BMJ Open

Reduced Effects in Noninferiority Trials of Reduced Intensity Therapies

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019494
Article Type:	Research
Date Submitted by the Author:	06-Sep-2017
Complete List of Authors:	aberegg, Scott; University of Utah Hospital, Pulmonary and Critical Care Medicine Hersh, Andrew; University of Utah Hospital Samore, Matthew; University of Utah
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Epidemiology
Keywords:	Clinical trials < THERAPEUTICS, STATISTICS & RESEARCH METHODS, bio-creep, putative placebo effect, non-inferiority trials

Reduced Effects in Noninferiority Trials of Reduced Intensity Therapies

Scott K. Aberegg*

Andrew M. Hersh*

Matthew H. Samore*§

*The University of Utah, Salt Lake City, Utah

§Veterans Affairs Salt Lake City Health Care System, Salt Lake City, Utah

Corresponding Author:

Scott K Aberegg, MD, MPH

Associate Professor of Medicine

The University of Utah School of Medicine

Mailing address: 1321 South 600 East, SLC, UT 84105

scottaberegg@gmail.com; scott.aberegg@hsc.utah.edu

Telephone: 801-664-2180

Running Head: Noninferiority Trials

The dataset may be obtained by emailing the corresponding author.

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 manusctipt.

There was no funding support for this work.

The authors declare no competing interests or conflicts of interest.

ABSTRACT:

Objectives: To identify noninferiority 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 noninferiority trials of reduced intensity therapies have less favorable results than other noninferiority trials in the cohort. Such a finding would provide suggestive evidence of bio-creep in these trials.

Design: This meta-research study utilized a cohort of noninferiority 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.

Setting: None.

Participants: None.

Interventions: None.

Primary outcome measures: Proportions of trials with a declaration of superiority, noninferiority, and point estimates favoring the experimental therapy, and mean absolute risk differences for trials with outcomes expressed as a proportion.

Results: Our search yielded 163 trials reporting 182 noninferiority comparisons; 36 comparisons from 31 trials were between the same therapy at reduced and full intensity. Compared to trials not evaluating reduced intensity therapies, fewer comparisons of reduced intensity therapies demonstrated superiority (2.8% versus 18.5%; P=0.019) and noninferiority (58.3% versus 82.2%; P=0.002). Likewise, point estimates for reduced intensity therapies more often favored active control than those for other trials (77.8% versus 39.7%; P<0.001) as did mean absolute risk differences (+2.5% versus -0.7%; P=0.018).

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

Keywords: Noninferiority trials; reduced intensity therapies; trial design; trial analysis; trial interpretation; bio-creep; putative placebo effect; presumed superiority to placebo; active control

Strengths and limitations of this study:

1.) **Hypothesis driven and novel study addressing a topic for which there exist few empirical data**
2.) **Rigorous and transparent methods using a cross section of noninferiority trials from the 5 highest impact journals**
3.) **The cross section represents only a small subset of all journals**
4.) **The results provide only suggestive evidence of an increased rate of type 1 errors in noninferiority trials of reduced intensity therapies and are subject to the ecological fallacy**

Introduction:

As noninferiority trials become commonplace[1,2], concerns about their validity take on greater importance[3-5]. In a typical noninferiority trial, an experimental therapy of unknown efficacy is compared to an active control which previously has been compared to placebo in a superiority trial and found to be efficacious. One assumption inherent in noninferiority trials is that a new (experimental) therapy that is declared noninferior 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 noninferiority trial has a large margin of noninferiority and the results favor active control[7,8]. Few empirical data exist as to if and how often therapies declared noninferior have reduced effectiveness due to erosion of presumed superiority to placebo[8-10].

We recently observed that noninferiority trials have been used to compare therapies at a reduced intensity (in terms of cumulative dose or omission of a component of a multifaceted 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 TPA to standard dose TPA for ischemic stroke, omitted bleomycin from ABVD therapy for lymphoma, and tested continuous versus intermittent androgen deprivation for prostate cancer[11-13]. Noninferiority 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 noninferiority trials of novel experimental therapies, there is little evidence to suggest how the novel therapy will fare compared to 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 noninferiority 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 noninferior 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 noninferiority 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 confidence interval below the prespecified margin of noninferiority. In this scenario, reduced dose aspirin meets noninferiority criteria when compared to 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 confidence interval does not include the margin of noninferiority and it is declared noninferior. 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 noninferior to aspirin. In this hypothetical sequence, inferiority of reduced dose aspirin is obscured within the margin of noninferiority 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 noninferiority and greater reductions in therapy intensity. Though this phenomenon, called "bio-creep", could happen in any noninferiority 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 noninferiority trials, categorizing them based on whether they compared a reduced intensity therapy to a full intensity active control, or otherwise. We hypothesized that trials of reduced intensity therapies would have less favorable results (in terms of point estimates and declarations of superiority and noninferiority) than trials that were not testing a reduced intensity therapy as the experimental therapy. We also wanted to determine if the margin of noninferiority was more conservative in trails of reduced intensity therapies.

Methods:

This study used a dataset that was created for a different analysis of noninferiority trials (Aberegg et al, in press). We searched MEDLINE for iterations of noninferiority (e.g., non-inferiority, noninferior)[15] combined with the MEDLINE-recognized 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 randomized controlled trials using a test of noninferiority for the primary hypothesis. The trials were published between June, 2011 and October, 2016. (Our five-year retrospective search period began in June, 2016 and took until the end of October. Prior to analyzing 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 prior to closing it and beginning analysis.) We reviewed the resulting abstracts and manuscripts and excluded those that did not meet inclusion criteria, those that used a cluster randomized design or Bayesian methodology, those that did not use an active control (e.g., FDA-mandated safety trials comparing a new therapy to placebo) and those that reported data that were incomplete or could not be summarized. We abstracted data relating to design parameters and results into a standardized form. We categorized trials as testing a reduced intensity therapy if the new therapy utilized 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 multi-component active control. We cross-checked the data several times with redundant methods to ensure accuracy and one author (AMH) checked a random sample of the data for accuracy and found no errors.

We used raw data from the trials to calculate 2-sided 95% confidence intervals for all results and categorized them according to CONSORT recommendations[16]. We chose to do this to standardize the presentation of results to comport with Figure 1 of the CONSORT statement[16,17]. We coded a trial's results as favorable if they warranted a CONSORT declaration of noninferiority and/or superiority. For trials where the primary outcome was reported as a measure of risk (e.g., hazard ratio, odds ratio, or relative risk) we calculated the absolute risk difference for the primary outcome for use in quantitative analyses[18]. 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 (e.g., multiple doses of the same drug) were tested in independent groups, we considered these to be independent noninferiority comparisons. We used Chi Square and Student's t-tests where appropriate. All descriptive statistics and analyses were performed with STATA version 14 (College Station, Texas).

Results:

Figure 1 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 noninferiority 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 specialty orientations accounted for over half of the trials: infectious diseases, hematology/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 rate of a favorable result (a determination of noninferiority or superiority) was 58.3% for these comparisons versus 82.2 % for comparisons not testing a reduced intensity therapy (difference 23.9%; 95% CI 6.6%-41.1%, P=0.02). 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% - 24%, P=0.019). Significantly fewer comparisons of RIT showed noninferiority than comparisons not involving RIT (58.3% versus 82.2%; P=0.002).

Point estimates of 151 absolute differences in the primary outcome were more likely to favor the active control when the new therapy was a reduced intensity therapy compared to trials not testing a reduced intensity therapy (60.3% versus 22.2%; difference 38.1%; P=0.0013). These results are shown graphically in Figure 2 (black circles representing RIT comparisons, Xs representing all other comparisons). Examination of Figure 2 shows a paucity of point estimates favoring the active control for trials with small sample sizes, a finding that suggests possible publication bias; however, formal tests of publication bias (Begg[19] and Harbord[20]), 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 favoring active control. For these trials, the mean prespecified margin of noninferiority was nearly identical for trials of reduced intensity therapy versus all other trials (8.8% versus 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 to placebo as the new therapy. An example is the trial of perioperative bridging anticoagulation versus placebo in patients with atrial fibrillation.[21] 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 maximize separation of the trial populations and increase the likelihood of finding statistically significant outcome differences[22]. Conversely, inadequate dosing of the active

control in a noninferiority trial can bias the results towards the null and increase the probability of falsely declaring noninferiority when the experimental therapy is truly inferior[5,23,24]. We identified a unique subset of noninferiority 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 noninferiority trials. First, as in the case of under-dosing of active control, intentionally reducing the dose of the experimental therapy will increase the probability of a false conclusion of noninferiority (i.e., a type I error). Second, the presumed superiority to placebo of therapies meeting noninferiority criteria in trials of reduced intensity therapies is more tenuous than usual. If separation of trial populations is reduced with reduced intensity therapy as would be expected based on dose-response considerations, finding superiority of a reduced intensity therapy to placebo in a superiority trial would be increasingly difficult, posing significant problems for sample size and recruitment. For example, consider the following scenario. Suppose that investigators using a superiority design aim to show that a new drug is superior to placebo by a margin of 3%, with a baseline event rate of 10%, power of 90% and 2-sided alpha of 0.05 – a sample of 3600 patients is needed for such a trial. The therapy is found to be superior to placebo then later, a noninferiority trial of the therapy at half the original dose shows that the halved dose meets criteria for noninferiority when compared to the full dose. Suppose also that the dose response relationship is linear with a slope of one and the halved dose gives only half the effect, or 1.5%. If the original investigators had aimed to show that a lower dose of the drug was superior to placebo by a margin of 1.5%, almost 16,000 patients would have needed to be enrolled in a superiority trial. A trial of such size is often not possible for financial and logistical reasons; thus, the presumed superiority to placebo that is a necessary attendant of the noninferiority claim is inherently tenuous in this scenario. If the effect were halved iteratively in a succession of noninferiority trials as in the aspirin example in Figure 1, the result would be a "sample size tsunami" that would become unmanageable after just one or two dose reductions, if trials were designed to prove superiority to placebo of progressively lower doses of the therapy.

