College of Natural Sciences

Type M Error: Consequences of Selection Bias when the Null is False

Nathan Mitchell, *Ben Prytherch*

"Why Most Published Research Findings are False"

Ioannidis[1]: Highly influential paper, states research often cannot be replicated because studies are filled with false findings caused by selective reporting.

Small sample sizes, small true effect sizes, and increased design flexibility (among other factors) can lead to increased rates of false findings.

Why Not Type I and Type II?

Type I errors, False Positives, and Type II errors, False Negatives, rely on the assumption that the difference in means is truly zero, but this assumption is not always realistic[3].

Ioannidis[1] concludes most findings are due to an error being made, so what is of interest is how exaggerated our (incorrect) measurement is, not the probability it was an error.

Statistical Power

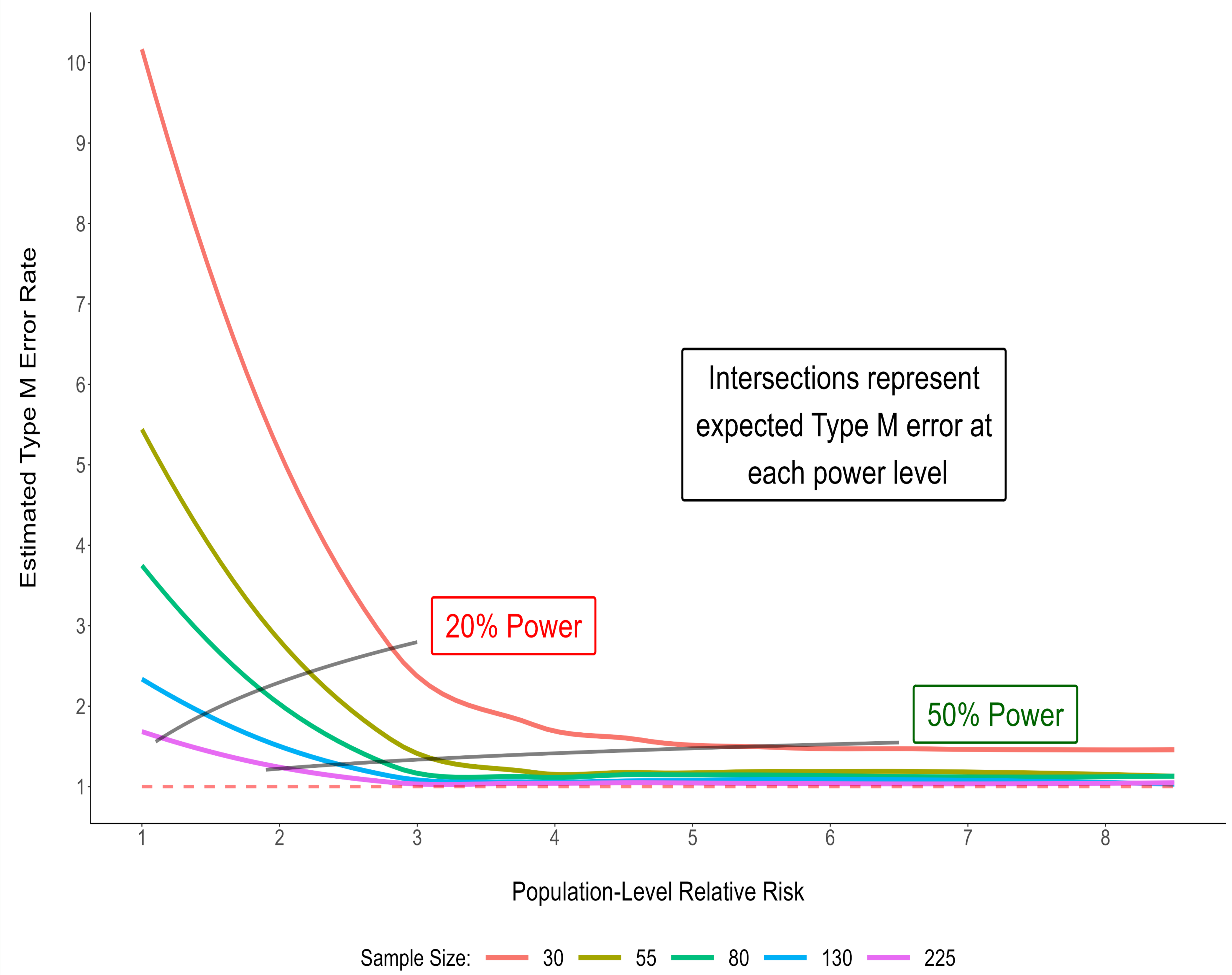
The Type M error is a function of power. As power increases, estimates become less exaggerated.

$$Power \;=\; P(Detecting\; Effect \mid Effect\; Exists)$$

Gelman & Carlin[2] note that power much below 50% will result in a large exaggeration of the estimate while power above 50% will not be as catastrophic.

Estimating Type M Error

The plot below features estimates of the Type M error for increasing sample sizes against the population-level relative risk. As the size of the real effect increases, the Type M error decreases because the effect becomes easier to detect and the estimate, in turn, becomes less exaggerated. When a median unbiased estimator of relative risk is used, the Type M error will approach one, as denoted by the dashed red line.



Note: The line representing the Type M error for a sample size of 30 doesn't converge to 1. This is because Relative Risk is asymptotically unbiased, meaning bias is reduced for larger sample sizes. A method was used to correct for this, known as the median unbiased estimator, but this method does not perform well when sample sizes are low. Because of this, the size of the error is exaggerated.

Relative Risk

A way to quantify effect size when a model has a binary outcome. Can be defined as:

$$RR = \frac{P(outcome \mid exposed)}{P(outcome \mid not\ exposed)}$$

Expressed as values greater than 0, with 1 representing no elevated risk for the exposed group and values greater than 1 representing an increased risk for the exposed group.

Type M Error

Introduced by Gelman & Carlin[2], this value measures how much a given estimate is expected to differ from the parameter value given the result found was significant.

$$Type\ M\ Error \;=\; \frac{RR_{expected\,\mid\,p\,<\,0.05}}{RR_{expected}}$$

Larger Type M errors indicate a more exaggerated (significant) effect size estimate.

Conclusions

When power is low and researchers select for significance, there can be dire consequences, though this selection isn't always done explicitly[4].

Regardless, effect size estimates will be exaggerated if significance is selected for. For larger sample sizes and higher power this exaggeration will not be as extreme, but the bias will still be present.

[1] Ioannidis J. P. (2005). Why most published research findings are false. *PLoS medicine*, *2*(8), e124
[2] Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on psychological science : a journal of the Association for Psychological Science*, *9*(6), 641–651.
[3] Meehl, P. E. (1967). Theory-Testing in Psychology and Physics: A Methodological Paradox. *Philosophy of Science*, *34*(2), 103–115.
[4] Gelman, A., & Loken, E. (2019). The garden of forking paths : Why multiple comparisons can be a problem , even when there is no " fishing expedition " or " p-hacking " and the research hypothesis was posited ahead of time.

CURC 2023 Celebrate Undergraduate Research and Creativity

COLORADO STATE UNIVERSITY