

# The Pitfalls of False Positives (and How to Avoid Them)

Tyson Holmes, PhD<sup>1</sup>

Imagine that you read a study report about a new treatment which is claimed to boost seedling growth and survival. You try the new treatment in your nursery but find that seedling performance is not improved.

Perhaps your first thought might be that you applied the treatment incorrectly; so you review your work but find that you had been quite careful and followed the study's method to the letter.

Then you consider the possibility that, though the treatment was applied correctly, conditions in your nursery differ in some important way from those in the study. Again you review the report. While you can't rule out the possibility that some differences exist, you see nothing in the report to indicate that conditions at your nursery were to blame.

Maybe then you begin to doubt the original study. Perhaps errors existed in their methods or in the analysis of their data. You talk to some colleagues but they concur that this research facility is well known for the quality of their work.

Another possibility remains...a statistical false positive.

Even in a well-controlled study, many factors of small to large effects may impact results; so it is quite possible that by pure chance alone the study found an effect of the new treatment when none truly exists.

Consider flipping a fair coin. The coin is fair because a flip is equally likely to land heads or tails. However, even a fair coin flipped 25 times has about a 5% chance of resulting in 8 or fewer tails.

In the same way, seedling growth for the old and the new treatments may be the same on average; but, by chance alone, the experiment may indicate that new treatment performs better than the old.

While few if any nursery experiments exist where one can completely eliminate false positive findings, measures can be taken to reduce the frequency of their occurrence.

- (1) Avoid small sample sizes. To understand why, consider the fair coin toss analogy again. While one has about a 5% chance receiving 8 or fewer tails when the fair coin is tossed 25 times, this chance balloons to about 25% if the fair coin is tossed only 20 times! A more intuitive explanation is that a smaller sample provides less information about how this type of seedling responds to the old and new treatments. With less information a mistaken conclusion is simply more likely.
- (2) Repeat studies. Suppose that the research design for comparing the old and new treatments has about a 5% chance of a false positive finding if old and new treatments truly have the same

---

<sup>1</sup> The author can be contacted at [thh.biostats@gmail.com](mailto:thh.biostats@gmail.com).

effect on seedling growth on average. If the study were independently repeated, the chance of having two false positive findings becomes  $(5 \times 5) / 100 =$  one quarter of one percent, which is getting to be quite small. In other words, one is much less likely to mistakenly conclude that the new treatment is effect, when it is not, if the study is independently repeated. While the costs of performing two studies is larger than one, this cost may be very tiny compared to the cost of widespread use of a new treatment that is truly no better than the old.

- (3) Choose statistical methods carefully. Statistical methods can be applied for comparing average seedling growth between the old and new treatments. Among the advantages of using statistical tests is that one can specify the allowable chance of a false positive. (Statisticians often refer to this as “Type I error.”) Problem is that the specified chances of a false positive does not always equal the actual chances. This disparity can arise for various technical reasons (e.g., the data do not satisfy the assumptions of the test).

In summary, keep in mind that positive findings can be false by chance alone. The basic safeguards discussed above can help you to minimize the chances of being misled by chance.