

BMJ Open Do non-inferiority trials of reduced intensity therapies show reduced effects? A descriptive analysis

Scott K Aberegg,¹ Andrew M Hersh,^{1,2} Matthew H Samore^{1,3}

To cite: Aberegg SK, Hersh AM, Samore MH. Do non-inferiority trials of reduced intensity therapies show reduced effects? A descriptive analysis. *BMJ Open* 2018;**8**:e019494. doi:10.1136/bmjopen-2017-019494

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-019494>).

Received 6 September 2017
Revised 30 December 2017
Accepted 6 February 2018

ABSTRACT

Objectives To identify non-inferiority trials within a cohort where the experimental therapy is the same as the active control comparator but at a reduced intensity and determine if these non-inferiority trials of reduced intensity therapies have less favourable results than other non-inferiority trials in the cohort. Such a finding would provide suggestive evidence of biocrep in these trials.

Design This metaresearch study used a cohort of non-inferiority trials published in the five highest impact general medical journals during a 5-year period. Data relating to the characteristics and results of the trials were abstracted.

Primary outcome measures Proportions of trials with a declaration of superiority, non-inferiority and point estimates favouring the experimental therapy and mean absolute risk differences for trials with outcomes expressed as a proportion.

Results Our search yielded 163 trials reporting 182 non-inferiority comparisons; 36 comparisons from 31 trials were between the same therapy at reduced and full intensity. Compared with trials not evaluating reduced intensity therapies, fewer comparisons of reduced intensity therapies demonstrated a favourable result (non-inferiority or superiority) (58.3%vs82.2%; $P=0.002$) and fewer demonstrated superiority (2.8%vs18.5%; $P=0.019$). Likewise, point estimates for reduced intensity therapies more often favoured active control than those for other trials (77.8%vs39.7%; $P<0.001$) as did mean absolute risk differences (+2.5% vs -0.7%; $P=0.018$).

Conclusions Non-inferiority trials comparing a therapy at reduced intensity to the same therapy at full intensity showed reduced effects compared with other non-inferiority trials. This suggests these trials may have a high rate of type 1 errors and biocrep, with significant implications for the design and interpretation of future non-inferiority trials.

INTRODUCTION

As non-inferiority trials become commonplace,^{1 2} concerns about their validity take on greater importance.³⁻⁵ In a typical non-inferiority trial, an experimental therapy of unknown efficacy is compared with an active control which previously has been compared with placebo in a superiority trial and found to be efficacious. One assumption inherent in non-inferiority trials is that a

Strengths and limitations of this study

- Hypothesis-driven and novel study addressing a topic for which there exist few empirical data.
- Rigorous and transparent methods using a cross-section of non-inferiority trials from the five highest impact journals.
- The cross-section represents only a small subset of all journals.

new (experimental) therapy that is declared non-inferior to an efficacious comparator would be superior to placebo if this hypothesis were tested in a superiority trial.^{5 6} This 'presumed superiority to placebo' may be incorrect if the non-inferiority trial has a large margin of non-inferiority and the results favour active control.^{7 8} The 'presumed superiority to placebo' may also be incorrect in the case where several iterations of non-inferiority trials occur, a phenomenon called 'biocrep' (see [figure 1](#)). Few empirical data exist as to if and how often therapies declared non-inferior have reduced effectiveness due to erosion of presumed superiority to placebo.⁸⁻¹⁰

We recently observed that non-inferiority trials have been used to compare therapies at a reduced intensity (in terms of cumulative dose or omission of a component of a multi-faceted therapy) to the same therapy at full intensity, with the aim of reducing costs or making the therapy more convenient or less toxic. For example, recent trials compared low-dose tissue plasminogen activator (TPA) to standard dose TPA for ischaemic stroke, omitted bleomycin from Adriamycin, bleomycin, vinblastine, dacarbazine therapy for lymphoma and tested intermittent versus continuous androgen deprivation for prostate cancer.¹¹⁻¹³ Non-inferiority trials of reduced intensity therapies present a unique opportunity to evaluate degradation of the presumed superiority to placebo of experimental therapies in these trials. In most non-inferiority



¹The University of Utah, Salt Lake City, Utah, USA

²Brooke Army Medical Center, San Antonio, Texas, USA

³Epidemiology, Veterans Affairs Salt Lake City Health Care System, Salt Lake City, Utah, USA

