Title: Publishing interim results of randomised clinical trials in peer-reviewed journals

Running Head: Comparing published interim and final results

Words: 3138

Authors:

Nicholas Counsell[1], Despina Biri[1], Joanna Fraczek[1], Allan Hackshaw[1]

[1]Cancer Research UK & UCL Cancer Trials Centre, University College London, London, UK

Corresponding Author: Nicholas Counsell

Address: CRUK & UCL CTC, 90 Tottenham Court Road (5th floor), London, W1T 4TJ, UK

E-mail: nicholas.counsell@ucl.ac.uk

Telephone: +44(0) 20 7679 9557

ABSTRACT

Background

Interim analyses of randomised controlled trials are sometimes published before the final results are available. In several cases, the treatment effects were noticeably different after patient recruitment and follow-up completed. We therefore conducted a literature review of peer-reviewed journals to compare the reported treatment effects between interim and final publications, and to examine the magnitude of the difference.

Methods

We performed an electronic search of MEDLINE 1990 to 2014 (keywords: 'clinical trial' OR 'clinical study' AND 'random*' AND 'interim' OR 'preliminary'), and we manually identified the corresponding final publication. Where the electronic search produced a final report in which the abstract cited interim results, we found the interim publication. We also manually searched every randomised controlled trial in eight journals, covering a range of impact factors and general medical and specialist publications (1996 to 2014). All paired articles were checked to ensure that the same comparison between interventions was available in both.

Results

Sixty-three studies are included in our review, and the same quantitative comparison was available in 58 of these. The final treatment effects were smaller than the interim ones in 39 (67%) trials and the same size or larger in 19 (33%). There was a marked reduction, defined as a ≥20% decrease in the size of the treatment effect from interim to final analysis, in 11 (19%) trials compared to a marked increase in 3 (5%), p=0.057. The magnitude of percentage change was larger in trials where commercial support was reported, and increased as the proportion of final events at the interim report decreased in trials where commercial support was reported (interaction p=0.023). There was no evidence of a difference between trials that stopped recruitment at the interim analysis where this was reported as being pre-specified versus those that were not pre-specified (interaction p=0.87).

Conclusions

Published interim trial results were more likely to be associated with larger treatment effects than those based on the final report. Publishing interim results should be discouraged, in order to have reliable estimates of treatment effects for clinical decision-making, regulatory authority reviews and health economic analyses. Our work should be expanded to include conference publications and manual searches of additional journal publications.

Keywords / short phrases:

Randomised controlled trials, treatment effects, interim analysis, final analysis, published results, clinical decision-making

INTRODUCTION

Randomised controlled trials usually take several years to conduct. Interim analyses of randomised controlled trials involve early looks at the data, usually by an Independent Data Monitoring Committee or Data and Safety Monitoring Boards, to examine safety and sometimes efficacy.[1,2] The data are examined confidentially by these review committees, with only major recommended changes to the trial communicated to the study investigators. Recommendations could be to continue the trial as planned, continue the trial but with modifications to the protocol, or to stop the trial because of safety concerns, a clear benefit has been found, or futility if the objectives are unlikely to be achieved.

There is often pressure on investigators to present trial findings sooner than later, which may involve oral or poster presentations at conferences, or publication in peer-reviewed journals. This is usually done for trials that do not have blinded interventions. Specifically, we observed examples where the final treatment effect was noticeably smaller than the one reported in the interim analysis.

Importantly, there have been several instances where the interim results were made available in the public domain, based on data while the trial was still recruiting or in follow-up (so has not yet reached the target sample size or number of events). Also, following the European Clinical Trial Regulation[3] which will apply from 2016, trial sponsors are required to submit a summary of intermediate results to the European Union database within one year of any analysis specified in the trial protocol if the trial

results are available (Article 37), for instance, when a trial that has closed early at a pre-specified interim analysis but patients remain in follow-up.

We therefore conducted a preliminary literature review to compare treatment effects that were reported in both interim and final publications, to examine the magnitude of the difference and key factors about the study or publication.

METHODS

We first performed an electronic search of MEDLINE 1990 to 2014 using the following keywords: 'clinical trial' OR 'clinical study' AND 'random*' AND 'interim' OR 'preliminary'. Titles and abstracts were examined, and we manually identified the subsequent final publication using citations, the trial name, the authors, and the regimen description. Where the electronic search produced a final report in which the abstract cited interim results, we found the corresponding interim paper using the reference. However, we were aware that our approach could only identify articles which specifically mentioned 'interim' or 'preliminary' in the abstract, and there are articles that do not. To address this we also manually searched eight journals (considered feasible), selected to cover a range of impact factors, and general medical and specialist publications associated with common disorders: American Heart Journal, British Medical Journal, Journal of Clinical Oncology, Journal of the American College of Cardiology,

Journal of the American Medical Association, Lancet, Neurology, and New England Journal of Medicine, beginning when many journals were available electronically (1996 to 2014). The manual search involved examining every article from a randomised trial in full, to identify interim reports (and the subsequent final report found), or to determine whether an interim analysis had been previously published. It would have been impractical to manually search every single medical journal in this detailed way. Therefore, our study should be a representative sample of randomised controlled trials that have reported both interim and final results, rather than a complete literature review of all trials ever published.

The search results are shown in a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (Figure 1), and although our study is not a standard meta-analysis we have reported our findings in accordance with the PRISMA statement where applicable.[4] All paired articles were reviewed independently (by NC and DB/JF) to ensure that they should be included in our review, and information about the study and the publication were extracted, where available (Table 1): disease area; sponsor; funding support; sample size and number of events; use of blinding; pre-specification, interpretation of finding and timing of interim analysis; effect size for the primary endpoint (e.g. relative risk, hazard ratio, or mean difference); and whether patients were allowed to crossover after interim results were reported.

The difference between published interim and final results was expressed as a percentage change, appropriate for effect sizes based on risk (e.g. relative risk or hazard ratio). This approach tended to give more conservative differences than calculating the difference based on the relative reductions (which

also could exaggerate small changes); for example a relative risk of 0.54 (interim) vs. 0.76 (final) represents a 41% reduction using the estimates themselves ([0.54-0.76]/0.54) but a 48% decrease using the relative effects ([0.46-0.24]/0.46). In some cases the treatment effect was estimated either using the number of events and participants or by combining effects across subgroups. We also investigated the association between the magnitude of percentage change in effect sizes and other study or publication factors using scatterplots, non-parametric correlations and tests, and multiple linear regression. Very few studies used continuous variables for endpoints and so were considered separately. Analyses were performed using SPSS version 22.

