Effects of publication bias and hidden multiplicity on reproducibility in biomedical discovery research

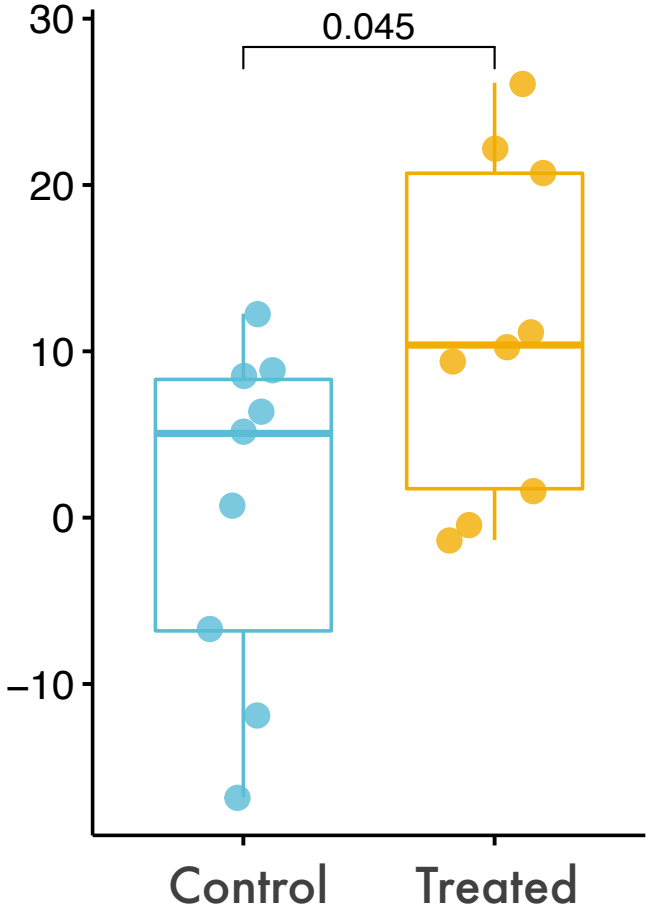
Ondrej Libiger

October 19th

NCS 2022

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

What is a good sample size for replicating a study?



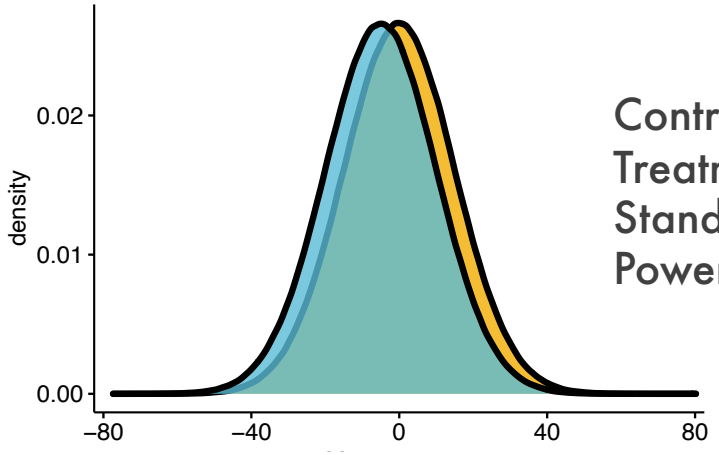
Published study:

- Compare 9 treated and 9 control mice
- Observed difference between means: 10 units
- Observed mean standard deviation: 10 units
- P-value: 0.049

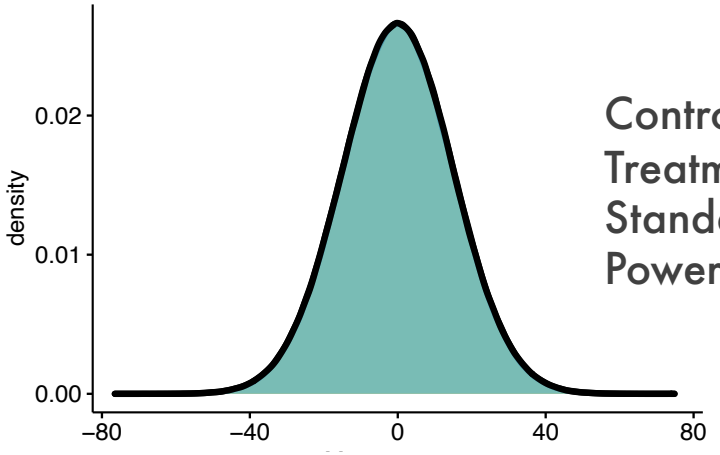
Problem:

How many animals should we use to achieve 80% power to detect a potential positive difference as statistically significant?

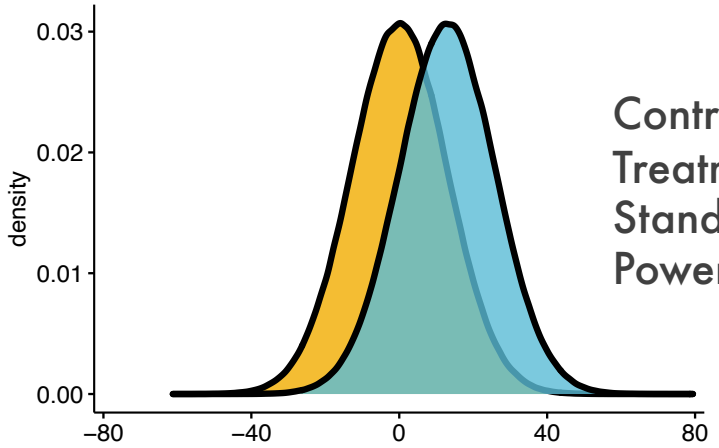
The observed results can arise from a variety of models