Correspondence to

Dr Scott K Aberegg;
scottaberegg@gmail.com

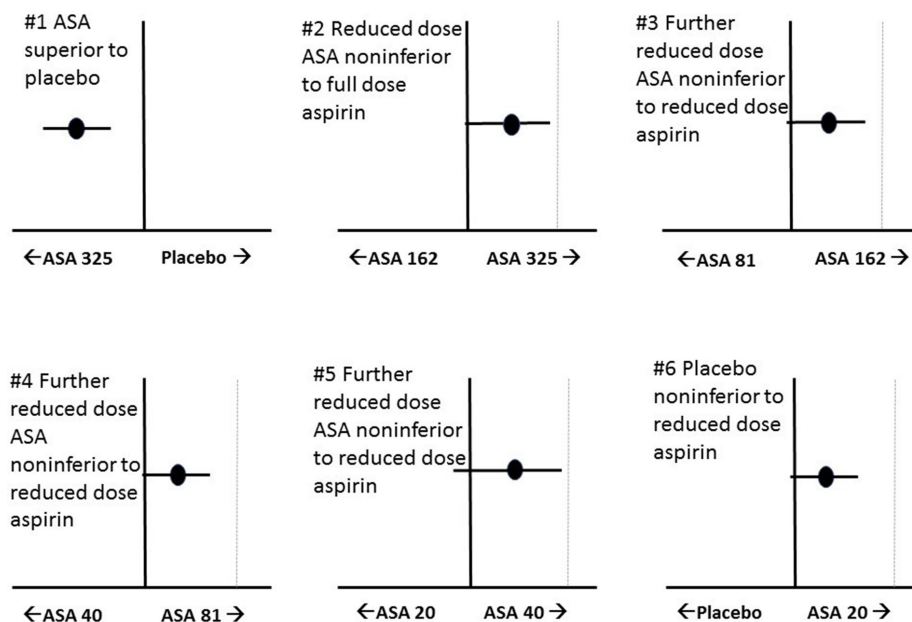


Figure 1 Diagram showing loss of presumed superiority to placebo with reduced intensity aspirin therapy in a hypothetical sequence of trials. The experimental therapy is on the left in each panel and the control is on the right; point estimates are represented as black ovals with bisecting horizontal lines representing 95% CI—point estimates on the left of the centre line favour the experimental therapy and point estimates on the right favour the active control. In panels 2–6, the vertical dashed line represents the margin of non-inferiority. In panel 1, aspirin 325 mg is superior to placebo control in a superiority trial. In panel 2, reduced dose aspirin at 162 mg as the experimental therapy is compared with full-dose aspirin as active control. The difference favours full-dose aspirin, but the reduced dose meets non-inferiority criteria because the upper bound of the 95% CI does not cross the non-inferiority margin. The dose of aspirin is successively reduced in panels 3–5, with the reduced dose from the previous panel serving as the active control in the subsequent panel. By panel 6, the dose of active control aspirin is 20 mg, and the experimental therapy is aspirin at a dose of 0 mg (ie, placebo) and placebo is non-inferior to aspirin—a highly paradoxical result compared with panel 1 where aspirin was superior to placebo. This result is obtained because in panels 2–6, reduced efficacy of the experimental therapy is concealed in the margin of non-inferiority. This phenomenon has been called ‘biocreep’. ASA, aspirin.

trials of novel experimental therapies, there is little evidence to suggest how the novel therapy will fare compared with the active control—it may be better, the same or worse. Because of dose–response effects, there is good a priori reason to suspect that reduced intensity therapies will be less efficacious than the full-intensity active control.¹⁴ If many reduced intensity therapies nonetheless meet non-inferiority criteria, this would constitute suggestive evidence of some loss of presumed superiority to placebo. An empirical demonstration of such an effect does not exist to date.