RESULTS

Sixty-three trials were identified, in which the same quantitative comparison was reported in 58 trials. More than half of the trials were in oncology (n=38, 60%), and 24 (38%) reported that subjects were blinded. Seventeen (27%) trials had a commercial company sponsor, and 30 (48%) received commercial company support (including academic initiated trials funded by a pharmaceutical company). Nineteen (30%) trials were still recruiting patients when the interim reports were published, with a median time to the final published report of 39 (range 5–228) months. Thirty (48%) of the interim reports were published in a journal with a current impact factor of >20, compared with 22 (35%) of the final reports. Thirty-five (56%) of the interim reports stated that the analysis was pre-specified in the trial protocol,

and 44 (70%) reported a difference existed between trial arms (i.e. a 'positive' finding); 10 (21%) trials reported patient crossover between arms.

Table 2 shows the interim and final treatment effect for each of the 58 trials in which the same quantitative comparison was reported by the investigators, and the magnitude of the difference between the reported effects. The final results were generally of a smaller magnitude than the interim ones, with a smaller effect size observed in 39 (67%) trials and the same or a larger effect size in 19 (33%). There was a marked reduction in 11 (19%) trials, defined as a ≥20% decrease in the size of the treatment effect from interim to final analysis, compared to a marked increase of ≥20% in 3 (5%) trials (p=0.057).

Figure 2a is a scatterplot of the effect size at interim and final results for the 54 studies which used risk as an endpoint, and Figure 2b is a Bland-Altman style plot. Thirty-seven data points are beyond the reference line indicating a reduction in effect size, 17 showed the same or an increase in effect size, the dashed lines indicate a change of ≥20%. The percentage change in effect size from interim to final results ranged from -120% to 26%, but there is a clear mean reduction of -10.3% (95% CI -16.9 to -3.7, p=0.002). If the 10 trials in which treatment crossover was allowed were excluded, the change in effect size reduced but remained statistically significant (-7.7%, 95% CI -15.0 to -0.3, p=0.041). Generally, the magnitude of the percentage change decreased as the effect size decreased, i.e. towards the no effect value of 1.

Scatterplots suggest that the magnitude of change in effect size from interim to final results increase as the proportion of final events at the interim report decreased (correlation=-0.330, p=0.023), and in trials where there was commercial company support, with a median of 14.5% versus 7.0% among those without such support (p=0.010); Figure 3a&b. Multiple linear regression showed that the (log-transformed) magnitude of percentage change was dependent on these two factors and that there was also a significant interaction between them (p=0.023). In trials with commercial support the (log-transformed) magnitude of percentage change increased more as the proportion of final events at the interim report decreased, i.e. larger changes in effect size were found in interim analyses reported on a smaller proportion of events (Figure 3c&d).

There was no material association with type of disease (p=0.72), study sponsor (p=0.83), use of blinding (p=0.77), pre-specifying the interim analysis (p=0.14), interpretation of finding (p=0.16), allowing crossover (p=0.41), proportion of final sample size at the interim report (p>0.99), year of interim publication (p=0.28), year of final publication (p=0.47), time between interim and final report (p=0.20), impact factor of the interim publication journal (p=0.26), or impact factor of the final publication journal (p=0.31); Figure 4. There was no evidence of a difference between trials that stopped recruitment at the interim analysis where the interim was reported as being pre-specified versus those that were not pre-specified (interaction p=0.87), although only 16 (28%) trials were still recruiting (Supplementary Figure).

Two examples of trials with noticeable differences between interim and final publications were: (i) examining Trastuzumab after adjuvant chemotherapy in patients with HER2-positive early-stage breast

cancer (HERA trial),[sup11,sup12] the interim hazard ratio for disease-free survival was 0.54 (95% CI 0.43 to 0.67), but the subsequent reported hazard ratio was 0.76 (95% CI 0.66 to 0.87), a 41% decrease in the estimate of treatment effect which was outside of the earlier confidence interval; and (ii) examining Sorafenib in patients with advanced clear-cell renal-cell carcinoma, [sup19,sup20] the interim hazard ratio for survival was 0.72 (95% CI 0.54 to 0.94), but the subsequent hazard ratio was 0.88 (95% CI 0.74 to 1.04), a 22% decrease in the estimate of treatment effect to a statistically non-significant result. In both cases, patients were allowed to change to the more effective treatment following the interim analysis (sensitivity analyses censoring these patients gave larger effect sizes but still of smaller magnitude than interim results).

There were also instances in which the authors' main conclusions differed between the interim and final reports. For example, the FinXX study investigators concluded that "the capecitabine-containing chemotherapy regimen reduced breast cancer recurrence compared with a control schedule of standard agent" at the planned interim analysis; however, at the time of the final analysis their conclusion was that "integration of capecitabine into a regimen that contains docetaxel, epirubicin, and cyclophosphamide did not improve recurrence-free survival significantly compared with a similar regimen without capecitabine".[sup25,sup26] The final publication was accompanied by an editorial that discussed when adjuvant breast cancer trials should be reported, and it stated that the FinXX trial was most likely reported prematurely in the earlier article[5], which highlights the importance of carefully considering the timing of any interim analyses and the associated data maturity in a pre-specified monitoring plan.

DISCUSSION

There have already been discussions in the literature about the potential problems of stopping randomised controlled trials early for benefit based on published interim results, including the issue of biased efficacy estimates,[6-8] although it has been argued that this bias may be small when considering all trials that use stopping guidelines rather than only trials that are stopped early.[9-11] We are the first to quantify the difference between interim and final results, using a relatively large representative number of randomised controlled trials identified in a systematic way, and have examined this in relation to key factors.

Among the 58 trial pairs for which the same quantitative comparison was available, we found a clear negative association overall, where the size of the treatment effect was reduced in 67% of the trials between the interim and final report, and 19% showed a marked reduction. We found that commercial company funding, and the proportion of final events at the interim report, were associated with noticeable changes in the effect size; in trials with commercial support, larger changes in effect size were found in interim analyses reported on a smaller proportion of events. These findings, from readily searchable and peer-reviewed journals, highlight the issue of disseminating interim results and can be considered hypothesis-generating with regards to other publication sources. There was no evidence of a difference when considering trials that were stopped early following a pre-planned interim analysis compared to those that were not pre-planned. Although it is useful to consider whether interim

analyses were pre-specified (which might then be considered as the 'final' analysis), our point is that there is a clear trend for the treatment effect found at this stage to be, on average, larger than one based on the original (full) target follow-up and number of events.