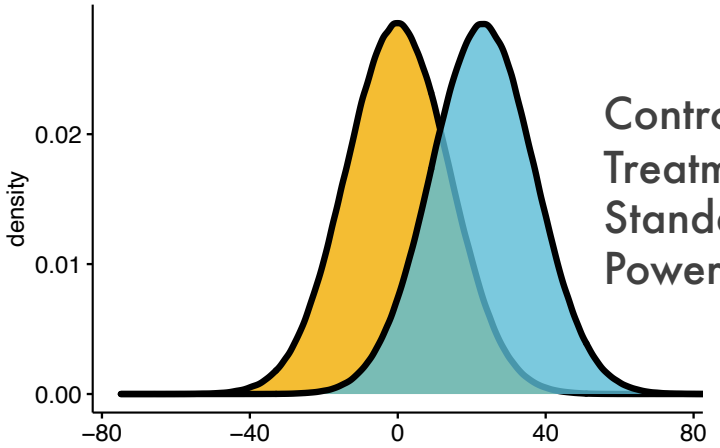
Control: N(0, 15)
Treatment: N(-5, 15)
Standard effect: -0.33
Power (N=12). 0.7%



Control: N(0, 15)
Treatment: N(0, 15)
Standard effect: 0
Power (N=12): 5%



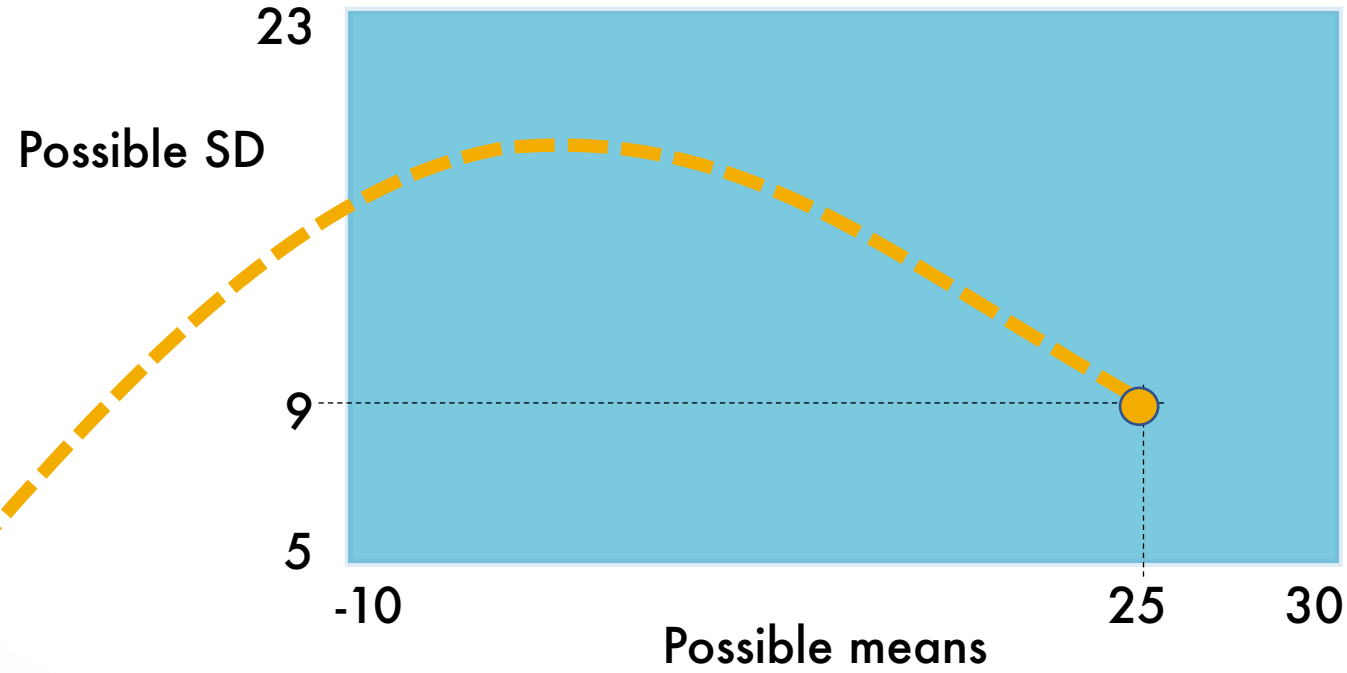
Control: N(0, 13)
Treatment: N(13, 13)
Standard effect: 1.0
Power (N=12): 77%



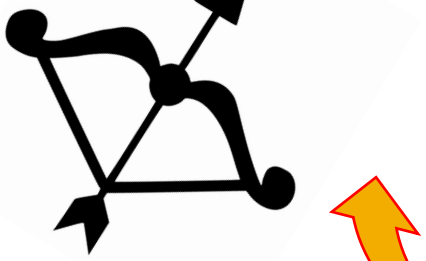
Control: N(0, 14)
Treatment: N(23,14)
Standard effect: 1.64
Power (N=12): 99%

