Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT
forward together
sonke siya phambili
saam vorentoe

Disparity Between Clinical Importance and Statistical Significance in Published Randomized Controlled Trials

Tonya Esterhuizen
Division of Epidemiology and Biostatistics
Faculty of Medicine and Health Sciences
Stellenbosch University
Cape Town
South Africa

4th February 2026

Acknowledgements

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT
forward together
sonke siya phambili
saam vorentoe

This work was conducted as part of the academic requirements of the PhD in Biostatistics, Stellenbosch University

Supervisors Profs L Thabane and L Mbuagbaw, McMaster University

This study is partially funded by the African Centre for Biostatistical Excellence, supported by the Fogarty International Centre of the National Institutes of Health under Award Number D43 TW010547

Open access Original research

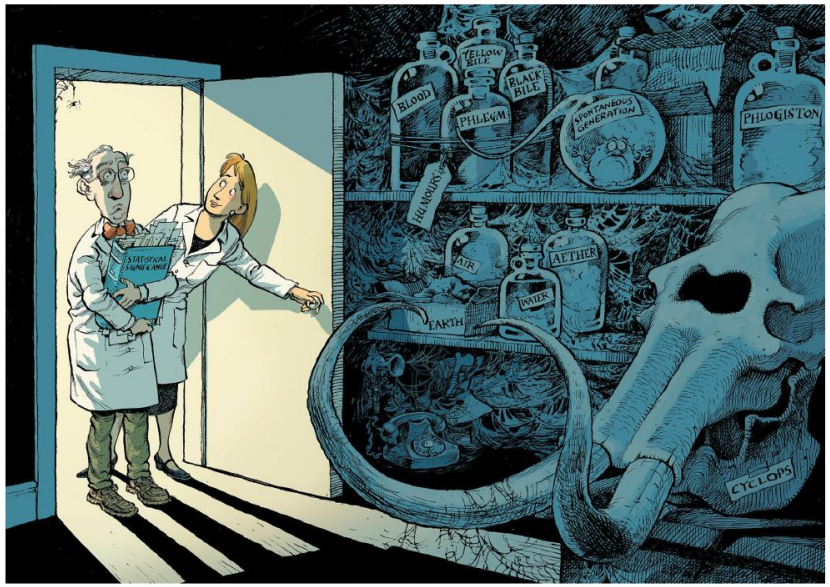
BMJ Open

Disparity between statistical significance and clinical importance in published randomised controlled trials: a methodological study

Tonya Marianne Esterhuizen ,[1] Lawrence Mbuagbaw ,[1,2,3,4] Nadia Rehman ,[2] Nathan Yanwou,[5] Devron J Swaby,[6] Esme Kittle,[7] Johann-Christoph Licht,[7] Lehana Thabane [1,2,3,4]

Introduction

ILLUSTRATION BY DAVID PARKINS



Retire statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

When was the last time you heard a seminar speaker claim there was 'no difference' between two groups because the difference was

How do statistics so often lead scientists to deny differences that those not educated in statistics can plainly see? For several generations, researchers have been warned that a

literature with overstated claims and, less famously, led to claims of conflicts between studies where none exists.

We have some proposals to keep scientists

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

It can be proven that most claimed research findings are false.

should be interpreted based only on *p*-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g.,

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R+1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, α. Assuming that c relationships are being probed in the field, the expected values of the 2×2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al

Disparity between statistical and clinical significance in published RCTs

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT
forward together
sonke siya phambili
saam vorentoe

Background

- Design, analysis and reporting of RCTs lack rigour
- Often analysed using null hypothesis significance testing (NHST) paradigm
- P-value limitations
- Threshold for rejection is arbitrary
- “Absence of evidence is not evidence of absence” and “winner’s curse”

Disparity between statistical and clinical significance in published RCTs

Background continued

- Clinical trials need decisions
- Clinical importance of effect needs to be considered
- Minimum clinically important difference (MCID)
- False positive and negative results (clinical vs statistical significance)

Objectives

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT
forward together
sonke siya phambili
saam vorentoe

1. To examine the prevalence of disparity between statistical and clinical significance in a methodological study of published RCTs
2. To explore factors associated with disparity

Methods

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

forward together
sonke siya phambili
saam vorentoe

Protocol publication

Open access Protocol

BMJ Open

Disparity between statistical and clinical significance in published randomised controlled trials indexed in PubMed: a protocol for a cross-sectional methodological survey

Tonya Marianne Esterhuizen [1], Lawrence Mbuagbaw [1,2,3,4,5,6], Lehana Thabane [1,2,5,7]

To cite: Esterhuizen TM, Mbuagbaw L, Thabane L. Disparity between statistical and clinical significance in published randomised controlled trials indexed in PubMed: a protocol for a cross-sectional methodological survey. *BMJ Open* 2024;**14**:e084375. doi:10.1136/bmjopen-2024-084375

► Prepublication history for this paper is available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2024-084375).

