**Preliminary findings for “An evaluation of reproducibility in statistical power calculations performed using G\*Power” (**[**https://osf.io/msz24/**](https://osf.io/msz24/)**) [DO NOT SHARE]**

Authors: Robert Thibault, Emmanuel Zavalis, Hugo Pedder

**Main findings**

1. **About 47,000 articles indexed by PubMed use G\*Power** (published between 2017–2021)**.** This includes about 32,000 articles that use G\*Power to calculate the sample size for their study and about 7,000 articles that clearly calculate the sample size for a repeated measures ANOVA.
2. **Two-thirds of published sample size calculations that use G\*Power are not reproducible**. We assessed 92 calculations randomly selected from the published Open Access literature indexed on PubMed Central. We could reproduce 2 G\*Power calculations without making any assumptions. We likely reproduced 28 after making assumptions. We could not reproduce 62 calculations. Of these 92 published calculations, 60 did not report the statistical test powered for, 51 did not report the type of effect size, and 16 did not report the value of the effect size.
3. **Few repeated measures ANOVA calculations select a non-default option.** We assessed 30 articles that used G\*Power to calculate sample size for a repeated measures ANOVA. Two indicated that they did not use the default option, 12 did not provide enough information to identify whether they used the default option, and 16 used the default option. In most of these cases, it appears the authors would have intended to select a non-default option.

**Main recommendations.** Given the findings above, we have suggestions for two features that could be added to G\*Power 4. These features could improve the quality of sample size calculations across thousands of published articles each year.

1. **Provide a button that outputs meta-data** **or a screenshot of G\*Power.** G\*Power 4 could include a prompt to use this button and recommend to users that they submit the meta-data as supplementary material alongside their manuscript submission. The meta-data could include a timestamp—It appears that many researchers run their sample size calculation after conducting their experiment. This feature would make many more calculations reproducible.
2. **Remove the default option for repeated measures ANOVAs.** Information could be added to help users select the appropriate option. At least for medical research, I believe it would make sense to put Cohen’s method or the SPSS methods as the first option.

Thank you for considering our input. We would be happy to discuss further and provide your team with the final detailed results when they are ready.