*Commentary*

# Probability of Replication Revisited

## Comment on "An Alternative to Null-Hypothesis Significance Tests"

**Gheorghe Doros and Andrew B. Geier**

*University of Pennsylvania*

In a recent article, Killeen (2005) proposed a measure of replicability named $p_{\text{rep}}$. In The Framework section, he defined replication as "an effect of the same sign as that found in the original experiment" (p. 346). He defined the probability of a replication (see the middle panel of his Fig 1.) as

$$g(\delta) = P(d'_2 > 0|\delta) \qquad (1)$$

Because $\delta$ is not known, Killeen proposed using $P(d'_2 > 0|d'_1)$ for calculating probability of replication (see the caption of Fig. 1 in Killeen, 2005).

Killeen proposed using this measure as a substitute for the classical $p$ value in reporting experimental results. However, we believe a mistake was made in deriving the main formula of the article:

$$p_{\text{rep}} = P(d'_2 > 0|d'_1) = \int_0^{\infty} n(d'_1,\ 2\sigma^2_{d_1}) \qquad \text{(K1)}$$

(see Equation 4, p. 347, from Killeen, 2005). As Killeen's notation leaves room for varying interpretations as to what $P(d'_2 > 0|d'_1)$ means, we present here an effort to replicate this formula by carrying out calculations under different assumptions. We also point out some mistakes in Killeen's article and end with a discussion about how these mistakes impact the results he reported.

The notations we use are the same as the ones in the original article. We parameterize the normal density as $n(\mu, \sigma^2)$ , where $\mu$ is the mean and $\sigma^2$ is the variance.

Address correspondence to Andrew B. Geier, Department of Psychology, University of Pennsylvania, 3720 Walnut St., Philadelphia, PA 19104-6241; e-mail: andrewbg@psych.upenn.edu.

### CALCULATIONS

First, assume $\delta$ is a constant and $d'_1$ and $d'_2$ are two independent normal replicates with mean $\delta$ and standard deviation $\sigma_d$ (which is Killeen's assumption throughout his Eliminating $\delta$ section).

We consider two proposals for calculating $P(d'_2 > 0|d'_1)$, given this assumption.

- A1: Calculate $P(d'_2 > 0|d'_1)$ as $g(d'_1)$; that is, compute $g(\delta)$ as a function of the unknown $\delta$ and then plug in $d'_1$ for $\delta$. In this scenario,

$$g(\delta) = P(d'_2 > 0|\delta) = P\left(\frac{d'_2 - \delta}{\sigma_d} > \frac{-\delta}{\sigma_d}|\delta\right) = \int_{-\infty}^{\delta/\sigma_d} n(0,1). \qquad (2)$$

By plugging in the observed value of $d'_1$ for $\delta$, we obtain

$$\int_{-\infty}^{d'_1/\sigma_d} n(0,1), \qquad (3)$$

which is exactly $1-p$, where $p$ is the $p$ value for testing $H_0$: $\delta = 0$ versus $H_A$: $\delta > 0$.

- A2: Calculate $P(d'_2 > 0|d'_1)$ as the conditional probability of $\{d'_2 > 0\}$ given the random variable $d'_1$. Under the assumption that the replicates $d'_1$ and $d'_2$ are independent,

$$P(d'_2 > 0|d'_1) = P(d'_2 > 0) = \int_{-\infty}^{\delta/\sigma_d} n(0,1),$$

which is identical to Equation 2. Plugging in the observed value of $d'_1$ for $\delta$, we obtain Equation 3.

A second set of assumptions considered is that $\delta$ is a random variable with mean $a$ and standard deviation $\sigma_\delta$ and that $d'_1|\delta$

and $d'_2|\delta$ are independent normal, random variables with mean $\delta$ and standard deviation $\sigma_d$.

Again we consider two proposals for calculating $P(d'_2 > 0|d'_1)$.

- B1: Calculate $P(d'_2 > 0|d'_1)$ as $g(d'_1)$. The derivations go through as in Proposal A1, with no change because all the calculations are conditional on $\delta$.
- B2: Calculate $P(d'_2 > 0|d'_1)$ as the conditional probability of $\{d'_2 > 0\}$ given the random variable $d'_1$. To make derivations easier, we further assume that $\delta$ is normally distributed. Under these assumptions it can be proven that

$$d'_2|d'_1 \sim N\left(a + \frac{\sigma_\delta^2}{\sigma_\delta^2 + \sigma_d^2}(d'_1 - a), \frac{\sigma_d^4 + 2\sigma_\delta^2\sigma_d^2}{\sigma_\delta^2 + \sigma_d^2}\right),$$

and thus

$$P(d'_2 > 0|d'_1) = \int_{-\infty}^{\mu_c/\sigma_c} n(0,1), \qquad (4)$$

where $\mu_c = a + \frac{\sigma_\delta^2}{\sigma_\delta^2 + \sigma_d^2}(d'_1 - a)$ and $\sigma_c^2 = \frac{\sigma_d^4 + 2\sigma_\delta^2\sigma_d^2}{\sigma_\delta^2 + \sigma_d^2}$. Estimating $a$ with $d'_1$, we get the estimate