Our results show that when a reduced intensity therapy is compared to a full intensity active control in noninferiority trials, the results disfavor reduced intensity therapies in absolute terms and when compared to noninferiority trials that do not compare two essentially identical therapies at different intensities. This observation is not entirely inconsistent with the general goal of a noninferiority trial which is to exclude differences greater than a prespecified margin. Nonetheless, our results emphasize that caution is warranted in the interpretation of results and conclusions of noninferiority trials of reduced intensity therapies. Clinicians may be advised to carefully inspect the results with an emphasis on the delta margin utilized and the 95% confidence interval of the results to determine it includes clinically important values[25,26]. 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 recognize the inherent threat of bio-creep and design them with a suitably conservative margin of noninferiority. Notably, trials of reduced intensity therapies in our cohort did not utilize a more conservative margin of noninferiority than other trials, perhaps because the enhanced threat to their validity has heretofore gone unrecognized. While our focus was on the specific vulnerability of trials of reduced intensity therapies, all noninferiority trials are susceptible to loss of presumed superiority to placebo and bio-creep.

To our knowledge, no prior investigations have evaluated the effects of reduced intensity therapies in noninferiority trials, nor has there been an empirical demonstration of bio-creep which remains a theoretical concept. This is because a demonstration of bio-creep or loss of some of the presumed superiority to placebo (sometimes called the putative placebo effect) would require the experimental therapy to be compared to placebo, which is usually ethically infeasible and the very reason a noninferiority design was selected[4,27]. We recognized that noninferiority 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 utilized simulations to evaluate the propensity for bio-creep in noninferiority trials depending upon different underlying assumptions[8-10]. Two of these studies showed significant risk of bio-creep[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 noninferiority 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 bio-creep.

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 noninferiority 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 favor 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 bio-creep but cannot confirm that these phenomena are operative. Doing so would require comparing reduced intensity therapies directly to placebo which is usually ethically infeasible[27]. Nonetheless the results provide a cautionary tale for noninferiority trials of reduced intensity therapies and indeed all noninferiority trials.

Conclusions:

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

1. Murthy VL, Desai NR, Vora A, Bhatt DL. Increasing Proportion of Clinical Trials Using Noninferiority End Points. *Clin Cardiol.* 2012;35(9):522-523.
2. Suda KJ, Hurley AM, McKibbin T, Motl Moroney SE. Publication of noninferiority clinical trials: changes over a 20-year interval. *Pharmacotherapy.* 2011;31(9):833-839.
3. Fleming TR. Current issues in non-inferiority trials. *Stat Med.* 2008;27(3):317-332.
4. D'Agostino RB, Sr., Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Stat Med.* 2003;22(2):169-186.
5. Administration F, Drug. Non-Inferiority Clinical Trials to Establish Effectiveness. 2016; https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm202140.pdf. Accessed March 7, 2017.
6. Rothmann M, Li N, Chen G. Design and analysis of non-inferiority mortality trials in oncology. *Stat Med.* 2003;22.
7. Lange S, Freitag G. Choice of delta: requirements and reality--results of a systematic review. *Biometrical journal Biometrische Zeitschrift.* 2005;47(1):12-27; discussion 99-107.
8. 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(7):e103616.
9. Odem-Davis K, Fleming TR. A simulation study evaluating bio-creep risk in serial non-inferiority clinical trials for preservation of effect. *Statistics in biopharmaceutical research.* 2015;7(1):12-24.
10. Everson-Stewart S, Emerson SS. Bio-creep in non-inferiority clinical trials. *Stat Med.* 2010;29(27):2769-2780.
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(25):2419-2429.
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. *The Lancet.*385(9976):1418-1427.
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(10):895-903.
14. Hill AB. The environment and disease: association or causation? 1965. *J R Soc Med.* 2015;108(1):32-37.
15. Gladstone BP, Vach W. About half of the noninferiority trials tested superior treatments: a trial-register based study. *J Clin Epidemiol.* 2013;66(4):386-396.
16. Piaggio G, Elbourne DR, Pocock SJ, Evans SW, Altman DG, f CG. Reporting of noninferiority and equivalence randomized trials: Extension of the consort 2010 statement. *JAMA.* 2012;308(24):2594-2604.
17. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA.* 2006;295(10):1152-1160.
18. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ.* 1999;319(7223):1492-1495.
19. Begg CB, Mazumdar M. Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics.* 1994;50(4):1088-1101.
20. Harbord RM, Egger M, Sterne JAC. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med.* 2006;25(20):3443-3457.
21. Douketis JD, Spyropoulos AC, Kaatz S, et al. Perioperative Bridging Anticoagulation in Patients with Atrial Fibrillation. *N Engl J Med.* 2015;373(9):823-833.

22. 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(14):1393-1401.
23. Aberegg SK, O'Brien JM, Jr. Anidulafungin and fluconazole for candidiasis. *N Engl J Med.* 2007;357(13):1347; author reply 1348.
24. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ.* 1996;313(7048):36-39.
25. Goodman S. A dirty dozen: twelve p-value misconceptions. *Semin Hematol.* 2008;45(3):135-140.
26. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol.* 2016;31:337-350.
27. Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet.* 2007;370(9602):1875-1877.
28. 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(24):2313-2323.
29. Sherman KE, Flamm SL, Afdhal NH, et al. Response-Guided Telaprevir Combination Treatment for Hepatitis C Virus Infection. *N Engl J Med.* 2011;365(11):1014-1024.
30. 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. *The Lancet.*386(9999):1156-1164.
31. 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. *The Lancet.*385(9971):875-882.
32. 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. *The Lancet.*383(9917):603-613.
33. van Herwaarden N, van der Maas A, Minten MJM, 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 : British Medical Journal.* 2015;350.
34. 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(23):2510-2522.
35. Rahman NM, Pepperell J, Rehal S, et al. Effect of opioids vs nsaids 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(24):2641-2653.
36. 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. *The Lancet.*386(9988):56-62.

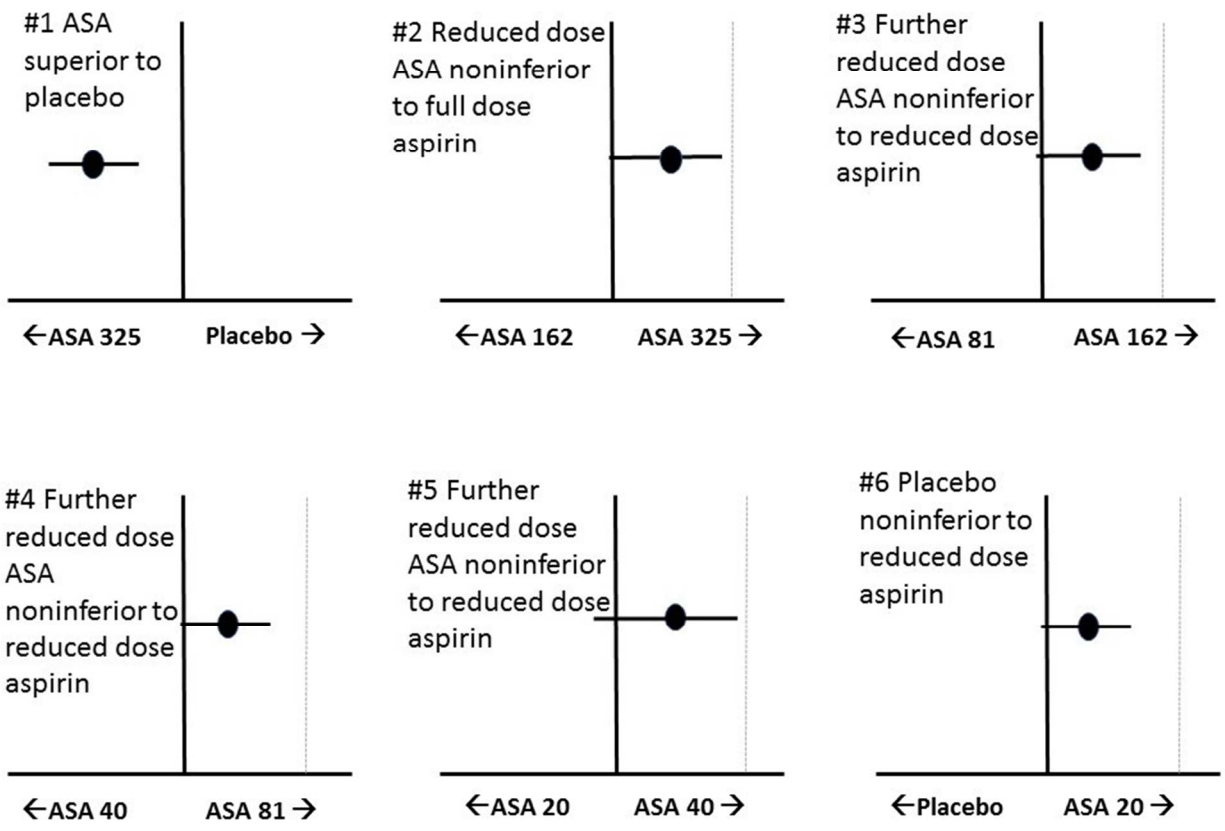

Figure 1. Diagram showing loss of presumed superiority to placebo with reduced intensity aspirin therapy. The experimental therapy is on the left in each panel and the control is on the right; point estimates are represented as black ovals with bisecting horizontal lines representing 95% confidence intervals – point estimates on the left of the center line favor the experimental therapy and point estimates on the right favor the active control. In panels #2-6, the vertical dashed line represents the margin of noninferiority. 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 to full dose aspirin as active control. The difference favors full dose aspirin, but the reduced dose meets noninferiority criteria because the upper bound of the 95% confidence interval does not cross the noninferiority 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 (i.e., placebo) and placebo is noninferior to aspirin – a highly paradoxical result compared to panel #1 where aspirin was superior to placebo. This result obtains because in panels #2-6, reduced efficacy of the experimental therapy is concealed in the margin of noninferiority. This phenomenon has been called “bio-creep.”

