

Statistical methods, data access, and reproducibility checklist

Dissect a familiar research article in terms of statistical analysis and reporting rigor.

Based on MBoC Author Submission Checklist (Updated August 4, 2016)

I. Data Presentation

Ensure that the following information is available in all relevant figure legends (or has been placed in the Materials and Methods section to avoid excessively long legends):

1. The exact sample size (n) for each experimental group/condition, given as a number, not a range;
2. A description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.);
3. A statement of how many times the experiment shown was replicated in the laboratory;
4. Definitions of statistical methods and measures:
 - a. Data from small samples ($n < 5$), for which descriptive statistics are not appropriate, have been plotted as individual points.
 - b. Complex techniques are described in the Materials and Methods section (common tests, such as t test, simple χ^2 tests, Wilcoxon and Mann-Whitney tests can be unambiguously identified by name only).
 - c. Tests are identified as one-sided or two-sided.
 - d. There are adjustments for multiple comparisons.
 - e. Statistical test results, e.g., P values, are given.
 - f. “Center values” are defined as median or mean.
 - g. Error bars are defined as s.d. or s.e.m. or c.i.
 - h. Requirement and description of adjustment for multiple comparisons are provided.
 - i. Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated.

II. Methodology and Statistics

Ensure that, where appropriate, the following information is included in the Materials and Methods section:

1. How the sample size was chosen to ensure adequate power to detect a pre-specified effect size;
2. Inclusion/exclusion criteria if samples or animals were excluded from the analysis;
3. Description of a method of randomization to determine how samples/animals were allocated to experimental groups and processed;
4. The extent of blinding if the investigator was blinded to the group allocation during the experiment and/or when assessing the outcome;
5. Justification for statistical tests that address the following questions (as appropriate):
 - a. Do the data meet the assumptions of the tests (e.g., normal distribution)?
 - b. Is there an estimate of variation within each group of data?
 - c. Is the variance similar between the groups that are being statistically compared?
6. If we assume that the raw experimental data are available, are there enough details for you to be able to perform the statistical/data analysis and completely reproduce the result presented in the paper?

III. Data Accessibility

Ensure the following are true about the data associated with the paper:

1. Accession codes are provided for deposited data. Data deposition in a public repository is mandatory for:
 - a. Protein, DNA, and RNA sequences
 - b. Macromolecular structures
 - c. Crystallographic data for small molecules
 - d. Microarray data
2. If computer code was used to generate results that are central to the paper's conclusions, a statement is included in the Materials and Methods section under "Code availability" to indicate whether and how the code can be accessed. Information on programming language (as necessary) and any restrictions on availability are provided.
3. Are the data underlying the analyses at all stages – original/raw data and intermediate results – provided for someone to reproduce the final statistical analysis?