In the most extreme case, one or more dose reductions could result in a reduced intensity therapy that approximates a placebo but is nonetheless considered non-inferior to a higher dose. Figure 1 shows how this could happen. In the first panel, full-dose aspirin is shown to be superior to placebo in a superiority trial. In the second panel, a non-inferiority trial compares reduced dose aspirin (as experimental therapy) to full-dose aspirin (as active control), and the reduced dose is

found to be numerically but not statistically worse with the upper bound of the CI below the prespecified margin of non-inferiority. In this scenario, reduced dose aspirin meets non-inferiority criteria when compared with full-dose aspirin even though there is a strong trend towards statistical inferiority of reduced dose aspirin. In the next panel, a further reduction in aspirin dose is again numerically worse than the previous reduced dose, but the CI does not include the margin of non-inferiority and it is declared non-inferior. This sequence culminates in the paradoxical result in panel 6, where the dose of the experimental therapy is reduced to zero, making it a placebo which is non-inferior to aspirin. In this hypothetical sequence, inferiority of reduced dose aspirin is obscured within the margin of non-inferiority in panels 2–5. However, the process need not be iterative—some loss of efficacy and thus presumed superiority to placebo occurs with just one dose reduction in panel 2. This problem will be exacerbated with larger margins of non-inferiority and greater reductions in therapy intensity. Though this

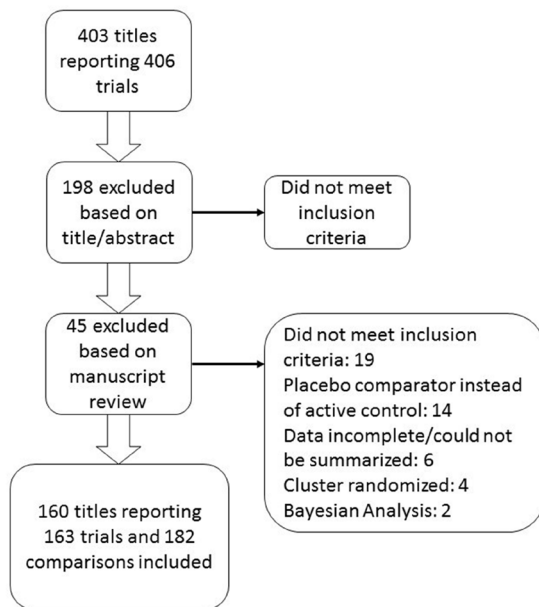


Figure 2 Flow diagram showing selection of trials.

phenomenon, called ‘biocreep’, could happen in any non-inferiority trial, the likelihood would appear to be greater in trials of reduced intensity therapies because of fundamental dose–response considerations.

We compiled a cohort of non-inferiority trials, categorising them based on whether they compared a reduced intensity therapy to a full-intensity active control or otherwise. We hypothesised that trials of reduced intensity therapies would have less favourable results (in terms of point estimates and declarations of superiority and non-inferiority) than trials that were not testing a reduced intensity therapy as the experimental therapy. We also wanted to determine if the margin of non-inferiority was more conservative in trials of reduced intensity therapies.

METHODS

This study used a dataset that was created for a different analysis of non-inferiority trials.¹⁵ We searched Medline for iterations of non-inferiority (eg, non-inferiority, non-inferior)¹⁶ combined with the Medline-recognised names of the five highest impact general medical journals (*New England Journal of Medicine*, *Lancet*, *JAMA*, *British Medical Journal*, *Annals of Internal Medicine*) to identify manuscripts reporting the results of prospective parallel group randomised controlled trials using a test of non-inferiority for the primary hypothesis published between June 2011 and October 2016 (inclusion criteria). Our 5-year retrospective search period began in June 2016 and took until the end of October. Prior to analysing the results, we elected to include articles published during the period of our search from June through October to make the dataset as contemporary as possible. We reviewed the resulting abstracts and manuscripts and excluded those that did not meet inclusion criteria, those that used a

cluster randomised design or Bayesian methodology, those that did not use an active control (eg, Food and Drug Administration-mandated safety trials comparing a new therapy to placebo) and those reported data that were incomplete or could not be summarised. We extracted data relating to design parameters and results into a standardised form. We categorised trials as testing a reduced intensity therapy if the new therapy used the exact same agents as the comparator but with a reduced dose, duration, an increased dosing interval at the same dose or the removal of one or more of the components of a multicomponent active control. We cross-checked the data several times with redundant methods to ensure accuracy, and one author (AMH) checked a 10% random sample of the data for accuracy and found no errors.

