All about the p-value

https://aeon.co/essays/it-s-time-for-science-to-abandon-the-term-statistically-significant

> "Tests of statistical significance proceed by calculating the probability of making our observations if there were no real effect. This isn't an assertion that there is no real effect, but rather a calculation of *what would be expected if* there were no real effect. The postulate that there is no real effect is called the null hypothesis, and the probability is called the *p*-value.
>
> The problem is that the *p*-value gives the right answer to the wrong question. What we really want to know is *not* the probability of the observations given a hypothesis about the existence of a real effect, but rather the probability that there *is* a real effect, given the observations."

i.e.: we want *P(Real effect | observations)* not *P(Observations | Real effect).* Assume for simplicity that in the event space of effects, *real effect* and *no effect* are mutually exclusive.

The *p*-value is *not*:

- the probability that your observations occurred by chance.
- A measure of the odds of H_0 to H_1
- A direct frequentist error rate

Simply speaking: a 5% significance threshold will flag out an immense number of false positives if the probability of the real effect, e.g. prevalence of some rare disease, is very rare. In general, we won't know the real prevalence of real effects (i.e. the prior) – so even if we can compute a *p*-value, we can't compute the number of false positives. E.g. let P(Disease) = 0.01, in a population of 10,000. Specificity (false negative rate) = 0.95, and sensitivity = 0.8. Then:

	Test positive	Test negative	
Disease	80	20	100
No disease	495	9,405	9,900

FDR = "false positive rate" ~ 86%

Admittedly, though Prof. Colquhoun says that these numbers are "disastrous", I think that these are pretty good numbers in practice – given that false positives are less important than false negatives.

Overall, he mentions controlling for FDR, instead of just using the p-value – i.e. report the chances that you've found a false positive, rather than the chance that you observe what the data tells you, given a hypothesis (effect). I'm guessing that Benjamini-Hochberg does this.

The American Statistical Association's Statement on p-values (2016)

The 6 principles of the *p*-value:

1. The *p*-value can indicate how compatible the data are with a specified statistical model
2. *P*-values do *not* measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
3. Scientific conclusions should not be based only on whether the *p*-value passes a specific threshold.
4. Proper inference requires full reporting and transparency.
5. A *p*-value, or statistical significance, does not measure the size of an effect or the importance of a result
6. By itself, a *p*-value does not provide a good measure of evidence regarding a model or a hypothesis.