Week 05: Genetic variation & Quantitative genetics

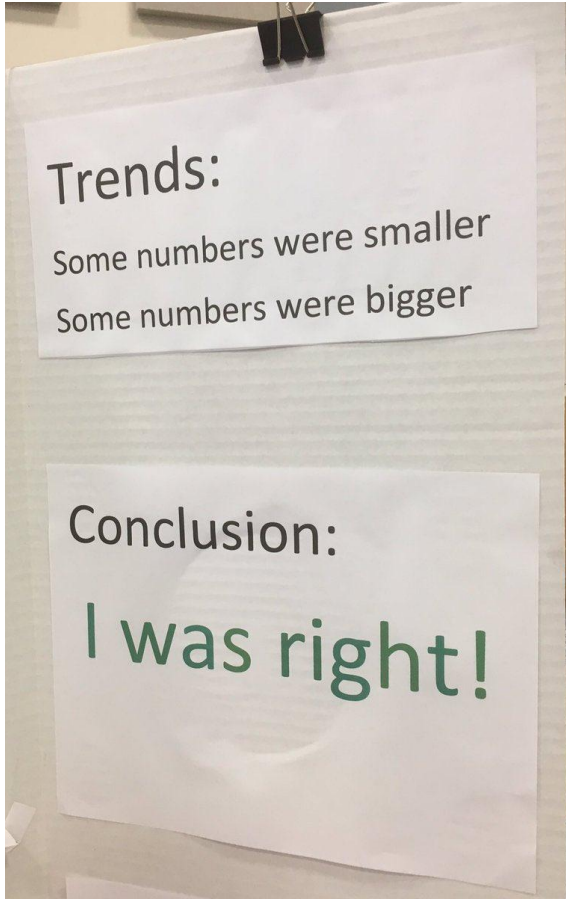
- Genome-wide association studies
  - Regularized linear regression
  - Polygenic risk score
  - Statistical inference, P-values, & Multiple hypothesis testing

Statistical hypothesis testing

Consider two competing hypotheses for a given SNP:

- Null hypothesis: the frequency of the SNP in the cases is the same as that in controls.
- Alternative hypothesis: the frequencies are different.

There's always some difference → Is it statistically significant difference?



Statistical hypothesis testing

1. **Decide on the effect** that you are interested in, design a suitable experiment or study, pick a data summary function and test statistic.
2. **Set up a null hypothesis**, a computationally tractable model of reality that lets you compute the null distribution, i.e., the possible outcomes of the test statistic and their probabilities under the assumption that the null hypothesis is true.
3. **Decide on the rejection region**, i.e., a subset of possible outcomes whose total probability is small.
4. **Do the experiment** and collect the data, compute the test statistic.
5. **Make a decision**: reject the null hypothesis – i.e. conclude that it is unlikely to be true – if the test statistic is in the rejection region.

Test statistic

	$X_n = 0$	$X_n = 1$	$X_n = 2$	**Totals**
$Y = 0$	O_{00}	O_{01}	O_{02}	$O_{0.}$
$Y = 1$	O_{10}	O_{11}	O_{12}	$O_{1.}$
Totals	$O_{.0}$	$O_{.1}$	$O_{.2}$	S

Pearson's χ^2 test

$$X^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

One-sample test

$$t = \frac{\bar{x} - \mu_0}{\text{SEM}}$$

Two-sample test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

P-value

The next step is to perform a statistical hypothesis test and get a **p-value**.

The p-value is:

a) The amount of evidence that there is an effect?

b) The probability that the observed outcome is important?

c) The probability that the SNP is not associated with the trait/disease?

The p-value is the probability that the experiment would have produced the observed outcome (or something more extreme) even if the SNP is not associated with the trait/disease.

How would you write code to simulate two distributions and calculate p-values using a permutation test.

P-value – Calculation using a permutation test

The p-value is the probability that the experiment would have produced the observed outcome (or something more extreme) even if the SNP is not associated with the trait/disease.

1. Given SNP & phenotype data, calculate the real test statistic.
2. Repeat the following 100,000 times to set up the null hypothesis for this test statistic:
    - Randomly assign phenotype measures to individuals.
    - Record the test statistic of the permuted assignments.
3. Calculate the p-value of the real test statistic.

