for analysis (Figure 3). The characteristics of the trials are reported in Table 1. Eighty-one trials (49.4%) were conducted in the adjuvant setting and 83 (50.6%) evaluated experimental therapy for women with

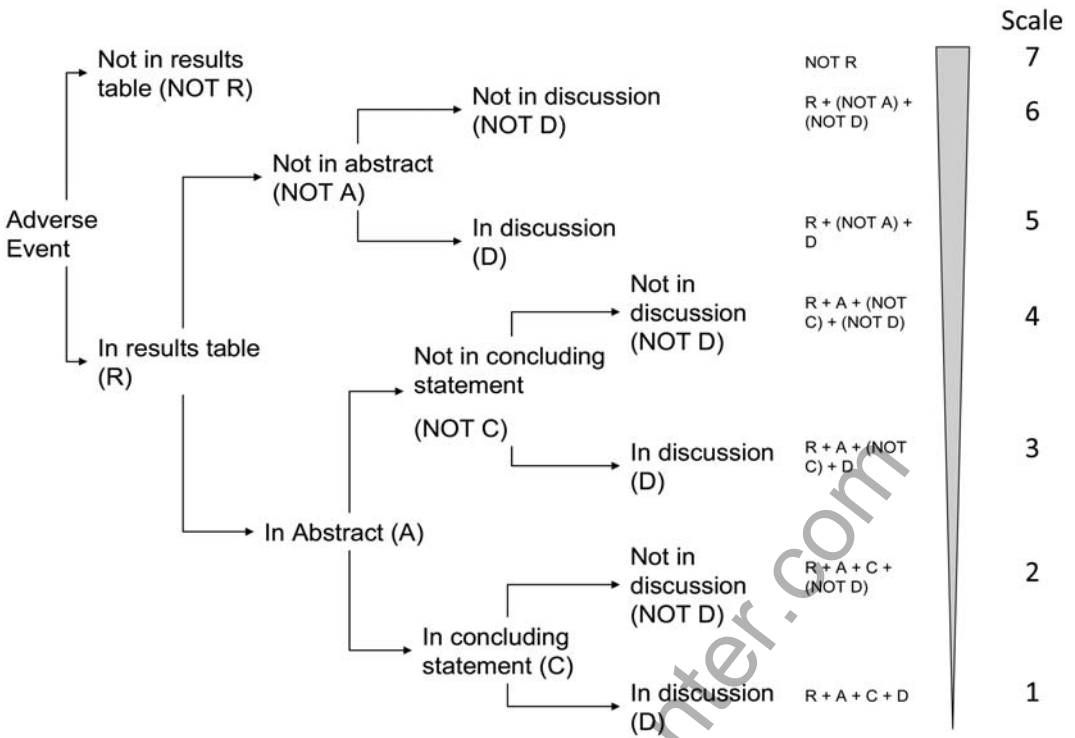

Figure 2. Hierarchy scale for reporting of adverse events. Not R, not eported in results table; Not D, not reported in discussion; Not A, not reported in abstract; Not C, not reported in concluding statement; R, reported in results table; D, reported in discussion; A, reported in abstract; C, reported in concluding statement.

metastatic breast cancer. OS was the PE in only 27 trials (16.5%) and DFS or PFS was the PE in 137 studies (83.5%). Only 30 trials (18%) were identified as included in ClinicalTrials.gov. Among these studies, the PE was changed in the final report in seven (23.3%) studies. Seventy-two (43.9%) studies were positive with a significant *P*-value for the difference in primary endpoint favoring the experimental arm compared with 92 (56.1%) with a non-significant *P*-value. The majority, 150 trials (91.4%), were published in medium or high impact journals; the calculated median impact factor was 19. Bias in reporting efficacy (Spearman's rho = 0.27 and chi-squared $P = 0.30$) and toxicity (Spearman's rho = 0.46 and chi-squared $P = 0.06$) were not influenced by date of publication.