$$\int_{-\infty}^{d'_1/\sigma_c} n(0,1). \qquad (5)$$

## TWO ERRORS IN KILLEEN'S ARTICLE

Throughout the Eliminating $\delta$ section, Killeen's assumptions are identical with our first set of assumptions. In Killeen's notation, $d'_2 - d'_1 = \Delta_2 - \Delta_1$, thus $d'_2 = d'_1 + Z$, where $Z$ $(Z = \Delta_2 - \Delta_1)$ is a normal variable that has mean 0 and variance $2\sigma_d^2$. We point out that $Z$ is not independent of $d'_1$; in fact, the correlation between $Z$ and $d'_1$ is $-\sigma_d^2$. Equation K1 makes use of the equality $d'_2 = d'_1 + Z$ and wrongly assumes that $d'_1$ and $Z$ are independent.

The second mistake that we want to point out is in the Parametric Variance section, where Killeen derived his Equation 7. Throughout this section, the assumptions are the same as our second set of assumptions. Under this set of assumptions, unconditionally, $d'_2$ and $d'_1$ are not independent and hence $\sigma_{d_R}^2 = \text{var}(d'_2 - d'_1) \neq \text{var}(d'_2) + \text{var}(d'_1)$. In fact, the correct calculations are

$$\begin{aligned}\sigma_{d_R}^2 &= \text{var}(d'_2 - d'_1) = \text{var}(d'_2) + \text{var}(d'_1) - 2\text{cov}(d'_2, d'_1) \\ &= 2(\sigma_d^2 + \sigma_\delta^2) - 2\sigma_\delta^2 = 2\sigma_d^2.\end{aligned}$$

## DISCUSSION

Killeen (2005) did not clearly spell out mathematically what $P(d'_2 > 0|d'_1)$ means. One could give $P(d'_2 > 0|d'_1)$ the interpretation given in any of our four proposed calculations.

The only assumptions we adopted are elucidated at the beginning of the proposed calculations. We are more inclined to interpret $P(d'_2 > 0|d'_1)$ as a conditional probability than as $g(d'_1)$. With this interpretation, under the first set of assumptions, $P(d'_2 > 0|d'_1)$ should not depend on $d'_1$. Any result of probability calculations involving a function of $d'_2$ conditional on $d'_1$ should not depend on $d'_1$, because $d'_2$ and $d'_1$ are independent (see Dudewicz & Mishra, 1988). The second set of assumptions allows $P(d'_2 > 0|d'_1)$ to be interpreted as a conditional probability and still depend on $d'_1$ (see Equation 4). Neither Equation 2 nor Equation 4 coincides with the formulas that Killeen proposed for $P(d'_2 > 0|d'_1)$.

In an effort to replicate Killeen's formulas, we also looked for estimates. We estimated $\delta$ and $a$ by $d'_1$ in Equations 2 and 4 and obtained Equations 3 and 5, respectively. Again, neither Equation 3 nor Equation 5 coincides with the formulas that Killeen proposed for $P(d'_2 > 0|d'_1)$.

Last, we considered asymptotics, small-sigma asymptotics $(\sigma_\delta \to 0)$ and large-sigma asymptotics $(\sigma_\delta \to \infty)$. By letting $\sigma_\delta \to 0$, Equation 4 reduces to Equation 2, and Equation 5 reduces to Equation 3. By letting $\sigma_\delta \to \infty$, Equations 4 and 5 reduce to $\int_0^\infty n(d'_1, 2\sigma_d^2)$, which is Killeen's proposed formula for the probability of replication when $\sigma_\delta = 0$. For the case $\sigma_\delta \neq 0$, Killeen proposed a formula analogous to K1:

$$P(d'_2 > 0|d'_1) = \int_0^\infty n(d'_1, \sigma_{d_R}^2), \qquad (\text{K2})$$

where $\sigma_{d_R}^2 = 2(\sigma_d^2 + \sigma_\delta^2)$. As we pointed out in the previous section, the variance formula is incorrect, the correct formula being $\sigma_{d_R}^2 = 2\sigma_d^2$. With this correction, Equation K2 is identical to the formula for the conditional probability of $\{d'_2 > 0\}$, given the random variable $d'_1$ as $\sigma_\delta \to \infty$. The asymptotics $\sigma_\delta \to \infty$ correspond to the situation when there exists infinite variability in the effects across the replicated experiments, and this situation does not seem to have been in Killen's thinking because the values of $\sigma_\delta$ considered in his article are smaller than 1.

We appreciate Killeen's discussion about the shortcomings of classical hypothesis testing. However, any measure that is no more than a simple transformation of the classical $p$ value (see Killeen's appendix) will inherit the shortcomings of that $p$ value.

## REFERENCES


Dudewicz, E.J., & Mishra, S.N. (1988). *Modern mathematical statistics* (Probability and Statistics Series). New York: John Wiley and Sons.

Killeen, P.R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science*, *16*, 345–353.



(Received 7/6/05; Accepted 8/8/05; Final materials received 8/18/05)