Balancing Errors as an Approach Towards Better Use of Larger Samples in Psychological Research

Isabel Thielmann & Benjamin E. Hilbig

University of Koblenz-Landau

Open Science Symposium, Trier

March 2019

Replicability Crisis in Psychology

Sample Sizes

- cornerstone argument raised during replicability crisis: small sample sizes and **low statistical power**

    → repeated calls for larger samples

- e.g., "it cannot be stressed enough that **researchers should collect bigger sample sizes**, and editors, reviewers, and readers should insist on them" (Asendorpf et al., p. 110)

Two Types of Errors

- calls for larger samples remain incomplete
  - β error is decreased
  - BUT: α error is usually held constant at 5%

- may result in **imbalanced errors**, which is problematic because…
  - greater importance assigned to one type of error (viz. β)
  - statistical test cannot achieve consistency

Balancing Errors

- **balancing errors** allows exploiting the advantages of large samples whenever possible

- Method: simulation study
  - comparison of fixed $\alpha = .05$ vs. balanced errors ($\alpha = \beta$)
  - evaluation criteria
    - Positive Predictive Value (PPV)
    - proportion of correct inferences p(correct)
  - independent samples t-test, with medium-sized effect ($d = 0.50$)
 - $5 \leq N \leq 350$ (per group)
  - $.01 \leq p(\text{H1}) \leq .99$

Positive Predictive Value

$$PPV = \frac{(1 - \beta) * p(\mathrm{H1})}{(1 - \beta) * p(\mathrm{H1}) + \alpha * \big(1 - p(\mathrm{H1})\big)}$$

$$\lim_{\beta \to 0} (PPV) = \frac{p(\mathrm{H1})}{p(\mathrm{H1}) + \alpha * \big(1 - p(\mathrm{H1})\big)}$$

Balancing Errors

- Effect on PPV

Once the sample size is sufficiently large to render β < α, adjusting α corresponding to β **always** results in a higher PPV than holding α fixed at .05 – irrespective of the probability p(H1) that the alternative hypothesis is true
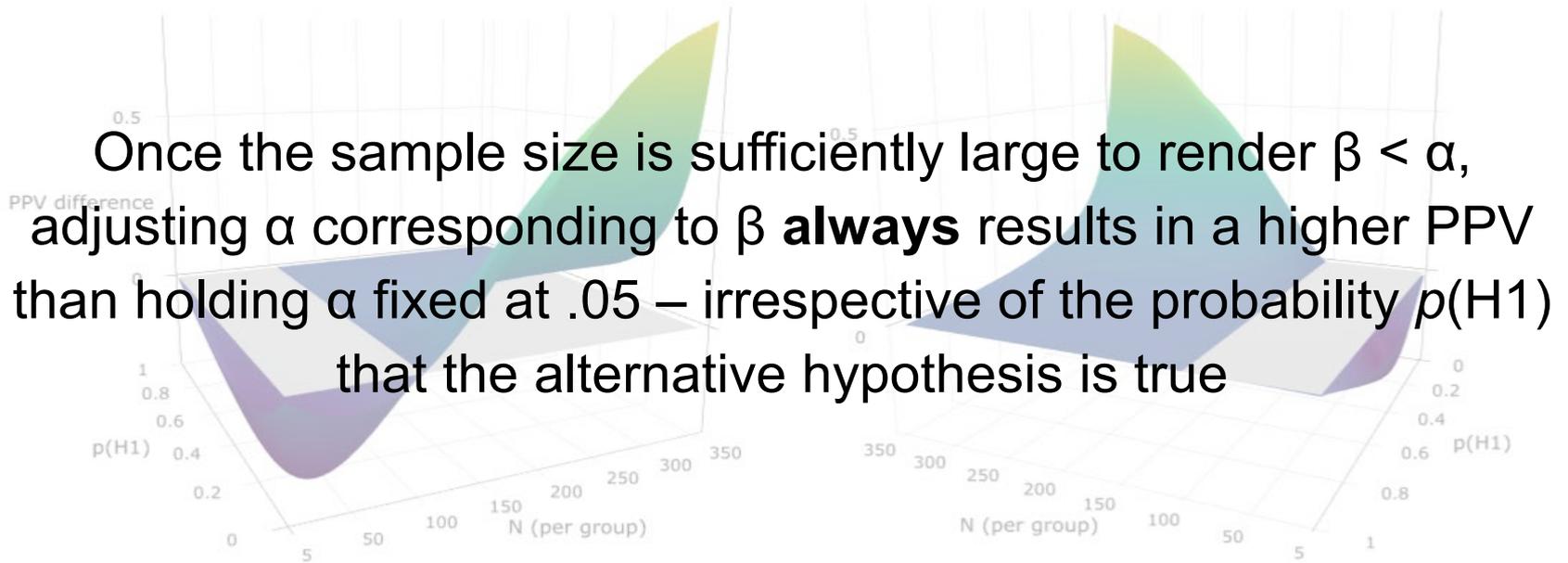
Proportion of Correct Inferences

$$p(correct) = (1 - \beta) * p(\text{H1}) + (1 - \alpha) * \big(1 - p(\text{H1})\big)$$

$$\lim_{\beta \to 0} (p(correct)) = 1 + \alpha * (p(\text{H1}) - 1)$$

Balancing Errors

- Effect on p(correct)

