

Statistical guidelines

All statistical analyses used must be described comprehensively in the paper. The Methods must include a statistics section with the following information.

Data sets should be summarized with descriptive statistics, which should include the n value for each data set, a clearly labelled measure of centre (such as the mean or the median), and a clearly labelled measure of variability (such as standard deviation or range). Ranges are more appropriate than standard deviations or standard errors for small data sets. Graphs should include clearly labelled error bars. Authors must state whether a number that follows the \pm sign is a standard error (s.e.m.) or a standard deviation (s.d.).

When plotting data, we strongly encourage authors to use violin plots, box-and-whisker plots, or dot plots in order to fully show data distribution. If using a bar chart, please note that individual data points must be overlaid on the bars.

Every article that contains statistical testing should state the name of the statistical test, the n value for each statistical analysis, the comparisons of interest, and a justification for the use of that test (including, for example, a discussion of the normality of the data when the test is appropriate only for normal data), the alpha level for all tests, whether the tests were one-tailed or two-tailed, and the actual P value for each test (unless $p < 0.001$).

All statements or interpretations of the results should be supported by appropriate, fully reported statistics. In detail, frequentist inferential statistics should be reported as follows wherever they occur (main text, Figure captions, Tables, SI): statistics (degrees of freedom) = value, p = value, effect size statistic = value, % Confidence Intervals = values.

Comparisons of relationships between variables – for example, interpretations pertaining to a difference in differences – must also be supported by appropriate statistics, rather than rhetoric comparisons (see Gelman, A. & Stern, H. [The difference between “significant” and “not significant” is not itself statistically significant](#). *Am. Stat.* 60, 328–331 (2006)).

Specific policies apply to the interpretation of null results:

To be able to interpret null results, there must be statistical evidence that the study is sufficiently powered to detect the smallest theoretically or pragmatically meaningful effect. It is not permissible to interpret the absence of evidence as evidence of absence (in the case of null results). Support for the null hypothesis over the alternative hypothesis cannot therefore be inferred through a null finding using NHST. Instead, appropriate statistical tests (e.g., Bayes Factors or equivalence tests) must form the basis of any interpretation. Statements such as ‘There is no difference between x and y .’ or ‘ X does not affect Y .’ must be revised to read ‘We found [no/little] credible evidence of a difference between x and y .’ or ‘We found [no/little] credible evidence that X affects Y .’

Marginally significant statistical results can be mentioned as such, but should not be discussed as theoretically informative for the hypotheses tested.