bias in the reporting of the primary end point

Fifty-four trials (32.9% of the total sample) were reported as positive, based on a non-PE, despite not finding a statistically significant difference in the PE. These reports were biased and used spin in an attempt to conceal that bias. When assessing only those reports with a non-significant difference in the PE between the arms ($N = 92$), the incidence of this bias increased to 59.0%. Compared with studies with a statistically significant difference between arms in the PE, studies with a non-significant difference showed a statistically significant association with not reporting the PE in the concluding statement of the abstract (27% versus 7%, OR = 5.15, 95% CI = 1.86–14.26, $P = 0.001$. Compared with studies where the PE did not change, there was a trend for trials with a change of PE to report a statistically significant difference for (the new) PE (OR = 2.29, 95% CI = 0.37–14.32, $P = 0.47$). There was no association between the JIF and bias (Spearman's rho = −0.10, $P = 0.20$).

There were no apparent differences in the probability of bias in trials conducted in the adjuvant or metastatic settings (Chi-squared $P = 0.146$). There was also no association between bias and the type of PE, (OS versus DFS or PFS, chi-squared $P = 0.23$).

bias in reporting of toxicity

A total of 110 (67.1%) papers met our definition of biased reporting of toxicity. Distribution of bias according to the hierarchy scale is reported in Table 2. There was a statistically significant association between biased reporting of toxicity and observation of a statistically significant difference in the arms for the PE (OR = 2.00, 95% CI = 1.02–3.94, $P = 0.044$). There was no association between biased reporting of toxicity and biased reporting of efficacy (chi-squared $P = 0.43$), or with change of the PE (OR = 0.58, 95% CI = 0.1–3.2, $P = 0.17$). The JIF was not associated with biased reporting of toxicity (Spearman's rho = −0.153 Chi squared $P = 0.73$). Bias in the reporting of toxicity was significantly associated with the use of OS as the PE (OR = 3.30, 95% CI = 1.1–10.1, $P = 0.028$). Reporting of toxicity was not influenced by the setting of the trial (adjuvant versus metastatic, OR = 1.68, 95% CI = 0.9 to 3.3, $P = 0.12$).

