The positive predictive value

John Ioannides (2005) writes about how most published research findings are false. At the same time, we have learned that if you set your alpha at 5%, the Type 1 error rate (or false positive rate) will not be higher than 5% (in the long run). How are these two statements related? Why aren't 95% of published research findings true?

The trick to understanding this is that two different probabilities are calculated. The Type 1 error rate is the maximum probability saying there is something when there is nothing (e.g., 5%). Ioannides calculates the *positive predictive value* (PPV), which is the post-study probability that a significant finding is true, or true positives/all positive results. Its complement is the *false discovery rate* (FDR), or the false positives/all positive results.

The difference between both these probabilities is that the Type 1 error rate is calculated as a percentage of **all results** (significant and not significant), while the positive predictive value is calculated as a percentage of **all positive results**, ignoring non-significant results. When you perform all studies, you will be aware of all outcomes (both significant and non-significant findings). When you read the literature, where there is publication bias and you often only have access to significant results, keeping the positive predictive value in mind can be useful.

Being post-study probabilities, the PPV and FDR depend on the percentage of studies you do where there is an effect (H1 is true), and when there is no effect (H0 is true), the statistical power, and the alpha level. If you perform 200 tests with 80% power, and 50% (i.e., 100) of the tests examine a true effect, you'll find the following results (in the long run):

	H0 True (50%)	**H1 True (50%)**
Significant Finding (Positive result) α = 5%, 1-β=80%	**False Positive 5%*50%=2.5% (5 studies)**	**True Positive 80%*50%=40% (80 studies)**
Non-Significant Finding (Negative result) 1-α = 95%, β=20%	**True Negative 95%*50%=47.5% (95 studies)**	**False Negative 20%*50%=10% (20 studies)**

For the 85 positive results (80 + 5), the false discovery rate is 5/85 = 0.0588, or approximately 6%. At the same time, the alpha of 5% guarantees that not more than 5% of all the 200 studies are Type 1 errors. This is also true. Of the 200 studies, at most 0.05*200 = 10 will be false positives. In the 200 studies, the Type 1 error rate is only 2.5%. You can redo these calculations by hand (try them for a scenario where the null hypothesis is true in 40% of the studies, and the alternative hypothesis is true in 60% of the studies).

Q1: We see that we control the Type 1 error rate at 5% by using an alpha of 0.05. Still, the Type 1 error rate turns out to be much lower, namely 2.5%, or 0.025. Why?

A) The Type 1 error rate is a variable with a distribution around the true error rate – sometimes it’s higher, sometimes it’s lower, due to random variation.

B) The Type 1 error rate is only 5% when H0 is true for all 200 studies.

C) The Type 1 error rate is only 5% when you have 50% power – if power increases above 50%, the Type 1 error rate becomes smaller.

D) The Type 1 error rate is only 5% when you have 100% power, and it becomes smaller if power is lower than 100%.

We can do these calculations by hand, but there is also a great app, made by Felix Schönbrodt, that calculates these probabilities for us. Go to http://shinyapps.org/apps/PPV/.

Let’s recreate the example we discussed above. On the left, you see some sliders. Set the “% of a priori true hypotheses” slider to 50%. Leave the ‘α level’ slider at 5%. Set the ‘Power’ slider to 0.8 (or 80%). Leave the ‘% of p-hacked studies’ slider at 0.

We get the following results summary:

true positives: 40%; false negatives: 10%; true negatives: 47.5%; false positives: 2.5%

Positive predictive value (PPV): 94.1% of claimed findings are true

False discovery rate (FDR): 5.9% of claimed findings are false

Q2: First, let's just look at the probability that you will find a true positive (which is often a goal in research). What will make the biggest difference in improving the probability that you will find a true positive? Check your ideas by shifting the sliders

A) Increase the % of a-priori true hypotheses

B) Decrease the % of a-priori true hypotheses

C) Increase the alpha level

D) Decrease the alpha level

E) Increase the power

F) Decrease the power

Increasing the power requires bigger sample sizes, or studying larger effects. Increasing the % of a-priori true hypotheses can be done by making better predictions – for example building on reliable findings, and relying on strong theories. These are useful recommendations if you want to increase the probability of performing studies where you find a statistically significant result.

Q3: Set the "% of a priori true hypotheses" slider to 50%. Leave the 'α level' slider at 5%. Leave the '% of p-hacked studies' slider at 0. The title of Ioannidis' paper is 'why most published research findings are false'. One reason might be that studies often have low power. At which value for power is the PPV 50%. In other words, at which level of power is a significant result just as likely to be true, as that it is false?

A) 80%

B) 50%

C) 20%

D) 5%

It seems low power alone is not the best explanation for why most published findings are false. Ioannidis (2005) discusses some examples where it becomes likely that most published research findings are false. Some of these assume that 'p-hacked studies', or studies that show a significant result due to bias, enter the literature. There are good reasons to believe this happens, as we discuss in the section on flexibility in the data analysis. In the 'presets by Ioannidis' dropdown menu, you can select some of these situations. Explore all of them, and pay close attention to the ones where the PPV is smaller than 50%.

Q4: In general, when are most published findings false? Interpret 'low' and 'high' in the answer options below in relation to the values in the first example in this assignment of 50% probability H1 is true, 5% alpha, 80% power, and 0% bias.

A) When the probability of examining a true hypothesis is low, combined with either low power or substantial bias (e.g., p-hacking).

B) When the probability of examining a true hypothesis is high, combined with either low power or substantial bias (e.g., p-hacking).

C) When the alpha level is high, combined with either low power or substantial bias (e.g., p-hacking).

D) When power is low and p-hacking is high (regardless of the % of true hypotheses one examines).

Q5: Set the "% of a priori true hypotheses" slider to 0%. Set the "% of p-hacked studies" slider to 0%. Set the "α level" slider to 5%. Play around with the power slider. Which statement is true? Without *p*-hacking, when the alpha level is 5%, and when 0% of the hypotheses are true, ____

A) the Type 1 error rate is 100%.

B) the PPV depends on the power of the studies.

C) regardless of the power, the PPV equals the Type 1 error rate.

D) regardless of the power, the Type 1 error rate is 5%, and the PPV is 0%.

Conclusion

People often say something like: "*Well, 1 in 20 results in the published literature are Type 1 errors*". After this assignment, you should be able to understand this is not true in practice. When in 100% of the studies you perform, the null hypothesis is true, and all studies are published, only then 1 in 20 studies, in the long run, are Type 1 errors (and the rest correctly reveals no statistically significant difference). In the scientific literature, the positive predictive value (the probability that given that a statistically significant result is observed, the effect is true) can be quite low, and under specific circumstances, it might even be so low that most published research findings are false. This will happen when researchers examine mostly studies where the null-hypothesis is true, with low power, or when the Type 1 error rate is inflated due to p-hacking or other types of bias.

You should not try to directly translate your Type 1 error rate (e.g., 0.05) into the probability that the alternative hypothesis is true, when a significant result has been observed. To make a statement about the probability that a theory is true, you need Bayesian statistics. *P*-values do not tell you the probability that the alternative hypothesis is true.

The probability of saying there is an effect, when there is no effect (or the Type 1 error rate), is not directly related to the probability that a significant *p*-value means a true effect (or the PPV). The Type 1 error rate and the PPV can be completely unrelated. So don't think that a 5% Type 1 error rate means that it is 95% likely that a significant result is a true effect. Publication bias, power, and Type 1 error rates together determine the probability that significant results in the literature reflect true effects.