

PRIMARY CARE INSIGHTS

Effect Size, P Values, and More in Clinical Trials

Clinicians may encounter *P* values, effect sizes, and other statistical metrics in prescribing information, clinical guidelines, and educational content. It is important to have a good understanding of what each value does and does not mean.¹

-  **P value**
-  **Effect Size**
-  **NNT/NNH**

	What it does tell you...	What it does not tell you...	Quick Check
P value	Compares the observed data to the null hypothesis (typically, no change) ² A <i>P</i> value of 0.05 means there is a 5% chance that the null hypothesis is correct; in other words, there is a 95% chance that there is a difference ²	The magnitude or clinical relevance of the effect for patients ³	→ Interpret alongside effect size
Effect Size	Measures the magnitude of the difference between two treatment groups ⁴	The probability that the difference is statistically significant ⁵	→ Consider within the context of the MCID
MCID	Minimum change on a symptom scale score that is considered clinically perceptible ⁶	Whether an individual patient will notice a substantial change ⁶	→ Look at the NNT for patient-level context of the potential benefit
NNT	An estimate of the required number of patients receiving 1 treatment (instead of an alternative) before you can expect 1 additional patient to benefit; NNT is an effect size ^{7,9}	How a given benefit weighs against potential adverse effects or other risks ⁷	→ Confirm the time frame and outcome assessed
NNH	An estimate of the required number of patients receiving 1 treatment (instead of an alternative) before you can expect 1 additional patient to experience the adverse event; NNH is an effect size ^{8,9}	How a given risk weighs against the potential treatment benefit ⁸	→ Compare with the NNT for the same treatment
LHH	The likelihood of experiencing the benefit (NNT) vs the harm (NNH); LHH is a ratio ⁹	If the measured outcomes are relevant to the patient ⁹	→ Assess which outcomes are being contrasted

LHH, likelihood of being helped or harmed; MCID, minimal clinically important difference; NNH, number needed to harm; NNT, number needed to treat.

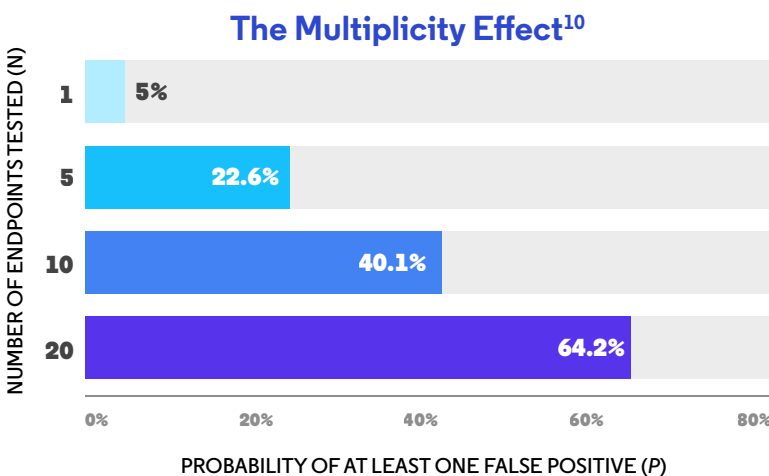
P Values and Multiplicity

P values depend on the sample size and variability in the data.^{1,3}

When multiple secondary endpoints are tested within the same trial without adjustment for multiple comparisons, the probability of at least one false-positive result increases with each additional test. In such instances, **you may conclude that a real effect or difference exists when it actually does not,** a phenomenon known as type I error inflation.¹¹