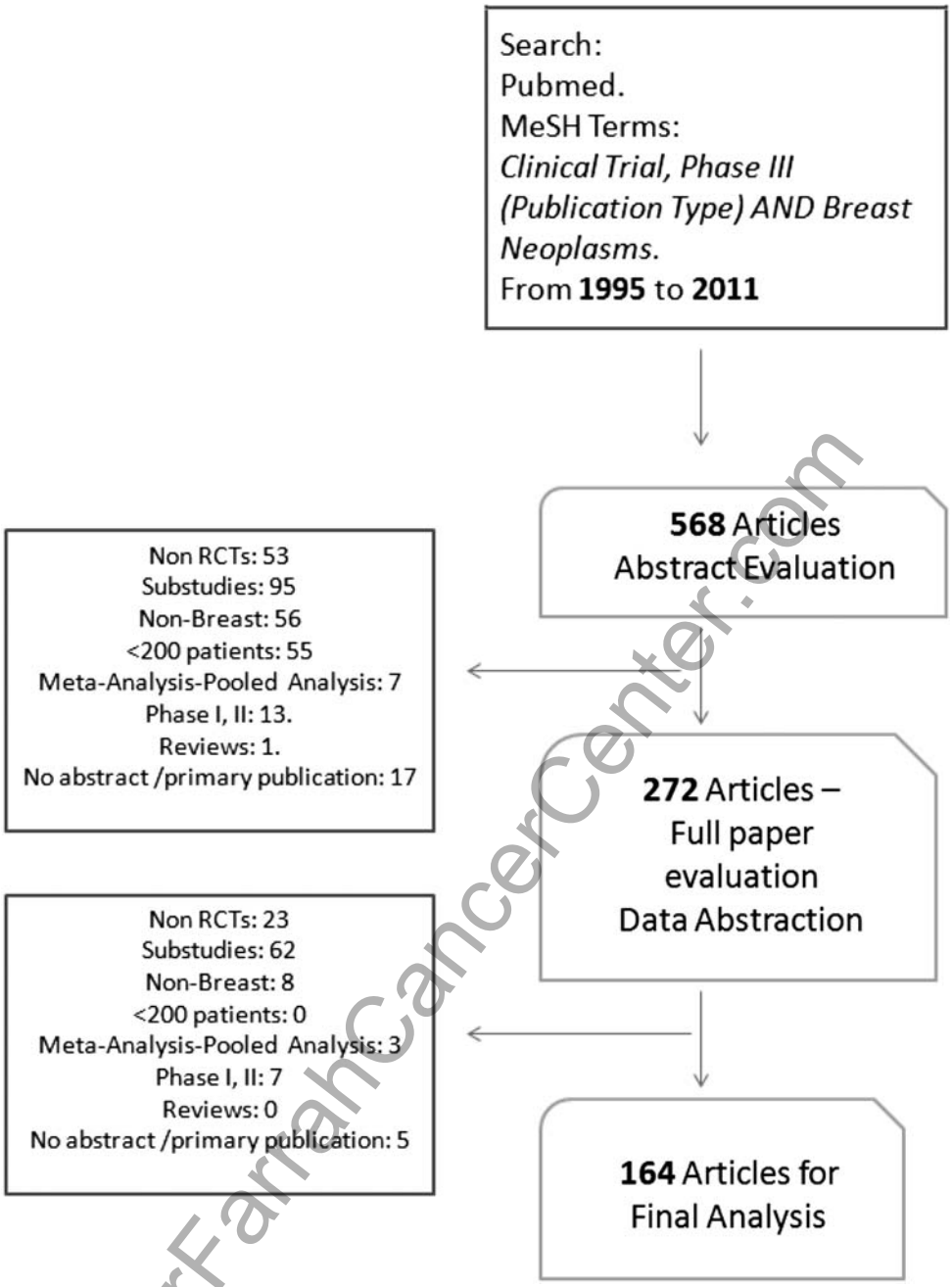

Figure 3. Flow diagram illustrating selection of articles for analysis.

influence of funding on results

Funding from industry partners was reported in 103 (62.8%) studies, 32 (19.5%) studies were funded by academic or governmental grants and in 29 (17.7) studies the source of funding was not stated. Three studies reported this in the abstract [15–17], while all others reported this information in the body of the manuscript. Success in finding a significant difference between the arms for the PE and bias in the reporting of this end point were not influenced by source of funding (chi-squared $P = 0.78$ and $P = 0.71$, respectively). Similarly, industry funding was not associated with biased reporting of toxicity (chi-squared $P = 0.71$). There was no effect of industry funding on odds of change in the PE (OR = 3.20, CI = 0.3–31.4, $P = 0.41$). There was a significant, but weak association between funding from the industry and higher JIF of the published study (Spearman's rho = 0.39 and chi-squared $P = 0.05$).

discussion

Several papers have evaluated the frequency and characteristics of bias in the reporting of efficacy [3, 4, 18–21], but these reports have tended to focus on heterogeneous medical conditions and not on cancer clinical trials. Furthermore, there are limited data in the literature about bias in the reporting of

Table 1. Characteristics of trials

	All reports ($n = 164$)
Difference in primary end point between arms	
Significant	72 (43.90)
Non–significant	92 (56.10)
PE	
DFS or PFS	137 (83.5)
OS	27 (16.5)
Setting	
Adjuvant	81 (49.4)
Metastatic	83 (50.6)
Funding	
Not industry or not stated	61 (37.20)
Industry	103 (62.80)
ClinicalTrials.gov	
PE changed	7 (4.2)
PE not changed	23 (14)
Protocol N/A	134 (81.8)

PE, primary end point; N/A, not available; DFS, disease-free survival; PFS, progression-free survival; OS, overall survival.

Table 2. Distribution of bias in reporting toxicity

Toxicity hierarchy scale	Number of trials $N = 164$	%	Positive PE (%)	Negative PE (%)
1	17	10	7 (4)	10 (6)
2	7	4	2 (1)	5 (3)
3	20	12	4 (2)	16 (10)
4	10	6	5 (3)	5 (3)
5	21	13	11 (7)	10 (6)
6	55	34	31 (19)	24 (15)
7	34	21	12 (7)	22 (13)

Higher numbers and shaded area refer to poor reporting of toxicity.
PE, primary end point.

toxicity [11]. Here, we have explored the frequency of bias in reporting of efficacy and toxicity in randomized trials evaluating treatments of breast cancer. We focus our research on breast cancer, given that it is the most common malignancy in women, has substantial mortality [22] and is a cancer site with a large number of trials.