Type I & type II errors

P-value captures if there is “sufficient" inconsistency with the null hypothesis.

Choosing $p < \alpha$ controls type I error at α.

- Type I error: False-positive rate (α)
- Type II error: False-negative rate (β)
- Remember the story of the boy that cried wolf!

David Robinson
@drob

Follow

Remember, mixing up Type I and Type II errors is called a Type III error

David Robinson
@drob

Follow

Giving mistakes numbers instead of names was a real Type IV error

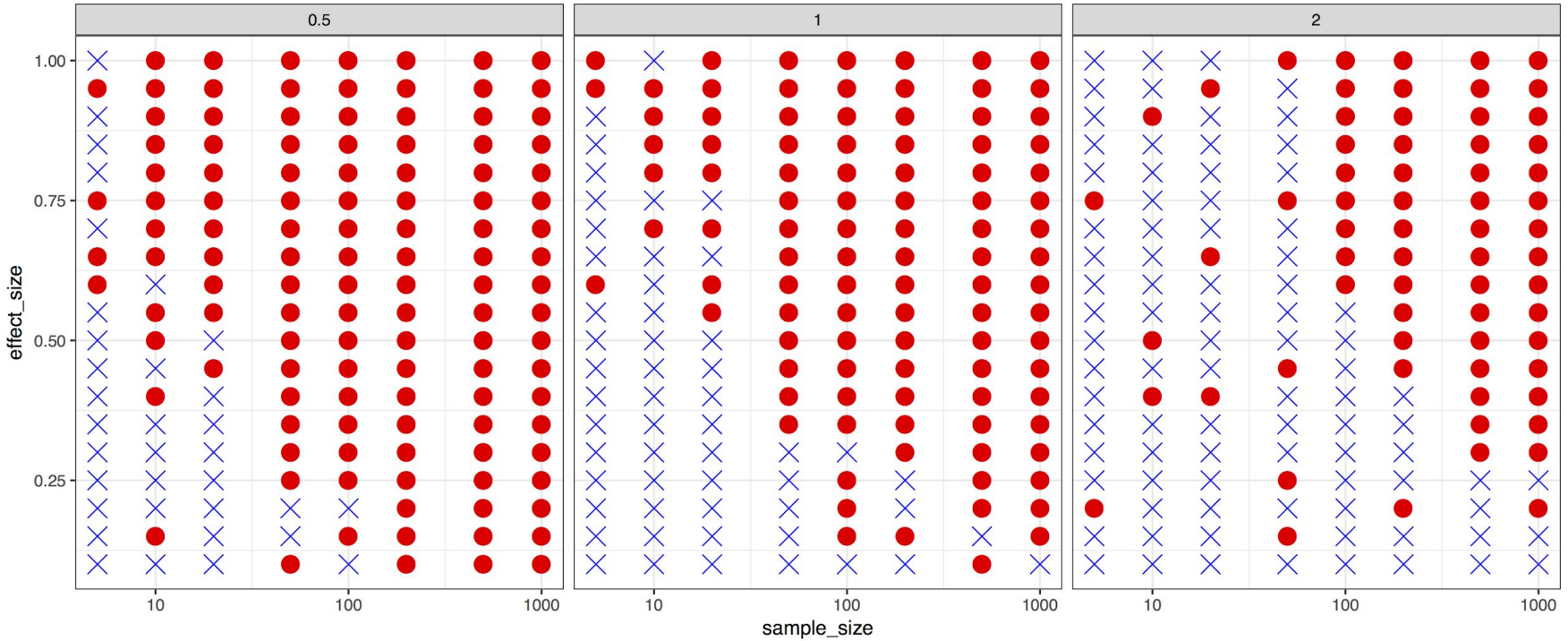
P-value

P-values are dependent on:

- Size of the effect (effect size)
- Sample size
- Variance within each group
- The underlying experimental design & the null hypothesis (need not always be random chance).
    a. Conversely, two completely different experiments can give same data but end up very different p-values.
        - 3 out of 9: Binomial p-value = 0.073; Neg. Binomial p-value = 0.033.