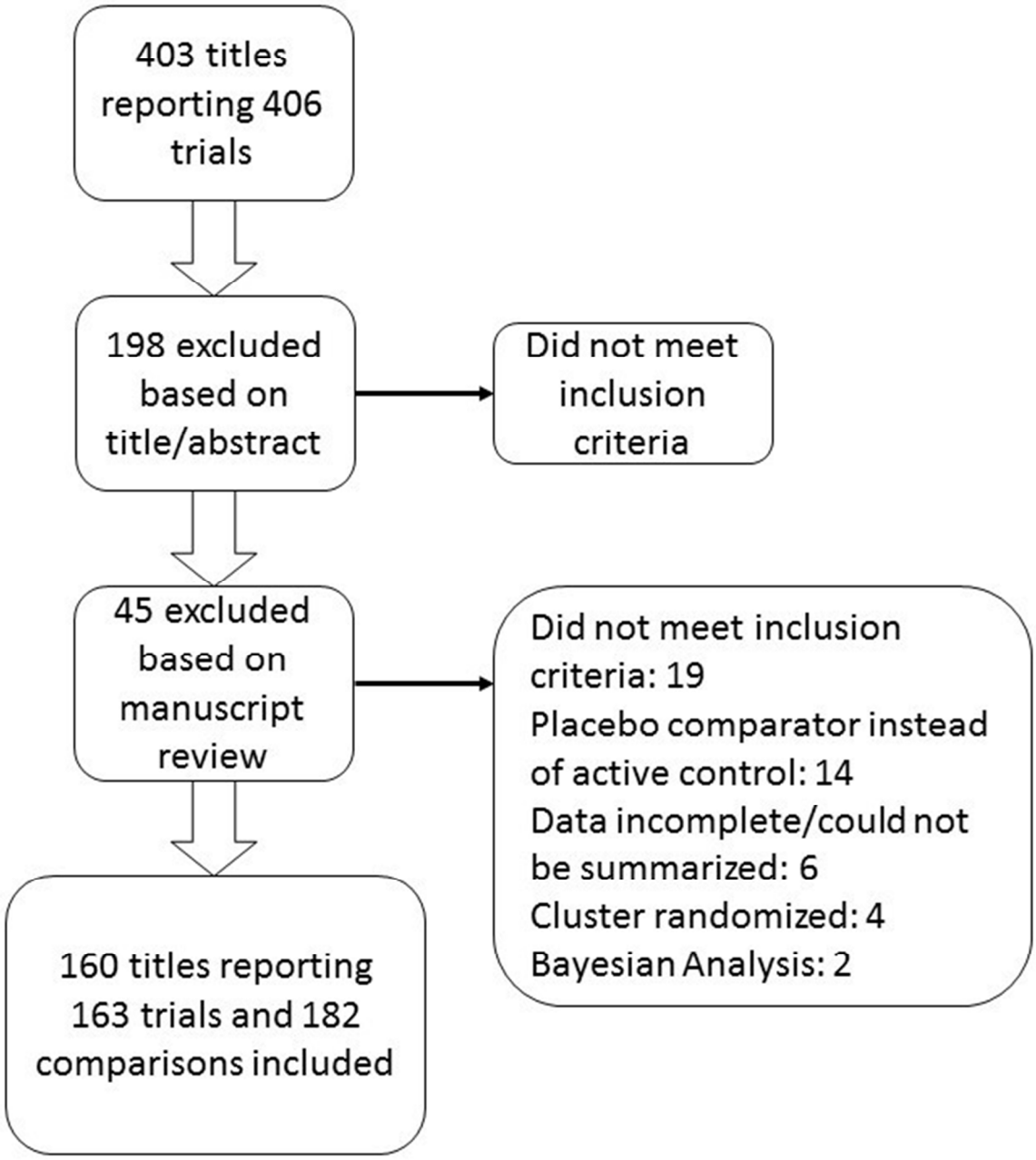

Figure 2. Flow diagram showing selection of trials.

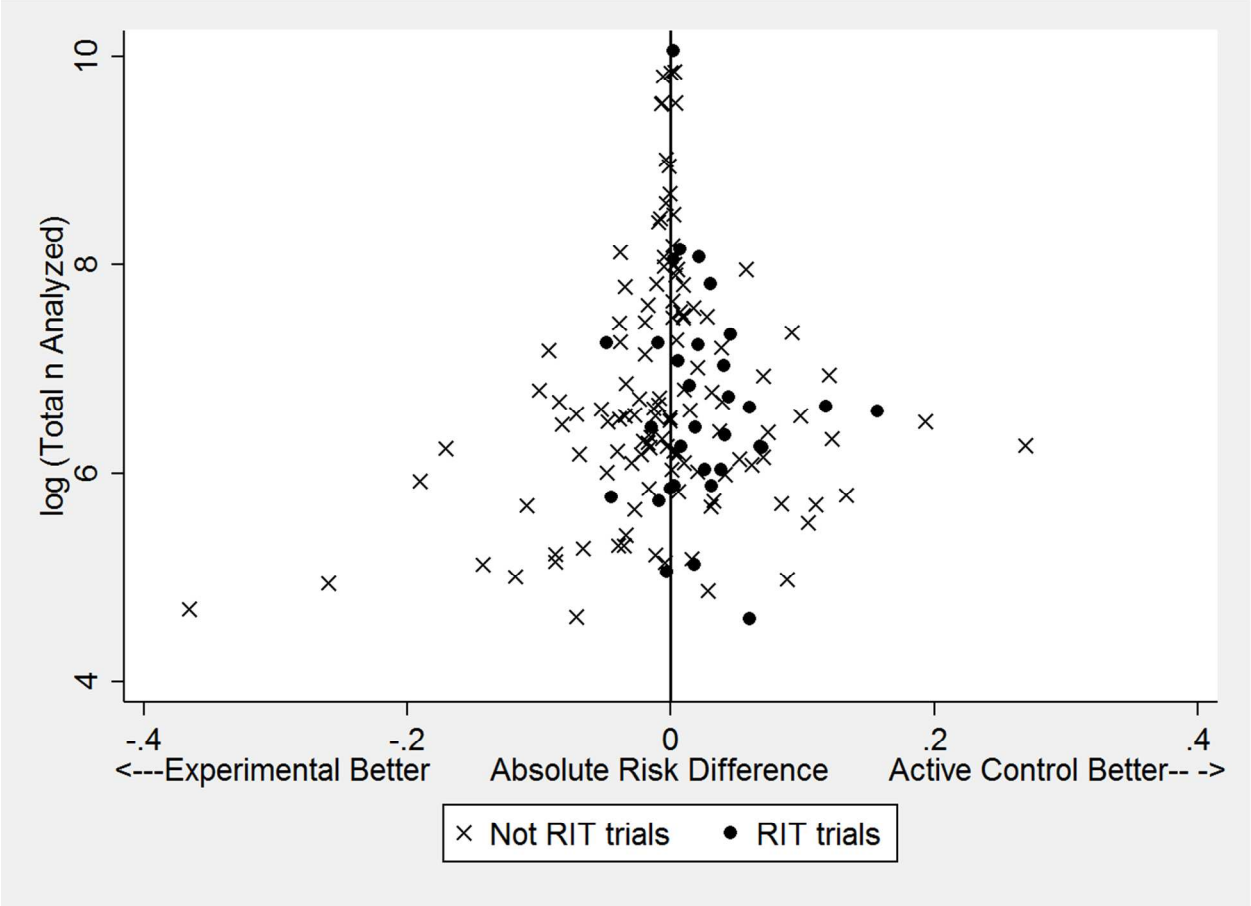

Figure 3. The log of the total number of patients analyzed in the trials plotted against the absolute risk differences for the primary outcome among 151 comparisons where a proportion could be calculated. Trials of reduced intensity therapies (black circles) tend to have absolute risk differences that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

		n (%)
Journal	**NEJM**	**64 (39%)**
	Lancet	**63 (39%)**
	JAMA	**23 (14%)**
	BMJ	**8 (5%)**
	Annals	**5 (3%)**
Year*	**2011**	**12(7%)**
	2012	**25 (15%)**
	2013	**34 (21%)**
	2014	**22 (14%)**
	2015	**43(26%)**
	2016	**27 (17%)**
Top Specialties	**Infectious Diseases**	**(26%)**
	Hematology/Oncology	**(25%)**
	Cardiology	**(17%)**
	Pulmonary/Critical Care	**(15%)**
	Endocrine	**(8%)**
Primary outcome measured as:	**Absolute Risk Difference**	**114 (70%)**
	Mean	**26 (16%)**
	Hazard Ratio	**13 (8%)**
	Relative Risk Difference	**8 (5%)**
	Odds Ratio	**2 (1%)**

Table 1. Characteristics of included trials.

*2011 and 2016 were incomplete years

First Author	Disease	Experimental Therapy	Active Control	Outcome
Anderson[28]	Ischemic 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[29]	Hepatitis C virus infection	24 weeks telaprevir	48 weeks telaprevir	sustained virologic response
Pritchard-Jones[30]	Wilms' tumor	omission of doxarubicin	inclusion of doxarubicin	event-free survival 2 years after diagnosis
Bernard[31]	Pyogenic vertebral osteomyelitis	6 weeks of antibiotics	12 weeks of antibiotics	clinical cure rate
Vaidya[32]	Breast cancer	targeted radiotherapy	whole breast radiotherapy	local recurrence rate
van Herwaarden[33]	Rheumatoid arthritis	withdrawal of adalimumab or etanercept	continuation of adalimumab or etanercept	rate of major flare at 18 months
Feres[34]	Coronary stenting	3 months antiplatelet therapy	12 months antiplatelet therapy	net adverse clinical and cerebral events
Rahman[35]	Malignant pleural effusions	12 French tube	24 French tube	pleurodesis efficacy
Barone[36]	Genital fistula	7 days postoperative bladder catheterization	14 days postoperative bladder catheterization	repair breakdown rate

Table 2. Examples of noninferiority trials of reduced intensity therapies included in the analysis.