We used raw data from the trials to calculate two-sided 95% CIs for all results and categorised them according to Consolidated Standards of Reporting Trials (CONSORT) recommendations.¹⁷ We chose to do this to standardise the presentation of results to comport with figure 1 of the CONSORT statement.^{17,18} We coded a trial’s results as favourable if they warranted a CONSORT declaration of non-inferiority (the upper bound of the 95% CI excluded the prespecified margin of non-inferiority) and/or superiority (the upper bound of the 95% excluded zero difference). For trials where the primary outcome was reported as a measure of risk (eg, HR, OR or relative risk), we calculated the absolute risk difference for the primary outcome for use in quantitative analyses.¹⁹ For trials that reported multiple primary outcomes, we considered the first outcome mentioned in the manuscript to be the primary outcome. For trials where multiple interventions (eg, multiple doses of the same drug) were tested in independent groups, we considered these to be independent non-inferiority comparisons. We used χ^2 and Student’s t-tests where appropriate. All descriptive statistics and analyses were performed with STATA V.14.

RESULTS

Figure 2 shows the results of our search strategy. From 403 manuscripts reporting 406 independent trials, 198 were excluded based on review of the abstract because inclusion criteria were not met, and 45 were excluded after manuscript review because inclusion criteria were not met or exclusion criteria were met. This left 160 manuscripts reporting 163 trials and 182 non-inferiority comparisons.

Table 1 shows basic characteristics of the trials. The two highest impact journals (*New England Journal of Medicine* and *Lancet*) published 127 (78%) of the trials. Four speciality orientations accounted for over half of the trials: infectious diseases, haematology/oncology, cardiology and pulmonary/critical care (see table 1).

There were 31 trials and 36 comparisons of a reduced intensity therapy as the experimental therapy to a full-intensity active control. A selection of these trials and the therapies they evaluated is listed in table 2. The

Table 1 Characteristics of 163 included trials

| | All trials n (%) | Non-RIT trials n (%) | RIT trials n (%) |
|--|------------------|----------------------|------------------|
| | Total n=163 | Total n=132 | Total n=31 |
| Journal | | | |
| <i>New England Journal of Medicine</i> | 64 (39) | 53 (40) | 11 (35) |
| <i>Lancet</i> | 63 (39) | 49 (36.7) | 14 (45) |
| <i>JAMA</i> | 23 (14) | 21 (16) | 2 (6) |
| <i>BMJ</i> | 8 (5) | 6 (5) | 2 (6) |
| <i>Annals of Internal Medicine</i> | 5 (3) | 3 (2) | 2 (6) |
| Year* | | | |
| 2011 | 12 (7) | 10 (8) | 2 (6) |
| 2012 | 25 (15) | 18 (14) | 7 (23) |
| 2013 | 34 (21) | 31 (23) | 3 (10) |
| 2014 | 22 (14) | 14 (11) | 8 (26) |
| 2015 | 43 (26) | 36 (27) | 7 (23) |
| 2016 | 27 (17) | 23 (17) | 4 (13) |
| Top specialties | | | |
| Infectious diseases | (25) | (24) | (26) |
| Haematology/Oncology | (21) | (17) | (39) |
| Cardiology | (17) | (19) | (6) |
| Pulmonary/Critical care | (13) | (14) | (6) |
| Endocrine | (6) | (7) | (3) |
| Primary outcome measured as: | | | |
| Absolute risk difference | 114 (70) | 92 (70) | 22 (71) |
| Mean | 26 (16) | 23 (17) | 3 (10) |
| HR | 13 (8) | 9 (7) | 4 (13) |
| Relative risk difference | 8 (5) | 7 (5) | 1 (3) |
| OR | 2 (1) | 1 (1) | 1 (3) |

Additional characteristics of the trials can be found in Aberegg *et al.*¹⁵

*2011 and 2016 were incomplete years.

RIT, reduced intensity therapy.

proportion of favourable results (a determination of non-inferiority or superiority) was 58.3% (95% CI 41% to 74%) for these comparisons versus 82.2% (95% CI 75% to 88%) for comparisons not testing a reduced intensity therapy (difference 23.9%; 95% CI 6.6% to 41.1%, $P=0.002$). Among comparisons involving reduced intensity therapies, 2.8% warranted a declaration of superiority versus 18.5% of the remainder of comparisons (difference 15.7%; 95% CI 7.4% to 24%, $P=0.019$).

