



# BMJ Open Testing for non-inferior mortality: a systematic review of non-inferiority margin sizes and trial characteristics

Sandra Pong <sup>1</sup>, Martin Urner,<sup>2,3</sup> Robert A Fowler,<sup>2,4</sup> Nicholas Mitsakakis,<sup>5</sup> Winnie Seto,<sup>1,6</sup> James S Hutchison,<sup>7</sup> Michelle Science,<sup>8</sup> Nick Daneman <sup>9</sup>

**To cite:** Pong S, Urner M, Fowler RA, *et al.* Testing for non-inferior mortality: a systematic review of non-inferiority margin sizes and trial characteristics. *BMJ Open* 2021;**11**:e044480. doi:10.1136/bmjopen-2020-044480

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-044480>).

Received 03 September 2020

Revised 30 March 2021

Accepted 06 April 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Sandra Pong;  
[sandra.pong@sickkids.ca](mailto:sandra.pong@sickkids.ca)

## ABSTRACT

**Objective** To describe the size and variability of non-inferiority margins used in non-inferiority trials of medications with primary outcomes involving mortality, and to examine the association between trial characteristics and non-inferiority margin size.

**Design** Systematic review.

**Data sources** Medline, Medline In Process, Medline Epub Ahead of Print and Embase Classic+Embase databases from January 1989 to December 2019.

**Eligibility criteria** Prospective non-inferiority randomised controlled trials comparing pharmacological therapies, with primary analyses for non-inferiority and primary outcomes involving mortality alone or as part of a composite outcome. Trials had to prespecify non-inferiority margins as absolute risk differences or relative to risks of outcome and provide a baseline risk of primary outcome in the control intervention.

**Results** 3992 records were screened, 195 articles were selected for full text review and 111 articles were included for analyses. 82% of trials were conducted in thrombosis, infectious diseases or oncology. Mortality was the sole primary outcome in 23 (21%) trials, and part of a composite primary outcome in 88 (79%) trials. The overall median non-inferiority margin was an absolute risk difference of 9% (IQR 4.2%–10%). When non-inferiority margins were expressed relative to the baseline risk of primary outcome in control groups, the median relative non-inferiority margin was 1.5 (IQR 1.3–1.7). In multivariable regression analyses examining the association between trial characteristics (medical specialty, inclusion of paediatric patients, mortality as a sole or part of a composite primary outcome, presence of industry funding) and non-inferiority margin size, only medical specialty was significantly associated with non-inferiority margin size.

**Conclusion** Absolute and relative non-inferiority margins used in published trials comparing medications are large, allowing conclusions of non-inferiority in the context of large differences in mortality. Accepting the potential for large increases in outcomes involving mortality while declaring non-inferiority is a challenging methodological issue in the conduct of non-inferiority trials.

## INTRODUCTION

The premise of non-inferiority trials is to demonstrate that a new treatment is no

## Strengths and limitations of this study

- There have been no previous reviews or studies that describe the size and variability of non-inferiority margins used in trials with high-stake outcomes such as mortality.
- Our comprehensive and sensitive search for non-inferiority trials spanned a 30-year period to ensure that virtually all non-inferiority trials with primary outcome involving mortality would be captured.
- We were reliant on authors to provide the values of non-inferiority margins and estimated risks of outcomes in their sample size calculations.

worse than a standard intervention by a prespecified non-inferiority margin chosen by researchers.<sup>1</sup> Yet proving that drugs, devices and other medical treatments are no worse than a comparison is challenging.<sup>2 3</sup> The acceptable width of the margin of non-inferiority is a controversial aspect in the design of these studies. It is a determinant of the required sample size of a trial and has a large influence on the interpretation of ‘not unacceptably worse.’ Wide margins allow smaller sample sizes to conclude non-inferiority, but if a margin is too wide, a conclusion of non-inferiority could be clinically irrelevant or ethically inappropriate. This would be especially disturbing if the implications of accepting a truly inadequate treatment as non-inferior involves death as an outcome.<sup>2</sup>

Design and analytical challenges, and the deficits in adherence to reporting standards of non-inferiority trials have been described in multiple studies and reviews.<sup>4–12</sup> Much attention has been focused on how non-inferiority margins are selected, whether they are justified<sup>10 13</sup> and how they affect the validity of trial results and conclusions.<sup>11 12</sup> The size of non-inferiority margins could also be influenced by the effectiveness of the standard treatment. A highly effective standard

treatment could allow researchers to tolerate higher thresholds for decreased effectiveness with a new treatment.<sup>14</sup> However, prior research has not described the size and variability of non-inferiority margins used in trials with high-stake outcomes, such as mortality, nor examined whether certain trial characteristics such as the type of patients, medical conditions studied, choice of outcomes and baseline risks of outcomes are associated with the selection of smaller or larger non-inferiority margins. There is a need to establish standards for the design and analyses of non-inferiority trials to promote consistent quality of these trials. An important step, therefore, is to identify the range of non-inferiority margins used in non-inferiority trials and determine whether trial characteristics influence the selection of margin sizes.

