

## Standard deviation versus standard error

*Last updated: 6 February 2025*

### Standard deviation

**Standard deviation (SD)** is a statistical measure that describes the variability of numerical observations of a variable around the mean of these observations.

When an entire population is measured, the population standard deviation is calculated  $\sigma$  is calculated as

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}}$$

where  $x_i$  is the value of each observation,  $\bar{x}$  is the mean of each observation ( $\bar{x} = \sum x_i / N$ ) and  $N$  is the number of observations (i.e., the sample size). A lower standard deviation indicates that the data is more closely clustered around its mean, whereas a higher standard deviation indicates that the data is more spread out from its mean.

Note that standard deviation is calculated differently depending on whether you are measuring standard deviation for a population versus just a sample from that population. When only a sample is taken from a population, the sample standard deviation  $s$  is calculated as

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N - 1}}.$$

See [this explanation by Dr. Paul Savory](#) for why this correction is made. Most research does not survey a whole population, and thus “standard deviation” usually denotes sample standard deviation  $s$  and not population standard deviation  $\sigma$ .

Imagine that we took a sample of ten individuals and measured their height in centimeters, obtaining ten observations: 176, 171, 159, 168, 185, 193, 174, 171, 168 and 189. The mean of this sample would be 175.4 and the sample standard deviation would be approximately 10.6.

### Standard error

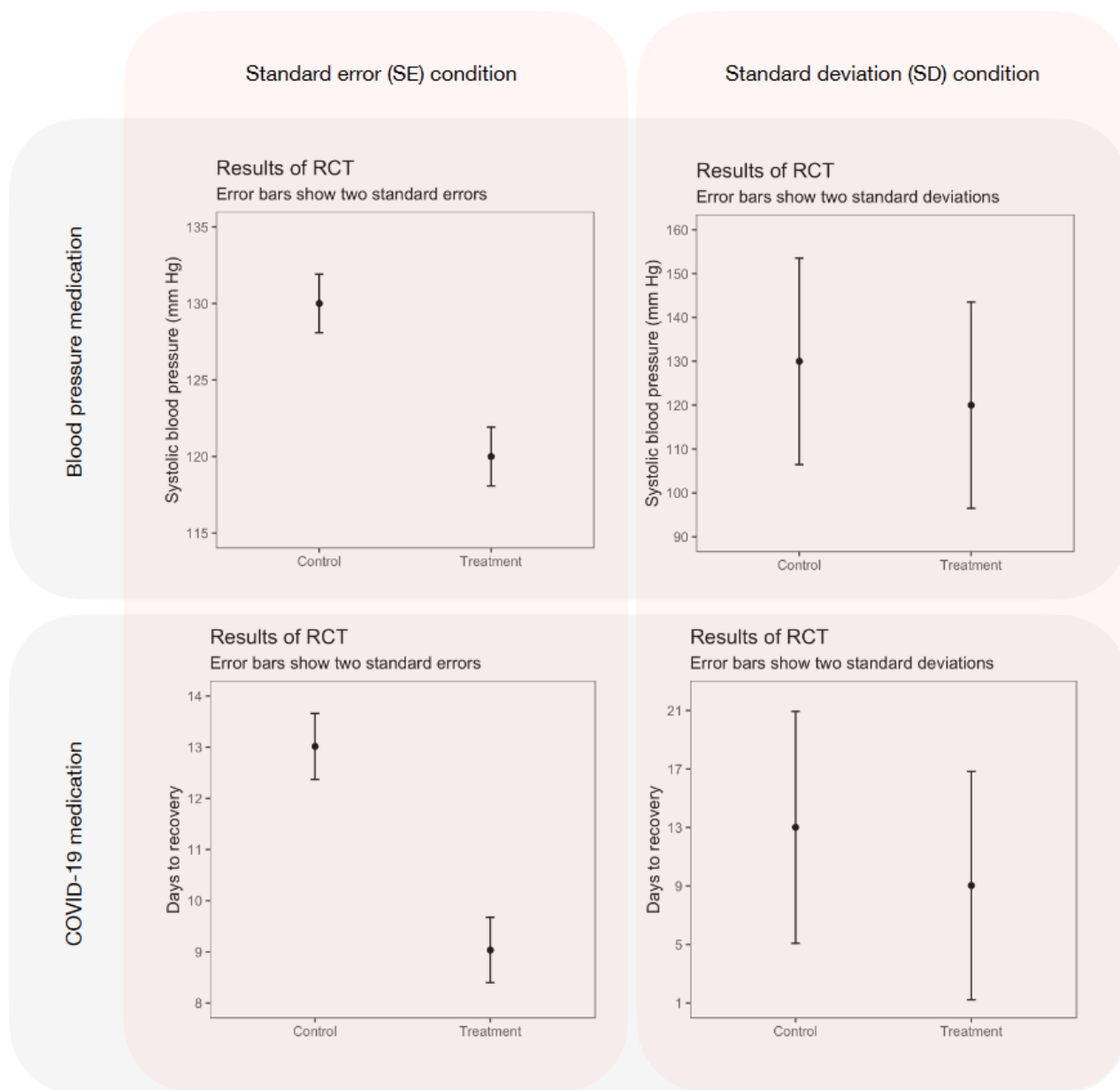
**Standard error (SE)** is a measure of the variability of an estimate. For instance, for the previous scenario, a different sample of ten individuals would likely yield a slightly different mean height. The standard error  $SE$  is calculated as