Point estimates of 151 absolute differences in the primary outcome were more likely to favour the active

control when the new therapy was a reduced intensity therapy compared with trials not testing a reduced intensity therapy (60.3% vs 22.2%; difference 38.1%; $P<0.001$). These results are shown graphically in [figure 3](#) (black circles representing reduced intensity therapies comparisons, Xs representing all other comparisons). Examination of [figure 3](#) shows a paucity of point estimates favouring the active control for trials with small sample sizes, a finding that suggests possible publication bias; however, formal tests of publication bias (Begg and Mazumdar²⁰ and Harbord *et al.*²¹), which are known to be insensitive, were not statistically significant. For the 151 comparisons where the outcome could be calculated as a proportion, the mean absolute risk difference between trials testing reduced intensity therapy versus trials not testing reduced intensity therapy was +2.5% versus -0.7% (difference 3.2%; $P=0.018$), with positive values favouring active control. For these trials, the mean prespecified margin of non-inferiority was nearly identical for trials of reduced intensity therapy versus all other trials (8.8% vs 8.4%; difference 0.4%, $P=0.73$).

As a sensitivity analysis, we coded other trials as reduced intensity therapies to determine if a different definition of reduced intensity therapy influenced the results. There were six trials where the active control was the standard of care but for which there was inadequate evidence of superiority to placebo, and it was compared with placebo as the new therapy. An example is the trial of perioperative bridging anticoagulation versus placebo in patients with atrial fibrillation.²² When these trials were coded as reduced intensity therapies, the results of all our analyses were materially unchanged (data not shown).

DISCUSSION

In placebo-controlled superiority trials, researchers generally use the highest tolerable dose of an experimental therapy to maximise separation of the trial populations and increase the likelihood of finding statistically significant outcome differences.²³ Conversely, inadequate dosing of the active control in a non-inferiority trial can bias the results towards the null and increase the probability of falsely declaring non-inferiority when the experimental therapy is truly inferior.^{5 24 25} We identified a unique subset of non-inferiority trials where investigators compared a reduced intensity therapy to the same therapy at full intensity. This arrangement invites errors in the interpretation of these trials, even while it creates an opportunity to evaluate theoretical underpinnings of non-inferiority trials. Our results show that when a reduced intensity therapy is compared with a full intensity active control in non-inferiority trials, the results disfavour reduced intensity therapies in absolute terms and when compared with non-inferiority trials that do not compare two essentially identical therapies at different intensities. This observation is not entirely inconsistent with the general goal of a non-inferiority trial which is to exclude differences greater than a prespecified

Table 2 Examples of non-inferiority trials of reduced intensity therapies included in the analysis. See online appendix 1 for a full bibliography of all 31 trials

| First author | Disease | Experimental therapy | Active control | Outcome |
|-------------------------------|----------------------------------|--|---|---|
| Anderson ²⁹ | Ischaemic stroke | Low-dose alteplase | Standard dose alteplase | Death or disability at 90 days |
| Johnson ¹¹ | Hodgkin's lymphoma | ABV | ABVD | 3-year progression free survival |
| Sherman ³⁰ | Hepatitis C virus infection | 24 weeks telaprevir | 48 weeks telaprevir | Sustained virological response |
| Pritchard-Jones ³¹ | Wilms' tumour | Omission of doxorubicin | Inclusion of doxorubicin | Event-free survival 2 years after diagnosis |
| Bernard ³² | Pyogenic vertebral osteomyelitis | 6 weeks of antibiotics | 12 weeks of antibiotics | Clinical cure rate |
| Vaidya ³³ | Breast cancer | Targeted radiotherapy | Whole breast radiotherapy | Local recurrence rate |
| van Herwaarden ³⁴ | Rheumatoid arthritis | Withdrawal of adalimumab or etanercept | Continuation of adalimumab or etanercept | Rate of major flare at 18 months |
| Feres ³⁵ | Coronary stenting | 3 months antiplatelet therapy | 12 months antiplatelet therapy | Net adverse clinical and cerebral events |
| Rahman ³⁶ | Malignant pleural effusions | 12 French tube | 24 French tube | Pleurodesis efficacy |
| Barone ³⁷ | Genital fistula | 7 days postoperative bladder catheterisation | 14 days postoperative bladder catheterisation | Repair breakdown rate |

