

## Correspondence

**Cite this article:** Rief W, Hofmann SG (2018). The limitations of equivalence and non-inferiority trials. *Psychological Medicine* **49**, 349–350. <https://doi.org/10.1017/S0033291718002891>

Received: 31 August 2018  
Revised: 11 September 2018  
Accepted: 12 September 2018  
First published online: 8 October 2018

**Key words:**

Clinical trials; equivalence trials; non-inferiority trials

**Author for correspondence:**

Winfried Rief, E-mail: [riefw@staff.uni-marburg.de](mailto:riefw@staff.uni-marburg.de)

# The limitations of equivalence and non-inferiority trials

Winfried Rief and Stefan G. Hofmann

Philipps-Universität Marburg, Gutenbergstrasse 18, Marburg 35032, Germany

**Abstract**

Equivalence and non-inferiority trials are becoming more and more popular. Typically, they compare the effects of a treatment of interest with the current gold-standard treatment as the comparator. However, for this approach, the definition of equivalence or non-inferiority margins (NIM) is crucial, and no clear rules for their definition exist. We criticized the practice of these trials of being over-inflationary in favor of (erroneous) equivalence, and we outlined our critique with some study examples comparing psychodynamic treatments with current first-line treatments for mental disorders. Here we answer to a commentary of Leichsenring *et al.* to our paper. Although focusing on our commentary, these authors are less arguing against our conclusions, but they address issues of study conduct, and lack of appreciation of our examples. However, the crucial question is: What is the risk of erroneous equivalence conclusions that we want to accept as responsible clinicians and scientists? We conclude that the scientific community has to define better and clearer criteria for NIMs. We do not believe that it is ethically justifiable to recommend a treatment that is 10 or 20% less effective than the current gold standard interventions.

Equivalence and non-inferiority trials typically compare the effects of a treatment of interest with the current gold-standard treatment as the comparator. For this approach, the definition of equivalence or non-inferiority margins (NIM) is crucial, as we had noted earlier (Rief and Hofmann, 2018). Leichsenring *et al.* (2018) opposed our arguments, based on other examples and their own trials advocating for psychodynamic treatments.

First, we want to thank Leichsenring and colleagues for their thorough report and interest on our paper. The authors present an impressive variety of NIMs that have been used in prior publications, confirming our argument that there is no clear consensus for defining NIMs. We also accept their critique that we did not differentiate between non-inferiority and equivalence trials; our critique applies to both types and does not require this differentiation.

The reasons for erroneous non-inferiority results of two treatments can be manifold. Poor study quality, insensitive ascertainment or statistical analyses procedures, poor implementation of (comparator) treatments are just a few examples, as outlined in our original commentary. Even if Leichsenring *et al.* do not appreciate our examples, they do not provide a convincing argument against these conclusions. We continue to encourage the scientific community to consider the pivotal question with regard to NIMs: What is the risk of erroneous equivalence conclusions that we want to accept as responsible clinicians and scientists in the era of the replicability crisis?

We explicitly reject Leichsenring's notion that we misrepresented the results of the Steinert *et al.* (2017) study. Therefore, we present the original results in Fig. 1 here again. While the authors concluded that results indicate equality between two treatments, we noted that for some variables, significant differences are reported. The overall effect as depicted by the diamond of Fig. 1 clearly indicates a significant disadvantage at post-treatment of psychodynamic treatments compared with other psychological interventions [ $g = -0.153$ ; 90% confidence intervals (CI)  $-0.23$  to  $-0.08$ ], or compared with cognitive behavioral therapy (CBT) ( $g = -0.16$ ; 90% CI  $-0.24$  to  $-0.08$ ; please note: CI are similar if tested for equivalence compared with superiority). These results are of special interest because the financial sponsor was a German psychodynamic society, and the principal investigators are well-known for their advocacy in favor of psychodynamic treatments (despite one CBT-coauthor). This example illustrates one major point we wanted to make: the very same dataset can confirm the superiority of one treatment, and the equivalence of both treatments, depending on the chosen methodological approach. Again, Leichsenring *et al.* do not reject this conclusion, which is mainly built on methodological reasoning independent of this specific example, and which we consider as relevant to be discussed in the scientific community. For more details and critical aspects of non-inferiority/equivalence trials, we refer the reader to our original commentary.