In this systematic review, our primary objective was to describe the size and variability of non-inferiority margins used in non-inferiority trials of medications with primary outcomes involving mortality. Our secondary objective was to assess whether selected trial characteristics were associated with non-inferiority margin size. We hypothesised that non-inferiority margins in these trials will be large and variable; and the size of non-inferiority margins will be related to the type of patients and medical conditions studied, as well as availability of industry funding and how mortality has been included in the outcome.

## METHODS

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement to report this systematic review.<sup>15</sup>

### Search strategy

We searched Medline, Medline In Process, Medline Epub Ahead of Print and Embase Classic+Embase databases (OvidSP) (search performed 8 February 2019, updated 12 December 2019) to identify randomised controlled non-inferiority trials published between 1989 and 2019. Our decision to start our search from 1989 was informed by a review that described the changes in publication rate of non-inferiority trials between 1989 and 2009, and found 583 published non-inferiority trials but only one that was published prior to 1998.<sup>3</sup>

Subject heading and text-word terms for 'equivalence trials or non-inferiority or inferiority studies' and mortality were used with the Cochrane sensitive trials filter. Of note, 'non-inferiority trial' and 'inferiority trial' terms are indexed together with 'equivalence trial' in Ovid and the term 'equivalence trial' was only introduced as a Medical Subject Heading in 2018. Results were restricted to the English language and trials performed in humans. The complete electronic database search strategies are presented in online supplemental appendix A. To ensure that all relevant trials were captured, the electronic database search was supplemented with a manual search by scanning the reference lists of included trials and relevant

reviews, in addition to a search of the reviewers' personal files.

### Eligibility criteria

We included all prospective non-inferiority randomised controlled trials involving human subjects that compared pharmacological therapies, where the primary analysis was for non-inferiority and the primary outcome included mortality, either alone or as part of a composite outcome. All trials had to prespecify a non-inferiority margin (as an absolute risk difference or relative to the risk of outcome) and provide a baseline estimate of the risk of primary outcome in the control intervention in a sample size calculation. In cases where these variables changed during the course of the trial, the initial values used in the original trial design were used for analyses. No distinction between paediatric or adult populations was made.

We excluded trials that did not provide a sample size calculation based on a prespecified non-inferiority margin and estimated baseline risk of outcome. To enable comparisons of non-inferiority margins across different trials, we also excluded trials that used non-inferiority margins expressed as incidence rate ratios, ORs or HRs because incidence and HRs are relative to an outcome event rate that changes with time and with ORs, the baseline risk of outcome in the control group cannot be determined to convert the ratio to a relative non-inferiority margin unless it was explicitly stated by the authors. We also excluded articles that described substudies, post hoc analyses or follow-up studies of randomised trials.

### Selection of trials

One review author (SP) screened titles and abstracts of all retrieved records for obvious exclusions. Two review authors (SP and MU) independently assessed potentially eligible trials based on full text review. Disagreements were resolved by arbitration by a third review author (ND).

### Data collection

One review author (SP) extracted data from the included trials using a standardised form to collect information on: year of publication, medical specialty area, inclusion of paediatric patients (age less than 18 years), mortality as a single or part of a composite primary outcome, estimated risk of primary outcome in the control group, non-inferiority margin, industry funding (disclosures in the publication about funding or sponsorship by a pharmaceutical company) and conclusion about non-inferiority.

### Statistical analyses

Trial characteristics were summarised using counts and proportions. To enable comparisons of non-inferiority margins across different trials as either absolute or relative margins, we converted non-inferiority margins expressed as absolute risk differences in percentages into relative non-inferiority margins relative to the estimated risk of outcome for each trial's control group. The reverse was also done to convert relative non-inferiority margins into equivalent margins in terms of absolute differences.