BMJ Open

Reduced Effects in Noninferiority Trials of Reduced Intensity Therapies

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019494.R1
Article Type:	Research
Date Submitted by the Author:	15-Nov-2017
Complete List of Authors:	aberegg, Scott; University of Utah Hospital, Pulmonary and Critical Care Medicine Hersh, Andrew; University of Utah Hospital Samore, Matthew; University of Utah
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Epidemiology
Keywords:	Clinical trials < THERAPEUTICS, STATISTICS & RESEARCH METHODS, bio-creep, putative placebo effect, non-inferiority trials

Do Non-inferiority Trials of Reduced Intensity Therapies Show Reduced Effects? A Descriptive Analysis

Scott K. Aberegg*

Andrew M. Hersh*#

Matthew H. Samore*§

*The University of Utah, Salt Lake City, Utah

#Brooke Army Medical Center, San Antonio, Texas

§Veterans Affairs Salt Lake City Health Care System, Salt Lake City, Utah

Corresponding Author:

Scott K Aberegg, MD, MPH

Associate Professor of Medicine

The University of Utah School of Medicine

Mailing address: 1321 South 600 East, SLC, UT 84105

scottaberegg@gmail.com; scott.aberegg@hsc.utah.edu

Telephone: 801-664-2180

The dataset may be obtained by emailing the corresponding author.

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 manusctipt.

There was no funding support for this work.

The authors declare no competing interests or conflicts of interest.

ABSTRACT:

Objectives: To identify noninferiority 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 noninferiority trials of reduced intensity therapies have less favorable results than other noninferiority trials in the cohort. Such a finding would provide suggestive evidence of bio-creep in these trials.

Design: This meta-research study utilized a cohort of noninferiority 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.

Setting: None.

Participants: None.

Interventions: None.

Primary outcome measures: Proportions of trials with a declaration of superiority, noninferiority, and point estimates favoring the experimental therapy, and mean absolute risk differences for trials with outcomes expressed as a proportion.

Results: Our search yielded 163 trials reporting 182 noninferiority comparisons; 36 comparisons from 31 trials were between the same therapy at reduced and full intensity. Compared to trials not evaluating reduced intensity therapies, fewer comparisons of reduced intensity therapies demonstrated a favorable result (noninferiority or superiority) (58.3% versus 82.2%; P=0.002) and fewer demonstrated superiority (2.8% versus 18.5%; P=0.019). Likewise, point estimates for reduced intensity therapies more often favored active control than those for other trials (77.8% versus 39.7%; P<0.001) as did mean absolute risk differences (+2.5% versus -0.7%; P=0.018).

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

Keywords: Noninferiority trials; reduced intensity therapies; trial design; trial analysis; trial interpretation; bio-creep; putative placebo effect; presumed superiority to placebo; active control

Strengths and limitations of this study:

1.) **Hypothesis driven and novel study addressing a topic for which there exist few empirical data**
2.) **Rigorous and transparent methods using a cross section of noninferiority trials from the 5 highest impact journals**
3.) **The cross section represents only a small subset of all journals**
4.) **The results demonstrate a correlation between reduced intensity therapies and reduced effects, and suggest but cannot prove causation**

Introduction:

As noninferiority trials become commonplace[1,2], concerns about their validity take on greater importance[3-5]. In a typical noninferiority trial, an experimental therapy of unknown efficacy is compared to an active control which previously has been compared to placebo in a superiority trial and found to be efficacious. One assumption inherent in noninferiority trials is that a new (experimental) therapy that is declared noninferior 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 noninferiority trial has a large margin of noninferiority and the results favor active control[7,8]. The "presumed superiority to placebo" may also be incorrect in the case where several iterations of noninferiority trials occur, a phenomenon called "bio-creep" (see Figure 1). Few empirical data exist as to if and how often therapies declared noninferior have reduced effectiveness due to erosion of presumed superiority to placebo[8-10].

We recently observed that noninferiority trials have been used to compare therapies at a reduced intensity (in terms of cumulative dose or omission of a component of a multifaceted 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 TPA (tissue plasminogen activator) to standard dose TPA for ischemic stroke, omitted bleomycin from ABVD therapy (Adriamycin Bleomycin, Vinblastine, Dacarbazine) for lymphoma, and tested intermittent versus continuous androgen deprivation for prostate cancer[11-13]. Noninferiority 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 noninferiority trials of novel experimental therapies, there is little evidence to suggest how the novel therapy will fare compared to 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 noninferiority 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 noninferior 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 noninferiority 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 confidence interval below the

prespecified margin of noninferiority. In this scenario, reduced dose aspirin meets noninferiority criteria when compared to 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 confidence interval does not include the margin of noninferiority and it is declared noninferior. 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 noninferior to aspirin. In this hypothetical sequence, inferiority of reduced dose aspirin is obscured within the margin of noninferiority 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 noninferiority and greater reductions in therapy intensity. Though this phenomenon, called "bio-creep", could happen in any noninferiority 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 noninferiority trials, categorizing them based on whether they compared a reduced intensity therapy to a full intensity active control, or otherwise. We hypothesized that trials of reduced intensity therapies would have less favorable results (in terms of point estimates and declarations of superiority and noninferiority) than trials that were not testing a reduced intensity therapy as the experimental therapy. We also wanted to determine if the margin of noninferiority was more conservative in trials of reduced intensity therapies.

Methods:

This study used a dataset that was created for a different analysis of noninferiority trials[15].. We searched MEDLINE for iterations of noninferiority (e.g., non-inferiority, noninferior)[16] combined with the MEDLINE-recognized 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 randomized controlled trials using a test of noninferiority for the primary hypothesis. The trials were published between June, 2011 and October, 2016. (Our five-year retrospective search period began in June, 2016 and took until the end of October. Prior to analyzing 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 randomized design or Bayesian methodology, those that did not use an active control (e.g., FDA-mandated safety trials comparing a new therapy to placebo) and those that reported data that were incomplete or could not be summarized. We excted data relating to design parameters and results into a standardized form. We categorized trials as testing a reduced intensity therapy if the new therapy utilized 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 multi-component 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 2-sided 95% confidence intervals for all results and categorized them according to CONSORT recommendations[17]. We chose to do this to standardize the

presentation of results to comport with Figure 1 of the CONSORT statement[17,18]. We coded a trial's results as favorable if they warranted a CONSORT declaration of noninferiority (the upper bound of the 95% confidence interval excluded the prespecified margin of noninferiority) 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 (e.g., hazard ratio, odds ratio, 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 (e.g., multiple doses of the same drug) were tested in independent groups, we considered these to be independent noninferiority comparisons. We used Chi Square and Student's t-tests where appropriate. All descriptive statistics and analyses were performed with STATA version 14 (College Station, Texas).

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 noninferiority 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 specialty orientations accounted for over half of the trials: infectious diseases, hematology/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 proportion of favorable results (a determination of noninferiority or superiority) was 58.3% for these comparisons versus 82.2 % for comparisons not testing a reduced intensity therapy (difference 23.9%; 95% CI 6.6%-41.1%, P=0.02). 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% - 24%, P=0.019).

Point estimates of 151 absolute differences in the primary outcome were more likely to favor the active control when the new therapy was a reduced intensity therapy compared to trials not testing a reduced intensity therapy (60.3% versus 22.2%; difference 38.1%; P<.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 favoring the active control for trials with small sample sizes, a finding that suggests possible publication bias; however, formal tests of publication bias (Begg[20] and Harbord[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 favoring active control. For these trials, the mean prespecified margin of noninferiority was nearly identical for trials of reduced intensity therapy versus all other trials (8.8% versus 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 to 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 maximize 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 noninferiority trial can bias the results towards the null and increase the probability of falsely declaring noninferiority when the experimental therapy is truly inferior[5,24,25]. We identified a unique subset of noninferiority 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 noninferiority trials. First, as in the case of under-dosing of active control, intentionally reducing the dose of the experimental therapy will increase the probability of a false conclusion of noninferiority (i.e., a type I error). Second, the presumed superiority to placebo of therapies meeting noninferiority criteria in trials of reduced intensity therapies is more tenuous than usual. If separation of trial populations is reduced with reduced intensity therapy as would be expected based on dose-response considerations, finding superiority of a reduced intensity therapy to placebo in a superiority trial would be increasingly difficult, posing significant problems for sample size and recruitment. For example, consider the following scenario. Suppose that investigators using a superiority design aim to show that a new drug is superior to placebo by a margin of 3%, with a baseline event rate of 10%, power of 90% and 2-sided alpha of 0.05 – a sample of 3600 patients is needed for such a trial. The therapy is found to be superior to placebo then later, a noninferiority trial of the therapy at half the original dose shows that the halved dose meets criteria for noninferiority when compared to the full dose. Suppose also that the dose response relationship is linear with a slope of one and the halved dose gives only half the effect, or 1.5%. If the original investigators had aimed to show that a lower dose of the drug was superior to placebo by a margin of 1.5%, almost 16,000 patients would have needed to be enrolled in a superiority trial. A trial of such size is often not possible for financial and logistical reasons; thus, the presumed superiority to placebo that is a necessary attendant of the noninferiority claim is inherently tenuous in this scenario. If the effect were halved iteratively in a succession of noninferiority trials as in the aspirin example in Figure 1, the result would be a “sample size tsunami” that would become unmanageable after just one or two dose reductions, if trials were designed to prove superiority to placebo of progressively lower doses of the therapy.

