

# Promoting Transparency and Replicability in Psychological Science: Tips and Challenges

Steve Lindsay  
University of Victoria

CSBBCS 2024 Edmonton, AB

"The University of Alberta respectfully acknowledges that we are situated on Treaty 6 territory, traditional lands of First Nations and Métis people."

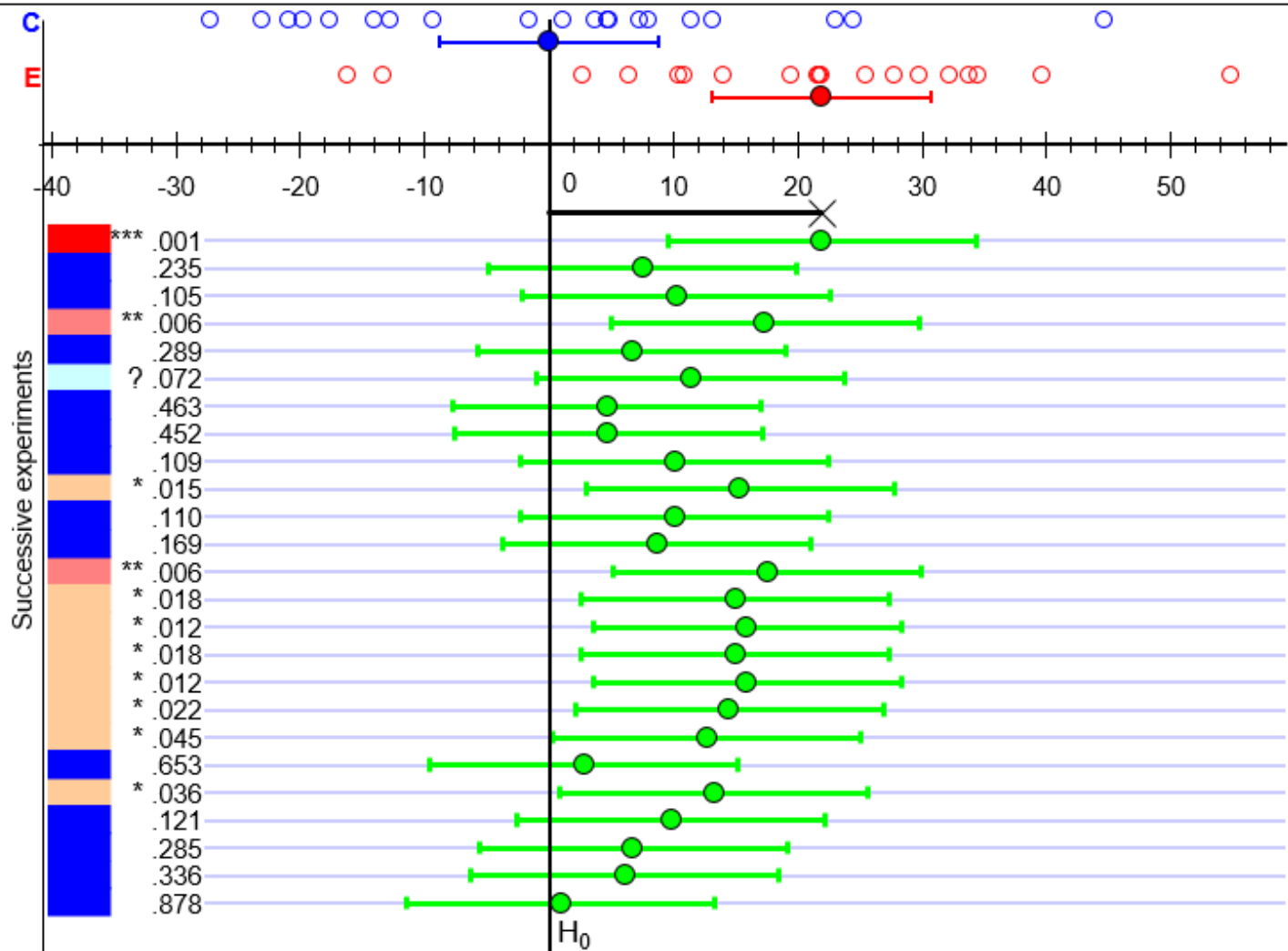
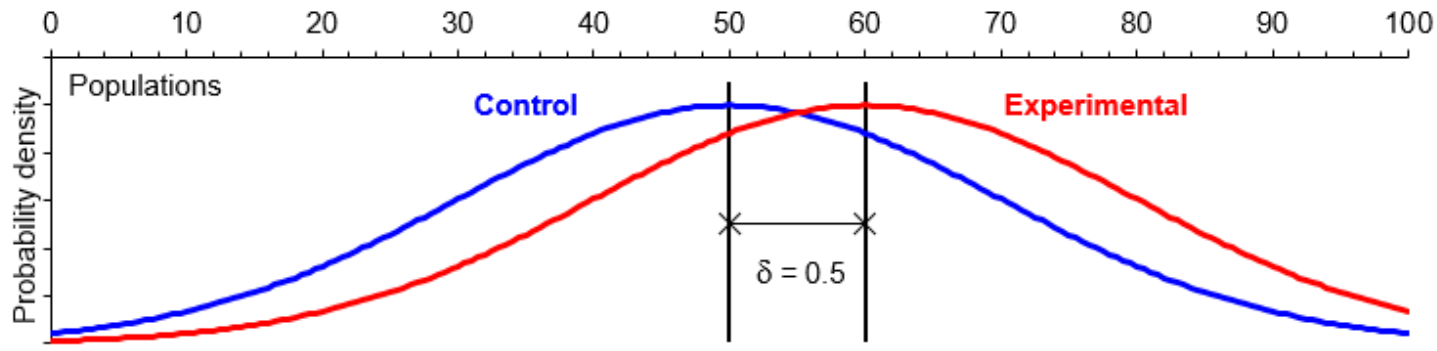
**What are you trying to do when you do science?**



