Notes on a paper of Ioannidis

John D. Cook
cook@mdanderson.org

February 26, 2010

Abstract

This note makes explicit a few calculations that are implicit in the essay *Why Most Published Research Findings Are False* by John P. A. Ioannidis[1].

Ioannidis develops three equations for positive predictive value (PPV) of a finding, the probability that the result is correct, in his essay *Why Most Published Research Findings Are False*. Suppose researchers select their hypotheses for investigation from a pool with a ratio of true hypotheses to false hypotheses equal to R. These hypotheses are tested in experiments with a type I error rate α and a type II error rate β. In the simplest case, one investigator testing one hypothesis with no bias, the PPV is given by

$$\frac{(1-\beta)R}{R - \beta R + \alpha}. \tag{1}$$

If there is bias, some proportion u of results that would otherwise have been found negative are published as positive results. In that case,

$$\frac{(1-\beta)R + u\beta R}{R - \beta R + \alpha + u(1 - \alpha + \beta R)}. \tag{2}$$

If there is no bias, but a hypothesis is investigated by n independent researchers, the PPV becomes

$$\frac{(1-\beta^n)R}{R + 1 - (1-\alpha)^n - \beta^n R}. \tag{3}$$

We only consider equations (2) and (3) since equation (1) is a special case of equation (3) with $n = 1$.

[1] Ioannidis JPA (2005) Why Most Published Research Findings Are False. PLoS Med 2(8): e124. doi:10.1371/journal.pmed.0020124

Ioannidis examines factors that lower PPV, except possibly in the strange case that $1-\beta < \alpha$. (If $1-\beta \geq \alpha$, either the type I or type II error rate is enormous.) He concludes that PPV decreases as a function of u, β, and n. We establish these claims below. We will find the following lemma useful.

Lemma 1. *Define*

$$f(x) = \frac{a+bx}{c+dx}. \tag{4}$$

Then $f(x)$ is a decreasing function if $ad > bc$.

Proof. The derivative

$$f'(x) = \frac{bc-ad}{(c+dx)^2}$$

is negative when its denominator is negative. □

Claim 1. *PPV decreases as a function of u provided $1-\beta > \alpha$.*

Proof. The PPV

$$\frac{(1-\beta)R + u\beta R}{R - \beta R + \alpha + u(1-\alpha+\beta R)}$$

can be put in the form $f(u)$ where f is defined in equation (4) if

$$\begin{aligned} a &= (1-\beta)R \\ b &= \beta R \\ c &= R - \beta R + \alpha \\ d &= (1-\alpha+\beta R). \end{aligned}$$

The function $f(u)$ is decreasing if $ad > bc$. Substitute the definitions of a, b, c, and d and you'll find that $ad > bc$ if and only if $1-\beta > \alpha$.

□

Claim 2. *PPV decreases as a function of β.*

Proof. First consider the case of single testing ($n = 1$) with bias. Then the PPV

$$\frac{(1-\beta)R + u\beta R}{R - \beta R + \alpha + u(1-\alpha+\beta R)}$$

can be put in the form $f(\beta)$ as in equation (4) if

$$\begin{aligned} a &= R \\ b &= (u-1)R \\ c &= R + \alpha + u(1-\alpha) \\ d &= (u-1)R. \end{aligned}$$

The condition $ad > bc$ holds for all values of the parameters.

Next consider the case of multiple testing without bias. Then the PPV

$$\frac{(1-\beta^n)R}{R+1-(1-\alpha)^n-\beta^n R}$$

can be written in the form $f(\gamma)$ where

$$\begin{aligned}
\gamma &= \beta^n \\
a &= R \\
b &= -R \\
c &= R+1-(1-\alpha)^n \\
d &= -R.
\end{aligned}$$

The condition $ad > bc$ always holds, and so f is a decreasing function of γ. Since β^n is an increasing function of β, it follows $f(\beta^n)$ is a decreasing function of β.

Note that we did not need to assume $1-\beta > \alpha$.

□

Claim 3. *PPV decreases as a function of n for $n > 0$ provided $1-\beta > \alpha$.*

Proof. We will prove that the expression for PPV is a decreasing function n as a continuous variable.

Define

$$g(n) = \frac{1-\beta^n}{R+1-(1-\alpha)^n-\beta^n R}.$$

The function $g(n)$ is PPV/R. Since R is a positive constant, it is sufficient to prove $g(n)$ is decreasing.

$$g'(n) = \frac{(1-\beta^n)(1-\alpha)^n\log(1-\alpha)-(1-(1-\alpha)^n)\beta^n\log\beta}{(R+1-(1-\alpha)^n-\beta^n R)^2}.$$

The denominator is positive and so we need only show that the numerator, call it $h(n)$, is negative. Let $x = (1-\alpha)^n$ and $y = \beta^n$. Since we assume $1-\beta > \alpha$, we have $x > y$.

Then

$$h(n) = \frac{(1-y)x\log x-(1-x)y\log y}{n}.$$

To show that $h(n)$ is negative, it is enough to show that

$$\frac{x\log x}{1-x} < \frac{y\log y}{1-y}$$

for $x > y$, or equivalently, that $\varphi(x) = (x \log x)/(1-x)$ is decreasing for $0 < x < 1$.

To show $\varphi(x)$ is decreasing, we show its derivative

$$\varphi'(x) = \frac{\log(x) - x + 1}{(1-x)^2}$$

is negative. The the numerator is negative on $(0, 1)$ because $\log(x) - x + 1$ is an increasing function equal to zero at $x = 1$. The denominator is positive and so $\phi'(x)$ is negative for $0 < x < 1$. $\square$