janssen | PHARMACEUTICAL COMPANIES OF Johnson&Johnson

We used simulations to explore all possible model parameters



1

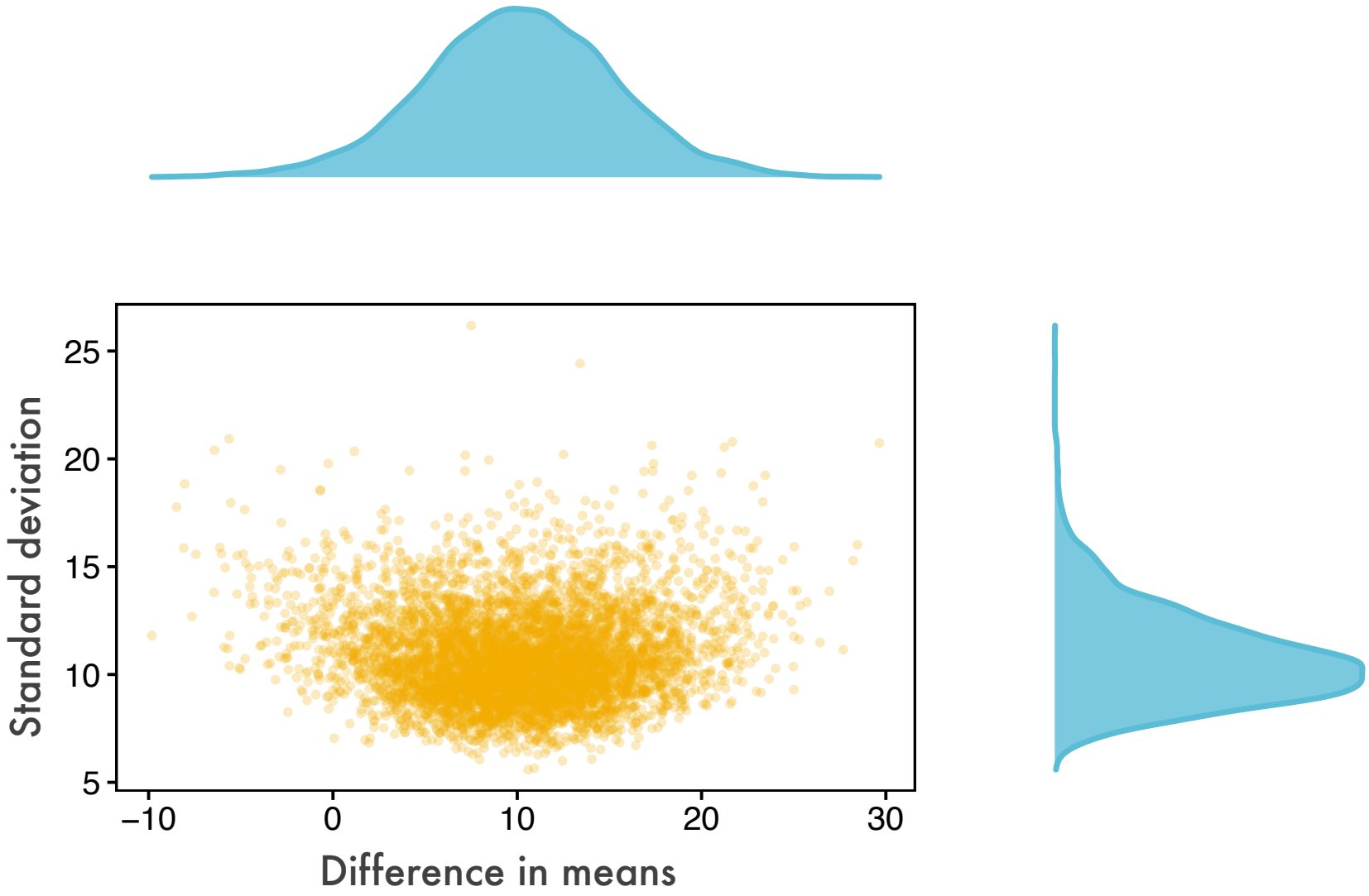


2

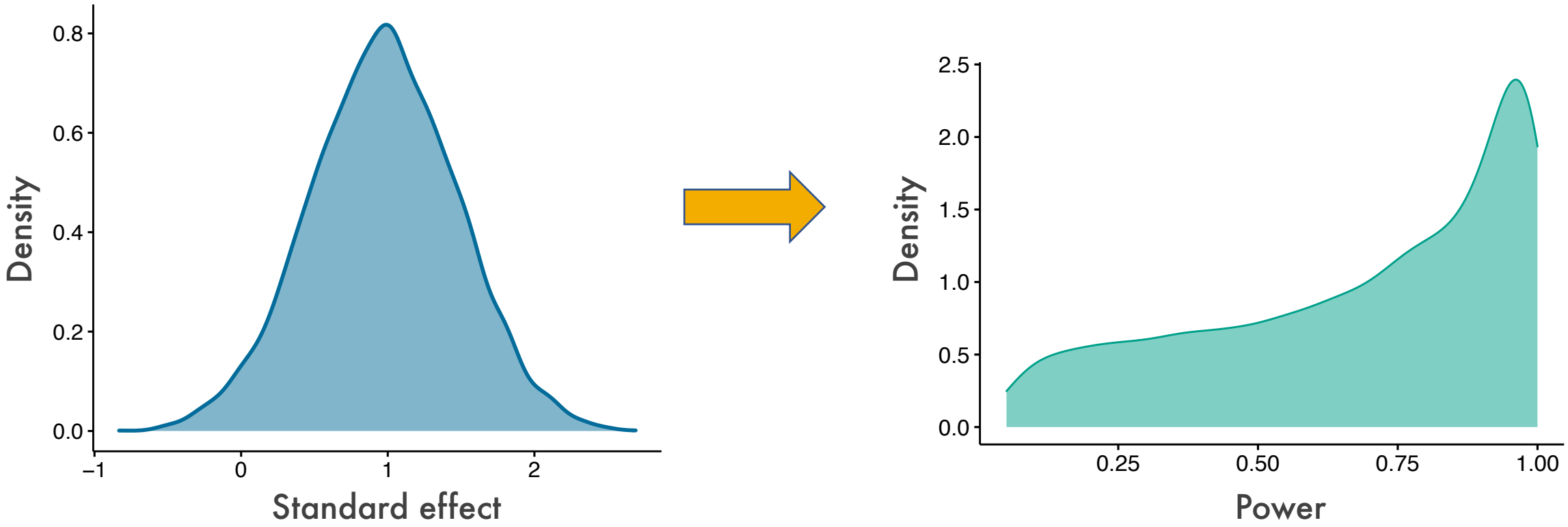
Draw a sample from:
1. N(25, 9) Treated
2. N(0, 9) Controls
Compare with prior results