$$SE = \frac{s}{\sqrt{N}}.$$

Note that as sample size  $N$  increases, the standard error decreases. For instance, the mean heights calculated from two different samples of 100 from the same population would likely be closer together than the mean heights calculated from two different samples of only 10. This is why studies with a larger sample size are generally considered more rigorous; their estimates will be more precise. For the previously-listed sample of heights, the standard error of the mean (SEM) would be  $10.6/\sqrt{10} \approx 3.3$ .

## **Reporting of standard deviation versus standard error**

When reading a scientific publication, one should take note of whether the authors are reporting variability in their data as standard deviation or standard error. Misidentifying which measure is being reported can give the reader an unrealistic picture of the variability in data and the uncertainty of estimates. There is some evidence that using one measure or the other graphically can lead to this same misinterpretation, even when the reader is aware of which measure is being used (see [Zhang et al., 2023](#)).



Within each row, both charts demonstrate the same hypothetical data from a randomized controlled trial (for blood pressure medication, top row, or COVID-19 medication, bottom row). However, the left charts use error bars to display standard error of the mean and the right charts use error bars to display standard deviation. Adapted from Figure S1 of [Zhang et al. \(2023\)](#).

## Incorporating variability into a meta-analysis

A [meta-analysis](#) is a study that synthesizes the results of independent studies on the same topic to obtain more precise estimates. For instance, one might perform a meta-analysis of randomized controlled trials that all tested the same drug to more precisely determine how effective the drug is at preventing intensive care unit admission or mortality due to COVID-19.

Because larger studies give more precise estimates, they are usually given more weight in meta-analyses so that their findings contribute more to the combined outcome estimate. The most common method for weighting studies for meta-analysis is [inverse variance weighting](#). With inverse variance weighting, each study is weighted by

$$\frac{1}{SE^2}$$

where  $SE$  is the standard error of that study's estimate. Thus, studies that yield a smaller standard error have a greater weight. Software for performing meta-analyses like [RevMan](#) typically allow the user to enter estimates from included studies alongside the reported standard deviation or standard error in those estimates, automatically performing weighting.