**Geoff Cumming**  
**La Trobe University**

<https://thenewstatistics.com/itns/>

I then switched to a demo in Excel using the “Dance p” tab in “ESCi 64 bit chapters 5-6 Jul 4 2011.xlsm” downloaded from <https://thenewstatistics.com/itns/esci/esci-for-utns/>



- To have high  $p$  of detecting a medium-sized effect in between-SS designs need largish  $N$ . For small effect, need large  $N$ .
- $p$  is noisy; a small  $p$  does not promise effect will replicate; a big  $p$  doesn't prove the null.
- When statistical power is low, a test cannot be sig *unless* the study yields a larger-than-average effect.

What is the shape of the distribution of  $p$  values if a true null is tested thousands of times?

<http://rpsychologist.com/d3/pdist/>

[I followed that link and showed how distribution of  $p$  values changes with ES and with  $N$ .]



Kristoffer Magnusson

Selective reporting of successful underpowered experiments >  
exaggerated effect size > failures to replicate

Ditto post-hoc p-hacking (e.g., dropping subjects, trimming trials,  
adding covariates, etc.).

Ditto Hypothesizing after Results are Known (HARKing; Kerr 1989).

It is fine to explore your data post hoc. In fact, it is good.

As long as you are transparent about it.

If you want to do a rigorous hypothesis test with NHST it is helpful to make a detailed plan of exactly what you intend to test and how, including fully specified statistical analyses.

You are always free to deviate from that plan.

But if you do, you will know about it. And if you are honest so will others.

When possible, give other scientists direct access to fine-grained details of your materials, procedures, analyses, and results.

Turns out to be surprisingly difficult!

In 2019 Psychological Science published its first-ever issue in which all 14 empirical articles had a data badge.

Crüwell et al. (2023) attempted to reproduce the analyses w/o additional input from authors. Exactly reproduced results for one, close for three.

Attaining procedural and computational reproducibility is difficult and can only be ascertained by empirical test.

[Orally also mentioned evidence that preregistrations are often incomplete and that deviations from them are not always disclosed. E.g., <https://link.springer.com/article/10.3758/s13428-023-02277-0> . Coulda shoulda mentioned tools for improving e.g., <https://www.nature.com/articles/s41562-019-0772-6> and <https://doi.org/10.1177/09567976231221573>



## Take Home Messages

1. Foster awareness/understanding of statistical/methodological/theoretical issues.
2. Shift culture to value quality over quantity, accuracy over speed, realness over surprisingness, substance over appearance, transparency over sleight of hand.
3. Improve theories (see van Rooij, Devezer, et alia).

See also Lindsay, D. S. (2020). [Seven Steps toward Transparency and Replicability in Psychological Science](#) *Canadian Psychology/Psychologie Canadienne*, 61, 310–317.