



## Description

Research, reproducibility and reliability. These three words are key when it comes to good science. As scientists, we strive to prove that our results are significantly different. My colleague often wished for a “Journal of Negative Results”, one which published the experiments that didn’t work. This would save her time trying experiments that other researchers had already discovered to have negative results. Sadly, we often do not publish our “non-significantly” different results and we endeavor to find a significant difference in our experiments for our work to be publishable.

This is where researchers sometimes go wrong. They confuse basic statistical rules, that make their results look reproducible and more significant. However, often these methods [are not statistically correct](#). Make sure you are not making any of the [following statistical errors](#) in your manuscript:

## Common Statistical Errors

- **Control Group**

When studying the effect of an intervention at multiple time points, it is crucial to have a control group. Many other variables creep into experiments that you may not have thought of, such as experimenters becoming accustomed to the study and reporting more efficiently than when they started. Subjects also learn as they partake in an experiment and become more efficient at tasks over time. This needs to be controlled. Control groups need to be the same size as the experimental groups and they need to be sampled randomly at the same time as the experimental groups.

- **Comparing Two Effects Without Comparing Them Directly**

This happens when two effects are compared to a control group, but not to each other. For example, you have two groups and you want to compare two different interventions. It may look as if intervention A had a greater effect than intervention B, but your statistical analysis proves they are not significantly different. However, you find that each intervention is significantly different to the control group. The only conclusion you can draw is that both interventions showed a statistical difference compared to the

control group. You cannot state that they were statistically different to each other.

- **Number of Observations vs Number of Subjects**

The more repetitions you have in your experiments, the more reliable your statistical analysis. Don't be tempted to artificially inflate your units of analysis by using the number of observations rather than the number of subjects that you tested. To check yourself, look at the aim of your experiment. If you aimed to test the effect of an intervention on a group, then the unit of analysis is the number of subjects.

- **Dealing with Outliers**

You may think you should remove extreme outliers because you assume something went wrong with that data point. You may think it is not possible for one or two of your data points to be so different from your other observations. However, the outliers may be genuine observations. If you do remove a datapoint, you must mention it as well as justify it with good reasons. This kind of transparency is essential for good science.

- **Small Sample Sizes**

In cases where samples are rare and therefore limited, only large effects will be detectable statistically. In addition, a [genuine effect](#) may not be detected. If you find yourself in this situation, you may be able to get around the small sample size by performing replications within and between samples and including enough controls.

- **Double Dipping**

You complete a set of experiments where you compare the effect of an intervention before and after that intervention. You find the results are not significantly different between the groups. However, during your analysis you notice a change in behavior of your samples. You are tempted to sub-divide your groups and do away with the baseline data to prove your assumption. This is known as a circular analysis, or double dipping, and could merely be an observation of background noise. In order to prove your assumption, you would have to re-run the experiment with newly defined analysis criteria and the appropriate controls.

- **P-Hacking**

This "flexibility of analysis" occurs when researchers artificially increase the probability of achieving a significant p-value by adding covariates, excluding subjects, or switching outcome parameters. These may work in giving you a statistically significant result, because the more tests you run, the more likely

you'll find a false positive result. This is how probability works. If you find yourself in this position, rather be transparent and use your existing data to justify additional research.

- **Failing to Correct for Multiple Comparisons**

Multiple comparisons occur when two groups are compared using more than one variable. For example, you study the effect of a drug on volunteers and measure the result of multiple symptoms. In these types of tests, the larger the number of factors, the more likely a false-positive for one of the variables is probable. Ensure you correct for multiple comparisons of a group with a large set of variables using the relevant statistics for your experiment.

- **Over-Interpretation of Non-Significant Results**

The significance of the p-value of 0.05 has been described as arbitrary. It only confirms the high probability of a result occurring consistently within a population. For the p-value to be meaningful, its needs to be reported with the effect size and the 95 % confidence interval. This will enable readers to understand the scale of the effect and decide whether the results can be extrapolated to the relevant population, based on the results of the sample population.

- **Confusing Correlation and Causation**

The relationship between two variables is often analyzed using correlations. If a correlation is found between the two variables, it does not mean that one effect causes another. Further tests should be performed to prove causation.

## Considerations

A [few tips](#) to help you:

## Do

**Plan** your research and objectives before you carryout your experiments.

Define your **population**.

Consider all possible **sources of variation** and control for them.

**Select your statistical tests** before you start your experiments.

Include complete **details** of your data and analysis thereof in your report.

## Don't

Use the **same data** set to formulate your hypothesis and test it.

Sample the **wrong population** or fail to specify your population.

**Neglect** to select random and representative samples.

Use **inappropriate** statistical methods.

Try to make you data fit your hypothesis by "**tweaking**" your observations or trying a different statistical test.

Your statistical analysis aims to tell a story of an intervention or effect that you are researching. You need to ensure you have all the parts of the story for you to draw meaningful conclusions. Which common statistical errors have you come across when reading journal articles?

## Category

1. Publishing Research
2. Understanding Ethics

## Date Created

2020/02/28

## Author

editor