

# Six percent power and barely selective inference

Art B. Owen  
Stanford University

WHOA-PSI 2019

It is now a little controversial to test a sharp null hypothesis like  $H_0 : \theta = 0$ . For one thing, we are almost always quite sure that  $\theta \neq 0$  before gathering any data. A classical justification for testing the sharp null anyway, is that once  $H_0$  is rejected we are likely to know the sign of  $\theta$  about which there may have been genuine doubt. By studying what really happens in low power settings, such as 6% power on a test at 5% significance, Andrew Gelman showed that  $H_0$  can frequently be rejected with the wrong value for  $\text{sign}(\theta)$ , as well as grossly exaggerated magnitude  $|\theta|$ . His numerics were not quite right, because sign errors of 24% stem from about 5.5% power but the larger point is valid. A simple, though subtle, remedy is to choose two significance levels  $\alpha_1$  and  $\alpha_S < 1/2$ . After testing  $H_0$  at level  $\alpha_1$ , test it again at the more stringent level  $\alpha_1\alpha_S$ . If the second test rejects, conclude that  $\text{sign}(\theta) = \text{sign}(\hat{\theta})$ , with type I error at most  $\alpha_S$ . If only the first test rejects, conclude that  $\theta \neq 0$  but make no conclusion about  $\text{sign}(\theta)$ . See <https://arxiv.org/abs/1610.10028>.