As already mentioned, this interpretation of a SGPV is very similar to the Neyman-Pearson interpretation of an equivalence test and a minimum effect tests (which are complementary). The difference is that where a SGPV of 1 can be interpreted as $p < .025$, equivalence tests provide exact p-values, and they continue to differentiate between for example $p = 0.024$ and $p = 0.002$. Whether this is desirable depends on the perspective that is used. From a Neyman-Pearson perspective on statistical inferences the main conclusion is based on whether or not $p < \alpha$, and thus an equivalence test and SGPV can be performed by simply checking whether the confidence interval falls within the equivalence range, just as a null-hypothesis test can be performed by checking whether the confidence interval contains zero or not. At the same time, it is recommended to report exact p-values (American Psychological Association, 2010), and exact p-values might provide information of interest to readers about how precisely how surprising the data, or more extreme data, is under the null model. Some researchers might be interested in combining an equivalence test with a null-hypothesis significance test. This allows a researcher to ask whether there is an effect that is statistically different from zero, and whether effect sizes that are considered meaningful can be rejected. Equivalence tests combined with null-hypothesis tests classify results into four possible categories, and for example allow researchers to conclude an effect is significant *and* equivalent (i.e., statistically different from zero, but also too small to be considered meaningful; see Lakens et al., 2018).

An important issue when calculating the SGPV is its reliance on the "small sample correction", where the SGPV is set to 0.5 whenever the ratio of the confidence interval width to the equivalence range exceeds 2:1 and the CI overlaps with the upper and lower bounds. This exception to the normal calculation of the SGPV is introduced to prevent misleading values. Without this correction it is possible that a confidence interval is extremely wide, and an equivalence range is extremely narrow, which without the correction would lead to a very low value for the SGPV. Blume et al. (2018) suggest that under such a scenario "the data favor alternative hypotheses", even when a better interpretation would be that there is not enough data to accurately estimate the true effect compared to the width of the equivalence range. Although it is necessary to set the SGPV to 0.5 whenever the ratio of the confidence interval width to the equivalence range exceeds 2:1, it leads to a range of situations where the SGPV is set to 0.5, while the p-value from the TOST procedure continues to differentiate (see for example Figure 6). An important benefit of equivalence tests is that it does not need such a correction to prevent misleading results.

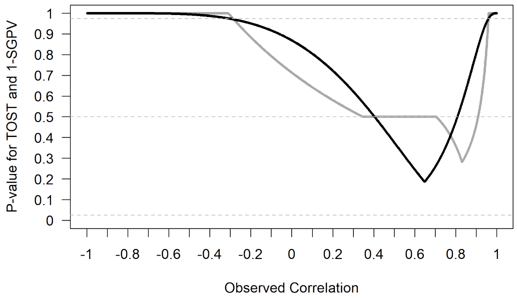

Figure 13. Comparison of p-values from TOST (black line) and 1-SGPV (grey curve) across a range of observed sample correlations (x-axis) tested against equivalence bounds of $r = 0.4$ and $r = 0.8$ with n = 10 and an alpha of 0.05.

As a more extreme example of the peculiar behavior of the "small sample correction" as currently implemented in the calculation of the SGPV, see Figure 13. In this figure observed correlations (from a sample size of 10) from -.99 to .99 are tested against an equivalence range from r = 0.4 to r = 0.8. We can see the SGPV has a peculiar shape because it is set to 0.5 for certain observed correlations, even though there is no risk of a "misleading" SGPV in this range. This example suggests that the current implementation of the "small sample correction" could be improved. If, on the other hand, the SGPV is mainly meant to be interpreted when it is 0 or 1, it might be preferable to simply never apply the "small sample correction".

Blume et al. (2018) claim that when using the SGPV "Adjustments for multiple comparisons are obviated" (p. 15). However, this is not correct. Given the direct relationship between TOST and SGPV highlighted in this manuscript (where a TOST $p = 0.025$ equals SGPV = 1, as long as the SGPV is calculated based on confidence intervals, and assuming data are sampled from a continuous bivariate normal distribution), not correcting for multiple comparisons will inflate the probability of concluding the absence of a meaningful effect based on the SGPV in exactly the same way as it will for equivalence tests. Whenever statistical tests are interpreted as support for a hypothesis (e.g., SPGV = 0 or SGPV = 1), it is possible to do so erroneously, and if researchers want to control error rates, they need to correct for multiple comparisons.