Our results show that when a reduced intensity therapy is compared to a full intensity active control in noninferiority trials, the results disfavor reduced intensity therapies in absolute terms and when compared to noninferiority trials that do not compare two essentially identical therapies at different intensities. This observation is not entirely inconsistent with the general goal of a noninferiority trial which is to exclude differences greater than a prespecified margin. Nonetheless, our results emphasize that caution is warranted in the interpretation of results and conclusions of noninferiority trials of reduced intensity therapies. Clinicians may be advised to carefully inspect the results with an emphasis

on the delta margin utilized and the 95% confidence interval 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 recognize the inherent threat of bio-creep and design them with a suitably conservative margin of noninferiority. Notably, trials of reduced intensity therapies in our cohort did not utilize a more conservative margin of noninferiority than other trials, perhaps because the enhanced threat to their validity has heretofore gone unrecognized. While our focus was on the specific vulnerability of trials of reduced intensity therapies, all noninferiority trials are susceptible to loss of presumed superiority to placebo and bio-creep.

To our knowledge, no prior investigations have evaluated the effects of reduced intensity therapies in noninferiority trials, nor has there been an empirical demonstration of bio-creep which remains a theoretical concept. This is because a demonstration of bio-creep or loss of some of the presumed superiority to placebo (sometimes called the putative placebo effect) would require the experimental therapy to be compared to placebo, which is usually ethically infeasible and the very reason a noninferiority design was selected[4,28]. We recognized that noninferiority 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 utilized simulations to evaluate the propensity for bio-creep in noninferiority trials depending upon different underlying assumptions[8-10]. Two of these studies including one modeled based upon empirical data[8] showed significant risk of bio-creep[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 noninferiority 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 bio-creep.

An alternative interpretation of our results was offered by two reviewers. The reviewers noted that since noninferiority 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 noninferiority trial. That is, the noninferiority trials were effective in filtering out truly noninferior therapies. We agree that it is reassuring that many noninferiority trials of reduced intensity therapies fail to demonstrate superiority or noninferiority but note that the majority do meet noninferiority criteria. This is concerning because any declaration of noninferiority is highly sensitive to the choice of delta – with a large enough delta any therapy can be declared noninferior.

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 noninferiority 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 favor 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 bio-creep 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 noninferiority trials of reduced intensity therapies and indeed all noninferiority trials.

Conclusions:

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

1. Murthy VL, Desai NR, Vora A, Bhatt DL. Increasing Proportion of Clinical Trials Using Noninferiority End Points. *Clin Cardiol.* 2012;35(9):522-523.
2. Suda KJ, Hurley AM, McKibbin T, Motl Moroney SE. Publication of noninferiority clinical trials: changes over a 20-year interval. *Pharmacotherapy.* 2011;31(9):833-839.
3. Fleming TR. Current issues in non-inferiority trials. *Stat Med.* 2008;27(3):317-332.
4. D'Agostino RB, Sr., Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Stat Med.* 2003;22(2):169-186.
5. Administration F, Drug. Non-Inferiority Clinical Trials to Establish Effectiveness. 2016; https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm202140.pdf. Accessed March 7, 2017.
6. Rothmann M, Li N, Chen G. Design and analysis of non-inferiority mortality trials in oncology. *Stat Med.* 2003;22.
7. Lange S, Freitag G. Choice of delta: requirements and reality--results of a systematic review. *Biometrical journal Biometrische Zeitschrift.* 2005;47(1):12-27; discussion 99-107.
8. 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(7):e103616.
9. Odem-Davis K, Fleming TR. A simulation study evaluating bio-creep risk in serial non-inferiority clinical trials for preservation of effect. *Statistics in biopharmaceutical research.* 2015;7(1):12-24.
10. Everson-Stewart S, Emerson SS. Bio-creep in non-inferiority clinical trials. *Stat Med.* 2010;29(27):2769-2780.

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(25):2419-2429.
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. *The Lancet.*385(9976):1418-1427.
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(10):895-903.
14. Hill AB. The environment and disease: association or causation? 1965. *J R Soc Med.* 2015;108(1):32-37.
15. Aberegg SK, Hersh AM, Samore MH. Empirical Consequences of Current Recommendations for the Design and Interpretation of Noninferiority Trials. *Journal of general internal medicine.* 2017.
16. Gladstone BP, Vach W. About half of the noninferiority trials tested superior treatments: a trial-register based study. *J Clin Epidemiol.* 2013;66(4):386-396.
17. Piaggio G, Elbourne DR, Pocock SJ, Evans SW, Altman DG, f CG. Reporting of noninferiority and equivalence randomized trials: Extension of the consort 2010 statement. *JAMA.* 2012;308(24):2594-2604.
18. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA.* 2006;295(10):1152-1160.
19. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ.* 1999;319(7223):1492-1495.
20. Begg CB, Mazumdar M. Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics.* 1994;50(4):1088-1101.
21. Harbord RM, Egger M, Sterne JAC. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med.* 2006;25(20):3443-3457.
22. Douketis JD, Spyropoulos AC, Kaatz S, et al. Perioperative Bridging Anticoagulation in Patients with Atrial Fibrillation. *N Engl J Med.* 2015;373(9):823-833.
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(14):1393-1401.
24. Aberegg SK, O'Brien JM, Jr. Anidulafungin and fluconazole for candidiasis. *N Engl J Med.* 2007;357(13):1347; author reply 1348.
25. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ.* 1996;313(7048):36-39.
26. Goodman S. A dirty dozen: twelve p-value misconceptions. *Semin Hematol.* 2008;45(3):135-140.
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;31:337-350.
28. Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet.* 2007;370(9602):1875-1877.
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(24):2313-2323.
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(11):1014-1024.
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. *The Lancet.*386(9999):1156-1164.
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. *The Lancet.*385(9971):875-882.

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. *The Lancet.*383(9917):603-613.
34. van Herwaarden N, van der Maas A, Minten MJM, 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 : British Medical Journal.* 2015;350.
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(23):2510-2522.
36. Rahman NM, Pepperell J, Rehal S, et al. Effect of opioids vs nsaids 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(24):2641-2653.
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. *The Lancet.*386(9988):56-62.

Captions to Figures 1-3.

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% confidence intervals – point estimates on the left of the center line favor the experimental therapy and point estimates on the right favor the active control. In panels #2-6, the vertical dashed line represents the margin of noninferiority. 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 to full dose aspirin as active control. The difference favors full dose aspirin, but the reduced dose meets noninferiority criteria because the upper bound of the 95% confidence interval does not cross the noninferiority 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 (i.e., placebo) and placebo is noninferior to aspirin – a highly paradoxical result compared to panel #1 where aspirin was superior to placebo. This result obtains because in panels #2-6, reduced efficacy of the experimental therapy is concealed in the margin of noninferiority. This phenomenon has been called “bio-creep.”

Figure 2. Flow diagram showing selection of trials.

Figure 3. The log of the total number of patients analyzed 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 favor active control. A paucity of datapoints in the bottom right of the figure may suggest publication bias.

		All Trials n (%)	Non-RIT trials n(%)	RIT trials n(%)
Journal	NEJM	64 (39%)	52 (40%)	11 (35%)
	Lancet	63 (39%)	49 (367%)	14 (45%)
	JAMA	23 (14%)	21 (16%)	2(6%)
	BMJ	8 (5%)	6 (5%)	2 (6%)
	Annals	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%
	Hematology/Oncology	21%	17%	39%
	Cardiology	17%	19%	6%
	Pulmonary/Critical	13%	14%	6%

	Care			
	Endocrine	**6%**	**7%**	**3%**
Primary outcome measured as:	**Absolute Risk Difference**	**114 (70%)**	**92 (70%)**	**22 (71%)**
	Mean	**26 (16%)**	**23 (17%)**	**3 (10%)**
	Hazard Ratio	**13 (8%)**	**9 (7%)**	**4 (13%)**
	Relative Risk Difference	**8 (5%)**	**7 (5%)**	**1 (3%)**
	Odds Ratio	**2 (1%)**	**1 (1%)**	**1 (3%)**

Table 1. Characteristics of 163 included trials. Additional characteristics of the trials can be found in reference 15, Aberegg et al. RIT = reduced intensity therapies.

*2011 and 2016 were incomplete years

First Author	Disease	Experimental Therapy	Active Control	Outcome
Anderson[29]	Ischemic 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 virologic response
Pritchard-Jones[31]	Wilms' tumor	omission of doxarubicin	inclusion of doxarubicin	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 catheterization	14 days postoperative bladder catheterization	repair breakdown rate

Table 2. Examples of noninferiority trials of reduced intensity therapies included in the analysis.

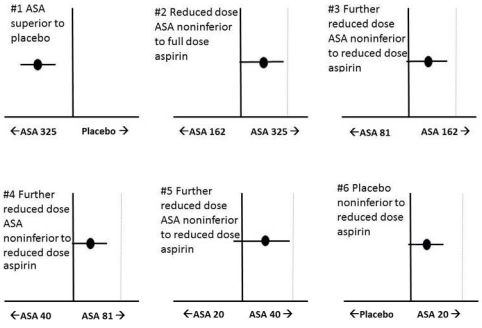

Figure 1. Diagram showing loss of presumed superiority to placebo with reduced intensity aspirin therapy in a hypothetical sequence of trials. The experimental therapy is on the left in each panel and the control is on the right; point estimates are represented as black ovals with bisecting horizontal lines representing 95% confidence intervals – point estimates on the left of the center line favor the experimental therapy and point estimates on the right favor the active control. In panels #2-6, the vertical dashed line represents the margin of noninferiority. 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 to full dose aspirin as active control. The difference favors full dose aspirin, but the reduced dose meets noninferiority criteria because the upper bound of the 95% confidence interval does not cross the noninferiority 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 (i.e., placebo) and placebo is noninferior to aspirin – a highly paradoxical result compared to panel #1 where aspirin was superior to placebo. This result obtains because in panels #2-6, reduced efficacy of the experimental therapy is concealed in the margin of noninferiority. This phenomenon has been called "bio-creep."