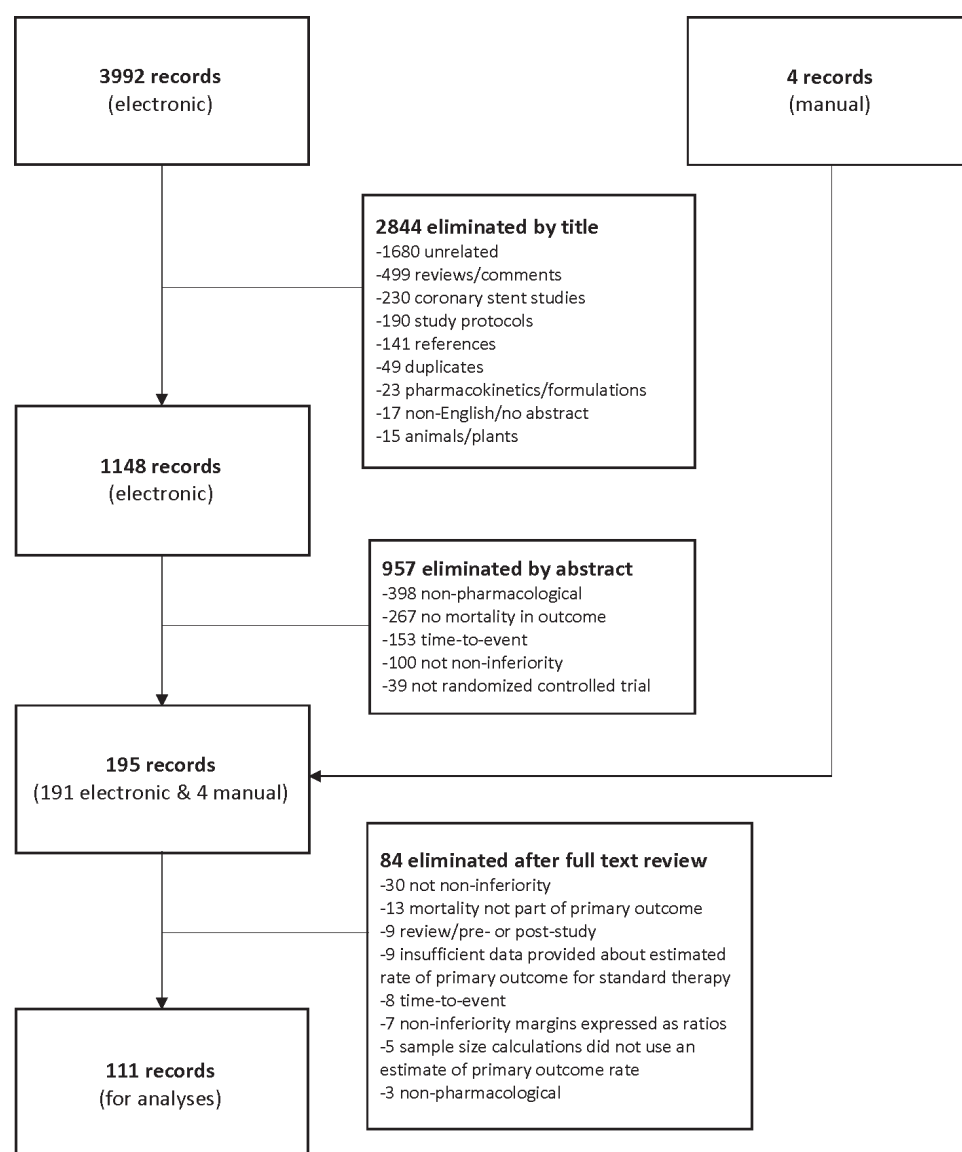
Graphical plots were used to explore an association between absolute non-inferiority margins and the estimated risks of outcome in control groups, and to describe the distribution of absolute and relative non-inferiority margins used in the trials.

For the primary objective, descriptive statistics (median, IQR, range) of absolute and relative non-inferiority margins were summarised for the overall cohort of trials included in the review. We also stratified these by trial characteristics: medical specialty, inclusion of paediatric patients, mortality as a single or composite outcome, industry funding and publication date pre-2010 or post-2010 release of the first Food and Drug Administration (FDA) draft guidance statement about non-inferiority trials. To investigate whether there was a difference in non-inferiority margins (absolute and relative) according to trial characteristics, we compared non-inferiority margins using Wilcoxon rank sum test (for two groups) and Kruskal-Wallis rank sum test (for >2 groups).

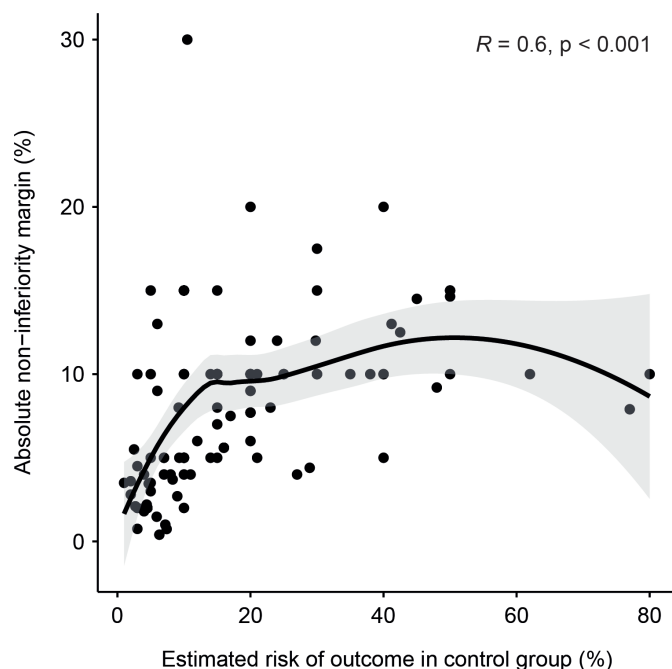
For the secondary objective, we used multivariable linear regression to examine the association between prespecified trial characteristics (medical specialty, inclusion of paediatric patients, mortality as single or composite outcome and industry funding) as independent variables and non-inferiority margin size as the outcome variable. Due to the skewed distribution of the absolute and relative non-inferiority margins, we applied a log-transformation to the outcome variable to improve the performance and diagnostics of the regression models. All comparisons were two sided and  $p < 0.05$  was considered statistically significant. Statistical analyses were conducted using R V.4.0.2.