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

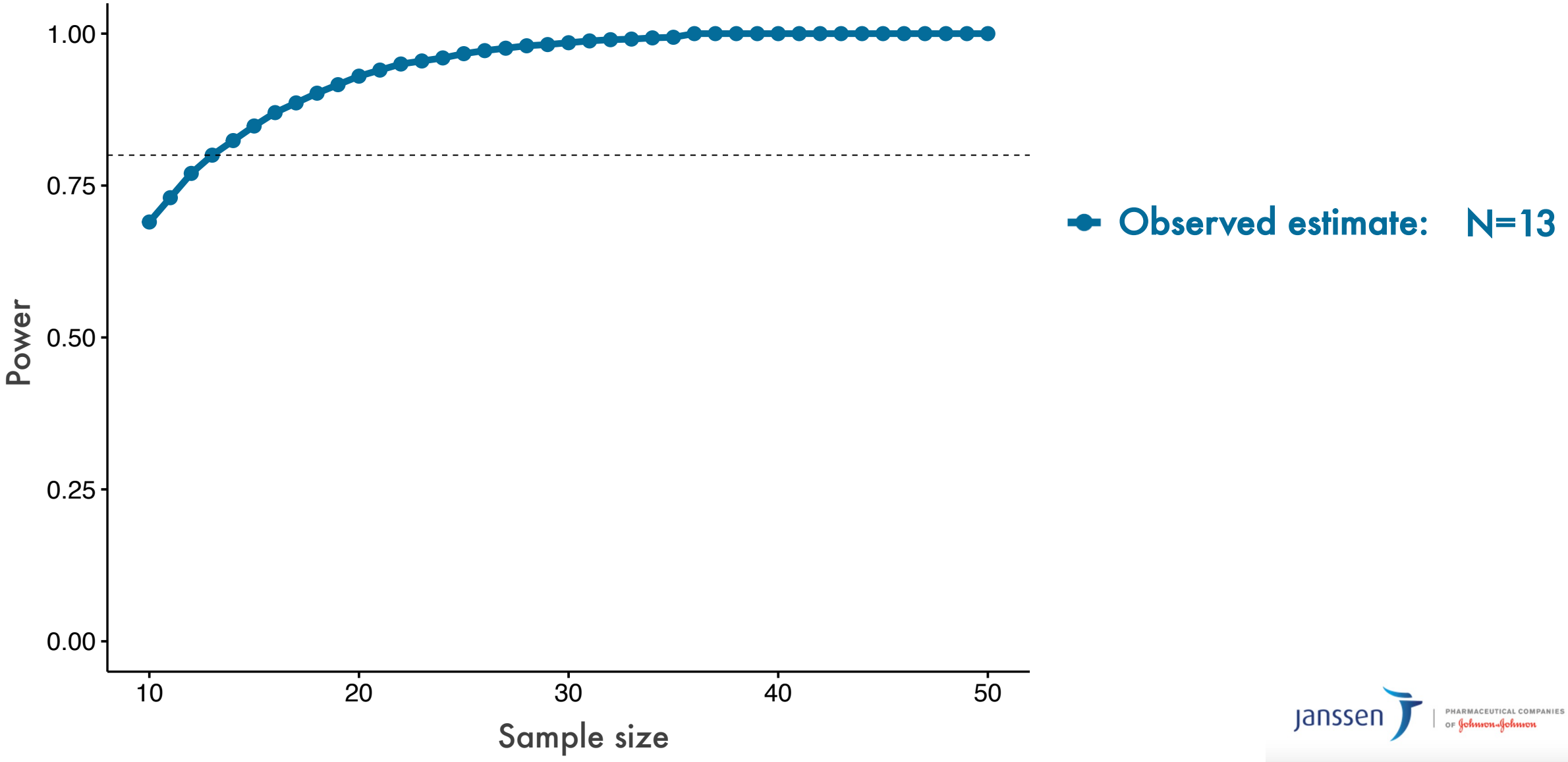
Different sets of model parameters can yield observed results



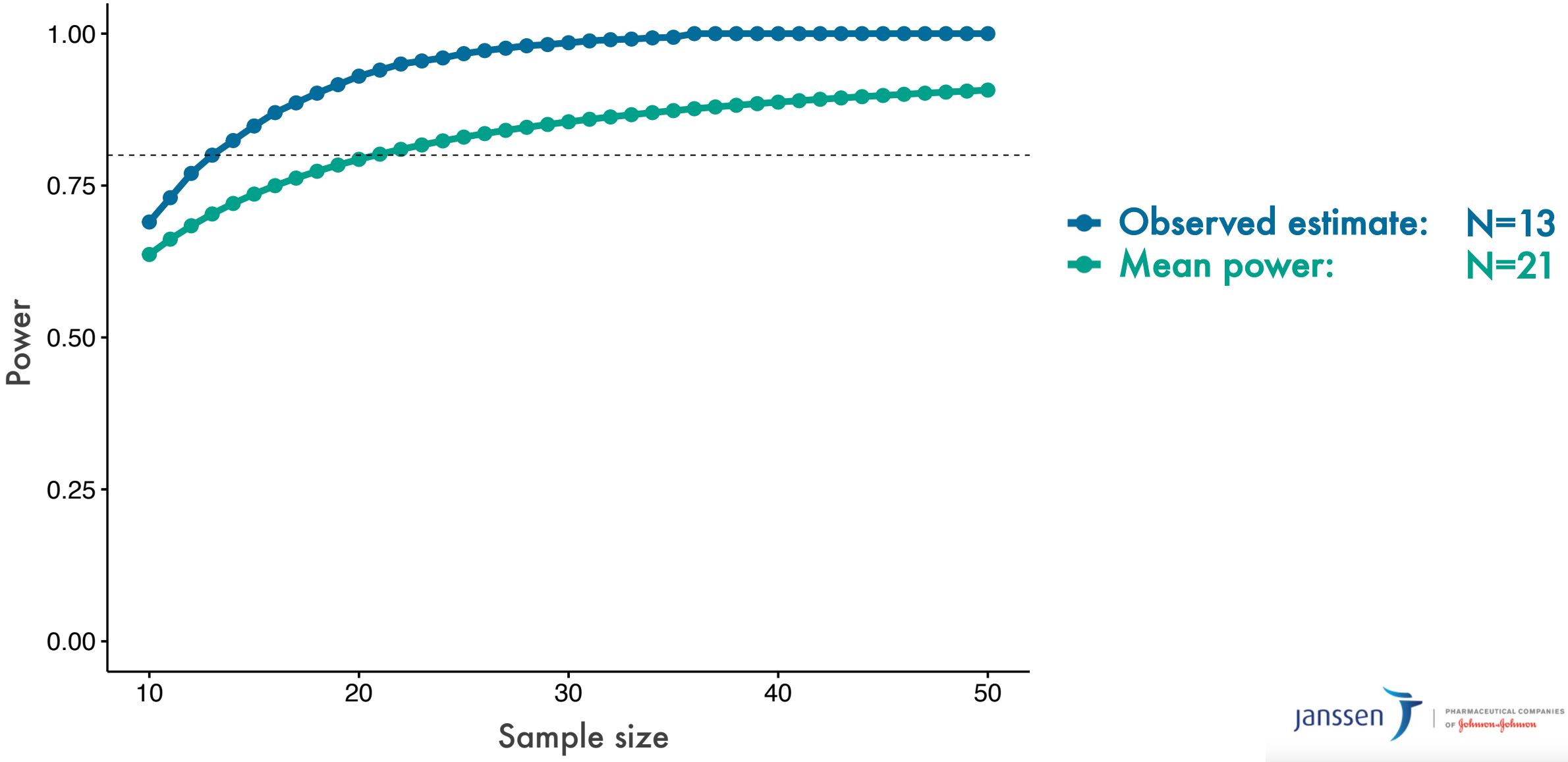
Treatment effects and power vary for each model



