

Figure 5. The relationship between p -values from the TOST procedure and the SGPV for the same scenario as in Figure 1.

the SGPV is either 0 or 1, p -values from the equivalence test fall between 0.975 and 1 or between 0 and 0.025. Where the SGPV is 1 as long as the confidence interval falls completely within the equivalence bounds, the p -value for the TOST continues to differentiate between results as a function of how far the confidence interval lies within the equivalence bounds (the further the confidence interval is from both bounds, the lower the p -value). The easiest way to see this is by plotting the SGPV against the p -value from the TOST procedure. The situations where the p -values from the TOST procedure continue to differentiate based on how extreme the results are, but the SGPV is a fixed value are indicated by the parts of the curve where there are vertical straight lines at second generation p -values of 0 and 1.

A third situation in which the SGPV remains stable across a range of observed effects, while the TOST p -value continues to differentiate, is whenever the CI is wider than the equivalence range, and the CI overlaps with the upper *and* lower equivalence bound. When the confidence interval is more than twice as wide as the equivalence range the SGPV is set to 0.5. Blume et al. (2018) call this the “small sample correction factor”. However, it is not a correction in the typical sense of the word, since the SGPV is not adjusted to any “correct” value. When the normal calculation would be “misleading” (i.e., the SGPV would be small, which normally would suggest support for the alternative hypothesis, but at the same time all values in the equivalence range are supported), the SGPV is set to 0.5 which according to Blume and colleagues signals that the SGPV is “uninformative”. Note that the CI can be twice as wide as the equivalence range whenever the sample size is small (and the confidence interval width is large) or when

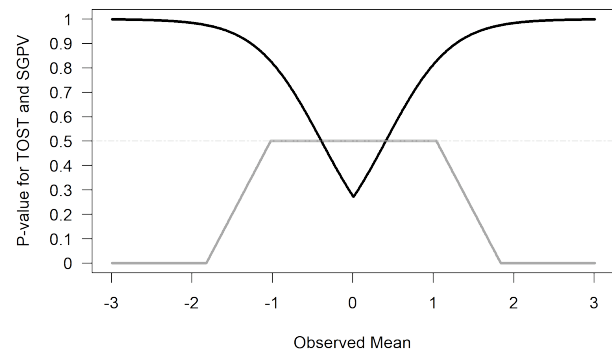


Figure 6. Comparison of p -values from TOST (black line) and SGPV (grey line) across a range of observed sample means (x-axis). Because the sample size is small ($n = 10$) and with a standard deviation of 2 the CI is more than twice as wide as the equivalence range (set to -0.4 to 0.4), the SGPV is set to 0.5 (horizontal light-grey line) across a range of observed means.

then equivalence range is narrow. It is therefore not so much a “small sample correction” as it is an exception to the typical calculation of the SGPV 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.

We can examine this situation by calculating the SGPV and performing the TOST for a situation where sample sizes are small and the equivalence range is narrow, such that the CI is more than twice as large as the equivalence range (see Figure 6). When the two statistics are plotted against each other we can see where the SGPV is the same while the TOST p -value still differentiates different observed means (indicated by straight lines in the curve, see Figure 7). We see the SGPV is 0.5 for a range of observed means where the p -value from the equivalence test still varies. It should be noted that in these calculations the p -values for the TOST procedure are *never* smaller than 0.05 (i.e., they do not get below 0.05 on the y-axis). In other words, we cannot conclude equivalence based on any of the observed means. This happens because we are examining a scenario where the 90% CI is so wide that it never falls completely within the two equivalence bounds.

As Lakens (2017) notes: “in small samples (where CIs are wide), a study might have no statistical power (i.e., the CI will always be so wide that it is necessarily wider than the equivalence bounds).” None of the p -values based on the TOST procedure are below 0.05, and thus, in the long run we have 0% power.