## RESULTS

We screened 3992 records for relevance using titles and abstracts and selected 195 articles for full-text review. After independent assessment of the full-text articles and discussion among reviewers, a total of 111 articles met



**Figure 1** PRISMA flow diagram.<sup>15</sup> PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



**Figure 2** Association between absolute non-inferiority margins and estimated risks of outcome in control group.

eligibility criteria to be included for analyses (figure 1). The agreement between reviewers was excellent (kappa statistic=0.86).

### Trial characteristics

Among the 111 trials included, 91 (82%) were trials conducted in thrombosis, infectious diseases or oncology. Mortality was the sole primary outcome in 23 (21%)

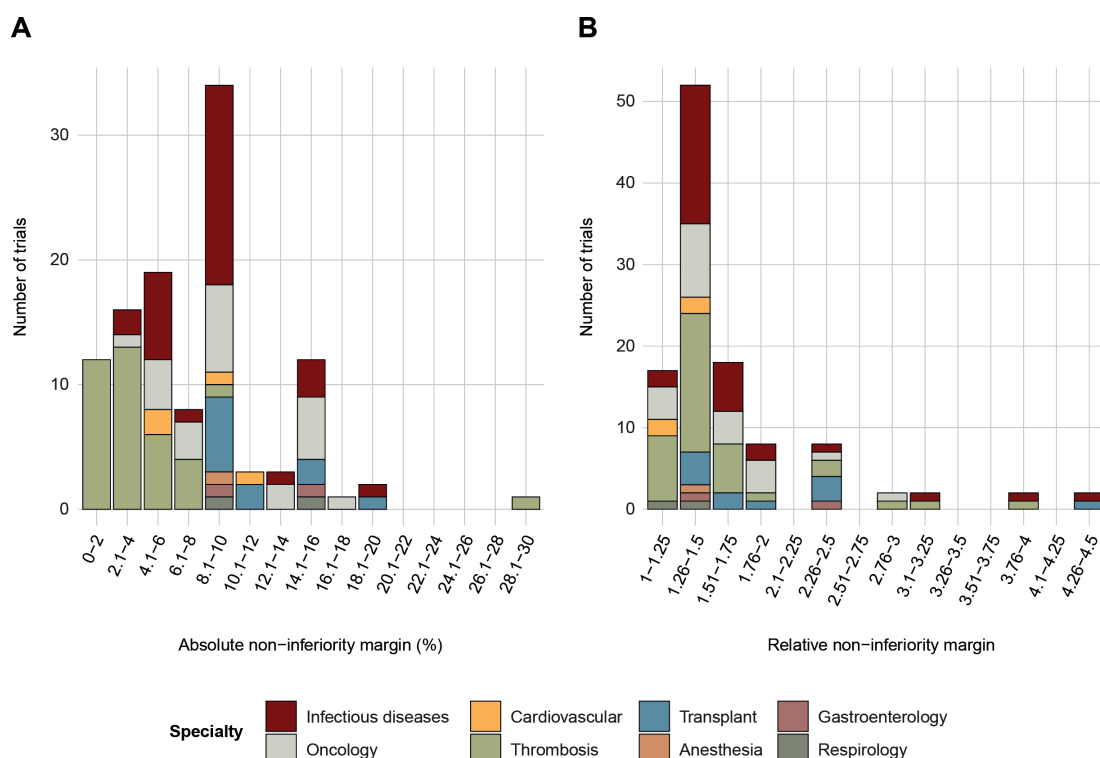
trials, and part of a composite primary outcome in 88 (79%) trials. Over half of the trials disclosed receiving some form of industry funding. Of the included trials, 82 (74%) concluded non-inferiority, 21 (19%) did not conclude non-inferiority and the remaining 8 (7%) were either inconclusive, stopped early or unclear about their conclusions. The non-inferiority margin was expressed as an absolute risk difference in 109 (98%) trials. A summary of the included trials is provided in online supplemental appendix B.