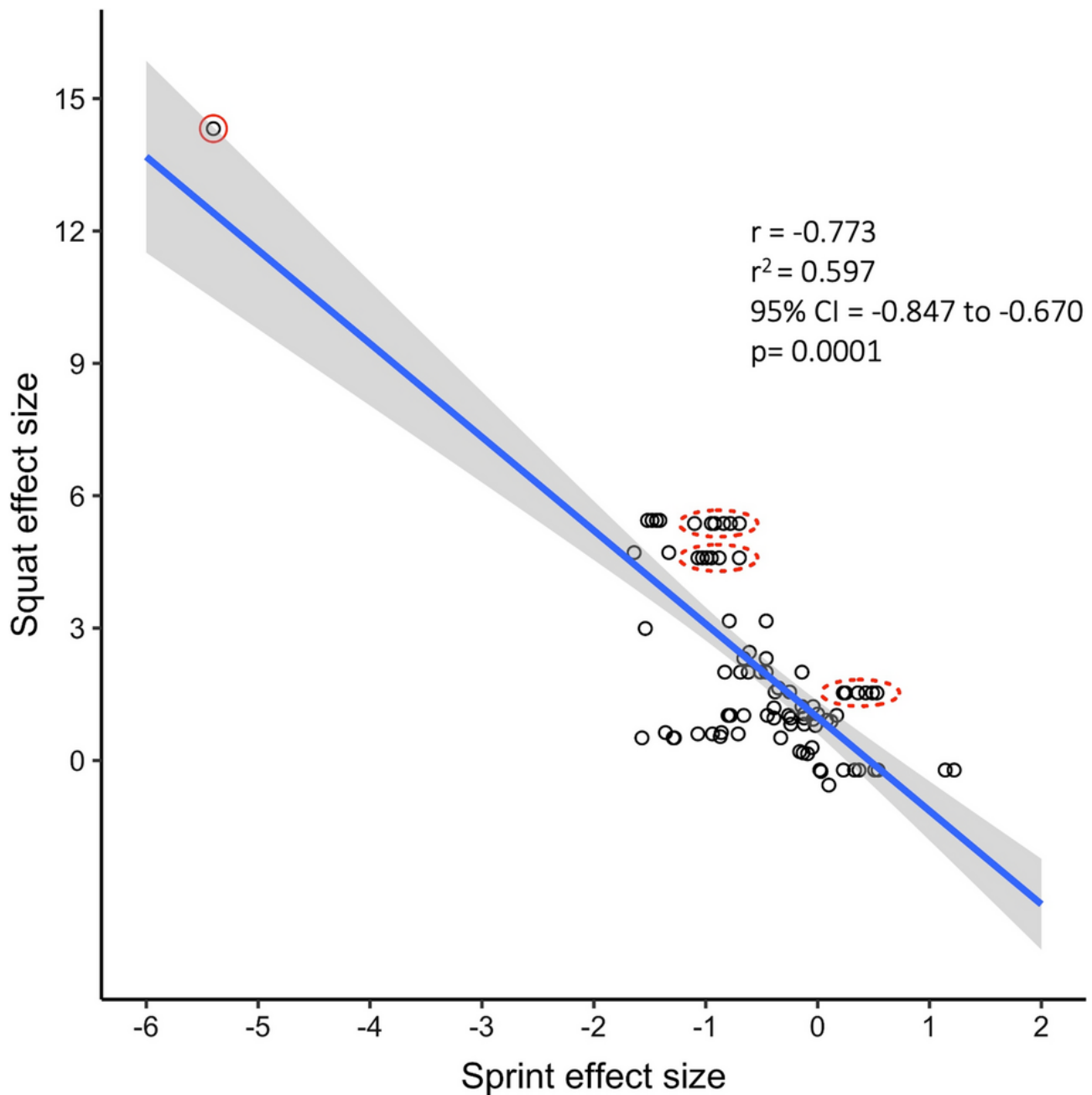
It is critical that those conducting a meta-analysis are cognizant of exactly which measure they are entering into the software. For instance, if the meta-analysis software expects the user to enter means, sample sizes and standard deviations, but the user misreads one study and enters the standard error reported by that study into the software instead of the standard deviation, that study will be weighted much more than it should be in the meta-analysis.

Popular methods for estimating the effect size of studies measuring continuous variables, such as [standardized mean difference \(SMD\)](#) / [Hedges'  \$g\$](#) , will also yield inaccurate results if standard errors are confused for standard deviations.

Confusion of standard errors with standard deviations can often be spotted in a meta-analysis just by looking for outliers. Consider looking closer if one included study appears to have a much larger effect size than the others included in the meta-analysis or if one included study appears to have much smaller standard deviation in outcome than the other included studies.

### Example 1: Overestimation of effect size in a meta-analysis due to using standard error instead of standard deviation

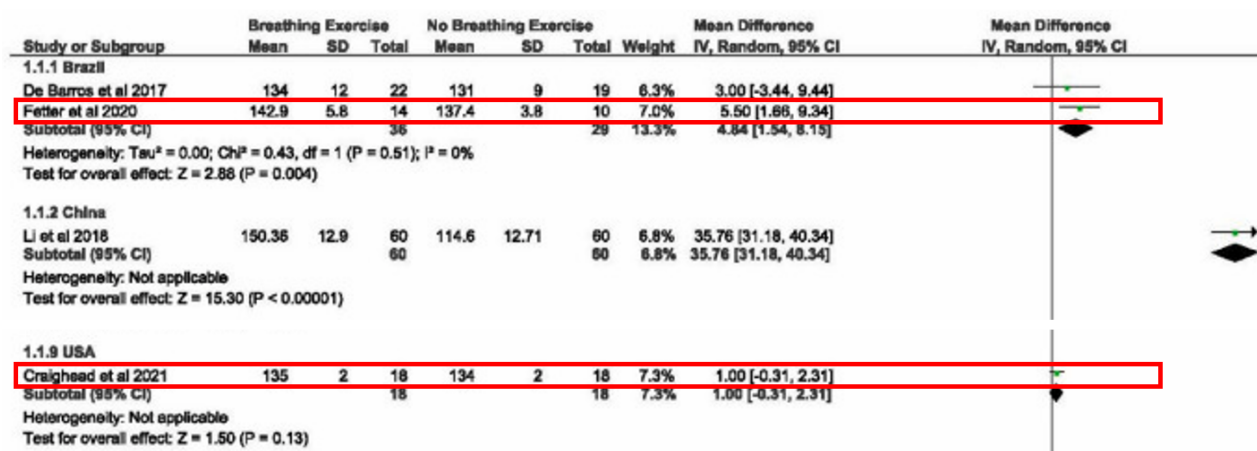
[Seitz et al. \(2014\)](#) performed a meta-analysis to determine whether exercises that increase lower body strength also improve sprinting performance. However, for three of their included studies, they calculated the effect size (in the form of Hedge's  $g$ ) using standard errors instead of standard deviations. As a result, the effect sizes were over-estimated for these studies, leading to an over-estimation of the overall effect of lower body strengthening on improvement in sprinting performance. The effects of this over-estimation were discussed in detail by [Kadlec et al. \(2022\)](#).