We acknowledge that there will be more than 63 randomised controlled trials where both the interim and final analyses have been made publicly available, but many of the interim results are expected to have been presented at conferences. In our study, we chose to examine only interim reports that were published in journals, to match the same proper peer-review process as the final report. A similar association is expected using conference publications, though given that trials showing beneficial effects are more likely to be selected for oral presentations, the magnitude of the difference between interim and final results could be greater than we found. Future work incorporating other publication sources is required to investigate this hypothesis. Another limitation of our review is that the electronic search strategy relied on 'interim' or 'preliminary' being stated in the abstract, but our detailed manual search verified that we were not systematically missing reports through misspecification of electronic search terms. Twelve trials were found through the manual search of the 8 journals, because either the earlier report was not referred to as an interim analysis, or it was only acknowledged in the main text rather than the abstract. However, the manual search was based on only 8 journals (because our project was unfunded), so our work could be expanded to incorporate additional manual searches. We did not examine methodological quality, because our aim was to look at within-trial comparisons of treatment effects, rather than to combine the effects as in a standard meta-analysis. Also, crossover to alternative treatments or other trial treatment could have diluted the effect sizes at the final analysis, creating more uncertainty over the true treatment effect, but this is a known issue[12]. However, a clear negative association remained after excluding such trials from our analysis. Indeed, other trial factors such as

patient monitoring and further therapy may also change after the results are published, such factors might directly influence any later analyses carried out, but the impact could not be assessed from available information in the articles. Despite these limitations, the articles we examined should be a representative subset of all interim/final paired reports, to provide a sufficiently reliable first analysis.

Finding an effect size that is smaller at the final analysis than at the interim stage is analogous to examining repeated measurements: relatively high or low observations are likely to be followed by a less extreme observation due to natural variability (i.e. regression towards the mean).[13] Also, as more patients (or events) are accrued, results should be more reliable. This was clearly illustrated in a study of candesartan for cardiovascular death, where successive interim analyses yielded increasingly smaller treatment effects (the analyses were examined by the Independent Data Monitoring Committee, and not published); the hazard ratio was 0.63 in March 2000 but 0.91 in March 2003.[14] Also, time-to-event outcome variables can be heavily biased if the proportional-hazards assumption is violated, which can occur if the censoring pattern has changed,[15] such as between the interim and final analysis because the data is not yet mature.

Interim analyses are important for interventional trials, by providing information for an Independent Data Monitoring Committee or Data and Safety Monitoring Board to confidentially review and to help inform whether it is appropriate for the trial to continue as planned (and our article is not about the pros and cons of such analyses).[1,2,16] However, there is often pressure on investigators to publish or present their research sooner rather than later, particularly when there appears to be a clear treatment

effect. In the European Clinical Trials Regulation, a measure intended to improve transparency is that trial sponsors should submit a summary of intermediate results to the EU database within one year of the analysis, if the analysis had been specified in the protocol and the trial results are available (Article 37).[3] Lilford *et al* (2001)[17] argued that all interim data should be publicly available as it accrues, to allow patients to make decisions, and that withholding this information is ethically dubious. Miller and Wender (2008)[18] counter that confidential data monitoring is acceptable, providing participants are no worse off than in the usual clinical setting; there are both individual and collective ethical considerations.[19] However, failure to obtain a reliable result could itself be considered unethical, in order to avoid future patients being exposed to interventions with uncertain effects.

The issue of reporting early results for primary or key secondary endpoints remains today, after we found several such trials published in high impact factor journals (e.g. Choueiri *et al*, Lonial *et al*, Ribas *et al*, Turner *et al*)[20-23], likely to have final analyses published in the future. Although the authors of many of the interim reports included in our review noted some caution in their discussion, once the results are published they are in the public domain, difficult to ignore, and open to interpretation. This should only be considered in the context of a pre-specified monitoring plan incorporated in the trial protocol; if the trial closes based on interim results, these could become the primary study results and later updated results could be influenced by post-closure factors influencing these results.[9-11]

Furthermore, when examining the health economic aspects of an intervention, an overestimate of the treatment effect would make the intervention appear more cost-effective than it really is. In our review,

30% of trials were still recruiting when the interim report was published. Geller and Pocock (1987)[24] stated "Once interim results are reported, no matter how carefully qualified, all trial participants are subject to external pressure.... if interim results are 'significant', there will be pressure to unblind the results". Also, it is questionable whether patients should still be randomised after the interim results are available and accepted.[6-9]

We conclude that while examining interim results is an important part of trial monitoring, *publishing* them need to be considered with caution, or better still avoided, particularly while patients are still being recruited. It is essential to have reliable estimates of treatment effects for both patients and clinicians to make proper informed choices, for regulatory authority reviews and reliable health economic analyses. We suggest further work be carried out to include conference abstracts and a larger number of journals to be searched manually.

Funding: This research was funded through the trials centre Cancer Research UK core programme grant (C444/A15953).

Declaration of conflicting interests: The Authors declare that there is no conflict of interest.

REFERENCES

1. Facey KM & Lewis JA. The management of interim analyses in drug development. *Stat Med* 1998; 17: 1801–9.

2. Guidance for Clinical Trial Sponsors. Establishment and Operation of Clinical Trial Data Monitoring Committees, www.fda.gov, 2006. Accessed 2013.

3. Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC Text with EEA relevance, Article 37, http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2014:158:FULL&from=EN, 2014. Accessed 2015.

4. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med* 2009; 6: e1000097.

5. Earl HM. Reporting of Adjuvant Breast Cancer Trials: When Is the Right Time? *J Clin Oncol* 2012; 30: 1-2.

6. Montori VM *et al.* Randomized trials stopped early for benefit: a systematic review. *JAMA* 2005; 294: 2203-9.

7. Bassler D *et al*. Early stopping of randomized clinical trials for overt efficacy is problematic. *J Clin Epidemiol* 2008; 61: 241-6.

8. Bassler D *et al*. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA* 2010; 303: 1180-7.

9. Goodman SN. Stopping at nothing? Some dilemmas of data monitoring in clinical trials. *Ann Intern Med* 2007; 146: 882-7.

10. Freidlin B & Korn EL. Stopping clinical trials early for benefit: impact on estimation. *Clin Trials* 2009; 6: 119-25.

11. Korn EL *et al*. Stopping or reporting early for positive results in randomized clinical trials: the National Cancer Institute Cooperative Group experience from 1990 to 2005. *J Clin Oncol* 2009;27: 1712-2.

12. Law MG. Problems with publishing results of interim analyses of randomized clinical trials. *HIV Clin Trials* 2000; 1: 30-6.

13. Bland JM & Altman DG. Statistic Notes: Regression towards the mean. *BMJ* 1994; 308: 1499.

14. Pocock S et al. The data monitoring experience in the Candesartan in Heart Failure Assessment of Reduction in Mortality and morbidity (CHARM) program. *Am Heart J* 2005; 149: 939-43.

15. Van Houwelingen HC *et al*. Interim analysis on survival data: Its potential bias and how to repair it. *Stat Med* 2005; 24: 2823-35.

16. Grant AM *et al*. Issues in data monitoring and interim analysis of trials. *Health Technol Assess* 2005; 9: 1-238, iii-iv.

17. Lilford RJ *et al*. Monitoring clinical trials-interim data should be publicly available. *BMJ* 2001; 323: 441-2.

18. Miller FG & Wendler D. Is it ethical to keep interim findings of randomized controlled trials confidential? *J Med Ethics* 2008; 34: 198-201.