### Association between absolute non-inferiority margins and estimated baseline risks of outcome (involving mortality) in control groups

Figure 2 is a scatterplot between absolute non-inferiority margins and estimated baseline risks of outcome (ie, mortality alone or a composite outcome that included mortality) in the control group for the trials included in this review. A Spearman's correlation shows a moderate, positive monotonic correlation ( $r_s=0.6$ ,  $p<0.001$ ) between the two. Variability in the absolute non-inferiority margins can be seen at both high and low estimates of baseline risks of outcome. There was also a strong correlation between the observed outcomes reported in the trials and the initial estimated risks of outcome in the control groups ( $r_s=0.81$ ,  $p<0.001$ , online supplemental appendix C).

### Distribution of non-inferiority margins for outcomes involving mortality

The distribution of absolute non-inferiority margins subdivided by medical specialty is shown in figure 3A.



**Figure 3** Distribution of absolute and relative non-inferiority margins for primary outcomes involving mortality.



**Table 1** Summary of characteristics of non-inferiority trials included

	n (%)	Absolute non-inferiority margin (%) for outcomes involving mortality			Relative non-inferiority margin for outcomes involving mortality		
		Median (IQR)	Range	P value	Median (IQR)	Range	P value
Overall	111	9 (4.2–10)	0.4–30	-	1.5 (1.3–1.7)	1.1–4.5	-
Medical specialty							
Thrombosis	37 (33.3)	3.6 (2–5)	0.4–30	<0.001	1.4 (1.3–1.7)	1.1–3.9	0.02
Infectious diseases	31 (27.9)	10 (5.5–10)	3–20		1.5 (1.5–1.7)	1.2–4.5	
Oncology	23 (20.7)	10 (7.5–13.8)	4–17.5		1.5 (1.3–1.9)	1.1–3	
Transplant	11 (9.9)	10 (10–13.5)	9–20		1.7 (1.5–2.5)	1.4–4.3	
Cardiology	4 (3.6)	7.5 (5–10.5)	5–12		1.3 (1.2–1.3)	1.1–1.4	
Gastroenterology	2 (1.8)	12.5 (11.3–13.8)	10–15		1.9 (1.6–2.2)	1.3–2.5	
Respirology	2 (1.8)	12.3 (11.1–13.4)	10–14.5		1.3 (1.27–1.3)	1.25–1.32	
Anaesthesia	1 (0.9)	10 (10–10)	10–10		1.4 (1.4–1.4)	1.4–1.4	
Paediatric patients included							
Yes	21 (18.9)	10 (5–10)	3.5–15	0.11	1.7 (1.5–2)	1.2–4.5	0.10
No	85 (76.6)	8 (4–10)	0.7–30		1.5 (1.3–1.7)	1.1–4.3	
Unclear/not explicitly stated	5 (4.5)	5.5 (2.2–5.6)	0.4–10		1.4 (1.1–1.5)	1.1–3.2	
Mortality outcome							
Single	23 (20.7)	10 (5–12.8)	0.4–15	0.24	1.3 (1.2–1.6)	1.1–2.5	0.03
Composite	88 (79.3)	7.8 (4–10)	0.8–30		1.5 (1.4–1.7)	1.1–4.5	
Industry funding							
Yes	61 (55)	5.6 (3.5–10)	0.4–30	0.01	1.5 (1.3–1.7)	1.1–4.5	0.41
No	42 (37.8)	10 (5–10)	0.8–15		1.5 (1.4–1.8)	1.1–4	
Unclear/not explicitly stated	8 (7.2)	10 (10–13.4)	5.5–15		1.4 (1.3–2.7)	1.3–4.3	
Pre-2010 and post-2010 release of draft FDA guidance statement							
Pre-2010	35 (31.5)	7.9 (3.8–10)	0.4–15	0.24	1.4 (1.2–1.7)	1.1–4	0.02
Post-2010	76 (68.5)	9.5 (4.5–10)	0.8–30		1.5 (1.4–1.7)	1.2–4.5	

FDA, Food and Drug Administration.

There was a wide range of non-inferiority margins for trial outcomes that involve mortality (0.4%–30%), with a skewed distribution and distinct peaks observed at 5%, 9% and 15%. Thrombosis trials used smaller non-inferiority margins more commonly than did other trials.

Figure 3B illustrates a similarly skewed distribution of relative non-inferiority margins subdivided by medical specialty. The most common relative non-inferiority margin observed was in the range of 1.26–1.5. Most relative non-inferiority margins clustered in the range of 1.3–1.7, however, there were also many relative non-inferiority margins that were greater than 2.