Received 17 January 2024
Accepted 24 June 2024

ABSTRACT

Introduction The commonly used frequentist paradigm of null hypothesis statistics testing with its reliance on the p-value and the corresponding notion of 'statistical significance' has been under ongoing criticism. Misinterpretation and misuse of the p-value have contributed to publication bias, unreliable studies, frequent false positives, fraud and mistrust in results of scientific studies. While p-values themselves are still useful, part of the problem may be the confusion between statistical and clinical significance. In randomised controlled trials of health interventions, this confusion could lead to erroneous conclusions about treatment efficacy, research waste and compromised patient outcomes.
The extent to which clinical and statistical significance of published randomised clinical trials do not match is not known. This is a protocol for a methodological study to understand the extent of the problem of disparities between statistical and clinical significance in published clinical trials, and to identify and assess the factors associated with discrepant results in these studies.

STRENGTHS AND LIMITATIONS OF THIS STUDY

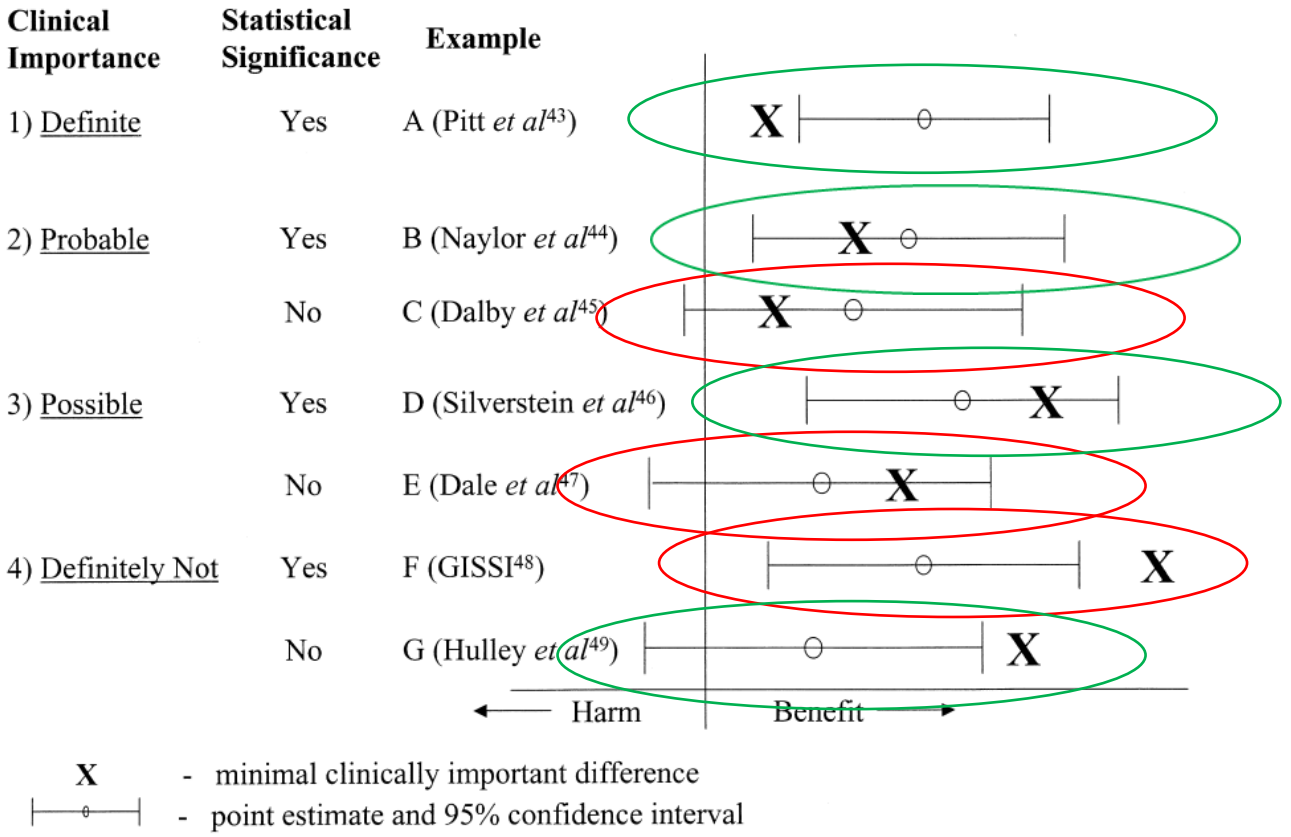
⇒ 500 randomised controlled trials of health interventions in humans from many different journals, across a wide range of disease and intervention conditions will be considered in this study.
⇒ Information required to assess clinical significance or discrepancy between statistical and clinical significance might not be available in some of the studies.
⇒ This study assumes that the minimal clinically important difference reported in the study has been correctly specified.
⇒ The disparity between clinical significance and p-values may be impacted by the scientific discipline of the trial, but we are not distinguishing between disciplines in this study.
⇒ We are searching only one database.

