# Brandt et al. example: equivalent & not significant

**Equivalence bounds based on critical effect sizes**

Another justifiable choice we would like to propose is to use the smallest observed effect size that could have been statistically significant in the original study. Based only on the alpha level and the sample size, we can calculate the criticial test value (e.g., $t$, $F$, $Z$). This critical test value can be transformed to a standardized effect size (e.g., $d_{crit} = t_{crit}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$), which can thus be interpreted as a *critical effect size*[^1]. Observed effect sizes smaller than the critical effect size would not have been statistically significant in the original study. Based on this we can guess that the authors were not interested in effects smaller than this critical effect size, and thus use it as the SESOI. An equivalence test with these bounds can reject all observed effect sizes that the original study had the power to detect.

Banerjee, Chatterjee, & Sinha (2012) reported that participants who had been asked to describe an unethical deed from their past judged the room to be darker than participants who had been asked to describe an ethical deed ($M_{unethical} = 4.71$, $SD_{unethical} = 0.85$, $M_{ethical} = 5.3$, $SD_{ethical} = 0.97$, $t(38) = 2.03$, $p = 0.049$, $d = 0.65$). A close replication by Brandt, IJzerman, & Blanken (2014) found no significant effect ($t(98) = 0.56$, $p = 0.574$, $d = 0.11$). Following the small telescopes approach, we can calculate the effect size the original study had 33% power to detect — $d = 0.49$ — and use this as our SESOI. If we run a TOST with Welch's $t$-test for independent samples and equivalence bounds of $\Delta_L = -0.49$ and $\Delta_U = 0.49$, we indeed find that the effect reported by the replication study is statistically equivalent, $t(97.78) = -1.9$, $p = 0.03$.