### Characteristics of non-inferiority margins

The characteristics of the non-inferiority margins in the trials included in this review are summarised in table 1. The median absolute non-inferiority margin was 9%

(IQR 4.2%–10%) and the median relative non-inferiority margin was 1.5 (IQR 1.3–1.7).

The differences in both absolute and relative non-inferiority margins used among medical specialties were significant. Thrombosis trials had the lowest median absolute non-inferiority margin of 3.6%. Although there was a wide range of absolute and relative non-inferiority margins used across trials, the absolute non-inferiority margins of at least one trial in every specialty was 10% or greater.

Trials with mortality as part of a composite primary outcome had significantly higher relative non-inferiority margins compared with those with mortality as a single primary outcome. In contrast, when the non-inferiority margin was expressed as an absolute risk difference, there was no significant difference in the margins between type of mortality outcome. Industry-funded trials had

**Table 2** Absolute non-inferiority margin regression analyses

Predictor	Adjusted $\beta$ coefficient*	95% CI	P value
<b>Specialty</b>			
Oncology	0.12	-0.19 to 0.43	0.45
Cardiovascular	-0.23	-0.84 to 0.39	0.46
Thrombosis	-1.14	-1.48 to -0.8	<0.001
Transplant	0.15	-0.3 to 0.6	0.5
Other†	0.42	-0.13 to 0.98	0.14
Infectious diseases	1 (reference)	-	-
<b>Paediatrics</b>			
Yes	-0.22	-0.54 to 0.11	0.19
No	1 (reference)	-	-
<b>Mortality outcome</b>			
Single	-0.27	-0.56 to 0.02	0.07
Composite	1 (reference)	-	-
<b>Industry funding</b>			
Yes	0.08	-0.19 to 0.36	0.55
No	1 (reference)	-	-

\*Omnibus F-test: 11.93 (8, 102),  $p < 0.05$ , adjusted R squared=0.44.

†Due to low number of trials, 'other' category combines trials in anaesthesia, gastroenterology and respiratory.

a significantly lower median absolute non-inferiority margin compared with those without industry funding.

In this review, 35 (32%) trials were published before 2010 when the first draft FDA guidance statement about non-inferiority trials was published. The relative non-inferiority margin sizes were significantly larger in trials that were published after 2010. A similar trend was seen with the absolute non-inferiority margin sizes, but the difference was not statistically significant.

### Association between trial characteristics and non-inferiority margin size

In our regression analyses of the association between trial characteristics and non-inferiority margin sizes, log-transformation of the non-inferiority margin (outcome variable) resulted in slight improvements to the performance of the regression models. The diagnostic plots of the regression models before and after log-transformation of the absolute and relative non-inferiority margins are provided in online supplemental appendix D. Table 2 shows the  $\beta$  coefficients with 95% CIs from our regression analysis of the association between trial characteristics and the log-transformed absolute non-inferiority margin (adjusted R squared=0.44). Thrombosis trials had significantly smaller log-absolute non-inferiority margins compared with trials in infectious diseases (reference group) when adjusted for paediatric patients, single or composite mortality outcome and industry funding.

When we analysed the same for log-transformed relative non-inferiority margins (table 3), trials related to transplant had significantly larger log-relative non-inferiority margins compared with infectious diseases (adjusted R squared=0.1).

### DISCUSSION

We conducted a systematic review of 111 non-inferiority trials that compared pharmacological therapies where mortality was included in the primary outcome. We found that the majority of non-inferiority trials focused on thrombosis, infectious diseases and oncology. There was a wide range of non-inferiority margins used in these trials, irrespective of whether they were expressed as a measure of absolute effect or when converted to a relative effect. Our results showed that in the design of at least half of the non-inferiority trials included in this review, all of which included mortality as part of their primary outcome, a 'new' drug therapy could have an absolute increase of 9% or relative increase of 50% in mortality outcomes compared with controls and still be accepted as non-inferior. Accepting the potential for this increase in mortality while declaring non-inferiority is a challenging methodological issue in the conduct of non-inferiority trials.

In our review, we also found that non-inferiority margins were more commonly expressed in terms of absolute risk differences than in relative terms. Whether to present absolute or relative non-inferiority margins is a source of

**Table 3** Relative non-inferiority margin regression analyses

Predictor	Adjusted $\beta$ coefficient*	95% CI	P value
<b>Specialty</b>			
Oncology	-0.08	-0.24 to 0.07	0.3
Cardiovascular	-0.23	-0.54 to 0.07	0.13
Thrombosis	0.01	-0.16 to 0.18	0.9
Transplant	0.24	0.01 to 0.46	0.04
Other†	-0.13	-0.41 to 0.15	0.35
Infectious diseases	1 (reference)	-	-
<b>Paediatrics</b>			
Yes	0.1	-0.06 to 0.26	0.23
No	1 (reference)	-	-
<b>Mortality outcome</b>			
Single	-0.11	-0.25 to 0.03	0.13
Composite	1 (reference)	-	-
<b>Industry funding</b>			
Yes	-0.13	-0.27 to 0.01	0.06
No	1 (reference)	-	-

\*Omnibus F-test: 2.56 (8, 102),  $p < 0.05$ , adjusted R squared=0.1.