The objective of a phase III RCT is to detect or exclude differences in end points that will reflect benefit to patients. The chosen end points should be measures of patient benefit (i.e. improved efficacy or better safety and tolerability). OS is the gold standard for the assessment of benefit: it is unambiguous and is not subject to investigator interpretation [23]. PFS or DFS may be suitable end points if these measures are valid surrogates for OS [24], or possibly in trials where there is a high rate of cross-over to the experimental arm which confounds interpretation of OS. For women with breast cancer, neither DFS nor PFS have been shown to be adequate surrogates for OS [25, 26] but 83.5% of our cohort of trials used these end points.

Bias in the reporting of the PE was prevalent, especially when statistical significance of the difference in the PE between the arms was not found. You et al. [27] evaluated reports of RCTs published between 2005 and 2009, and found that there was misinterpretation of the PE in 21.6% of the trials; this included non-significance in a superiority trial interpreted as showing treatment equivalence, study conclusion based on end points other than PE, study considered positive despite a non-significant *P*-value, and study conclusions based only on one end point when there were co-PEs. We found a higher incidence of inappropriate reporting of the PE in RCTs for breast cancer that increased dramatically when only the trials with a non-significant p-value were assessed. Consequently, spin was used frequently to influence, positively, the interpretation of negative trials, by emphasizing the apparent benefit of a secondary end point. We found bias in reporting efficacy and toxicity in 32.9% and 67.1% of trials, respectively, with spin and bias used to suggest efficacy in 59% of the trials that had no significant difference in their PE. These results are similar to those in other areas of medicine [3]. In contrast to those data where bias in the reporting of toxicity was less frequent when the PE was positive, we found that bias in the reporting of toxicity was higher when the trial had a significant *P*-value for the difference in the PE between experimental and control arms. A possible explanation for this finding may be that investigators and/or sponsors then focus on efficacy as the basis of registration and downplay toxicity to make the results more attractive.

To avoid selection for publication of positive trials, and/or publication of a subset of the original recorded outcomes on the basis of the results, registration of trials is now mandatory. Due to our period of evaluation (1995–2011), only 18% of our trials were registered in ClinicalTrials.gov. In some of these trials the PE was changed between the time of registration and reporting of their results. Among these trials, there was a trend towards change of the PE being associated with positive results, suggesting that it may be a strategy to make a negative trial appear positive. This may be on the basis of a low likelihood of observing enough events for this end point to be statistically significant or even a lack of effect of the experimental therapy to modify the original PE (usually OS). Trial registration does not necessarily remove bias in reporting outcome, although it does make it easier to detect [18].

The pharmaceutical industry is increasingly influential in clinical trial sponsorship with data showing an increase in industry sponsorship of phase III RCTs from 24 to 72% over a 30 year period [3, 19, 20]. In our cohort of trials, 67% were industry sponsored, but we found no association between industry sponsorship and biased reporting of either efficacy or toxicity, and no association of for-profit sponsorship with change of the PE between that listed in trial registries and the final publication.

There are some limitations to our study. First, we searched only breast cancer trials and we cannot extrapolate our findings to published reports of trials for other types of cancer. Second, including studies less than 200 patients would likely to increase the level of bias, but the clinical impact of such studies is low. Third, we utilized subjective measures for some of our outcome measures such as the presence of spin. Fourth, our scales used to assess bias in reporting of efficacy and toxicity were based on our interpretation of the characteristics that a paper has to accomplish to be considered unbiased, but they

have not been validated. Fifth, many of our included trials were not available at ClinicalTrials.gov. This database was established in 2002 [28] and many trials initiated before this date were not included. Furthermore, many European trials were not initially included in the US-based ClinicalTrials.gov database and European Clinical Trials Registries do not have easily searchable databases [29]. Our analysis of change in the PE should, therefore, be interpreted with caution.