Methods

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT
forward together
sonke siya phambili
saam vorentoe

- Design: Methodological survey of RCTs published between 2018 and 2022, indexed in PubMed.
- Inclusion: two arm, phase three, primary outcome publication, superiority RCTs of health interventions on humans
- Exclusion: protocol publications, pharmacokinetic studies, pilot trials, systematic reviews and duplicates
- Searching and sampling: Medline PubMed, validated search terms, top 300 best matches from each year screened. First 100 eligible studies from each year were selected.
- Data extracted: MCID, sample size, alpha level, effect size, CI, event numbers, p-values, conclusions, impact factor, allocation concealment
- Statistical analysis: assessment of clinical importance and disparity between clinical and statistical significance. Factors associated with disparity – study level multinomial logistic regression analysis

How we assessed clinical importance of trial results and disparity

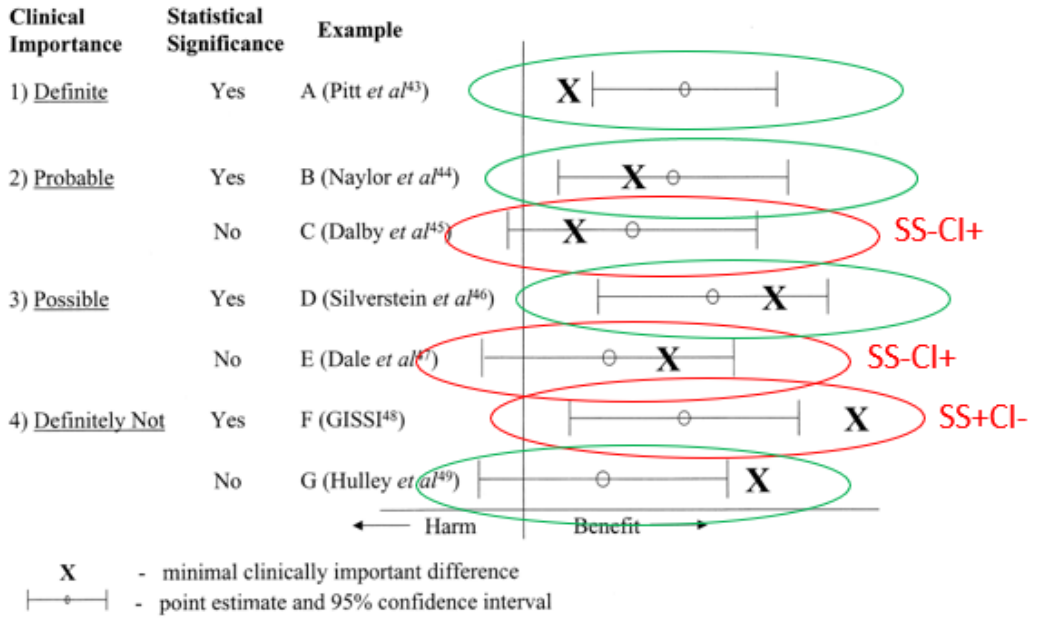


RE 1. Possible combinations for the relationship between clinical importance and statistical significance (see text for details).

M. Man-son-hing *et al.*, "Determination of the Clinical Importance of Study Results: A Review," *JGIM*, vol. 17, pp. 469–476, 2002.