338x190mm (300 x 300 DPI)

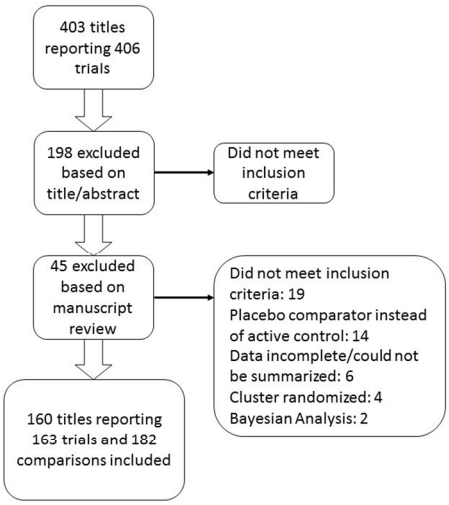

Figure 2. Flow diagram showing selection of trials.

338x190mm (300 x 300 DPI)

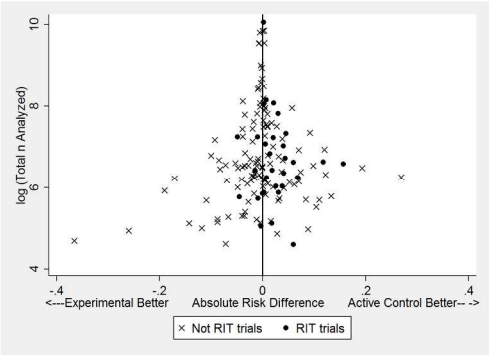

Figure 3. The log of the total number of patients analyzed in the trials plotted against the absolute risk differences for the primary outcome among 151 comparisons where a proportion could be calculated. Trials of reduced intensity therapies (black circles) tend to have absolute risk differences that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

338x190mm (300 x 300 DPI)

BMJ Open

Do Non-inferiority Trials of Reduced Intensity Therapies Show Reduced Effects? A Descriptive Analysis.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019494.R2
Article Type:	Research
Date Submitted by the Author:	30-Dec-2017
Complete List of Authors:	aberegg, Scott; University of Utah Hospital, Pulmonary and Critical Care Medicine Hersh, Andrew; University of Utah Hospital Samore, Matthew; University of Utah
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Epidemiology
Keywords:	Clinical trials < THERAPEUTICS, STATISTICS & RESEARCH METHODS, bio-creep, putative placebo effect, non-inferiority trials

Do Non-inferiority Trials of Reduced Intensity Therapies Show Reduced Effects? A Descriptive Analysis

Scott K. Aberegg*

Andrew M. Hersh*#

Matthew H. Samore*§

*The University of Utah, Salt Lake City, Utah

#Brooke Army Medical Center, San Antonio, Texas

§Veterans Affairs Salt Lake City Health Care System, Salt Lake City, Utah

Corresponding Author:

Scott K Aberegg, MD, MPH

Associate Professor of Medicine

The University of Utah School of Medicine

Mailing address: 1321 South 600 East, SLC, UT 84105

scottaberegg@gmail.com; scott.aberegg@hsc.utah.edu

Telephone: 801-664-2180

The dataset may be obtained by emailing the corresponding author.

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 manusctipt.

There was no funding support for this work.

The authors declare no competing interests or conflicts of interest.

ABSTRACT:

Objectives: To identify noninferiority 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 noninferiority trials of reduced intensity therapies have less favorable results than other noninferiority trials in the cohort. Such a finding would provide suggestive evidence of bio-creep in these trials.

Design: This meta-research study utilized a cohort of noninferiority 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.

Setting: None.

Participants: None.

Interventions: None.

Primary outcome measures: Proportions of trials with a declaration of superiority, noninferiority, and point estimates favoring the experimental therapy, and mean absolute risk differences for trials with outcomes expressed as a proportion.

Results: Our search yielded 163 trials reporting 182 noninferiority comparisons; 36 comparisons from 31 trials were between the same therapy at reduced and full intensity. Compared to trials not evaluating reduced intensity therapies, fewer comparisons of reduced intensity therapies demonstrated a favorable result (noninferiority or superiority) (58.3% versus 82.2%; P=0.002) and fewer demonstrated superiority (2.8% versus 18.5%; P=0.019). Likewise, point estimates for reduced intensity therapies more often favored active control than those for other trials (77.8% versus 39.7%; P<0.001) as did mean absolute risk differences (+2.5% versus -0.7%; P=0.018).

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

Keywords: Noninferiority trials; reduced intensity therapies; trial design; trial analysis; trial interpretation; bio-creep; putative placebo effect; presumed superiority to placebo; active control

Strengths and limitations of this study:

1.) **Hypothesis driven and novel study addressing a topic for which there exist few empirical data**
2.) **Rigorous and transparent methods using a cross section of noninferiority trials from the 5 highest impact journals**
3.) **The cross section represents only a small subset of all journals**

Introduction:

As noninferiority trials become commonplace[1,2], concerns about their validity take on greater importance[3-5]. In a typical noninferiority trial, an experimental therapy of unknown efficacy is compared to an active control which previously has been compared to placebo in a superiority trial and found to be efficacious. One assumption inherent in noninferiority trials is that a new (experimental) therapy that is declared noninferior 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 noninferiority trial has a large margin of noninferiority and the results favor active control[7,8]. The "presumed superiority to placebo" may also be incorrect in the case where several iterations of noninferiority trials occur, a phenomenon called "bio-creep" (see Figure 1). Few empirical data exist as to if and how often therapies declared noninferior have reduced effectiveness due to erosion of presumed superiority to placebo[8-10].

We recently observed that noninferiority trials have been used to compare therapies at a reduced intensity (in terms of cumulative dose or omission of a component of a multifaceted 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 TPA (tissue plasminogen activator) to standard dose TPA for ischemic stroke, omitted bleomycin from ABVD therapy (Adriamycin Bleomycin, Vinblastine, Dacarbazine) for lymphoma, and tested intermittent versus continuous androgen deprivation for prostate cancer[11-13]. Noninferiority 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 noninferiority trials of novel experimental therapies, there is little evidence to suggest how the novel therapy will fare compared to 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 noninferiority 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 noninferior 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 noninferiority 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 confidence interval below the prespecified margin of noninferiority. In this scenario, reduced dose aspirin meets noninferiority criteria when compared to 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 confidence interval does not include the margin of noninferiority and it is declared noninferior. 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 noninferior to aspirin. In this hypothetical sequence, inferiority of reduced dose aspirin is obscured within the margin of noninferiority 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 noninferiority and greater reductions in therapy intensity. Though this phenomenon, called "bio-creep", could happen in any noninferiority 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 noninferiority trials, categorizing them based on whether they compared a reduced intensity therapy to a full intensity active control, or otherwise. We hypothesized that trials of reduced intensity therapies would have less favorable results (in terms of point estimates and declarations of superiority and noninferiority) than trials that were not testing a reduced intensity therapy as the experimental therapy. We also wanted to determine if the margin of noninferiority was more conservative in trials of reduced intensity therapies.

Methods:

This study used a dataset that was created for a different analysis of noninferiority trials[15].. We searched MEDLINE for iterations of noninferiority (e.g., non-inferiority, noninferior)[16] combined with the MEDLINE-recognized 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 randomized controlled trials using a test of noninferiority for the primary hypothesis published between June, 2011 and October, 2016 (inclusion criteria). (Our five-year retrospective search period began in June, 2016 and took until the end of October. Prior to analyzing 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 randomized design or Bayesian methodology, those that did not use an active control (e.g., FDA-mandated safety trials comparing a new therapy to placebo) and those that reported data that were incomplete or could not be summarized. We excted data relating to design parameters and results into a standardized form. We categorized trials as testing a reduced intensity therapy if the new therapy utilized 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 multi-component 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 2-sided 95% confidence intervals for all results and categorized them according to CONSORT recommendations[17]. We chose to do this to standardize the presentation of results to comport with Figure 1 of the CONSORT statement[17,18]. We coded a trial's results as favorable if they warranted a CONSORT declaration of noninferiority (the upper bound of the

95% confidence interval excluded the prespecified margin of noninferiority) 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 (e.g., hazard ratio, odds ratio, 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 (e.g., multiple doses of the same drug) were tested in independent groups, we considered these to be independent noninferiority comparisons. We used Chi Square and Student's t-tests where appropriate. All descriptive statistics and analyses were performed with STATA version 14 (College Station, Texas).

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 noninferiority 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 specialty orientations accounted for over half of the trials: infectious diseases, hematology/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 proportion of favorable results (a determination of noninferiority or superiority) was 58.3% (95% CI 41%-74%) for these comparisons versus 82.2 % (95% CI 75%-88%) for comparisons not testing a reduced intensity therapy (difference 23.9%; 95% CI 6.6%-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% - 24%, P=0.019).

Point estimates of 151 absolute differences in the primary outcome were more likely to favor the active control when the new therapy was a reduced intensity therapy compared to trials not testing a reduced intensity therapy (60.3% versus 22.2%; difference 38.1%; P<.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 favoring the active control for trials with small sample sizes, a finding that suggests possible publication bias; however, formal tests of publication bias (Begg[20] and Harbord[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 favoring active control. For these trials, the mean prespecified margin of noninferiority was nearly identical for trials of reduced intensity therapy versus all other trials (8.8% versus 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 to 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 maximize 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 noninferiority trial can bias the results towards the null and increase the probability of falsely declaring noninferiority when the experimental therapy is truly inferior[5,24,25]. We identified a unique subset of noninferiority 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 noninferiority trials. Our results show that when a reduced intensity therapy is compared to a full intensity active control in noninferiority trials, the results disfavor reduced intensity therapies in absolute terms and when compared to noninferiority trials that do not compare two essentially identical therapies at different intensities. This observation is not entirely inconsistent with the general goal of a noninferiority trial which is to exclude differences greater than a prespecified margin. Nonetheless, our results emphasize that caution is warranted in the interpretation of results and conclusions of noninferiority trials of reduced intensity therapies. Clinicians may be advised to carefully inspect the results with an emphasis on the delta margin utilized and the 95% confidence interval 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 recognize the inherent threat of bio-creep and design them with a suitably conservative margin of noninferiority. Notably, trials of reduced intensity therapies in our cohort did not utilize a more conservative margin of noninferiority than other trials, perhaps because the enhanced threat to their validity has heretofore gone unrecognized. While our focus was on the specific vulnerability of trials of reduced intensity therapies, all noninferiority trials are susceptible to loss of presumed superiority to placebo and bio-creep.