In conclusion, bias in the reporting of efficacy and toxicity remains prevalent. Clinicians, reviewers, journal editors and regulators should apply a critical eye to trial reports and be wary of the possibility of biased reporting. Guidelines are necessary to improve the reporting of both efficacy and toxicity.

disclosure

The authors have declared no conflicts of interest.

references

1. Ocana A, Tannock IF. When are 'positive' clinical trials in oncology truly positive? J Natl Cancer Inst 2010; 103: 16–20.
2. Dwan K, Altman DG, Arnaiz JA et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS One 2008; 3: e3081.
3. 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.
4. Kirkham JJ, Altman DG, Williamson PR. Bias due to changes in specified outcomes during the systematic review process. PLoS One 2010; 5: e9810.
5. Ioannidis JP. Adverse events in randomized trials: neglected, restricted, distorted, and silenced. Arch Intern Med 2009; 169: 1737–1739.
6. Ioannidis JP, Evans SJ, Gotzsche PC et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 2004; 141: 781–788.
7. Williamson PR, Gamble C, Altman DG et al. Outcome selection bias in meta-analysis. Stat Methods Med Res 2005; 14: 515–524.
8. Krzyzanowska MK, Pintilie M, Tannock IF. Factors associated with failure to publish large randomized trials presented at an oncology meeting. JAMA 2003; 290: 495–501.
9. Ioannidis JP. Why most published research findings are false. PLoS Med 2005; 2: e124.
10. Cuervo LG, Clarke M. Balancing benefits and harms in health care. BMJ 2003; 327: 65–66.
11. Pitrou I, Boutron I, Ahmad N et al. Reporting of safety results in published reports of randomized controlled trials. Arch Intern Med 2009; 169: 1756–1761.
12. Ioannidis JP. Limitations are not properly acknowledged in the scientific literature. J Clin Epidemiol 2007; 60: 324–329.
13. Barry HC, Ebell MH, Shaughnessy AF et al. Family physicians' use of medical abstracts to guide decision making: style or substance? J Am Board Fam Pract 2001; 14: 437–442.
14. Burnand B, Kernan WN, Feinstein AR. Indexes and boundaries for 'quantitative significance' in statistical decisions. J Clin Epidemiol 1990; 43: 1273–1284.
15. Gianni L, Eiermann W, Semiglazov V et al. Neoadjuvant chemotherapy with trastuzumab followed by adjuvant trastuzumab versus neoadjuvant chemotherapy alone, in patients with HER2-positive locally advanced breast cancer (the NOAH trial): a randomised controlled superiority trial with a parallel HER2-negative cohort. Lancet 2010; 375: 377–384.
16. Cortes J, O'Shaughnessy J, Loesch D et al. Eribulin monotherapy versus treatment of physician's choice in patients with metastatic breast cancer (EMBRACE): a phase 3 open-label randomised study. Lancet 2011; 377: 914–923.
17. Martin M, Segui MA, Anton A et al. Adjuvant docetaxel for high-risk, node-negative breast cancer. N Engl J Med 2010; 363: 2200–2210.
18. 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.
19. Boutron I, Dutton S, Ravaud P et al. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA 2010; 303: 2058–2064.
20. 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.
21. Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. PLoS Med 2008; 5: e217; discussion e217.
22. Siegel R, Ward E, Brawley O et al. Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. CA Cancer J Clin 2011; 61: 212–236.
23. Pazdur R. Endpoints for assessing drug activity in clinical trials. Oncologist 2008; 13(Suppl 2): 19–21.
24. Booth CM, Eisenhauer EA. Progression-free survival: meaningful or simply measurable? J Clin Oncol 2012; 30(10): 1030–1033.
25. Ocana A, Tannock IF. When are 'positive' clinical trials in oncology truly positive? J Natl Cancer Inst 2011; 103: 16–20.
26. Amir E, Seruga B, Kwong R, Tannock IF, Ocaña A. Poor correlation between progression-free and overall survival in modern clinical trials: are composite endpoints the answer? Eur J Cancer 2012; 48(3): 385–388.
27. You B, Gan HK, Pond G et al. 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.
28. http://clinicaltrials.gov/ct2/info/about. (16 March 2012, date last accessed).
29. https://www.clinicaltrialsregister.eu/. (16 March 2012, date last accessed).