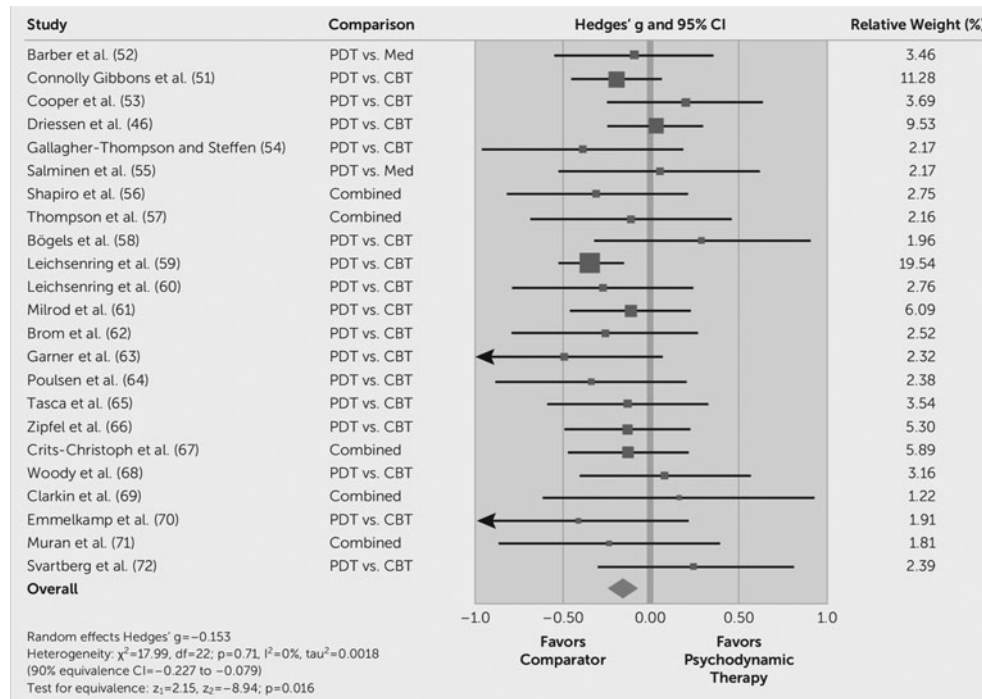


Fig. 1. Superior, inferior, or equivalent? The example of the Steinert *et al.* study.

The scientific community has to define better and clearer criteria for NIMs. We do not believe that it is ethically justifiable to recommend a treatment that is 10 or 20% less effective than the current gold standard interventions. For some of the examples of NIMs summarized by Leichsenring *et al.* the treatments may be even 30 or 50% less effective than comparator treatments, even though authors reported them as 'equally effective treatment alternatives'. This is not acceptable, neither ethically nor scientifically.

**Conflict of interest.** The authors declare no conflict of interest.

## References

- Leichsenring F, Abbass A, Driessen E, Hilsenroth M, Luyten P, Rabung S and Steinert C (2018) Equivalence and non-inferiority testing in psychotherapy research. *Psychological Medicine* **48**, 1917–1919.
- Rief W and Hofmann SG (2018) Some problems with non-inferiority tests in psychotherapy research: psychodynamic therapies as an example. *Psychological Medicine* **48**, 1392–1394.
- Steinert C, Munder T, Rabung S, Hoyer J and Leichsenring F (2017) Psychodynamic therapy: as efficacious as other empirically supported treatments? A meta-analysis testing equivalence of outcomes. *American Journal of Psychiatry* **174**, 943–953.