

Example language for writing up statistical results

Here are three examples of language one could use when reporting statistical results that avoid dichotomization based on p-values and language about “statistical significance”.

Example 1

Frank Harrell started a thread at *datamethods.org* specifically to discuss ways to report statistical results honestly and accurately (<https://discourse.datamethods.org/t/language-for-communicating-frequentist-results-about-treatment-effects/934>). His post is worth reading through in entirety, but here is one of his longer suggestions:

“Treatment B was observed in our sample of n subjects to have a 4mmHg lower mean SBP than treatment A with a 0.95 2-sided compatibility interval of $[-13, 5]$, indicating a wide range of plausible true treatment effects. The degree of evidence against the null hypothesis that the treatments are interchangeable is $p=0.11$. The smaller the p-value the greater the evidence against the null.”

Example 2

From “[Inferential statistics as descriptive statistics: there is no replication crisis if we don't expect replication](#)” by Amrhein et al. 2019, pg 266:

‘...A more correct summary of the results would have been: "Our estimate of the hazard-rate ratio was 1.61, and thus exposure could be associated with autism; however, possible hazard-rate ratios that are highly compatible with our data, given our model, ranged from 0.997 (essentially no association) to 2.59 (a relatively strong association)." If applicable, this could then be followed by a discussion of why the authors seem to think the exposure effect might be negligible despite the association, and how strong they judge their evidence not only based on the width of an interval estimate, but also in view of possible shortcomings of their study, of their prior knowledge about other studies on autism, and of possible costs of their interpretation for the health of the patients.’

Example 3

From the [Nature commentary](#) by Amrhein et al. 2019:

“...For example, the authors above could have written: ‘Like a previous study, our results suggest a 20% increase in risk of new-onset atrial fibrillation in patients given the anti-inflammatory drugs. Nonetheless, a risk difference ranging from a 3% decrease, a small negative association, to a 48% increase, a substantial positive association, is also reasonably compatible with our data, given our assumptions.’ Interpreting the point estimate, while acknowledging its uncertainty, will keep you from making false declarations of ‘no difference’, and from making overconfident claims.”

References

Amrhein, V., Greenland, S., and McShane, B. (2019), “Scientists rise up against statistical significance”, *Nature*, 567, 305-307.
 Amrhein, V., Trafimow, D. and Greenland, S. (2019), “Inferential Statistics as Descriptive Statistics: There Is No Replication Crisis if We Don't Expect Replication”, *The American Statistician*, 73:sup1, 262-270.