ABV, Adriamycin, bleomycin, vinblastine; ABVD, Adriamycin, bleomycin, vinblastine, dacarbazine.

margin. Nonetheless, our results emphasise that caution is warranted in the interpretation of results and conclusions of non-inferiority trials of reduced intensity therapies. Clinicians may be advised to carefully inspect the results with an emphasis on the delta margin used and the 95% CI of the results to determine it includes clinically

important values.^{26 27} In addition, careful evaluation of the purported and demonstrated benefits of the reduced dose, be they reduced cost, side effects or inconvenience, is warranted to provide assurance that any loss of efficacy is justified by these secondary factors. Likewise, investigators designing these trials should recognise the inherent threat of biocreep and design them with a suitably conservative margin of non-inferiority. Notably, trials of reduced intensity therapies in our cohort did not use a more conservative margin of non-inferiority than other trials, perhaps because the enhanced threat to their validity has heretofore gone unrecognised. While our focus was on the specific vulnerability of trials of reduced intensity therapies, all non-inferiority trials are susceptible to loss of presumed superiority to placebo and biocreep.

To our knowledge, no prior investigations have evaluated the effects of reduced intensity therapies in non-inferiority trials nor has there been an empirical demonstration of biocreep which remains a theoretical concept. This is because a demonstration of biocreep or loss of some of the presumed superiority to placebo (sometimes called the putative placebo effect) would require the experimental therapy to be compared with placebo, which is usually ethically infeasible and the very reason a non-inferiority design was selected.^{4 28} We recognised that non-inferiority trials of reduced intensity therapies constituted a natural experiment of sorts that could provide suggestive empirical evidence of loss of the presumed superiority to placebo. Several studies have

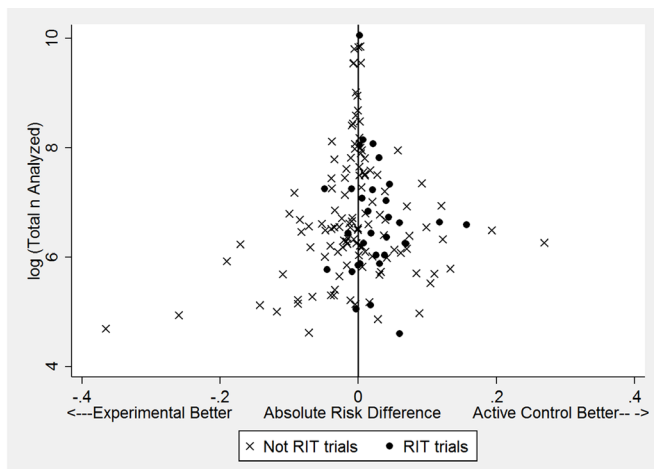


Figure 3 The log of the total number of patients analysed in the trials plotted against the absolute risk differences for the primary outcome among 151 comparisons where a proportion could be calculated. Trials of reduced intensity therapies (black circles) tend to have absolute risk differences that favour active control. A paucity of data points in the bottom right of the figure may suggest publication bias. RIT, reduced intensity therapy.

used simulations to evaluate the propensity for biocreep in non-inferiority trials depending on different underlying assumptions.^{8–10} Two of these studies including one modelled based on empirical data⁸ showed significant risk of biocreep,^{8,9} while one concluded that there was little risk if certain assumptions were met.¹⁰ The results of these simulations hinge critically on the underlying assumptions, particularly the distribution of true treatment effects that are selected for the simulation model. Our empirical data add to and compliment these results. In general, there is a concern for but not an expectation of reduced treatment effects of the experimental therapy in non-inferiority trials. In the case of reduced intensity therapies, there is an expectation of reduced effects based on dose–response considerations. The only situations in which a diminished effect would not be expected with a reduced intensity therapy are those in which there is no dose–response relationship between the therapy and its therapeutic effect or where superiority trials which established the efficacy of the active control used a dose so high as that the slope of a sigmoidal dose–response curve was zero. Thus, our results serve as a preliminary ‘proof of concept’ for the theoretical notion of biocreep.