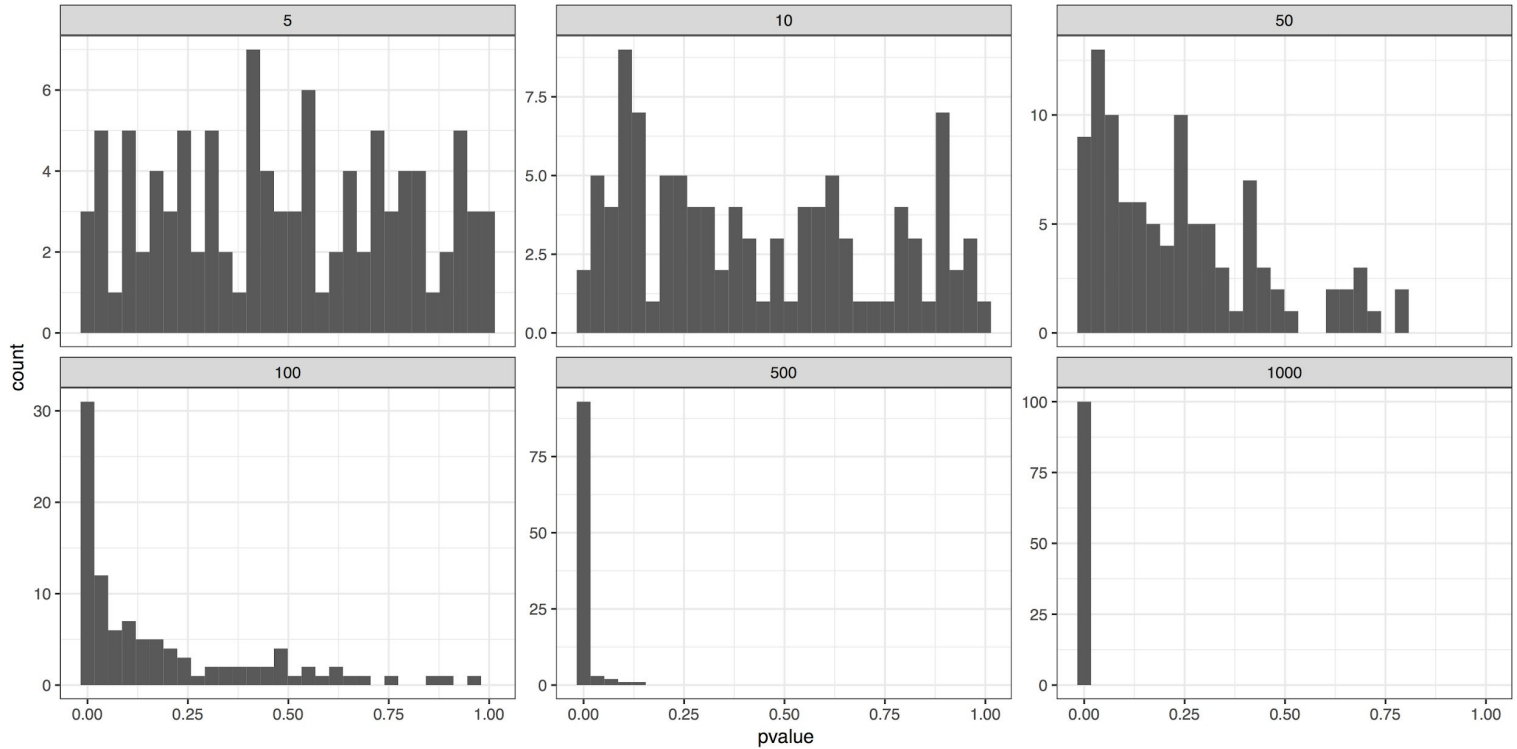
P-value

P-values are dependent on: sample size, effect size, within-group variance.



P-value

P-values are dependent on: sample_size (effect_size = 0.25, std_deviation = 1)



P-value – Significant or not?

- Fisher (1920s):
  - Informal method to help interpret the data along with prior experience, domain knowledge, size of the effect, etc.
- Neyman & Pearson:
  - Control false positive rate at α, set by the experimenter based on what can be tolerated.
  - Formulate null and alternative hypothesis.
  - Reject null when $p < \alpha$.
    - The threshold $\alpha = 0.05$ is merely a convention.

P-value – Significant or not?