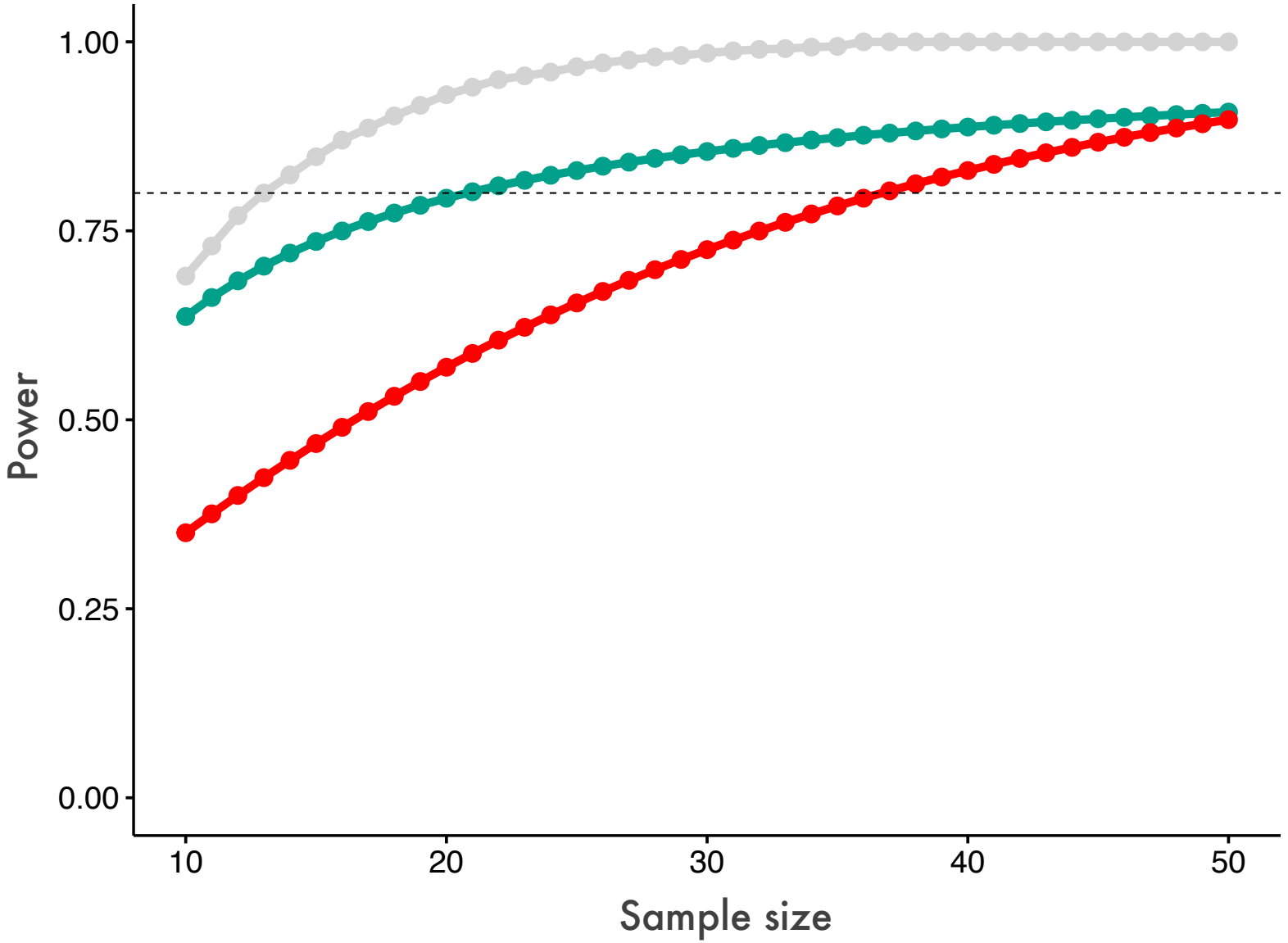
Most calculations assume observed effect is 'true' effect