19. Pocock SJ. When to stop a clinical trial. *BMJ* 1992; 305: 235-40.

20. Choueiri TK *et al*. Cabozantinib versus Everolimus in Advanced Renal-Cell Carcinoma. *N Engl J Med* 2015; 373: 1814-23.

21. Lonial S *et al*. Elotuzumab Therapy for Relapsed or Refractory Multiple Myeloma. *N Engl J Med* 2015; 373: 621-31.

22. Ribas A et al. Pembrolizumab versus investigator-choice chemotherapy for ipilimumab-refractory melanoma (KEYNOTE-002): a randomised, controlled, phase 2 trial. *Lancet Oncol* 2015. 16: 908-18.

23. Turner NC *et al*. Palbociclib in Hormone-Receptor–Positive Advanced Breast Cancer. *N Engl J Med* 2015; 373: 209-19.

24. Geller NF & Pocock SJ. Interim analyses in randomized clinical trials: Ramifications and guidelines for practitioners. *Biometrics* 1987; 43: 213-23.

TABLE 1.Selected characteristics of the 63 studies

Characteristic	No. of studies (%)
Disease area	
Oncology	38 (60)
Neurology	9 (14)
Cardiology	6 (10)
Osteoporosis	3 (5)
Gastroenterology	2 (3)
Gynaecology	2 (3)
Infectious diseases	2 (3)
Respiratory	1 (2)
Type of sponsor	
Commercial	17 (27)
Non-commercial	46 (73)
Commercial funding support received	
No	33 (52)
Yes	30 (48)
Subjects blinded	
No	39 (62)
Yes	24 (38)
Interim analysis reported as pre-specified	
No	28 (44)
Yes	35 (56)
Interpretation of finding at interim analysis	
Difference between trial arms	44 (70)
No difference	14 (22)
Ambiguous	5 (8)
Patients crossover between arms	
No	37 (79)
Yes	10 (21)
NA/missing	*16*
Sample size at interim analysis, as a proportion of the final sample size	
<25%	2 (3)
25-49	5 (8)
50-74	8 (13)
75-99	4 (6)
100%	44 (70)
Number of events at interim analysis, as a proportion of the final number of events	
<25%	2 (4)
25-49	10 (21)
50-74	24 (51)
≥75%	11 (23)
NA/missing	*16*
Year of publication of the interim analysis	
<1995	17 (27)
1995-2004	26 (41)
≥2005	20 (32)
Year of publication of the final analysis	
<1995	7 (11)
1995-2004	24 (38)
≥2005	32 (51)
Length of time between interim and final publication	
<2.5 years	21 (33)
2.5-4.9	26 (41)
≥5 years	16 (25)
Impact factor of interim report*	
Discontinued	1 (2)
<5	17 (27)
5-9.9	6 (10)

10-14.9	3 (5)
15-19.9	6 (10)
≥20	30 (48)
Impact factor of final report*	
<5	17 (27)
5-9.9	8 (13)
10-14.9	5 (8)
15-19.9	11 (17)
≥20	22 (35)

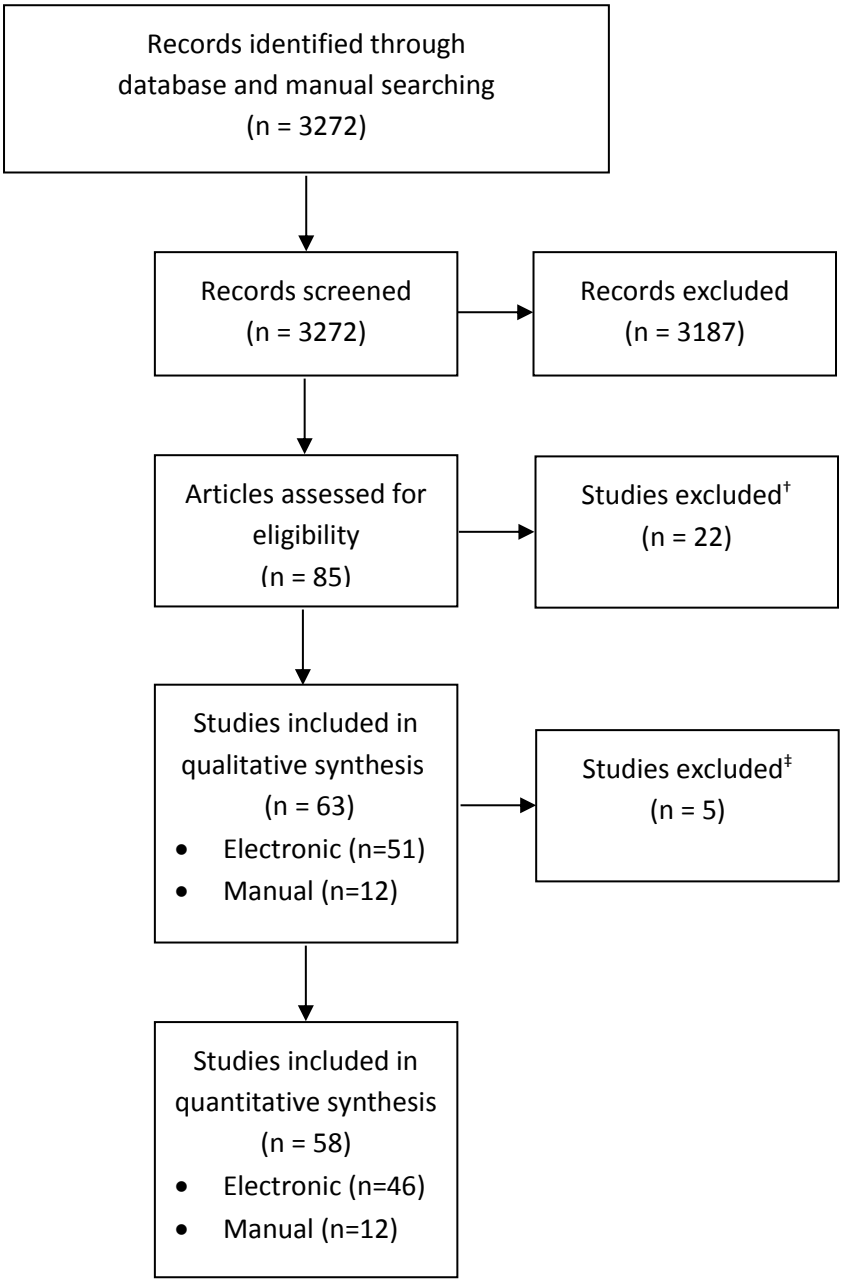
*ISI Web of Science, 2012

TABLE 2.Interim and final results of 58 randomised controlled trials included in the review