This list is culled from peer-reviewed journal articles in which:

a) the authors set themselves the threshold of 0.05 for significance,

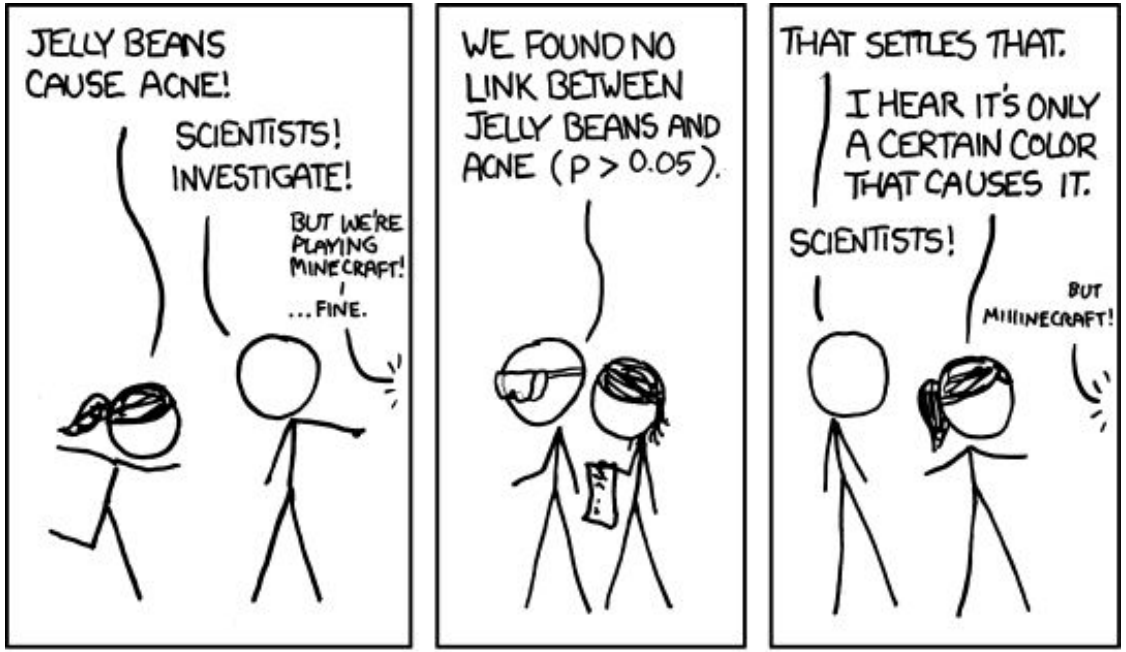
b) failed to achieve that threshold value for p and

c) described it in such a way as to make it seem more interesting.