Considering all possible models makes a difference



We can quantify the uncertainty of power calculations



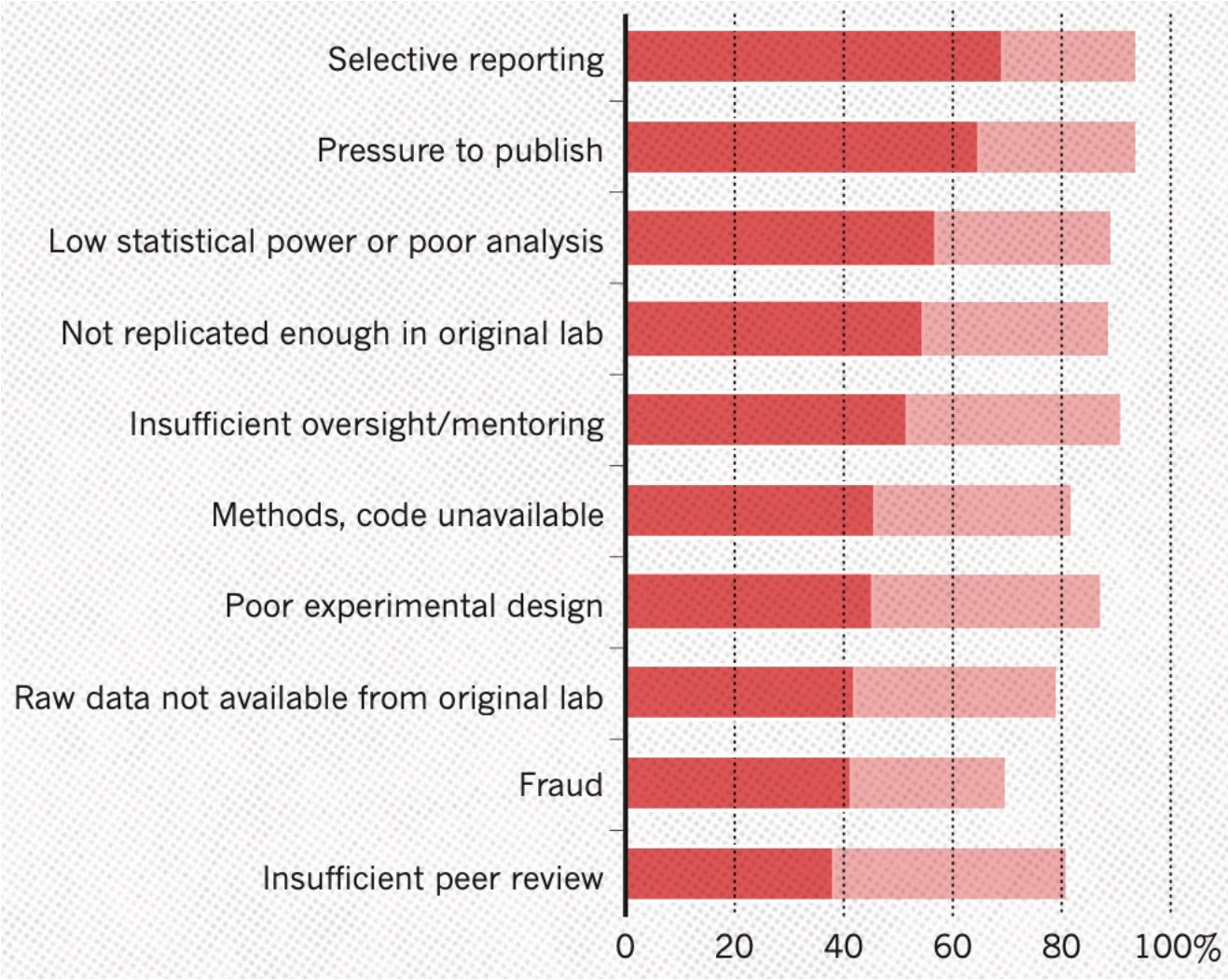
Observed estimate: N=13

Mean power: N=21

20th percentile: N=37

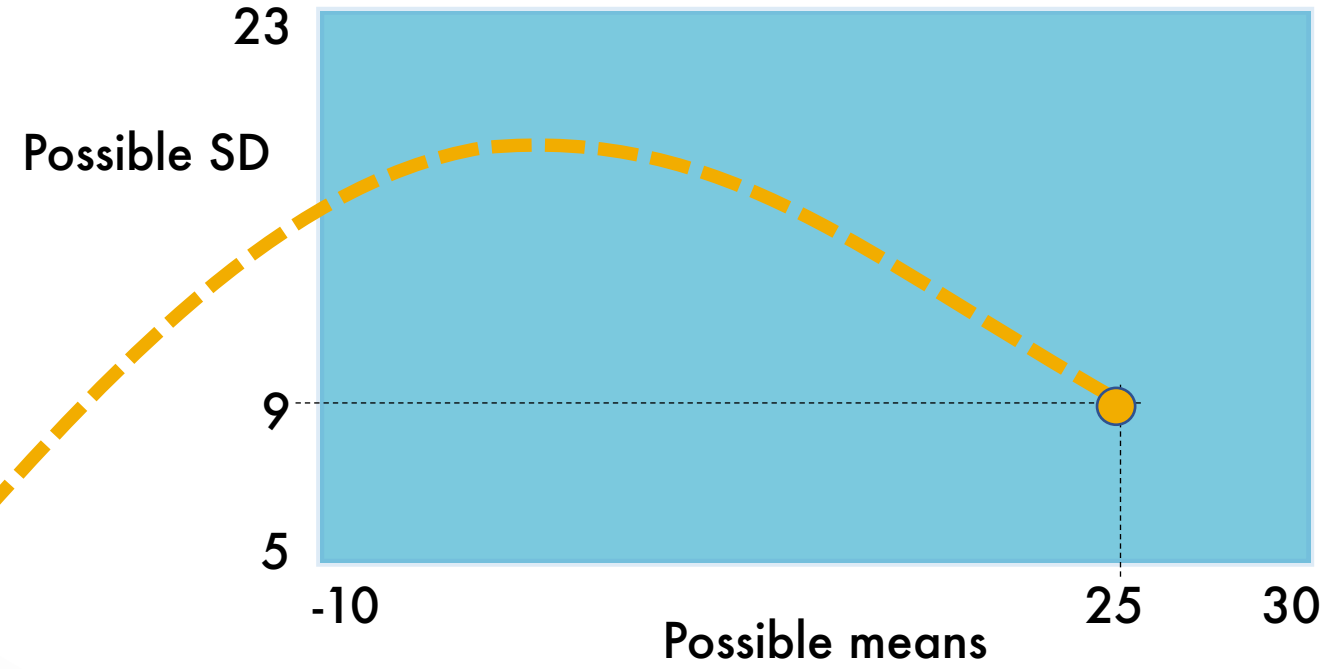
Publication bias and hidden multiplicity skew results