Interim report, year	Interim treatment effect	Final report, year	Final treatment effects	Percentage change
STEWART, 1989[sup1]	relative risk = 0.10	FORREST, 1996[sup2]	relative risk = 0.22	-120%
TONER, 2001[sup3]	hazard ratio = 0.22	GRIMISON, 2010[sup4]	hazard ratio = 0.38	-73%
RITA-2 TP., 1997[sup5]	relative risk = 0.52	HENDERSON, 2003[sup6]	relative risk = 0.85	-63%
RASCOL, 1998[sup7]	change from baseline = 12.0	RASCOL, 2000[sup8]	change from baseline = 4.48	-63%
LEES, 1995[sup9]	relative risk = 0.63	KATZENSCHLAGER,	relative risk = 0.93	-48%
PICCART-GEBHART,	hazard ratio = 0.54	GIANNI, 2011[sup12]	hazard ratio = 0.76	-41%
HAROUSSEAU, 1992[sup13]	hazard ratio = 0.63	HAROUSSEAU, 1997[sup14]	hazard ratio = 0.86	-36%
KAYE, 1992[sup15]	relative risk = 0.53	KAYE, 1996[sup16]	relative risk = 0.68	-28%
LAM, 2011[sup17]	relative risk = 0.21	LAM, 2012[sup18]	relative risk = 0.26	-24%
ESCUDIER, 2007[sup19]	hazard ratio = 0.72	ESCUDIER, 2009[sup20]	hazard ratio = 0.88	-22%
RIBEIRO, 1990[sup21]	hazard ratio = 0.44	RIBEIRO, 1993[sup22]	hazard ratio = 0.53	-20%
BARNETT, 2012[sup23]	relative risk = 0.66	MUKESH, 2013[sup24]	relative risk = 0.79	-20%
JOENSUU, 2009[sup25]	hazard ratio = 0.66	JOENSUU, 2012[sup26]	hazard ratio = 0.79	-20%
GOSS, 2003[sup27]	hazard ratio = 0.57	INGLE, 2008[sup28]	hazard ratio = 0.68	-19%
GEYER, 2006[sup29]	hazard ratio = 0.49	CAMERON, 2008[sup30]	hazard ratio = 0.57	-16%
TPSG., 1989[sup31]	hazard ratio = 0.43	TPSG., 1993[sup32]	hazard ratio = 0.50	-16%
ESCUDIER, 2007[sup33]	hazard ratio = 0.79	ESCUDIER, 2011[sup34]	hazard ratio = 0.91	-15%
DE BONO, 2011[sup35]	hazard ratio = 0.65	FIZAZI, 2012[sup36]	hazard ratio = 0.74	-14%
JOENSUU, 2006[sup37]	hazard ratio = 0.58	JOENSUU, 2009[sup38]	hazard ratio = 0.66	-14%
LARSEN, 1997[sup39]	hazard ratio = 0.58	LARSEN, 1999[sup40]	hazard ratio = 0.65	-12%
KURTH, 1984[sup41]	relative risk = 0.62	KURTH, 1997[sup42]	relative risk = 0.70	-12%
COOMBES, 2004[sup43]	hazard ratio = 0.68	COOMBES, 2007[sup44]	hazard ratio = 0.76	-12%
FELSENBERG, 1998[sup45]	change from baseline = 5.55	POLS, 1999[sup46]	change from baseline = 4.90	-12%
KOERTKE, 2003[sup47]	relative risk = 0.78	KOERTKE, 2007[sup48]	relative risk = 0.85	-9%
ROMOND, 2005[sup49]	hazard ratio = 0.48	PEREZ, 2011[sup50]	hazard ratio = 0.52	-8%
ESGOIBISPMS., 1998[sup51]	relative risk = 0.78	KAPPOS, 2001[sup52]	relative risk = 0.84	-8%
RITA TP., 1993[sup53]	relative risk = 0.88	HENDERSON, 1998[sup54]	relative risk = 0.94	-7%
BOCCARDO, 1990[sup55]	hazard ratio = 0.71	BOCCARDO, 1993[sup56]	hazard ratio = 0.75	-6%
BASELGA, 2012[sup57]	hazard ratio = 0.43	YARDLEY, 2013[sup58]	hazard ratio = 0.45	-5%
SAUNDERS, 1996[sup59]	hazard ratio = 0.75	SAUNDERS, 1999[sup60]	hazard ratio = 0.78	-4%
BANG, 2012[sup61]	hazard ratio = 0.56	NOH, 2014[sup62]	hazard ratio = 0.58	-4%
ELLIS, 1971[sup63]	relative risk = 0.92	WIERNIK, 1990[sup64]	relative risk = 0.95	-3%
BASELGA, 2012[sup65]	hazard ratio = 0.64	SWAIN, 2013[sup66]	hazard ratio = 0.66	-3%
DEGIULI, 2004[sup67]	relative risk = 0.65	DEGIULI, 2010[sup68]	relative risk = 0.67	-3%
MALLER, 1991[sup69]	relative risk = 0.96	MALLER, 1993[sup70]	relative risk = 0.99	-3%
DEARNALEY, 2007[sup71]	hazard ratio = 0.67	DEARNALEY, 2014[sup72]	hazard ratio = 0.69	-3%
PILEPICH, 1997[sup73]	relative risk = 0.57	LAWTON, 2001[sup74]	relative risk = 0.58	-2%
LUDWIG, 1991[sup75]	relative risk = 0.90	LUDWIG, 1995[sup76]	relative risk = 0.91	-1%
KARIM, 2002[sup77]	hazard ratio = 0.90	VAN DEN BENT, 2005[sup78]	hazard ratio = 0.90	-1%