An alternative interpretation of our results was offered by two reviewers. The reviewers noted that since non-inferiority or superiority criteria were met for only 58% of trials of reduced intensity therapies, the proposed sequence of biocreep illustrated in figure 1 was interrupted for 42% of the trials with the first non-inferiority trial. That is, the non-inferiority trials were effective in filtering out truly non-inferior therapies. (If publication bias leads to unfavourable results not being published differentially, the true proportion of favourable results may be lower than 58%.) We agree that it is reassuring that many non-inferiority trials of reduced intensity therapies fail to demonstrate superiority or non-inferiority but note that the majority do meet non-inferiority criteria. This is concerning because any declaration of non-inferiority is highly sensitive to the choice of delta—with a large enough delta any therapy can be declared non-inferior.

Strengths of our study are that it was conducted based on an a priori hypothesis and used explicit, replicable and transparent methods. Limitations include that we sampled only selected journals for a limited publication epoch. Since the highest impact journals appear to publish the bulk of non-inferiority trials, the impact of this limitation should be minimal. Confirmation and replication of the effects we report could be sought by extending our analysis to trials both before and after the period we studied, and with a more comprehensive array of journals. Even though we showed that reduced intensity therapies have effects that tend to favour full intensity, the comparison of these trials to those that do not compare therapies of differing intensities is subject to the ecological fallacy. Our findings can only suggest erosion of presumed superiority to placebo and early biocreep but cannot confirm that these phenomena are operative. Doing so would require comparing reduced intensity

therapies directly to placebo which is usually ethically infeasible.²⁸ Nonetheless, the results provide a cautionary tale for non-inferiority trials of reduced intensity therapies and indeed all non-inferiority trials.

CONCLUSIONS

Non-inferiority trials of reduced intensity therapies show reduced effects, yet the majority meet non-inferiority criteria. This finding is consistent with loss of some of the presumed superiority to placebo and early biocreep. The results justify caution in the interpretation of non-inferiority trials of reduced intensity therapies and highlight the critical importance of the prespecified margin of non-inferiority in all such trials to avoid false declarations of non-inferiority.

Contributors SKA and AMH designed the study and performed data abstraction and analysis and drafting and reviewing the manuscript. MHS provided critical analysis of the design and analysis of the study and assisted with drafting and reviewing and revising the manuscript.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The dataset used for this manuscript may be obtained by contacting the corresponding author.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Murthy VL, Desai NR, Vora A, *et al*. Increasing proportion of clinical trials using noninferiority end points. *Clin Cardiol* 2012;35:522–3.
- Suda KJ, Hurley AM, McKibbin T, *et al*. Publication of noninferiority clinical trials: changes over a 20-year interval. *Pharmacotherapy* 2011;31:833–9.
- Fleming TR. Current issues in non-inferiority trials. *Stat Med* 2008;27:317–32.
- D’Agostino RB, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Stat Med* 2003;22:169–86.
- Administration F, Drug. Non-Inferiority Clinical Trials to Establish Effectiveness. 2016 <https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm202140.pdf> (accessed 7 Mar 2017).
- Rothmann M, Li N, Chen G, *et al*. Design and analysis of non-inferiority mortality trials in oncology. *Stat Med* 2003;22:239–64.
- Lange S, Freitag G. Choice of delta: requirements and reality—results of a systematic review. *Biom J* 2005;47–12–27.
- Gladstone BP, Vach W. Choice of non-inferiority (NI) margins does not protect against degradation of treatment effects on an average: an observational study of registered and published NI trials. *PLoS One* 2014;9:e103616.
- Odem-Davis K, Fleming TR. A simulation study evaluating bio-creep risk in serial non-inferiority clinical trials for preservation of effect. *Stat Biopharm Res* 2015;7:12–24.
- Everson-Stewart S, Emerson SS. Bio-creep in non-inferiority clinical trials. *Stat Med* 2010;29:2769–80.