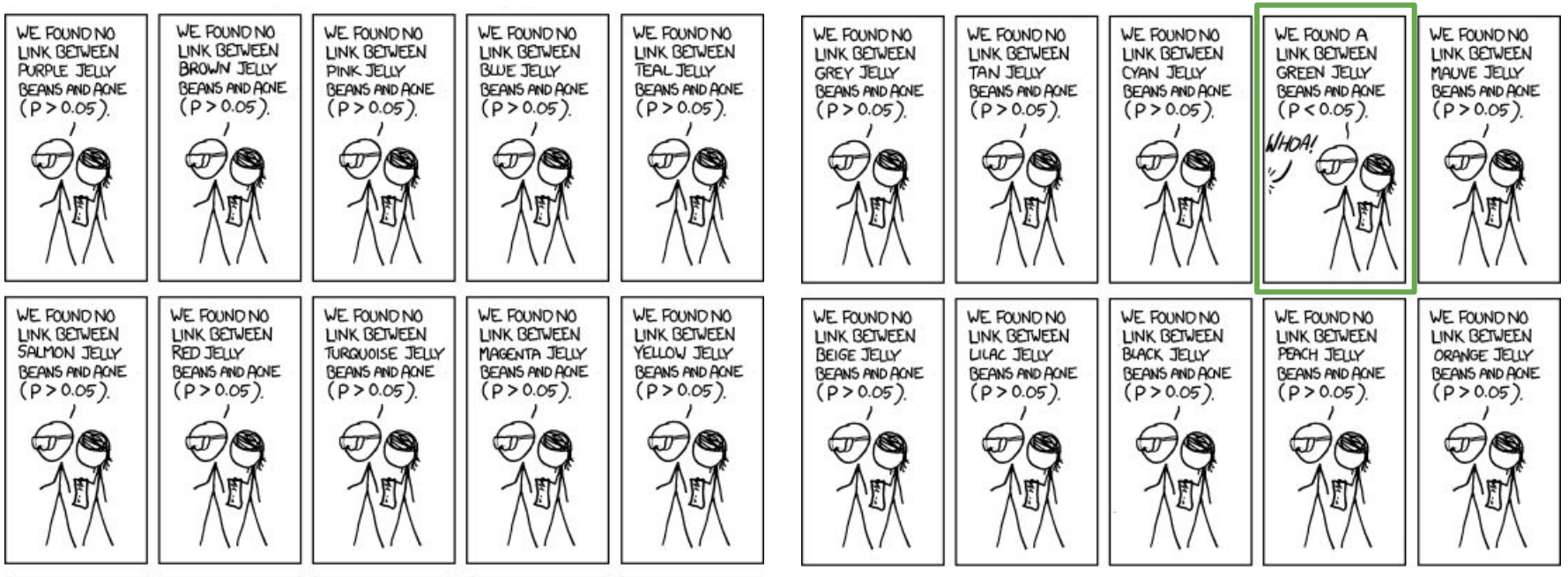
(barely) not statistically significant (p=0.052)
a barely detectable statistically significant difference (p=0.073)
a borderline significant trend (p=0.09)
a certain trend toward significance (p=0.08)
a clear tendency to significance (p=0.052)
a clear trend (p<0.09)
a clear, strong trend (p=0.09)
a considerable trend toward significance (p=0.069)
a decreasing trend (p=0.09)
a definite trend (p=0.08)
a distinct trend toward significance (p=0.07)
a favorable trend (p=0.09)

https://mchankins.wordpress.com/2013/04/21/still-not-significant-2/

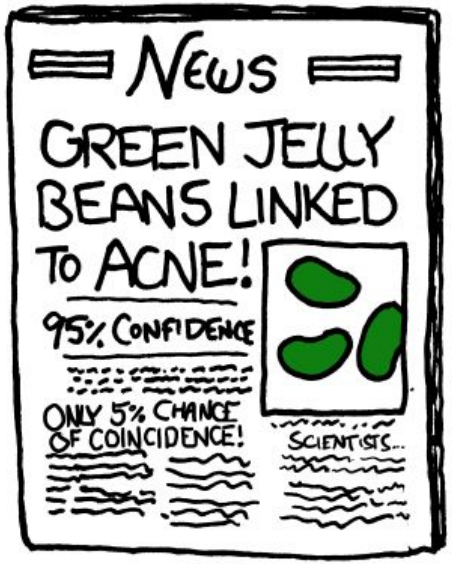
Multiple hypothesis testing



Multiple hypothesis testing



Multiple hypothesis testing



The more inferences are made, the more likely erroneous inferences are to occur.

Let α be the Type 1 error rate for a statistical test.

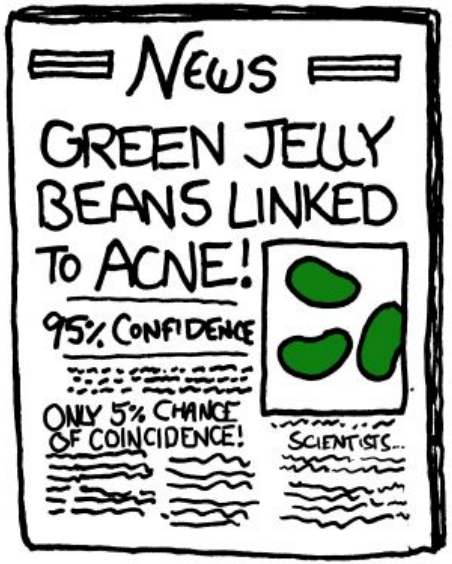
If the test is performed n times, what is the experimental-wise error rate α'? (Same as: What is the probability of obtaining at least 1 FP?)

$\alpha' = 1 - (1 - \alpha)^n$ (Check for $\alpha = 0.05$ & $n = 5$.)