Types of disparity

- The outcome of disparity was a three-level nominal variable:
  - No disparity
  - SS-CI+ (not statistically significant, clinically important)
  - SS+CI- (statistically significant, not clinically important)

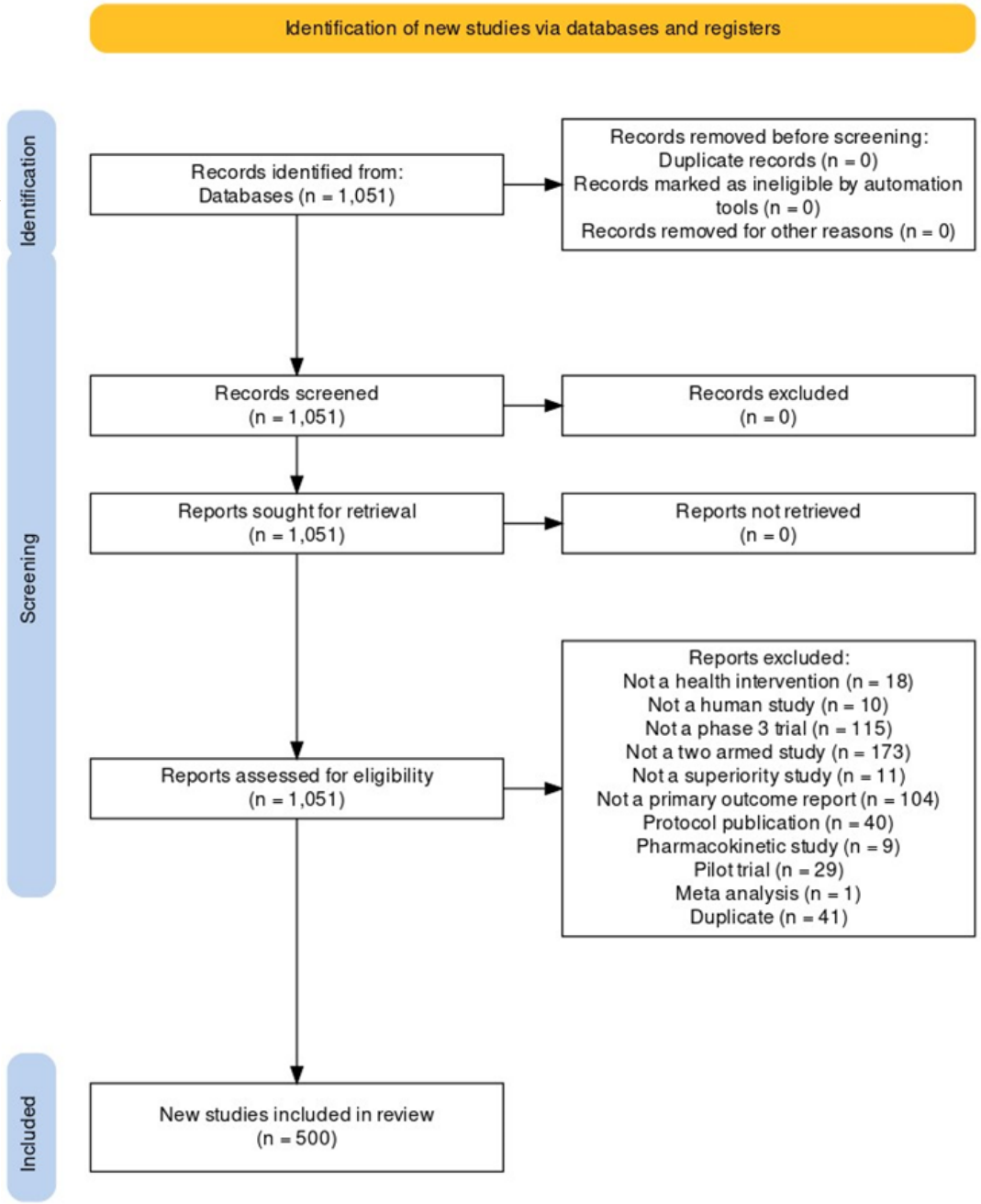


RE 1. Possible combinations for the relationship between clinical importance and statistical significance (see text for details).

M. Man-son-hing *et al.*, "Determination of the Clinical Importance of Study Results: A Review," *JGIM*, vol. 17, pp. 469–476, 2002.