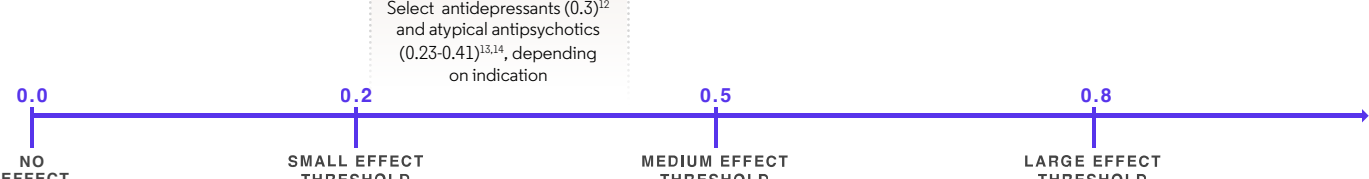
With five endpoints, the probability of at least one false-positive result is 22.6%—more than quadruple the 5% probability associated with a single endpoint.¹⁰

Secondary endpoints reported without multiplicity adjustment should be interpreted as exploratory rather than confirmatory.¹⁰



Effect Size (Cohen’s d) to Quantify Magnitude

Cohen’s *d* is commonly used with three conventional thresholds^{3,4}



Cohen’s *d*, a measure of standardized mean difference, represents the treatment-versus-placebo difference in standard deviation units, helping to facilitate comparisons across different measures and studies.⁴

Meaningful Change and Progress Assessment

A statistically significant difference may still fall below the threshold for clinically meaningful change, termed minimal clinically important difference (MCID).

The MCID represents the minimum change in score on a validated rating scale that is considered clinically perceptible. Published MCID values can help clinicians interpret clinical outcomes, although reported values may vary depending on the method used to derive them.⁶

Placebo-Adjusted MCID Values for Selected Rating Scales

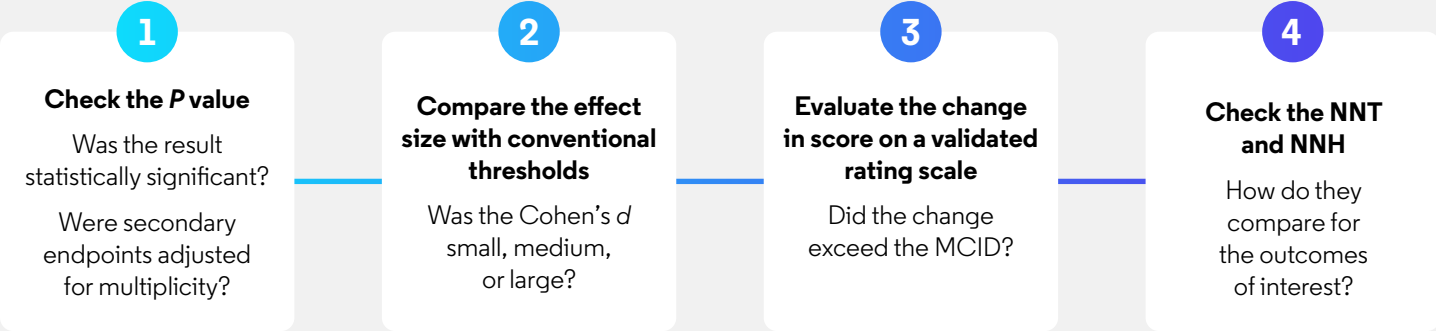
Scale	Placebo-Adjusted MCID
MADRS ^{15,16}	≥2
HAM-D ⁵	≈3

 The Montgomery-Åsberg Depression Rating Scale (MADRS) and Hamilton Depression Rating Scale (HAM-D) are clinician-rated scales frequently used in MDD clinical trials.^{17,18}

 A trial may report a statistically significant drug–placebo difference that does not meet the MCID threshold, underscoring the need to interpret multiple metrics collectively rather than relying on any single variable to determine treatment effectiveness.¹

How to Read Clinical Trial Statistics in Less Than One Minute

These skills help clinicians critically evaluate treatment evidence and translate it into a patient-facing context.^{2,4,6-8}



MCID, minimal clinically important difference; NNH, number needed to harm; NNT, number needed to treat.