The result may not be that significant even if its p-value $< \alpha$.

To solve this problem, the nominal p-value need to be corrected/adjusted.

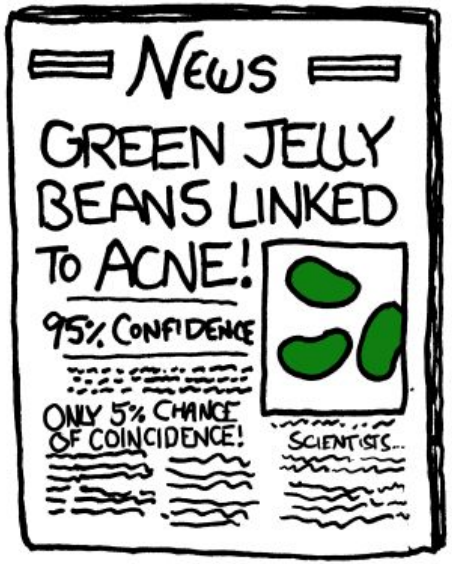
Correcting for multiple hypothesis testing



Controlling for **Family-wise Error Rate**
(FWER: the probability of at least 1 FP):

- Bonferroni correction:
  - $p'_i = p_i * n$ (permutation test)
- Permutation test:
  - Permute the data K times, each time calculate minimum p-value
  - p'_i = #{min_pvalue < p_i} / K

Correcting for multiple hypothesis testing

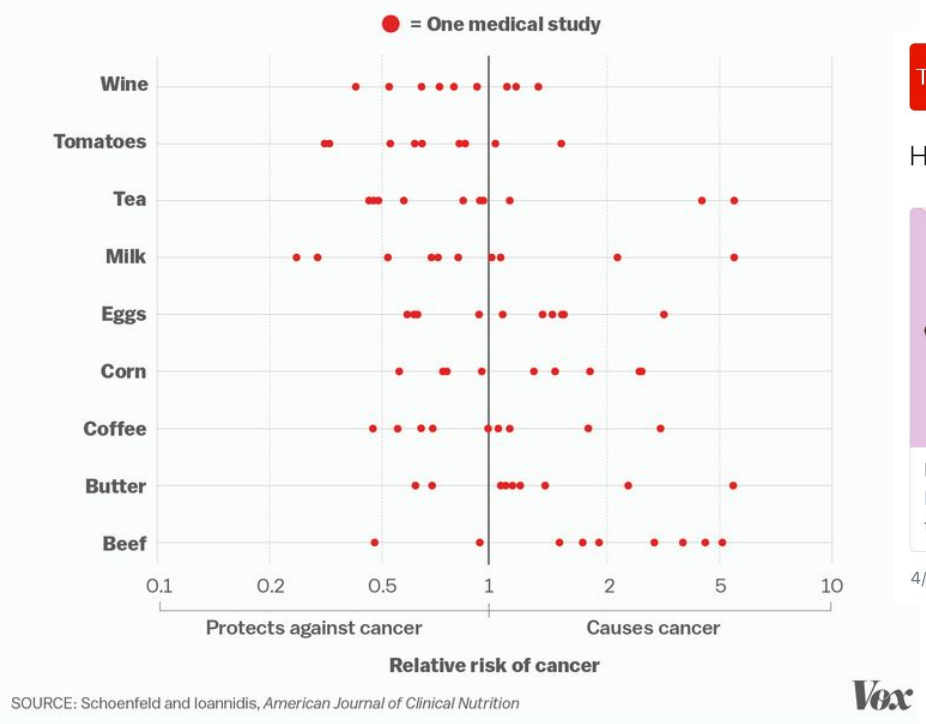


Controlling for **False Discovery Rate**
(FDR: proportion of FP among all significant hypotheses):

- Benjamini-Hochberg correction:
  - $p'_i = p_i * (n / i)$

Multiple hypothesis testing

Publication bias (studies with nonsignificant results have lower publication rates)



Questionable research practices

- Exclusively using p-values to determine the relevance and sanity of the results of a statistical test.
- Analyzing the data until the desired results are found.
- Collecting more data to reach smaller p-values.
- Trying many hypothesis until one of them gives a low p-value, and reporting just that final result.