To our knowledge, no prior investigations have evaluated the effects of reduced intensity therapies in noninferiority trials, nor has there been an empirical demonstration of bio-creep which remains a theoretical concept. This is because a demonstration of bio-creep or loss of some of the presumed superiority to placebo (sometimes called the putative placebo effect) would require the experimental therapy to be compared to placebo, which is usually ethically infeasible and the very reason a noninferiority design was selected[4,28]. We recognized that noninferiority 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 utilized simulations to evaluate the propensity for bio-creep in noninferiority trials depending upon different underlying assumptions[8-10]. Two of these studies including one modeled based upon empirical data[8] showed significant risk of bio-creep[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 noninferiority 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 bio-creep.

An alternative interpretation of our results was offered by two reviewers. The reviewers noted that since noninferiority 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 noninferiority trial. That is, the noninferiority trials were effective in filtering out truly noninferior therapies. (If publication bias leads to unfavorable results not being published differentially, the true proportion of favorable results may be lower than 58%.) We agree that it is reassuring that many noninferiority trials of reduced intensity therapies fail to demonstrate superiority or noninferiority but note that the majority do meet noninferiority criteria. This is concerning because any declaration of noninferiority is highly sensitive to the choice of delta – with a large enough delta any therapy can be declared noninferior.

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 noninferiority 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 favor 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 bio-creep 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 noninferiority trials of reduced intensity therapies and indeed all noninferiority trials.

Conclusions:

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

1. Murthy VL, Desai NR, Vora A, Bhatt DL. Increasing Proportion of Clinical Trials Using Noninferiority End Points. *Clin Cardiol.* 2012;35(9):522-523.
2. Suda KJ, Hurley AM, McKibbin T, Motl Moroney SE. Publication of noninferiority clinical trials: changes over a 20-year interval. *Pharmacotherapy.* 2011;31(9):833-839.
3. Fleming TR. Current issues in non-inferiority trials. *Stat Med.* 2008;27(3):317-332.
4. D'Agostino RB, Sr., Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Stat Med.* 2003;22(2):169-186.
5. Administration F, Drug. Non-Inferiority Clinical Trials to Establish Effectiveness. 2016; https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm202140.pdf. Accessed March 7, 2017.
6. Rothmann M, Li N, Chen G. Design and analysis of non-inferiority mortality trials in oncology. *Stat Med.* 2003;22.
7. Lange S, Freitag G. Choice of delta: requirements and reality--results of a systematic review. *Biometrical journal Biometrische Zeitschrift.* 2005;47(1):12-27; discussion 99-107.
8. 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(7):e103616.
9. Odem-Davis K, Fleming TR. A simulation study evaluating bio-creep risk in serial non-inferiority clinical trials for preservation of effect. *Statistics in biopharmaceutical research.* 2015;7(1):12-24.
10. Everson-Stewart S, Emerson SS. Bio-creep in non-inferiority clinical trials. *Stat Med.* 2010;29(27):2769-2780.
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(25):2419-2429.
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. *The Lancet.*385(9976):1418-1427.
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(10):895-903.
14. Hill AB. The environment and disease: association or causation? 1965. *J R Soc Med.* 2015;108(1):32-37.
15. Aberegg SK, Hersh AM, Samore MH. Empirical Consequences of Current Recommendations for the Design and Interpretation of Noninferiority Trials. *Journal of general internal medicine.* 2017.
16. Gladstone BP, Vach W. About half of the noninferiority trials tested superior treatments: a trial-register based study. *J Clin Epidemiol.* 2013;66(4):386-396.
17. Piaggio G, Elbourne DR, Pocock SJ, Evans SW, Altman DG, f CG. Reporting of noninferiority and equivalence randomized trials: Extension of the consort 2010 statement. *JAMA.* 2012;308(24):2594-2604.
18. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA.* 2006;295(10):1152-1160.
19. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ.* 1999;319(7223):1492-1495.

20. Begg CB, Mazumdar M. Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics.* 1994;50(4):1088-1101.
21. Harbord RM, Egger M, Sterne JAC. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med.* 2006;25(20):3443-3457.
22. Douketis JD, Spyropoulos AC, Kaatz S, et al. Perioperative Bridging Anticoagulation in Patients with Atrial Fibrillation. *N Engl J Med.* 2015;373(9):823-833.
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(14):1393-1401.
24. Aberegg SK, O'Brien JM, Jr. Anidulafungin and fluconazole for candidiasis. *N Engl J Med.* 2007;357(13):1347; author reply 1348.
25. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. *BMJ.* 1996;313(7048):36-39.
26. Goodman S. A dirty dozen: twelve p-value misconceptions. *Semin Hematol.* 2008;45(3):135-140.
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;31:337-350.
28. Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet.* 2007;370(9602):1875-1877.
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(24):2313-2323.
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(11):1014-1024.
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. *The Lancet.*386(9999):1156-1164.
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. *The Lancet.*385(9971):875-882.
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. *The Lancet.*383(9917):603-613.
34. van Herwaarden N, van der Maas A, Minten MJM, 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 : British Medical Journal.* 2015;350.
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(23):2510-2522.
36. Rahman NM, Pepperell J, Rehal S, et al. Effect of opioids vs nsaids 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(24):2641-2653.
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. *The Lancet.*386(9988):56-62.

Captions to Figures 1-3.

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% confidence intervals – point estimates on the left of the center line favor the experimental therapy and point estimates on the right favor the active control. In panels #2-6, the vertical dashed line represents the margin of noninferiority. 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 to full dose aspirin as active control. The difference favors full dose aspirin, but the reduced dose meets noninferiority criteria because the upper bound of the 95% confidence interval does not cross the noninferiority 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 (i.e., placebo) and placebo is noninferior to aspirin – a highly paradoxical result compared to panel #1 where aspirin was superior to placebo. This result obtains because in panels #2-6, reduced efficacy of the experimental therapy is concealed in the margin of noninferiority. This phenomenon has been called “bio-creep.”

Figure 2. Flow diagram showing selection of trials.

Figure 3. The log of the total number of patients analyzed 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 favor active control. A paucity of datapoints in the bottom right of the figure may suggest publication bias.

		All Trials n (%)	Non-RIT trials n(%)	RIT trials n(%)
		Total n 163	Total n 132	Total n 31
Journal	NEJM	64 (39%)	53 (40%)	11 (35%)
	Lancet	63 (39%)	49 (367%)	14 (45%)
	JAMA	23 (14%)	21 (16%)	2(6%)
	BMJ	8 (5%)	6 (5%)	2 (6%)
	Annals	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%
	Hematology/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%)
	Hazard Ratio	13 (8%)	9 (7%)	4 (13%)
	Relative Risk Difference	8 (5%)	7 (5%)	1 (3%)
	Odds Ratio	2 (1%)	1 (1%)	1 (3%)

Table 1. Characteristics of 163 included trials. Additional characteristics of the trials can be found in reference 15, Aberegg et al. RIT = reduced intensity therapies.

*2011 and 2016 were incomplete years

First Author	Disease	Experimental Therapy	Active Control	Outcome
Anderson[29]	Ischemic 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 virologic response
Pritchard-Jones[31]	Wilms' tumor	omission of doxarubicin	inclusion of doxarubicin	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 catheterization	14 days postoperative bladder catheterization	repair breakdown rate

Table 2. Examples of noninferiority trials of reduced intensity therapies included in the analysis. See Appendix 1 for a full bibliography of all 31 trials.

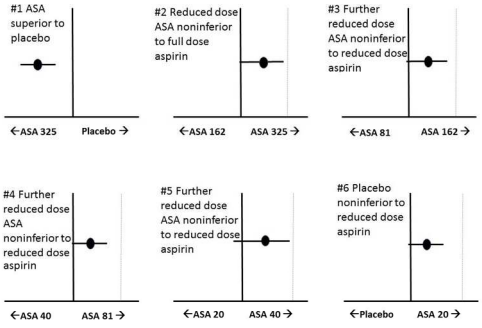

Figure 1. Diagram showing loss of presumed superiority to placebo with reduced intensity aspirin therapy in a hypothetical sequence of trials. The experimental therapy is on the left in each panel and the control is on the right; point estimates are represented as black ovals with bisecting horizontal lines representing 95% confidence intervals – point estimates on the left of the center line favor the experimental therapy and point estimates on the right favor the active control. In panels #2-6, the vertical dashed line represents the margin of noninferiority. 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 to full dose aspirin as active control. The difference favors full dose aspirin, but the reduced dose meets noninferiority criteria because the upper bound of the 95% confidence interval does not cross the noninferiority 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 (i.e., placebo) and placebo is noninferior to aspirin – a highly paradoxical result compared to panel #1 where aspirin was superior to placebo. This result obtains because in panels #2-6, reduced efficacy of the experimental therapy is concealed in the margin of noninferiority. This phenomenon has been called "bio-creep."