References

1. Hengartner MP, Plöderl M. Statistically significant antidepressant-placebo differences on subjective symptom-rating scales do not prove that the drugs work: effect size and method bias matter. *Front Psychiatry*. 2018;9:517. doi:10.3389/fpsy.2018.00517

2. Wasserstein RL, Lazar NA. The ASA’s statement on p-values: context, process, and purpose. *Am Stat*. 2016;70(2):129-133. doi:10.1080/000313105.2016.1154108

3. Sullivan GM, Feinn R. Using effect size— or why the *P* value is not enough. *J Grad Med Educ*. 2012;4(3):279-282. doi:10.4300/JGME-D-12-00156.1

4. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Lawrence Erlbaum Associates; 1988.

5. Hengartner MP, Plöderl M. Estimates of the minimal important difference to evaluate the clinical significance of antidepressants in the acute treatment of moderate-to-severe depression. *BMJ Evid Based Med*. 2022;27(2):69-73. doi:10.1136/bmjebm-2020-111600

6. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407-415. doi:10.1016/0197-2456(89)90005-6

7. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med*. 1988;318(26):1728-1733. doi:10.1056/NEJM198806303182605

8. Citrome L. Relative vs absolute measures of benefit and risk: what’s the difference? *Acta Psychiatr Scand*. 2010;121(2):94-102. doi:10.1111/j.1600-0447.2009.01449.x

9. Citrome L, Ketter TA. When does a difference make a difference? Interpretation of number needed to treat, number needed to harm, and likelihood to be helped or harmed. *Int J Clin Pract*. 2013;67(5):407-411. doi:10.1111/jcnp.12142

10. Bender R, Lange S. Adjusting for multiple testing—when and how? *J Clin Epidemiol*. 2001;54(4):343-349. doi:10.1016/s0895-4356(00)00314-0

11. Laporte S, Divin M, Girault D, et al. Clinical research and methodology: what usage and what hierarchical order for secondary endpoints? *Therapie*. 2016;71(1):27-41. doi:10.1016/j.therap.2016.01.002

12. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet*. 2018;391(10128):1357-1366. doi:10.1016/S0140-6736(17)32802-7

13. Yildiz A, Sifakis S, Mavridis D, Vieta E, Leucht S. Comparative efficacy and tolerability of pharmacological interventions for acute bipolar depression in adults: a systematic review and network meta-analysis. *Lancet Psychiatry*. 2023;10(9):693-705. doi:10.1016/S2215-0366(23)00199-2

14. Wang J, Li W, Li M, Wu H, Qiu Z. Comparative efficacy and safety of 4 atypical antipsychotics augmentation treatment for major depressive disorder in adults: a systematic review and network meta-analysis. *Medicine (Baltimore)*. 2023;102(38):e34670. doi:10.1097/MD.00000000000034670

15. McIntyre RS, Stahl SM, Shim SR, et al. Adjunctive antipsychotics in major depressive disorder: a systematic review and network meta-analysis. *JAMA Psychiatry*. 2026;83(7):741-750. doi:10.1001/jamapsychiatry.2026.0658

16. Montgomery SA, Möller HJ. Is the significant superiority of escitalopram compared with other antidepressants clinically relevant? *Int Clin Psychopharmacol*. 2009;24(3):111-118. doi:10.1097/YIC.0b013e32832a8eb2

17. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134:382-389. doi:10.1192/bjp.134.4.382

18. Hamilton M. A rating scale for depression. *J Neuro Neurosurg Psychiatry*. 1960;23(1):56-62. doi:10.1136/jnnp.23.1.56

This content has been created for US health care professionals and is intended for information and educational purposes only. It should not be used to replace clinical judgment or a discussion with the patient’s health care team. All decisions regarding patient diagnosis and care must be made by a qualified health care professional based on each patient’s individual needs.

Referenced tools are adjunctive aids and are not diagnostic on their own, and this resource does not endorse or recommend any specific drug, treatment, or manufacturer.