The most severe over-estimation of effect size made by [Seitz et al. \(2014\)](#) was for [Wong et al. \(2010\)](#) (solid red circle in the upper right corner of this plot). [Kadlec et al. \(2022\)](#) corrected the statistical errors made by Seitz et al., finding that Seitz et al. had dramatically overestimated the correlation between lower body strength training and improvement in sprinting performance as a result of these errors. Adapted from Figure 1 of Kadlec et al.

## Example 2: Over-weighting studies in a meta-analysis due to using standard error instead of standard deviation

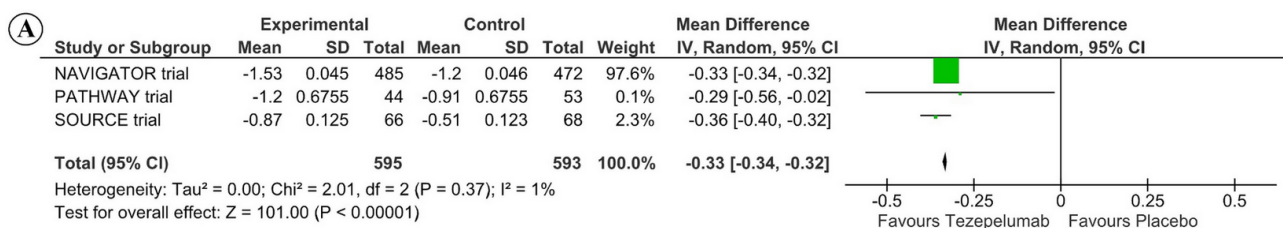
Garg et al. (2024) performed a meta-analysis to assess whether the systolic blood pressure was different for groups performing breathing exercises versus not performing breathing exercises at baseline (i.e. before any intervention was applied; note that there is no logical reason to perform this comparison, as it gives no information on the effectiveness of the intervention, but this meta-analysis has numerous issues that are beyond the scope of this guide). For at least two studies, Fetter et al. (2020) and Craighead et al. (2021), the authors include the standard error as reported by these studies instead of the standard deviation, resulting in these studies being weighted far too heavily.



Garg et al. (2024) overweighted Fetter et al. (2020) and Craighead et al. (2021) (highlighted in their forest plot in red boxes) as a result of confusing standard error for standard deviation. As a result, Fetter et al. and Craighead et al. are each weighted greater than Li et al. (2018) which had three to four times as many participants. Adapted from Figure 2 of Garg et al.

## Example 3: Over-weighting studies in a meta-analysis due to using standard error instead of standard deviation

Chagas et al. (2023) performed a meta-analysis to determine whether treatment with the drug tezepelumab reduced asthma patients' scores on the Asthma Control Questionnaire-6 (ACQ-6). For the NAVIGATOR trial, the authors entered the reported standard error as the standard deviation, leading to this trial being weighted to 97.6% for this outcome (meaning that this portion of the meta-analysis was almost entirely based on the results of the NAVIGATOR trial).



*Chagas et al. (2023)* overweighted the *NAVIGATOR trial* as a result of confusing standard error for standard deviation. Note that the “standard deviation” reported in the above forest plot is much lower for the *NAVIGATOR trial* than the *PATHWAY trial* or the *SOURCE trial*. Adapted from Figure 2 of *Chagas et al.*

## Additional resources

- “Points of significance: Error bars” (2013)
- “The Standard Error/Standard Deviation Mix-Up: Potential Impacts on Meta-Analyses in Sports Medicine” (2024)
- “With Great Power Comes Great Responsibility: Common Errors in Meta-Analyses and Meta-Regressions in Strength & Conditioning Research” (2022)
- *The Cochrane Handbook for Systematic Reviews of Interventions*
- *The Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*