11. Johnson P, Federico M, Kirkwood A, *et al.* Adapted treatment guided by interim PET-CT scan in advanced hodgkin's lymphoma. *N Engl J Med* 2016;374:2419–29.
12. Behringer K, Goergen H, Hitz F, *et al.* Omission of dacarbazine or bleomycin, or both, from the ABVD regimen in treatment of early-stage favourable Hodgkin's lymphoma (GHSG HD13): an open-label, randomised, non-inferiority trial. *Lancet* 2015;385:1418–27.
13. Crook JM, O'Callaghan CJ, Duncan G, *et al.* Intermittent androgen suppression for rising PSA level after radiotherapy. *N Engl J Med* 2012;367:895–903.
14. Hill AB. The environment and disease: association or causation? 1965. *J R Soc Med* 2015;108:32–7.
15. Aberegg SK, Hersh AM, Samore MH. Empirical consequences of current recommendations for the design and interpretation of noninferiority trials. *J Gen Intern Med* 2018;33.
16. Gladstone BP, Vach W. About half of the noninferiority trials tested superior treatments: a trial-register based study. *J Clin Epidemiol* 2013;66:386–96.
17. Piaggio G, Elbourne DR, Pocock SJ, *et al.* Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA* 2012;308:2594–604.
18. Piaggio G, Elbourne DR, Altman DG, *et al.* Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006;295:1152–60.
19. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ* 1999;319:1492–5.
20. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;50:1088–101.
21. Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med* 2006;25:3443–57.
22. Douketis JD, Spyropoulos AC, Kaatz S, *et al.* Perioperative bridging anticoagulation in patients with atrial fibrillation. *N Engl J Med* 2015;373:823–33.
23. Steinbrook R. How best to ventilate? Trial design and patient safety in studies of the acute respiratory distress syndrome. *N Engl J Med* 2003;348:1393–401.
24. Aberegg SK, O'Brien JM. Anidulafungin and fluconazole for candidiasis. *N Engl J Med* 2007;357:1347.
25. Jones B, Jarvis P, Lewis JA, *et al.* Trials to assess equivalence: the importance of rigorous methods. *BMJ* 1996;313:36–9.
26. Goodman S. A dirty dozen: twelve p-value misconceptions. *Semin Hematol* 2008;45:135–40.
27. Greenland S, Senn SJ, Rothman KJ, *et al.* Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 2016;31:337–50.
28. Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet* 2007;370:1875–7.
29. Anderson CS, Robinson T, Lindley RI, *et al.* Low-dose versus standard-dose intravenous alteplase in acute ischemic stroke. *N Engl J Med* 2016;374:2313–23.
30. Sherman KE, Flamm SL, Afdhal NH, *et al.* Response-guided telaprevir combination treatment for hepatitis C virus infection. *N Engl J Med* 2011;365:1014–24.
31. Pritchard-Jones K, Bergeron C, de Camargo B, *et al.* Omission of doxorubicin from the treatment of stage II-III, intermediate-risk Wilms' tumour (SIOP WT 2001): an open-label, non-inferiority, randomised controlled trial. *Lancet* 2015;386:1156–64.
32. Bernard L, Dinh A, Ghout I, *et al.* Antibiotic treatment for 6 weeks versus 12 weeks in patients with pyogenic vertebral osteomyelitis: an open-label, non-inferiority, randomised, controlled trial. *Lancet* 2015;385:875–82.
33. Vaidya JS, Wenz F, Bulsara M, *et al.* Risk-adapted targeted intraoperative radiotherapy versus whole-breast radiotherapy for breast cancer: 5-year results for local control and overall survival from the TARGIT-A randomised trial. *Lancet* 2014;383:603–13.
34. van Herwaarden N, van der Maas A, Minten MJ, *et al.* Disease activity guided dose reduction and withdrawal of adalimumab or etanercept compared with usual care in rheumatoid arthritis: open label, randomised controlled, non-inferiority trial. *BMJ* 2015;350:h1389.
35. Feres F, Costa RA, Abizaid A, *et al.* Three vs twelve months of dual antiplatelet therapy after zotarolimus-eluting stents: the OPTIMIZE randomized trial. *JAMA* 2013;310:2510–22.
36. Rahman NM, Pepperell J, Rehal S, *et al.* Effect of opioids vs nsaid and larger vs smaller chest tube size on pain control and pleurodesis efficacy among patients with malignant pleural effusion: the time1 randomized clinical trial. *JAMA* 2015;314:2641–53.
37. Barone MA, Widmer M, Arrowsmith S, *et al.* Breakdown of simple female genital fistula repair after 7 day versus 14 day postoperative bladder catheterisation: a randomised, controlled, open-label, non-inferiority trial. *Lancet* 2015;386:56–62.