Results

PRISMA diagram: Identification and screening of studies for inclusion



Results

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT
forward together
sonke siya phambili
saam vorentoe

Study characteristics

- 500 studies, median journal impact factor of 5.57 (IQR 3-12).
- Mostly grant funded (39.6%), drug interventions (45%) and placebo controlled (50%).
- Continuous primary outcome variables in 63.2% and binary in 32.6%.
- MCID was reported in 80.8%.
- High methodological quality apart from only 55% having allocation concealment

Results

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT
forward together
sonke siya phambili
saam vorentoe

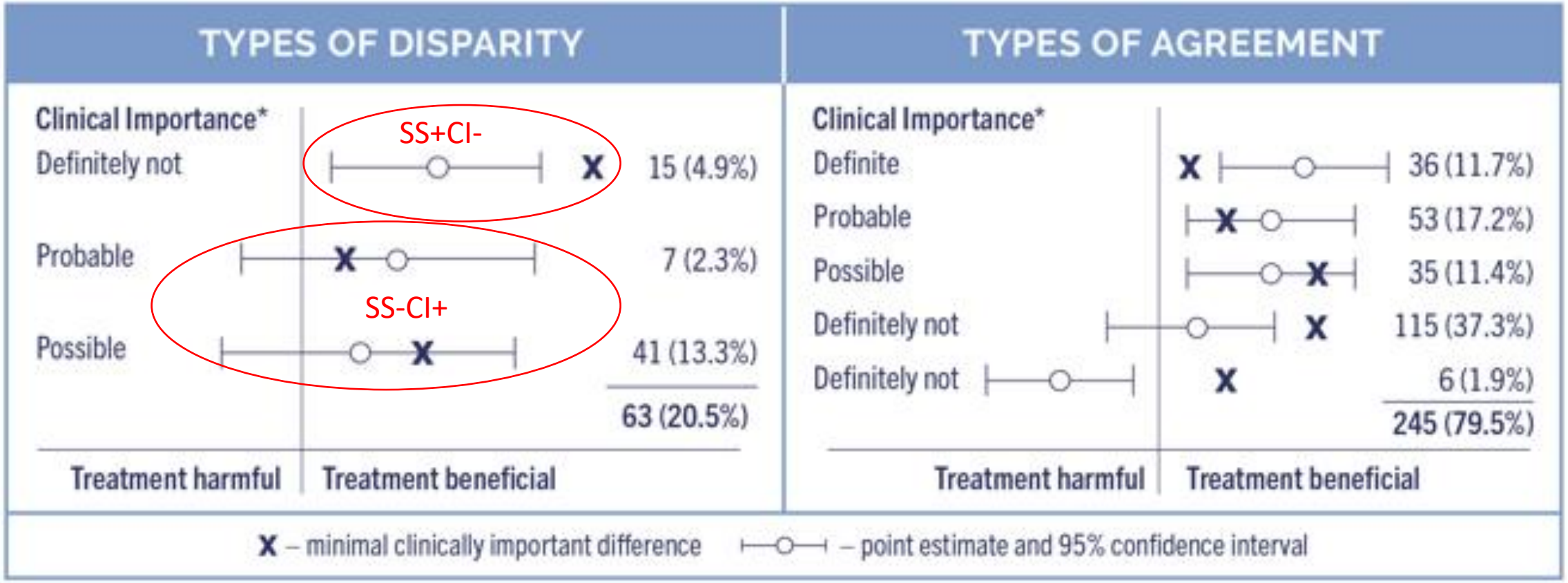
Prevalence of disparity between statistical and clinical significance

Table 2 Classification of clinical importance versus statistical significance in published RCTs, and disparity between clinical importance and statistical significance (n=500)

	Clinical importance[12]					
Statistical significance of primary outcome	**Definite**	**Probable**	**Possible**	**Definitely not**	**Not able to be determined**	**Total**
Yes (in favour of treatment)	36	53	35	15 (SS+CI–)	85	224
Yes (in favour of control)	0	0	0	6	4	10
No	0	7 (SS–CI+)	41 (SS–CI+)	115	96	259
Not able to be determined	0	0	0	0	7	7
Total	36	60	76	136	192	500

Agreement between statistical significance and clinical importance.
Disparity between statistical significance and clinical importance.

Esterhuizen TM, *et al. BMJ Open* 2025;**15**:e100411. doi:10.1136/bmjopen-2025-100411


TYPES OF DISPARITY
Clinical Importance*
Definitely not
15 (4.9%)
SS+CI-
Probable
7 (2.3%)
SS-CI+
Possible
41 (13.3%)
63 (20.5%)
Treatment harmful
Treatment beneficial
TYPES OF AGREEMENT
Clinical Importance*
Definite
36 (11.7%)
Probable
53 (17.2%)
Possible
35 (11.4%)
Definitely not
115 (37.3%)
Definitely not
6 (1.9%)
245 (79.5%)
Treatment harmful
Treatment beneficial
X – minimal clinically important difference
– point estimate and 95% confidence interval

Results

Factors associated with disparity

SS+CI− disparity

- Complementary/ alternative medicines (OR 4.85, 95% CI 1.08 to 21.70).