ANDRADA HAMER,	relative risk = 0.61	ANDRAD HAMER, 2013[sup80]	relative risk = 0.61	0%
KOSMIDIS, 2000[sup81]	hazard ratio = 0.98	KOSMIDIS, 2002[sup82]	hazard ratio = 0.97	1%
SCHRODER, 2009[sup83]	hazard ratio = 0.80	SCHRODER, 2012[sup84]	hazard ratio = 0.79	1%
MOLYNEUX, 2002[sup85]	relative risk = 0.77	MOLYNEUX, 2005[sup86]	relative risk = 0.76	1%
VERONESI, 1998[sup87]	relative risk = 0.87	VERONESI, 2007[sup88]	relative risk = 0.84	3%
NAKANISHI, 1988[sup89]	relative risk = 0.99	NAKANISHI, 1992[sup90]	relative risk = 0.94	5%
HAIOUN, 1994[sup91]	relative risk = 0.90	HAIOUN, 1997[sup92]	relative risk = 0.84	7%
VANDENBROUCKE,	hazard ratio = 0.90	BOEL, 1999[sup94]	hazard ratio = 0.84	7%
HERSKOVIC, 1992[sup95]	hazard ratio = 0.71	AL-SARRAF, 1997[sup96]	hazard ratio = 0.66	7%
NGUYEN-KHAC, 2008[sup97]	relative risk = 0.77	NGUYEN-KHAC, 2011[sup98]	relative risk = 0.71	8%
MORTON, 2006[sup99]	hazard ratio = 0.92	MORTON, 2014[sup100]	hazard ratio = 0.84	9%
DIEHL, 1998[sup101]	hazard ratio = 0.61	DIEHL, 2003[sup102]	hazard ratio = 0.55	9%
POWLES, 1998[sup103]	relative risk = 0.94	POWLES, 2007[sup104]	relative risk = 0.85	10%
KORCZYN, 1998[sup105]	change from baseline = 8.00	KORCZYN, 1999[sup106]	change from baseline = 9.00	13%
NETTRG., 2003[sup107]	relative risk = 0.99	NAUNHEIM, 2006[sup108]	relative risk = 0.85	14%
ENSRUD, 2004[sup109]	change from baseline = 2.00	BLACK, 2006[sup110]	change from baseline = 2.36	18%
WALLACK, 1995[sup111]	hazard ratio = 0.97	WALLACK, 1998[sup112]	hazard ratio = 0.77	21%
PAAVONEN, 2007[sup113]	(1 – vaccine efficacy) = 0.10	PAAVONEN, 2009[sup114]	(1 – vaccine efficacy) = 0.07	26%
LUNDELL, 2008[sup115]	hazard ratio = 0.69	GALMICHE, 2011[sup116]	hazard ratio = 0.51	26%

FIGURE 1.PRISMA flow diagram

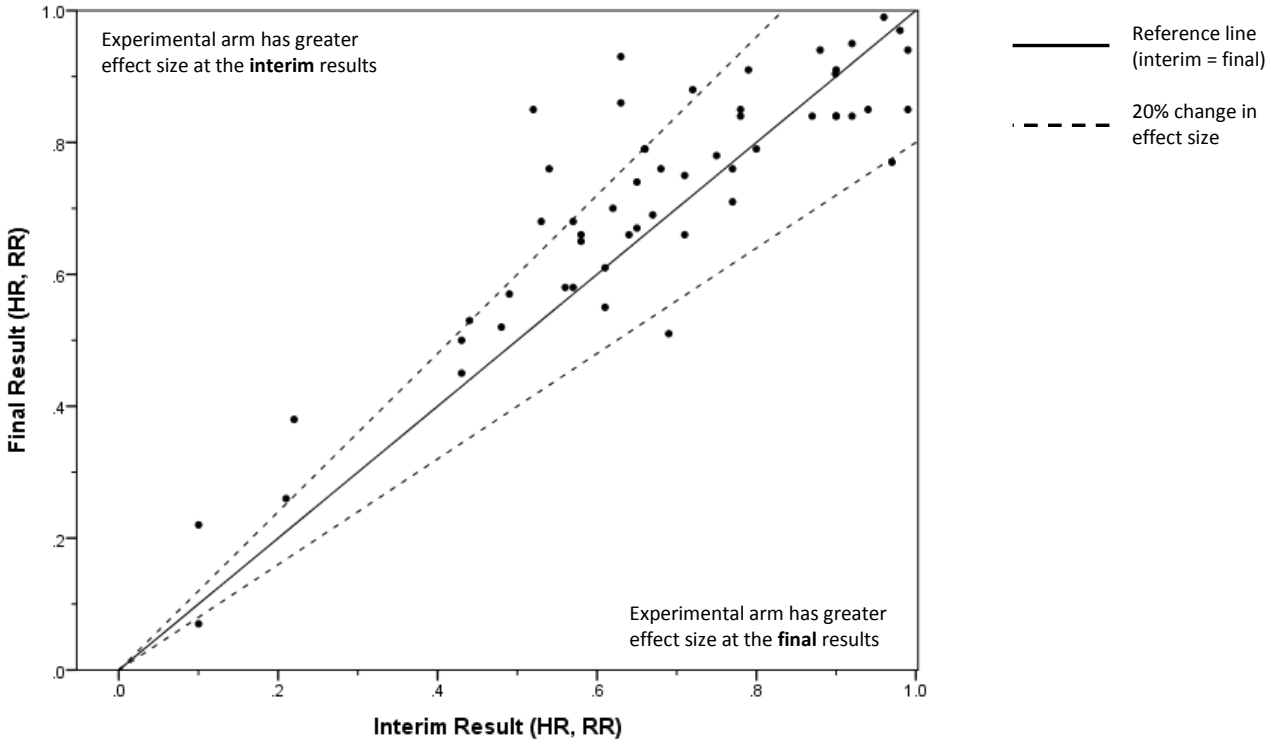


†A randomised comparison of the same efficacy endpoint was not reported in both articles; either different data was available in the two reports, or data was not provided by the randomised groups. Non-inferiority and equivalence designs were also excluded.

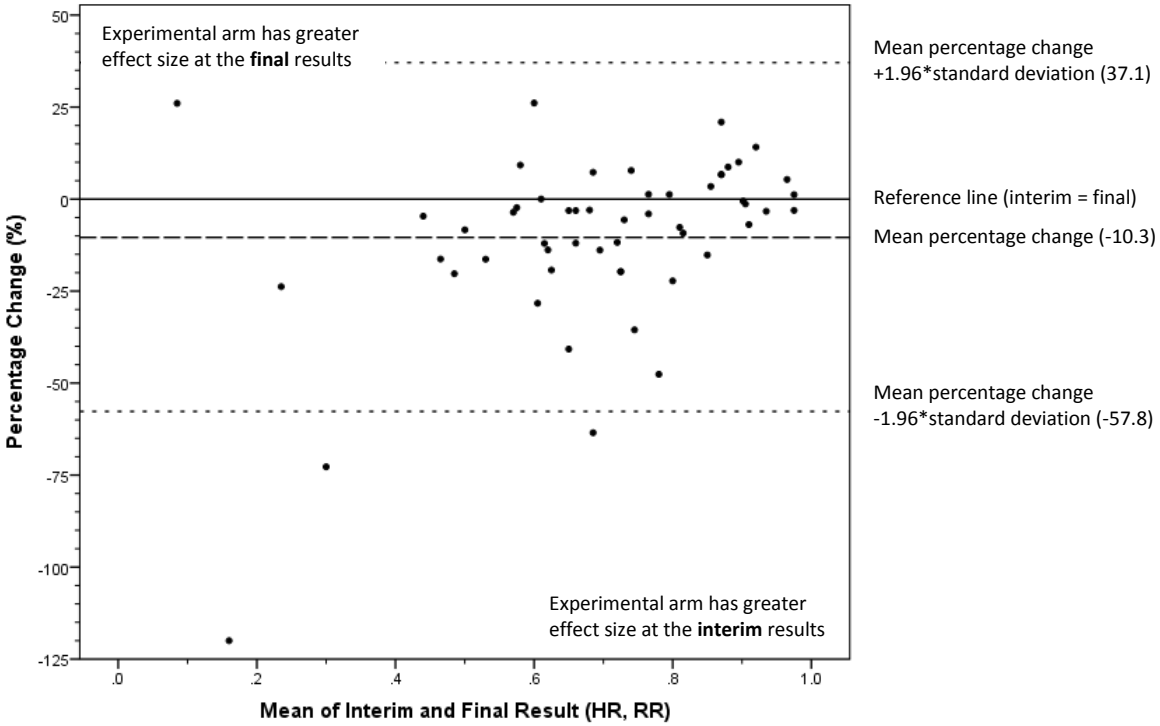
‡The interim and final results were not directly comparable: three reported multiple outcomes at different time points,[sup117-sup122] one reported using different methods and subgroups,[sup123,sup124] and one had no control arm.[sup125,sup126] The interim and final results were consistent in four of these trials.[sup117-sup124]