Balanced errors are to be preferred over fixed α:
(1) whenever β < α, practically irrespective of p(H1)
(2) whenever p(H1) > .50, practically irrespective of the absolute magnitude of α and β.
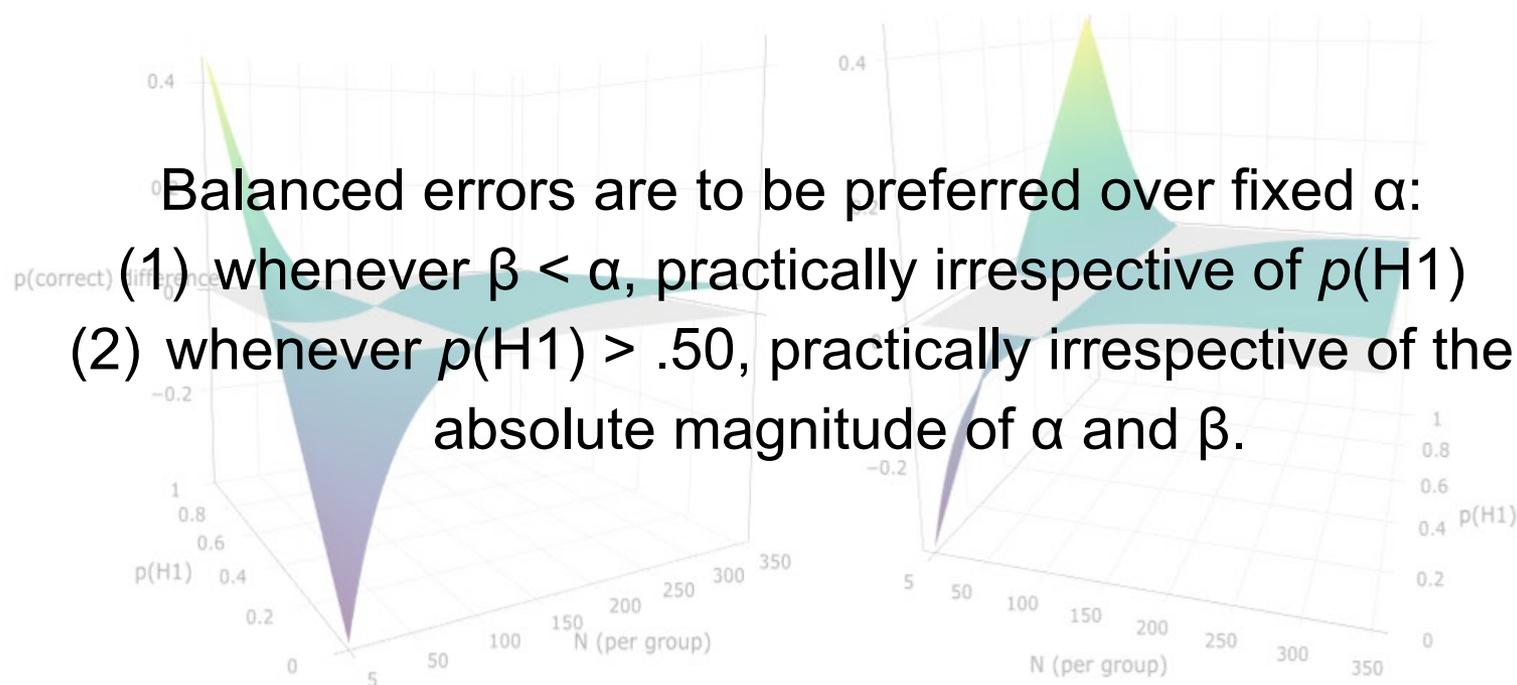
Compromise Power Analysis

- compromise power analysis allows to specify α as a function of N, effect size, and α/β ratio (Erdfelder, 1984; Erdfelder et al., 1996)

- e.g., implemented in G*Power (Faul et al., 2009)

- should be used in an **a priori fashion**

Summary & Conclusion

- balancing errors is advantageous in many situations

- whenever researchers have the sample size to **reduce both types of errors**, they should do so

  – "NP [Neyman-Pearson] testing can be made **consistent** by allowing Type I error rates to decrease toward zero as the sample size increases" (Rouder et al., 2009, p. 228)

If our field is to truly profit from the developments triggered by the replicability crisis, we should strive for smaller error probabilities **in general**

THANK YOU…
FOR YOUR ATTENTION!

…and to my collaborators and colleagues

Benjamin E. Hilbig

Morten Moshagen

Felix Henninger
Edgar Erdfelder

For more information: thielmann@uni-landau.de
www.cognition.uni-landau.de