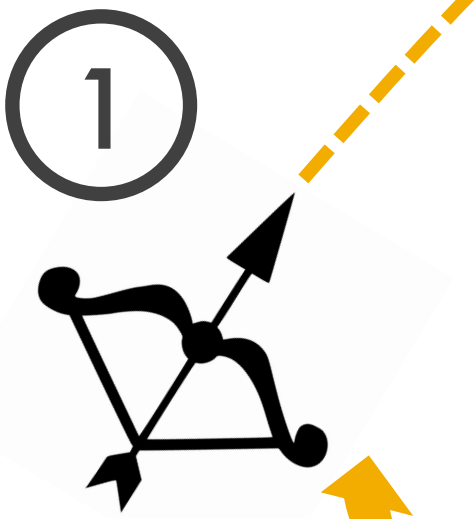
- 1M biomedical research papers published each year (2/minute)
- 7,600 medical research organizations in US alone
- Employ >100,000 people in US; >400,000 in Europe



Baker M., Is there a reproducibility crisis? Nature v. 533, 2016

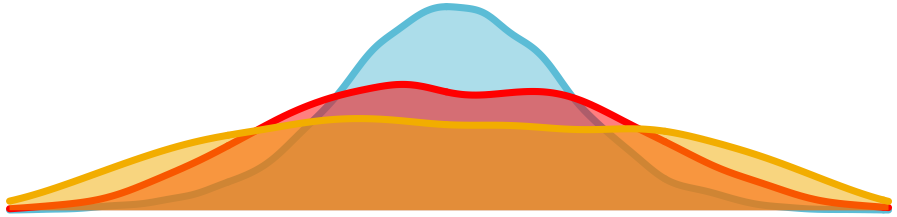
We modified our simulations to explore the effect of bias

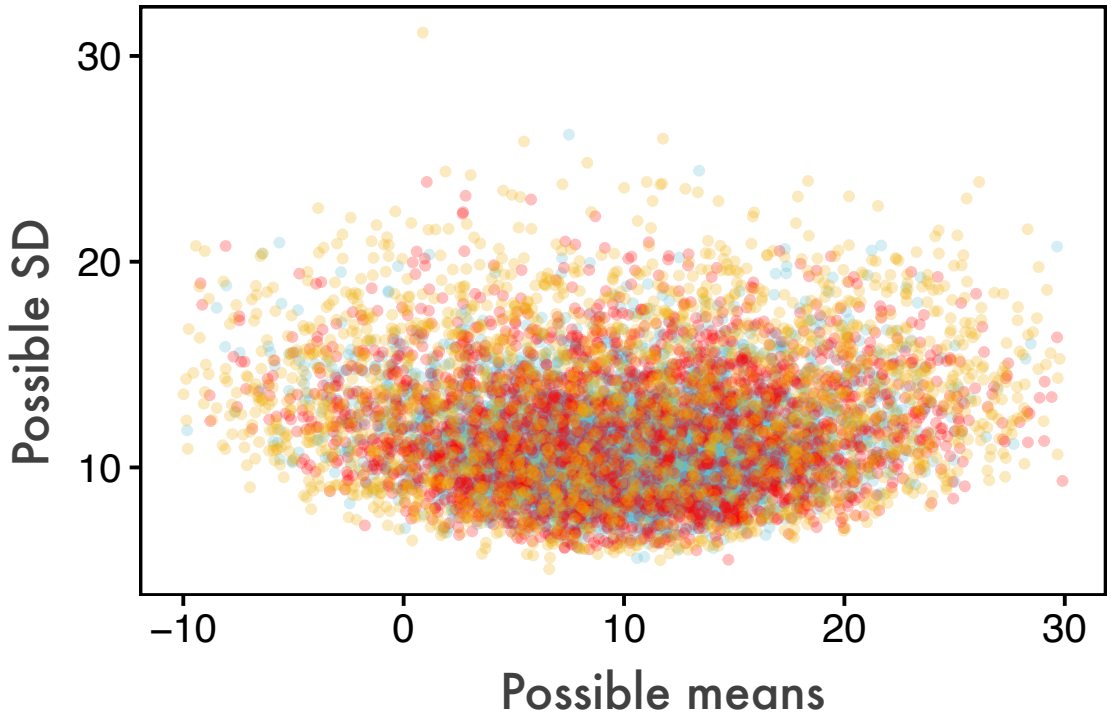


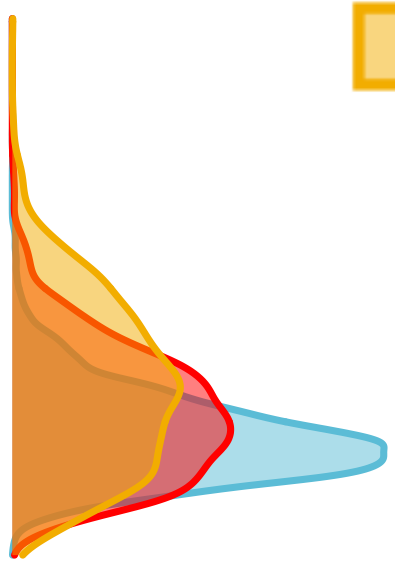


Draw many samples from:
1. N(25, 9) Treated
2. N(0, 9) Controls
Compare with prior results

