Figure 11. Comparison of *p*-values from TOST (black line) and 1-SGPV (grey curve) across a range of observed sample correlations (x-axis) tested against equivalence bounds of r = -0.45 and r = 0.45 with n = 30 and an alpha of 0.05.
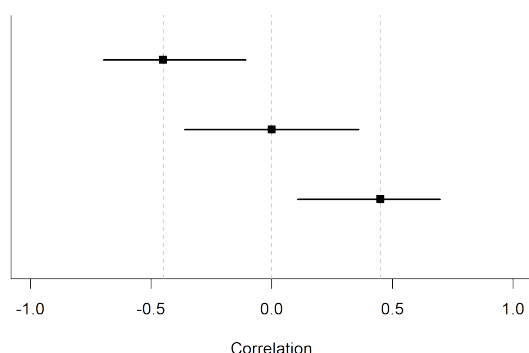

Figure 12. Three 95% confidence intervals for observed effect sizes of $r = -0.45$, $r = 0$, and $r = 0.45$ for n = 30. Only the confidence interval for $r = 0$ is symmetric.

than the true correlation. As can be seen in Figure 12, the proportion of the confidence interval that overlaps with the equivalence range is larger than 50% when the observed correlations are $r = -.45$ and $r = .45$, meaning that the two second generation p-values associated with these correlations are larger than 50%. Because the confidence intervals are asymmetric around the observed effect size of 0.45 (ranging from 0.11 to 0.70) according to Blume et al. (2018) 58.11% of the data-supported hypotheses are null hypotheses, and therefore 58.11% of the data-supported hypotheses are compatible with the null premise.

The further away from 0, the larger the SGPV when the observed mean falls on the equivalence bound. The

SGPV is the proportion of values in a 95% confidence interval that overlap with the equivalence range, but not the probability that these values will be observed. In the most extreme case (i.e., a sample size of 4, and equivalence bounds set to $r = -0.99$ and 0.99, with a true correlation of 0.99) 97.60% of the confidence interval overlaps with the equivalence range, even though in the long run only 36% of the correlations observed in the future will fall in this range.

It should be noted that in larger sample sizes the SGPV is closer to 0.5 whenever the observed correlation falls on the equivalence bound, but this extreme example nevertheless clearly illustrates the difference between question the SGPV answers, and the question a *p*-value answers. The conclusion of this section on asymmetric confidence intervals is that a SGPV of 1 or 0 can still be interpreted as a $p < 0.025$ or $p > 0.975$ in an equivalence test, since the SGPV and *p*-value for the TOST procedure are always directly related at the values $p = 0.025$ and $p = 0.975$. Although Blume et al. (2018) state that "the degree of overlap conveys how compatible the data are with the null premise" this definition of what the SGPV provides does not hold for asymmetric confidence intervals. Although a SGPV of 1 or 0 can be directly interpreted, a SGPV between 0 and 1 is not interpretable as "compatibility with the null hypothesis" under the assumption of a bivariate normal distribution, and the generalizability of this statement needs to be examined beyond normal bivariate distributions. Indeed, Blume and colleagues write in the supplemental material that "The magnitude of an inconclusive second-generation *p*-value can vary slightly when the effect size scale is transformed. However definitive findings, i.e. a *p*-value of 0 or 1 are *not* affected by the scale changes."

What are the Relative Strengths and Weaknesses of Equivalence Testing and the SGPV?

When introducing a new statistical method, it is important to compare it to existing approaches and specify its relative strengths and weaknesses. Here, we aimed to compare the SGPV against equivalence tests based on the TOST procedure. First of all, even though a SGPV of 1 or 0 has a clear interpretation (we can reject effects outside or inside the equivalence range), intermediate values are not as easy to interpret (especially for effects that have asymmetric confidence intervals). In one sense, they are what they are (the proportion of overlap), but it can be unclear what this number tells us about the data we have collected. This is not too problematic, since the main use of the SGPV (e.g., in all examples provided by Blume and colleagues) seems to be to examine whether the SGPV is 0, 1, or inconclusive.