**Supplemental material**

1. **Why p-values obtained from a study that expected to find a significant difference in one direction but observed an effect in the other direction were excluded**.

To explain why these results should not be included is key to define publication bias as the publishing behaviors that gives studies which find support for their tested hypotheses a higher chance of being published, as opposed to the publication of non-significant findings. Assuming there is publication bias, a study that predicted the effect of an intervention and does not find evidence in favor of the predicted effect (p > 0.05), this study could be rejected for publication. Consider the following example (Gidley et al., 2020), where the authors hypothesized that “the compliant treadmill will result in greater **leg stiffness** and a reduction in coordination variability”. However, the effect occurred in the other direction where leg stiffness (i.e., Kleg) was significantly lower for compliant treadmill than for rigid treadmill (Kleg on the rigid treadmill was significantly larger than on compliant treadmill (13.72 ± 3.2 kN•m−1 > 13.01 ± 3.0 kN•m−1, **p = 0.041,** ES = 0.707)). If we included this *p-*value in the z-curve, this significant *p*-value would represent evidence in favor of the tested research hypothesis, although what the study found is the opposite.

Alexis D. Gidley, D. Eli Lankford & Joshua P. Bailey (2020) The construction of common treadmills significantly affects biomechanical and metabolic variables, Journal of Sports Sciences, 38:19, 2236-2241, DOI: [10.1080/02640414.2020.1776929](https://doi.org/10.1080/02640414.2020.1776929)

1. **Secondary z-curve**

Results from the performed secondary z-curve which included p-values from studies that tested a hypothesis (n = 89) and studies that only stated a goal (n = 30) but lacked a hypothesis are displayed in the **Table 4** and **Figure 2b**.

**Table 4**. Results from the secondary z-curve

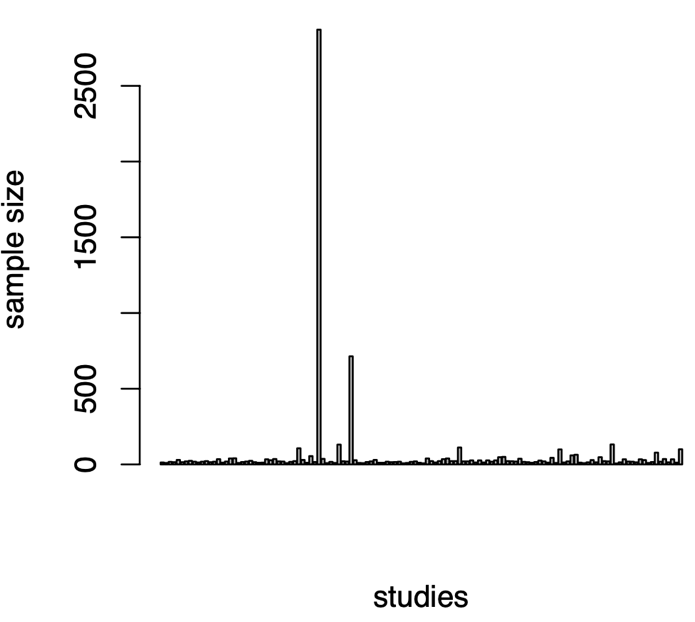
|  |  |  |
| --- | --- | --- |
| Parameter | Estimate | 95% CI |
| ODR | 0.69 | 0.60 – 0.77 |
| ERR | 0.60 | 0.42 – 0.74 |
| EDR | 0.48 | 0.11 – 0.69 |
| Soric FDR | 0.06 | 0.02 – 0.44 |
| File Drawer R | 1.11 | 0.44 ­– 8.33 |



**Figure 2b**. Distribution of z-scores over [0-6] interval. The vertical red line refers to a z-score of 1.96, the critical value for statistical significance when using a two-tailed alpha of 0.05. The dark blue line is the density distribution for the inputted *p*-values (represented in the histogram as z-scores). The dotted lines represent the 95% CI for the density distribution. Range represents the minimum and maximum values of z-scores used to fit the z-curve.