†Due to low number of trials, 'other' category combines trials in anaesthesia, gastroenterology and respiratory.

debate in the design of non-inferiority trials.<sup>16</sup> There is no clear consensus on the selection of the most appropriate effect measure but it has been demonstrated that different ways of expressing effect measures could result in different conclusions within the same non-inferiority trial.<sup>141617</sup> Since a relative non-inferiority margin accounts for the estimated baseline risk of outcome, it would be a more conservative choice over an absolute margin to conclude non-inferiority should the event rate in the control group be lower than expected.

We detected significant variations in absolute and relative non-inferiority margin size according to medical specialty, which could be partially explained by differences in acuity of diseases, patient age and life expectancy.<sup>18</sup> We also found that industry-funded trials had significantly lower median absolute non-inferiority margins compared with those without industry funding, presumably related to greater financial resources and higher capacity to support larger trials that are necessary when smaller non-inferiority margins are used. However, the difference was not significant when relative non-inferiority margins were compared between trials with and without industry funding.

When we compared non-inferiority margin size between trials published before and after the release of the 2010 draft FDA guidance document on non-inferiority studies, we found that the median non-inferiority margin in trials published after 2010 was increased rather than decreased. This was significant only for relative non-inferiority margins, but not for absolute non-inferiority margins. Perhaps future guidelines could generate reductions in non-inferiority margins used for randomised controlled trials involving mortality, if they recommend margins lower, rather than higher than the current median (<9% absolute mortality, <1.5 times relative mortality).

There is currently limited research in the predefined determinants of the size of non-inferiority margins used in non-inferiority trials. Gayet-Ageron *et al* conducted a survey among trialists to assess the association of predefined trial factors and non-inferiority margins. They found that lower non-inferiority margins were associated with mortality as a primary outcome, low baseline risk and lower costs of new treatments. In contrast, population age group and difficulties with patient recruitment did not appear to affect the choice of margin.<sup>19</sup> Because of the nature of a survey study, these results were based on self-report by respondents and were not necessarily reflective of actual practice when non-inferiority trials are designed and conducted.

In our review of published non-inferiority trials of drug therapies that included mortality as part of their primary outcome, we examined for an association between non-inferiority margin size and medical specialty, inclusion of paediatric patients, mortality as a single or composite outcome and presence of industry funding. Medical specialty was the only trial characteristic found to be significantly associated with the size of non-inferiority margins—specifically, thrombosis trials were associated with smaller absolute non-inferiority margins, while transplant trials were associated with larger relative non-inferiority margins.

Similar to Gayet-Ageron *et al*'s<sup>19</sup> results, we found a significant correlation between the size of absolute non-inferiority margins and estimated baseline risks of outcomes in the control group. While this association was moderate ( $r_s=0.6$ ), it suggests that larger absolute non-inferiority margins are used when estimated risks of outcome occurring in control groups are higher. As can be seen in figure 2, this relationship appears most evident for baseline outcome risks up to approximately 20%, beyond which larger absolute non-inferiority margins are no longer associated with higher baseline risk of outcome.

A strength of our review is the comprehensive and sensitive search for non-inferiority trials which spanned a 30-year period to ensure that virtually all non-inferiority trials with a primary outcome involving mortality would be captured. There were no limits placed on the type of medical specialty or patient population as long as a trial compared mortality between pharmacological therapies. However, we relied on authors to provide the values of non-inferiority margins and estimated risks of outcome in their sample size calculations within the publication or in their online supplemental materials. The accuracy of reporting these variables was taken at face value. To enable standardised comparisons of absolute and relative non-inferiority margins to be made consistently across all trials included in the review, we omitted non-inferiority trials that used HRs, ORs and event rate measures that either changed with time or would not allow us to determine the estimated risk of outcome in the control group required for analyses. Although there was a large amount of variability in the regression models with low adjusted R-squared values, the direction and significance of the independent variables adjusted for in the models indicated that there was an important effect of medical specialty on non-inferiority margin size.

The absolute and relative non-inferiority margins used in published trials comparing medications are large, allowing conclusions of non-inferiority in the context of large differences in mortality, and highly variable. Most trials use non-inferiority margins based on an absolute risk difference, which has only a moderate association with baseline estimates of risk for outcomes. With increasing popularity of non-inferiority trials, clinicians and other users of the medical literature should pay close attention to the size of non-inferiority margins used in these trials and consider the influence of study design parameters and inherent trial characteristics when interpreting the results. A collaborative effort to develop standards for the design and analyses of future non-inferiority trials would be beneficial to the scientific community.