FIGURE 2.Comparison of interim and final results for trial endpoints associated with risk* (hazard ratio, relative risk). Figure 2a is a scatterplot of the effect size at interim and final results, and 2b is a Bland-Altman style plot of the mean effect size against percentage change from interim to final results.

2a



2b



*Four trials are not shown because their results are based on continuous endpoints[sup7,sup8,sup45,sup46,sup105,sup106,sup109,sup110] (of which 1 had a ≥20% reduction in effect size)

FIGURE 3.Scatterplot of the magnitude of change in effect size from interim to final results (%) for: 3a proportion of final events at the interim report, and 3b funding source. Figures 3c and 3d show the data in Figure 3a according to funding source (i.e. commercial and non-commercial)

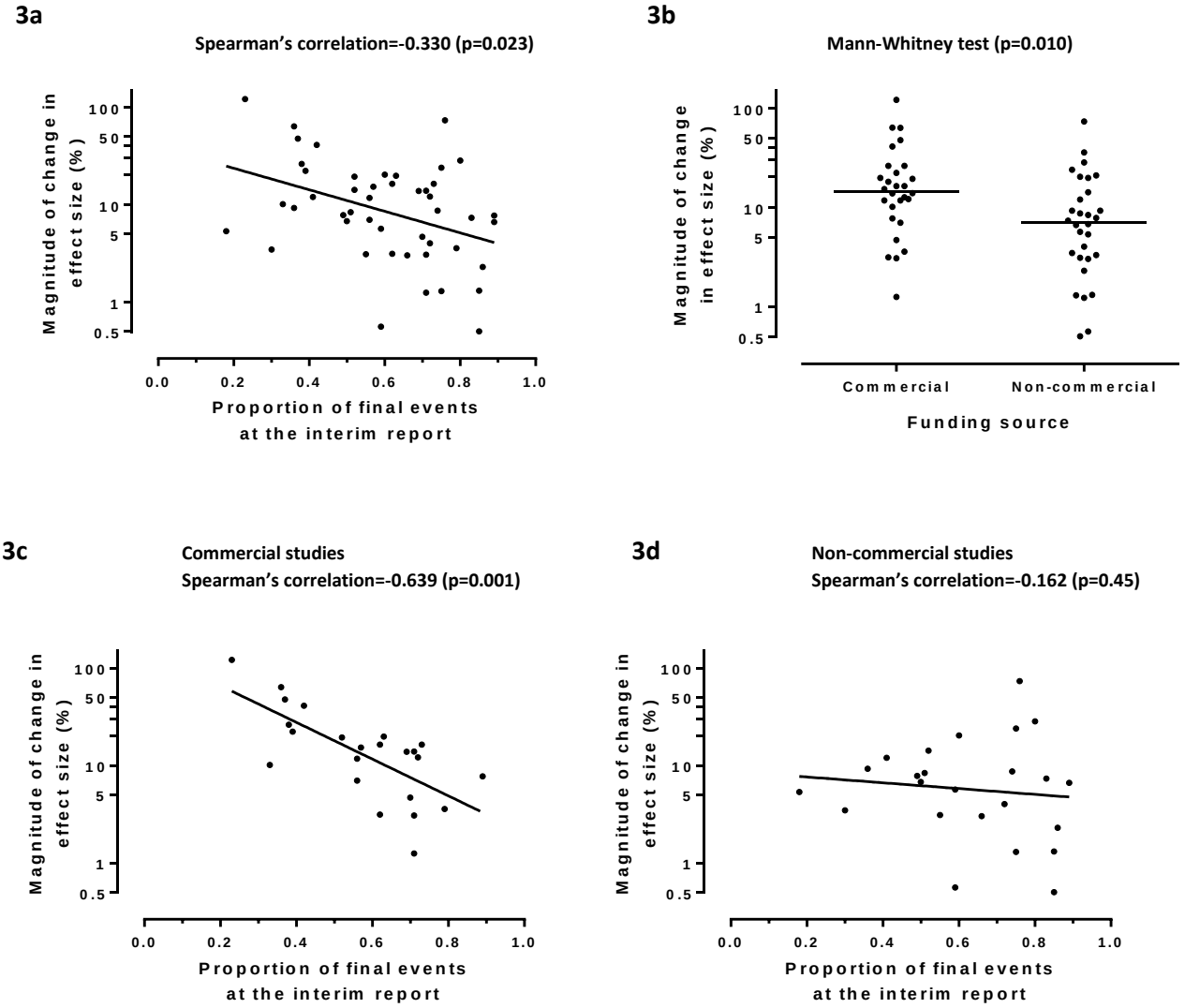
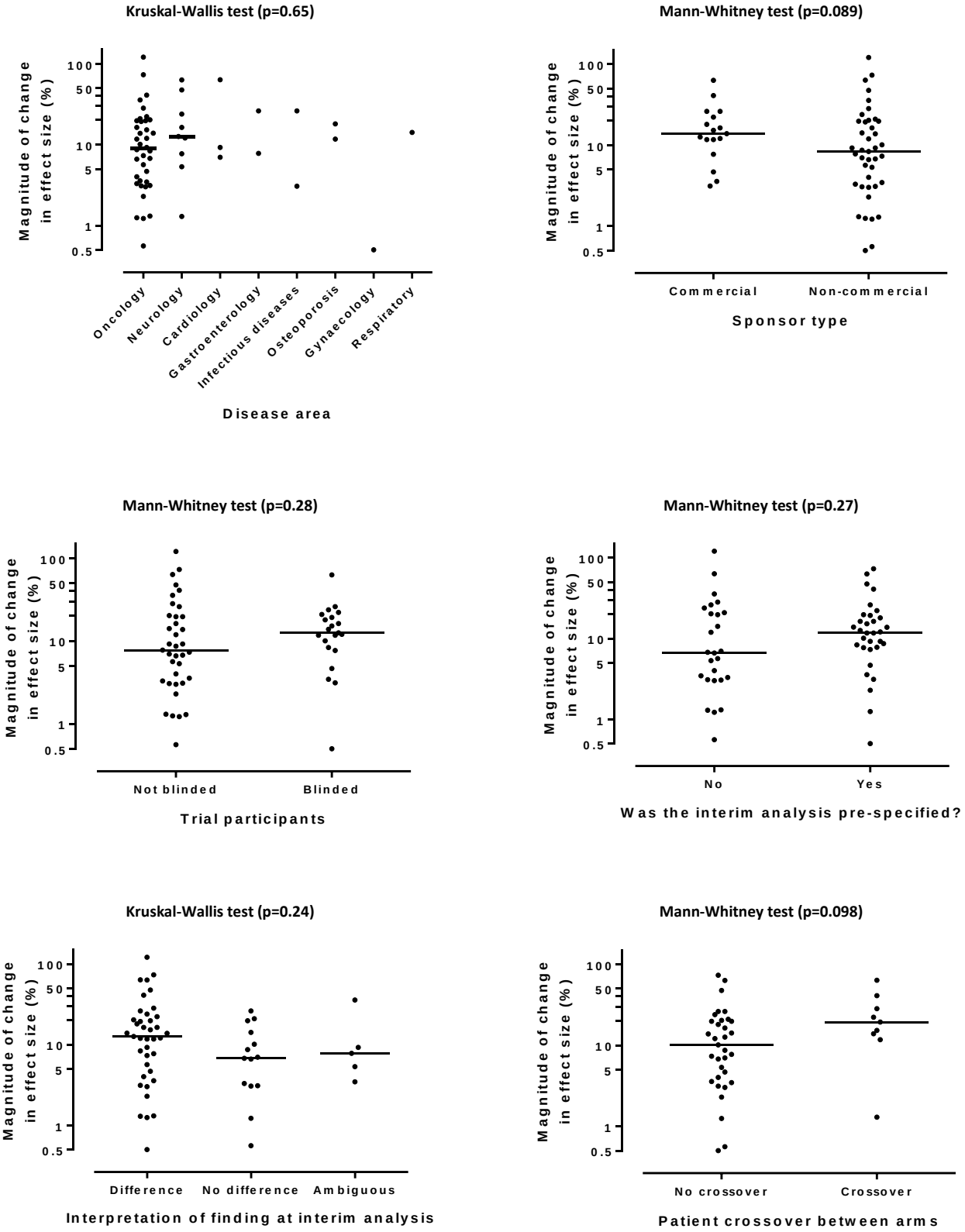
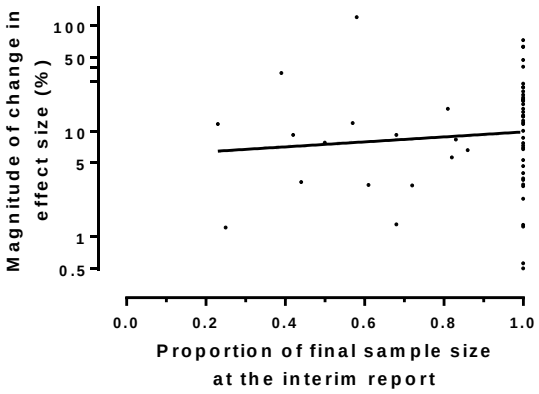