SS−CI+ disparity.

- Low journal impact factor (OR 3.6, CI 1.50 to 8.62),

- Small sample size (OR 2.75, CI 1.11 to 6.81),
- No funding source stated (OR 4.02, CI 1.16 to 13.85) or grant funding relative to industry funding (OR 3.46, CI 1.14 to 10.56)

- sample size not being the same as the calculated sample size (OR 2.28, CI 1.13 to 4.62)

- failure to mention allocation concealment (OR 2.04, CI 1.04 to 3.98)

Discussion & Conclusion

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT
forward together
sonke siya phambili
saam vorentoe

- We found a relatively high prevalence of disparity between clinical and statistical significance.
- Most disparity was SS-CI+.
- Many studies did not report adequate information to assess clinical importance.
- Stricter adherence to reporting guidelines and discussion of clinical importance should be enforced by all journals.

Implications of the findings

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT
forward together
sonke siya phambili
saam vorentoe

- Focus should be on real-world importance of results rather than statistical significance – update of reporting guidelines to include this
- Justification of chosen MCID. More research on specific MCIDs for outcomes.
- Transparent reporting of results and avoidance of “spin”
- Better statistical education of researchers to use statistical thinking
- Implications for GRADE in systematic reviews or guideline development?

Limitations

- We excluded studies where we were not able to asses clinical importance from the analysis resulting in a lower precision in estimates than planned.
- Assumption that delta value used in sample size calculation for the RCT was the MCID. MCIDs might not be known for some outcomes.

Acknowledgements

- Supervisors Profs Lehana Thabane and Lawrence Mbuagbaw
- Mrs Alvina Matthee, librarian at Faculty of Medicine and Health Sciences, Stellenbosch University.

Selected key references

- G. Li, S. D. Walter, and L. Thabane, “Shifting the focus away from binary thinking of statistical significance and towards education for key stakeholders: revisiting the debate on whether it’s time to de-emphasize or get rid of statistical significance,” *J. Clin. Epidemiol.*, vol. 137, pp. 104–112, 2021, doi: 10.1016/j.jclinepi.2021.03.033
- G. M. Allan *et al.*, “Are potentially clinically meaningful benefits misinterpreted in cardiovascular randomized trials ? A systematic examination of statistical significance , clinical significance , and authors ’ conclusions,” *BMC Med.*, vol. 15, no. 58, pp. 1–10, 2017, doi: 10.1186/s12916-017-0821-9.
- R. L. Wasserstein and N. A. Lazar, “The ASA’s Statement on *p* -Values: Context, Process, and Purpose,” *Am. Stat.*, vol. 70, no. 2, pp. 129–133, Apr. 2016, doi: 10.1080/00031305.2016.1154108.
- J. A. C. Sterne and G. D. Smith, “Sifting the evidence-what’s wrong with significance tests?,” *BMJ*, vol. 322, pp. 226–230, 2001.
- J. P. A. Ioannidis, “Why most published research findings are false,” *PLoS Med.*, vol. 2, no. 8, pp. 2–8, 2005, doi: 10.1371/journal.pmed.0020124.
- R. Jaeschke, J. Singer, and G. H. Guyatt, “Measurement of health status. Ascertaining the minimal clinically important difference,” *Control. Clin. Trials*, vol. 10, no. 4, pp. 407–415, 1989, doi: 10.1016/0197-2456(89)90005-6.
- K. B. Y. Chan, M. Man-Son-Hing, F. J. Molnar, and A. Laupacis, “How well is the clinical importance of study results reported? An assessment of randomized controlled trials,” *C. Can. Med. Assoc. J.*, vol. 165, no. 9, pp. 1197–1202, 2001.
- M. Man-son-hing *et al.*, “Determination of the Clinical Importance of Study Results A Review,” *JGIM*, vol. 17, pp. 469–476, 2002.

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

forward together
sonke siya phambili
saam vorentoe

Thank you
Enkosi
Dankie

Photo by Stefan Els

Medicine and Health Sciences| EyeNzululwazi ngezoNyango neMpilo| Geneeskunde en Gesondheidswetenskappe