#### Author affiliations

<sup>1</sup>Department of Pharmacy, The Hospital for Sick Children, Toronto, Ontario, Canada

<sup>2</sup>Interdepartmental Division of Critical Care Medicine, University of Toronto, Toronto, Ontario, Canada

<sup>3</sup>Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada

<sup>4</sup>Tory Trauma Program, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

<sup>5</sup>Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

<sup>6</sup>Faculty of Pharmacy, University of Toronto, Toronto, Ontario, Canada

<sup>7</sup>Department of Critical Care Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada

<sup>8</sup>Division of Infectious Diseases, The Hospital for Sick Children, Toronto, Ontario, Canada

<sup>9</sup>Division of Infectious Diseases, Department of Medicine, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

**Contributors** SP, ND and RF drafted the study protocol. SP designed the data collection form, screened titles and abstracts, performed statistical analyses and drafted the study manuscript. SP and MU reviewed full-text articles and performed data collection. ND, MU, RF, NM, WS, JH and MS critically reviewed the final manuscript.

**Funding** This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. SP is supported by a SickKids Clinician-Scientist Training Program Scholarship from The Hospital for Sick Children. MU is supported by a Vanier Canada Graduate Scholarship from the Canadian Institutes of Health Research.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data and statistical code are available on reasonable request to the corresponding author.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

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

## ORCID iDs

Sandra Pong <http://orcid.org/0000-0003-2963-9601>

Nick Daneman <http://orcid.org/0000-0001-8827-3764>

## REFERENCES

- Piaggio G, Elbourne DR, Altman DG, *et al.* Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006;295:1152–60.
- Mauri L, D'Agostino RB. Challenges in the design and interpretation of Noninferiority trials. *N Engl J Med* 2017;377:1357–67.
- Suda KJ, Hurley AM, McKibbin T, *et al.* Publication of noninferiority clinical trials: changes over a 20-year interval. *Pharmacotherapy* 2011;31:833–9.
- Wangge G, Klungel OH, Roes KCB, *et al.* Room for improvement in conducting and reporting non-inferiority randomized controlled trials on drugs: a systematic review. *PLoS One* 2010;5:e13550.
- Schiller P, Burchardi N, Niestroj M, *et al.* Quality of reporting of clinical non-inferiority and equivalence randomised trials—update and extension. *Trials* 2012;13:214.
- Lange S, Freitag G. Choice of delta: requirements and reality—results of a systematic review. *Biom J* 2005;47:12–27.
- Althunian TA, de Boer A, Klungel OH, *et al.* Methods of defining the non-inferiority margin in randomized, double-blind controlled trials: a systematic review. *Trials* 2017;18:107.
- Rehal S, Morris TP, Fielding K, *et al.* Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. *BMJ Open* 2016;6:e012594.
- Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. *Trials* 2011;12:106.
- D'Agostino RB, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues—the encounters of academic consultants in statistics. *Stat Med* 2003;22:169–86.
- Aberegg SK, Hersh AM, Samore MH. Empirical consequences of current recommendations for the design and interpretation of noninferiority trials. *J Gen Intern Med* 2018;33:88–96.
- Hersh AM, Walter RJ, Aberegg SK. Use of mortality as an endpoint in noninferiority trials may lead to ethically problematic conclusions. *J Gen Intern Med* 2019;34:618–23.
- Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. *Non-inferiority clinical trials to establish effectiveness—guidance for industry*. Silver Spring, MD: Food and Drug Administration, 2016. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/non-inferiority-clinical-trials>
- Head SJ, Kaul S, Bogers AJJC, *et al.* Non-inferiority study design: lessons to be learned from cardiovascular trials. *Eur Heart J* 2012;33:1318–24.
- Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.
- Abulizi X, Flandre P. Choice of treatment-effect measures when noninferiority margins originally defined in absolute difference translated into relative difference influenced the results of clinical trials. *J Clin Epidemiol* 2018;96:63–72.
- Althunian TA, de Boer A, Groenwold RHH, *et al.* Defining the noninferiority margin and analysing noninferiority: an overview. *Br J Clin Pharmacol* 2017;83:1636–42.
- Gladstone BP, Vach W. Choice of non-inferiority (NI) margins does not protect against degradation of treatment effects on an average—an observational study of registered and published NI trials. *PLoS One* 2014;9:e103616.
- Gayet-Ageron A, Agoritsas T, Rudaz S, *et al.* The choice of the noninferiority margin in clinical trials was driven by baseline risk, type of primary outcome, and benefits of new treatment. *J Clin Epidemiol* 2015;68:1144–51.