

data-supported hypotheses that are also null hypotheses” (Blume et al., 2018). The researcher specifies an equivalence range around a null hypothesis of values that are considered practically equivalent to the null hypothesis. The SGPV measures the degree to which a set of data-supported parameter values falls within the interval null hypothesis. If the estimation interval falls completely within the equivalence range, the SGPV is 1. If the confidence interval falls completely outside of the equivalence range, the SGPV is 0. Otherwise the SGPV is a value between 0 and 1 that expresses the overlap of data-supported hypotheses and the equivalence range. When calculating the SGPV the set of data-supported parameter values can be represented by a confidence interval (CI), although one could also choose to use credible intervals or Likelihood support intervals (SI). When a confidence interval is used, the SGPV and equivalence tests such as the Two One-Sided Tests (TOST) procedure (Lakens, 2017; Meyners, 2012; Quertemont, 2011; Schuirmann, 1987) appear to have close ties, because both tests compare a confidence interval against an equivalence range. Here, we aim to examine the similarities and differences between the TOST procedure and the SGPV. We limit our analysis to continuous data sampled from a bivariate normal distribution. The TOST procedure also relies on the confidence interval around the effect. In the TOST procedure the data are tested against the lower equivalence bound in the first one-sided test, and against the upper equivalence bound in the second one-sided test (Lakens, Scheel, & Isager, 2018). For an excellent discussion of the strengths and weaknesses of different frequentist equivalence tests, including alternatives to the TOST procedure, see Meyners (2012). If both tests statistically reject an effect as extreme or more extreme than the equivalence bound, you can conclude the observed effect is practically equivalent to zero from a Neyman-Pearson approach to statistical inferences. Because one-sided tests are performed, one can also conclude equivalence by checking whether the $1-2\times\alpha$ confidence interval (e.g., when the alpha level is 0.05, a 90% CI) falls completely within the equivalence bounds. Because both equivalence tests as the SGPV are based on whether and how much a confidence interval overlaps with equivalence bounds, it seems worthwhile to compare the behavior of the newly proposed SGPV to equivalence tests to examine the unique contribution of the SGPV to the statistical toolbox.

The relationship between p -values from TOST and SGPV when confidence intervals are symmetrical

The second generation p -value (SGPV) is calculated as:

$$p_{\delta} = \frac{|I \cap H_0|}{|I|} \times \max \left\{ \frac{|I|}{2|H_0|}, 1 \right\}$$

where I is the interval based on the data (e.g., a 95% confidence interval) and H_0 is the equivalence range. The first term of this formula implies that the second generation p -value is the width of the confidence interval that overlaps with the equivalence range, divided by the total width of the confidence interval. The second term is a “small sample correction” (which will be discussed later) that comes into play whenever the confidence interval is more than twice as wide as the equivalence range. To examine the relation between the TOST p -value and the SGPV we can calculate both statistics across a range of observed effect sizes. Building on the example by Blume et al. (2018), in Figure 1 p -values are plotted for the TOST procedure and the SGPV. The statistics are calculated for hypothetical one-sample t -tests for observed means ranging from 140 to 150 (on the x-axis). The equivalence range is set to 145 ± 2 (i.e., an equivalence range from 143 to 147), the observed standard deviation is assumed to be 2, and the sample size is 30. For example, for the left-most point in Figure 1 the SGPV and the TOST p -value is calculated for a hypothetical study with a sample size of 30, an observed standard deviation of 2, and an observed mean of 140, where the p -value for the equivalence test is 1, and the SGPV is 0.

Our conclusions about the relationship between TOST p -values and SGPV hold for second generation p -values calculated from confidence intervals, and assuming data is sampled from a bivariate normal distribution. Readers can explore the relationship between TOST p -values and SGPV for themselves in an online Shiny app: http://shiny.ieis.tue.nl/TOST_vs_SGPV/.

The SGPV treats the equivalence range as the null-hypothesis, while the TOST procedure treats the values outside of the equivalence range as the null-hypothesis. For ease of comparison we can plot 1 -SGPV (see Figure 2) to make the values more easily comparable. We see that the p -value from the TOST procedure and the SGPV follow each other closely. When we discuss the relationship between the p -values from TOST and the SGPV, we focus on their correspondence at three values, namely where the TOST $p = 0.025$ and SGPV is 1, where the TOST $p = 0.5$ and SGPV = 0.5, and where the TOST $p = 0.975$ and SGPV = 1. These three values are important for the SGPV because they indicate the values at which the SGPV indicates the data should be interpreted as compatible with the null hypothesis (SGPV =