1. **Outliers’ exclusion**

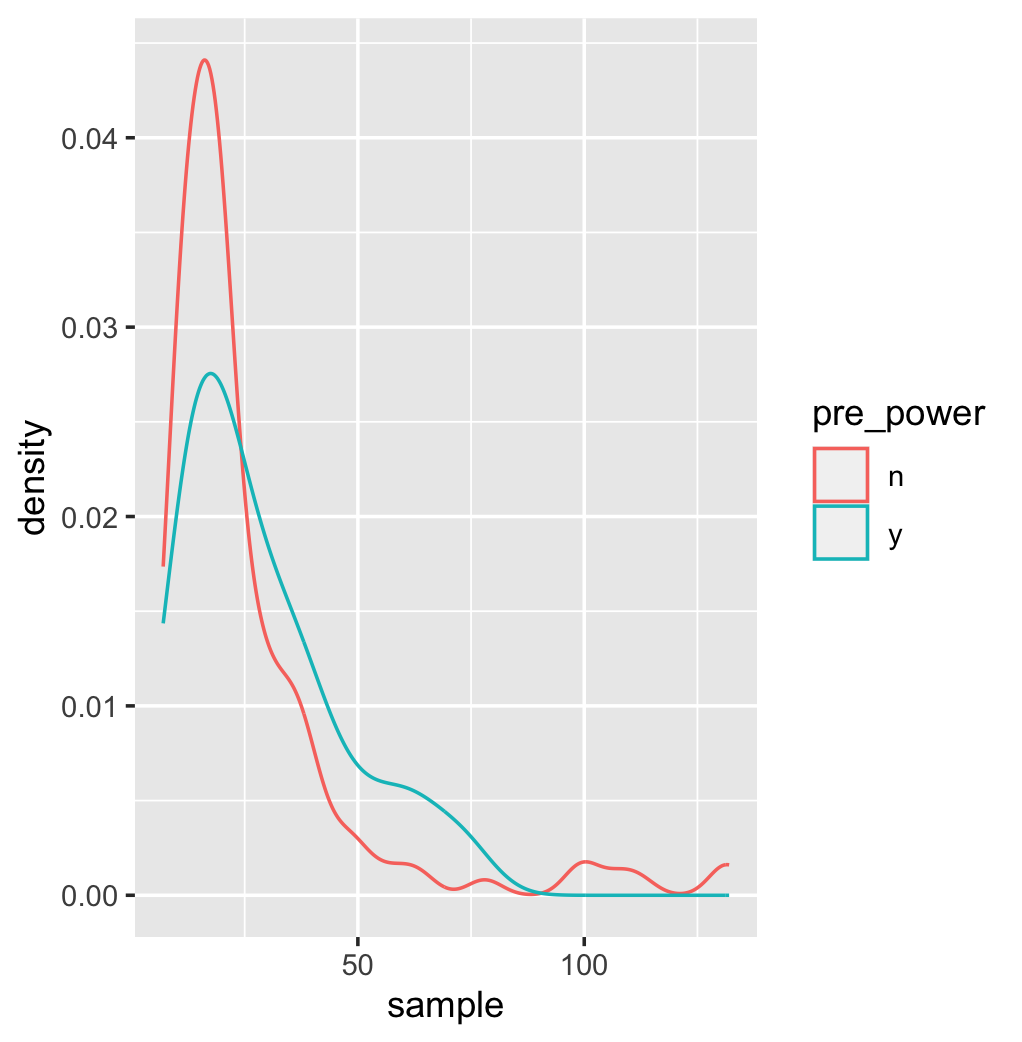
In the sample of studies without an a-priori power analysis, there are two studies with a sample size of 2871 and 714 participants, respectively (see **Figure 3**). With two these large studies, the mean ± SD for this sample is 47 ± 222, whereas without them the mean ± SD is 26 ± 23. Because these two studies clearly represent two outliers, we have removed these two studies from the final sample of studies. Thus, descriptive statistics and the regression analysis have been performed without these two large studies.



**Figure 3**. Distribution of sample sizes for studies that did not include an a-priori-power analysis.

1. **Poisson distribution**

We performed a Poisson simple regression because of the distribution of data (sample sizes), which followed a Poisson distribution (see **Figure 4**).



**Figure 4**. Distribution of sample sizes for studies that included an a-priori-power analysis (blue) and studies that did not include an a-priori power analysis (red).