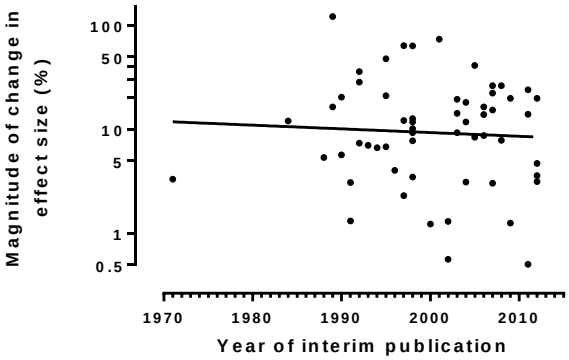
FIGURE 4.Scatterplot of the magnitude of change in effect size from interim to final results (%) according to several factors associated with the trial or publication



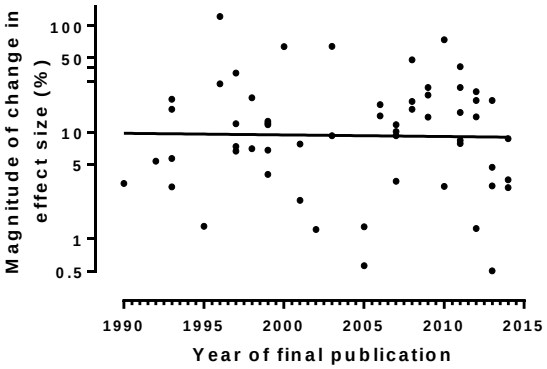
Spearman's correlation=0.167 (p=0.21)



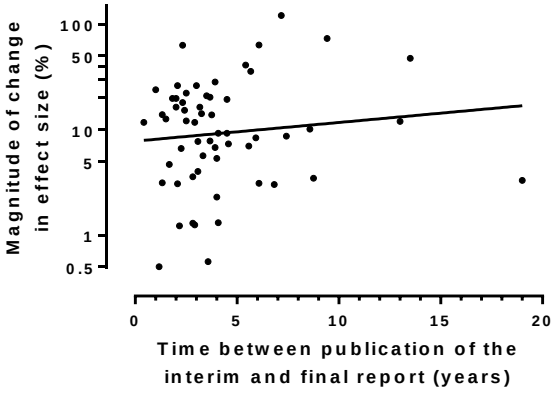
Spearman's correlation=-0.021 (p=0.87)



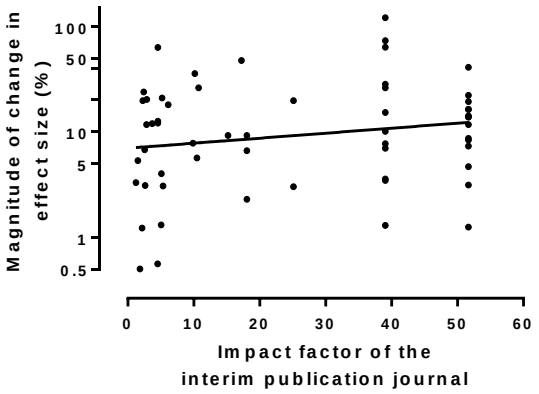
Spearman's correlation=0.001 (p=0.99)



Spearman's correlation=0.062 (p=0.64)



Spearman's correlation=0.181 (p=0.17)



Spearman's correlation=0.109 (p=0.41)