338x190mm (300 x 300 DPI)

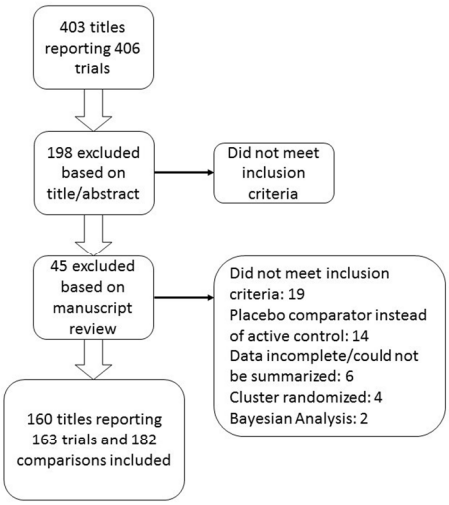

Figure 2. Flow diagram showing selection of trials.

338x190mm (300 x 300 DPI)

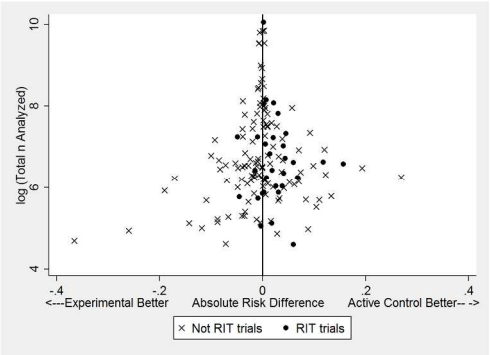

Figure 3. The log of the total number of patients analyzed in the trials plotted against the absolute risk differences for the primary outcome among 151 comparisons where a proportion could be calculated. Trials of reduced intensity therapies (black circles) tend to have absolute risk differences that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

338x190mm (300 x 300 DPI)

Appendix 1.

Bibliography of 31 noninferiority trials (reporting 36 comparisons) that used a reduced intensity therapy as the new therapy compared to the same therapy at full intensity as active control.[1-31]

1. Anderson CS, Robinson T, Lindley RI, et al. Low-Dose versus Standard-Dose Intravenous Alteplase in Acute Ischemic Stroke. *New England Journal of Medicine* 2016;374(24):2313-23. doi: 10.1056/NEJMoa1515510
2. Radford J, Illidge T, Counsell N, et al. Results of a Trial of PET-Directed Therapy for Early-Stage Hodgkin's Lymphoma. *New England Journal of Medicine* 2015;372(17):1598-607. doi: 10.1056/NEJMoa1408648
3. Magnussen H, Disse B, Rodriguez-Roisin R, et al. Withdrawal of Inhaled Glucocorticoids and Exacerbations of COPD. *New England Journal of Medicine* 2014;371(14):1285-94. doi: 10.1056/NEJMoa1407154
4. Paton NI, Kityo C, Hoppe A, et al. Assessment of Second-Line Antiretroviral Regimens for HIV Therapy in Africa. *New England Journal of Medicine* 2014;371(3):234-47. doi: 10.1056/NEJMoa1311274
5. Bwakura-Dangarembizi M, Kendall L, Bakeera-Kitaka S, et al. A Randomized Trial of Prolonged Co-trimoxazole in HIV-Infected Children in Africa. *New England Journal of Medicine* 2014;370(1):41-53. doi: 10.1056/NEJMoa1214901
6. Johnson P, Federico M, Kirkwood A, et al. Adapted Treatment Guided by Interim PET-CT Scan in Advanced Hodgkin's Lymphoma. *New England Journal of Medicine* 2016;374(25):2419-29. doi: 10.1056/NEJMoa1510093
7. Hussain M, Tangen CM, Berry DL, et al. Intermittent versus Continuous Androgen Deprivation in Prostate Cancer. *New England Journal of Medicine* 2013;368(14):1314-25. doi: 10.1056/NEJMoa1212299
8. Crook JM, O'Callaghan CJ, Duncan G, et al. Intermittent Androgen Suppression for Rising PSA Level after Radiotherapy. *New England Journal of Medicine* 2012;367(10):895-903. doi: 10.1056/NEJMoa1201546
9. Mallick U, Harmer C, Yap B, et al. Ablation with Low-Dose Radioiodine and Thyrotropin Alfa in Thyroid Cancer. *New England Journal of Medicine* 2012;366(18):1674-85. doi: 10.1056/NEJMoa1109589
10. Kim K, Kim YH, Kim SY, et al. Low-Dose Abdominal CT for Evaluating Suspected Appendicitis. *New England Journal of Medicine* 2012;366(17):1596-605. doi: 10.1056/NEJMoa1110734
11. Sherman KE, Flamm SL, Afdhal NH, et al. Response-Guided Telaprevir Combination Treatment for Hepatitis C Virus Infection. *New England Journal of Medicine* 2011;365(11):1014-24. doi: 10.1056/NEJMoa1014463
12. Rahman NM, Pepperell J, Rehal S, et al. Effect of opioids vs nsaids 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(24):2641-53. doi: 10.1001/jama.2015.16840
13. 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(23):2510-22. doi: 10.1001/jama.2013.282183
14. Leuppi JD, Schuetz P, Bingisser R, et al. Short-term vs conventional glucocorticoid therapy in acute exacerbations of chronic obstructive pulmonary disease: The reduce randomized clinical trial. *JAMA* 2013;309(21):2223-31. doi: 10.1001/jama.2013.5023

15. Strnad V, Ott OJ, Hildebrandt G, et al. 5-year results of accelerated partial breast irradiation using sole interstitial multicatheter brachytherapy versus whole-breast irradiation with boost after breast-conserving surgery for low-risk invasive and in-situ carcinoma of the female breast: a randomised, phase 3, non-inferiority trial. *The Lancet* 2016;387(10015):229-38. doi: https://doi.org/10.1016/S0140-6736(15)00471-7
16. 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. *The Lancet* 2015;386(9999):1156-64. doi: https://doi.org/10.1016/S0140-6736(14)62395-3
17. 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. *The Lancet* 2015;386(9988):56-62. doi: https://doi.org/10.1016/S0140-6736(14)62337-0
18. 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. *The Lancet* 2015;385(9976):1418-27. doi: https://doi.org/10.1016/S0140-6736(14)61469-0
19. 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. *The Lancet* 2015;385(9971):875-82. doi: https://doi.org/10.1016/S0140-6736(14)61233-2
20. Vain NE, Satragno DS, Gorenstein AN, et al. Effect of gravity on volume of placental transfusion: a multicentre, randomised, non-inferiority trial. *The Lancet* 2014;384(9939):235-40. doi: https://doi.org/10.1016/S0140-6736(14)60197-5
21. Efficacy of 400 mg efavirenz versus standard 600 mg dose in HIV-infected, antiretroviral-naive adults (ENCORE1): a randomised, double-blind, placebo-controlled, non-inferiority trial. *The Lancet* 2014;383(9927):1474-82. doi: https://doi.org/10.1016/S0140-6736(13)62187-X
22. Herpertz-Dahlmann B, Schwarte R, Krei M, et al. Day-patient treatment after short inpatient care versus continued inpatient treatment in adolescents with anorexia nervosa (ANDI): a multicentre, randomised, open-label, non-inferiority trial. *The Lancet* 2014;383(9924):1222-29. doi: https://doi.org/10.1016/S0140-6736(13)62411-3
23. Chetchotisakd P, Chierakul W, Chaowagul W, et al. Trimethoprim-sulfamethoxazole versus trimethoprim-sulfamethoxazole plus doxycycline as oral eradicative treatment for melioidosis (MERTH): a multicentre, double-blind, non-inferiority, randomised controlled trial. *The Lancet* 2014;383(9919):807-14. doi: https://doi.org/10.1016/S0140-6736(13)61951-0
24. Sandberg T, Skoog G, Hermansson AB, et al. Ciprofloxacin for 7 days versus 14 days in women with acute pyelonephritis: a randomised, open-label and double-blind, placebo-controlled, non-inferiority trial. *The Lancet* 2012;380(9840):484-90. doi: https://doi.org/10.1016/S0140-6736(12)60608-4
25. Kirchhof P, Andresen D, Bosch R, et al. Short-term versus long-term antiarrhythmic drug treatment after cardioversion of atrial fibrillation (Flec-SL): a prospective, randomised, open-label, blinded endpoint assessment trial. *The Lancet* 2012;380(9838):238-46. doi: https://doi.org/10.1016/S0140-6736(12)60570-4
26. Engert A, Haverkamp H, Kobe C, et al. Reduced-intensity chemotherapy and PET-guided radiotherapy in patients with advanced stage Hodgkin's lymphoma (HD15 trial): a randomised, open-label, phase 3 non-inferiority trial. *The Lancet* 2012;379(9828):1791-99. doi: https://doi.org/10.1016/S0140-6736(11)61940-5

27. Gülmezoglu AM, Lumbiganon P, Landoulsi S, et al. Active management of the third stage of labour with and without controlled cord traction: a randomised, controlled, non-inferiority trial. *The Lancet* 2012;379(9827):1721-27. doi: https://doi.org/10.1016/S0140-6736(12)60206-2
28. Tashima KT, Smeaton LM, Fichtenbaum CJ, et al. Hiv salvage therapy does not require nucleoside reverse transcriptase inhibitors: A randomized, controlled trial. *Annals of Internal Medicine* 2015;163(12):908-17. doi: 10.7326/M15-0949
29. Schulman S, Parpia S, Stewart C, et al. Warfarin dose assessment every 4 weeks versus every 12 weeks in patients with stable international normalized ratios: A randomized trial. *Annals of Internal Medicine* 2011;155(10):653-59. doi: 10.7326/0003-4819-155-10-201111150-00003
30. Mol GC, van de Ree MA, Klok FA, et al. One versus two years of elastic compression stockings for prevention of post-thrombotic syndrome (OCTAVIA study): randomised controlled trial. *BMJ* 2016;353 doi: 10.1136/bmj.i2691
31. van Herwaarden N, van der Maas A, Minten MJM, 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 : British Medical Journal* 2015;350 doi: 10.1136/bmj.h1389