Reporting

Reportinig Questionable Research Practices (Wicherts et al, 2016, Table 1)

1. Failing to assure reproducibility (verifying the data collection and data analysis).
2. Failing to enable replication (re-running of the study). Poor practices could include inadequate detail in methods sections, failure to make experimental material available, etc.
3. Failing to mention, misrepresenting, or misidentifying the study preregistration.
4. Failing to report “failed studies” that were originally deemed relevant to the research question.
5. Misreporting results and *p*-values.
6. Presenting exploratory analyses as confirmatory (Hypothesizing After Results Known).

Guidance:

Many of the above concerns can be mitigated simply by following the Tri-Agency recommendations for data sharing and management. Beginning in the fall of 2017, all Tri-Agency grants must include a data-management plan: “research data resulting from agency funding should normally be preserved in a publicly accessible, secure and curated repository or other platform for discovery and reuse by others.” Providing a codebook along with your data is a good idea, and see these practical tips for ethical data sharing.

As noted elsewhere, the American Statistical Association states that scientific conclusions should not be based on *p*-values alone. Consequently, effect sizes (raw or standardized) with confidence intervals should be reported for each hypothesis. According to Cumming & Finch (2001) a confidence interval can be interpreted as a plausible range of population effect sizes that could have produced the effect observed in the sample. Be sure to visualize your effect size before writing about it to avoid over interpretation of results (see links for d-values and correlations).

Exploratory analysis should be reported as exploratory. Because *p*-values are only meaningful with an a priori hypothesis, they should not be reported with exploratory analyses (see details in Gigerenzer (2004) and Wagenmakers et al., (2012)). Rather, exploratory analyses should be seen as an empirical process for generating hypotheses to be tested in a subsequent study. Confidence intervals are still appropriate with exploratory analyses.

Many psychology papers have reporting errors that substantially change interpretation of results. A review of 28 years of *p*-values (over 250,000 *p*-values) revealed that the test statistic (e.g., t(28) = 2.60) is inconsistent with the reported *p*-value (e.g., *p* = .0147) more than 50% of the time (Nuijten, Hartgerink, van Assen, Epskamp, & Wicherts, 2015). Interestingly, willingness to share data appears to be associated with fewer statistical reporting errors (Wicherts, Bakker, & Molenaar, 2011). Fortunately, errors can be detected easily with an automated process. A PDF of a thesis can be checked at http://statcheck.io/. Note that some journals (e.g., *Psychological Science*) are now using statcheck on all non-desk rejected articles as part of the review process.

You will likely find the article "Writing Empirical Articles: Transparency, Reproducibility, Clarity, and Memorability" a helpful resource.

New APA Reporting Standards for Quantitative Journal Articles:

In addition, the American Psychological Association has issued new Journal Article Reporting Standards (JARS). We strongly recommend that you read the article and follow its prescriptions.

Example Text:

Below is example text for correlations and *d*-values from the Cumming & Calin-Jageman (2016) textbook. We suggest using a similar style, but also adding *p*-values to the text (and of course ensuring APA-style).

Cumming, G., & Calin-Jageman, R. (2016). Introduction to the new statistics: Estimation, open science, and beyond. Routledge.

Correlation:

"The correlation between well-being and self-esteem was *r*= .35, 95% CI[.16, .53], N = 95. Relative to other correlates of well-being that have been reported, this is a fairly strong relationship. The CI, however, is somewhat long and consistent with anywhere from a weak positive to a very strong positive relationship." (pp. 324-325)

The correlation between well-being and gratitude was *r*= .35, 95% CI [-.11, .69], N = 20. The CI is quite long. These data are only sufficient to rule out a strong negative relationship between these variables." (p. 325)

d-value:

"Motor skill for men (M = 33.8%, s = 12.8%, n = 62) was a little higher than for women (M = 29.7% s = 15.8%, n = 89). The difference in performance may seem small in terms of raw scores (M_{mean}- M_{women}= 4.0%, 95% CI [-.7, 8.8]), but the standardized effect size was moderate ($d_{unbiased}$= 0.28, 95% Ci [-.05, 0.61]) relative to the existing literature. However, both CIs are quite long, and are consistent with anywhere from no advantage up to a large advantage for men. More data are required to estimate more precisely the degree to which ender might be related to motor skill."(p. 188)

Ethical Issues in Reporting

Considering and reporting power, positive predictive values, and confidence intervals for all hypothesis tests help to facilitate a method of reporting that draws attention to the statistical inference limitations of a study. In the past, it was common for authors to be in a scenario associated with low power, low positive predictive values, and wide confidence intervals, but not report these statistics and correspondingly draw overly strong conclusions. The process we have outlined in this document is designed to provide readers of your thesis with a realistic view of the inferential limitations of your study and allow them to fairly consider the informational value of the study and other hypotheses. This is consistent with the Canadian Psychological Association's Ethics Guidelines presented below.

III.8 Canadian Psychology Association Code of Ethics

Acknowledge the limitations, and not suppress disconfirming evidence, of their own and their

colleagues' methods, findings, interventions, and views, and acknowledge alternative hypotheses and explanations.

III.9 Canadian Psychology Association Code of Ethics

Evaluate how their own experiences, attitudes, culture, beliefs, values, individual differences, specific training, external pressures, personal needs, and historical, economic, and political context might influence their activities and thinking, integrating this awareness into their attempts to be as objective and unbiased as possible in their research, service, teaching, supervision, employment, evaluation, adjudication, editorial, and peer review activities.

Student Check List 5 of 5: Reporting

__ The student used http://statcheck.io/ on the thesis document and provided the committee with the resulting report.

__ The student reported confidence intervals and effect sizes.

__ Interpretation of results was based on the full range of the confidence interval – which conveys the uncertainty of the effect size estimate. Do recognize the center of the confidence interval is more likely than the extremes.

__ Applied recommendations were only made in a manner consistent with the full range of the confidence interval. In particular, keep in mind the lower-bound of the confidence interval.

__ Student has clearly identified exploratory analyses and avoided reporting *p*-values for these analyses.

__ Only a priori confirmatory hypotheses were reported as confirmatory.

__ All analyses conducted were reported.

__ All studies conducted were reported.

Source URL:https://www.uoguelph.ca/psychology/book-page/reporting