Bias changes the distribution of possible parameters







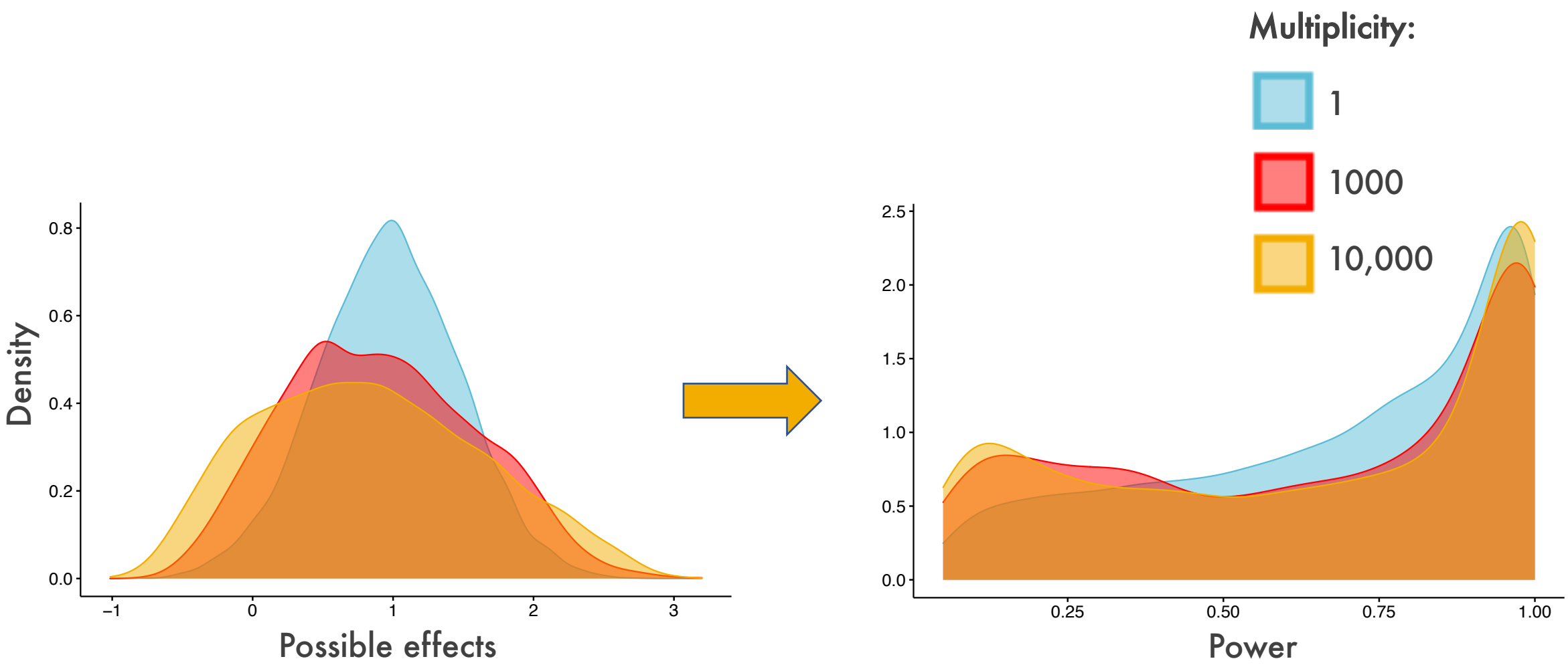
Multiplicity:

1

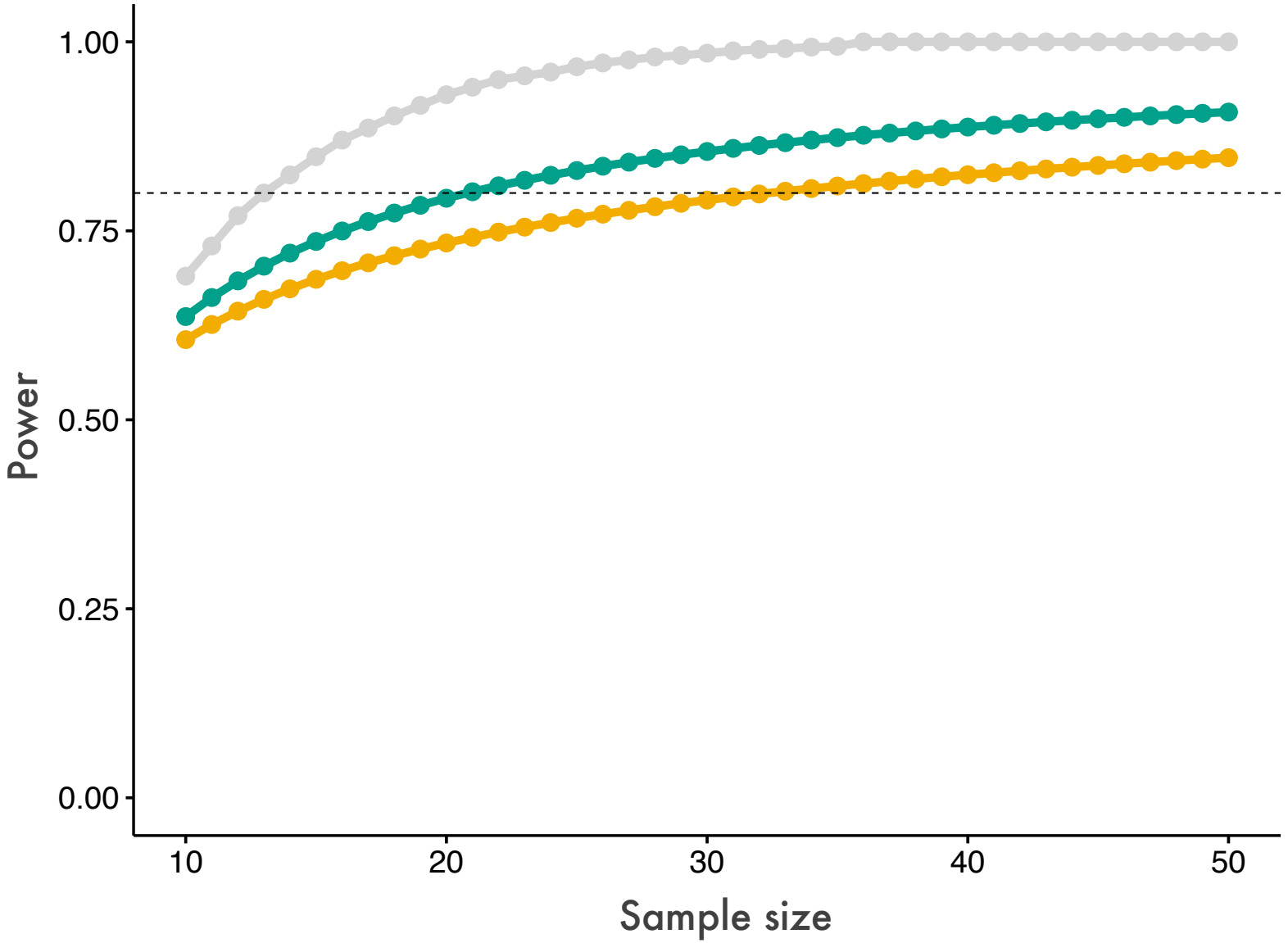
1000

10,000

Bias changes the distribution of possible effects and power



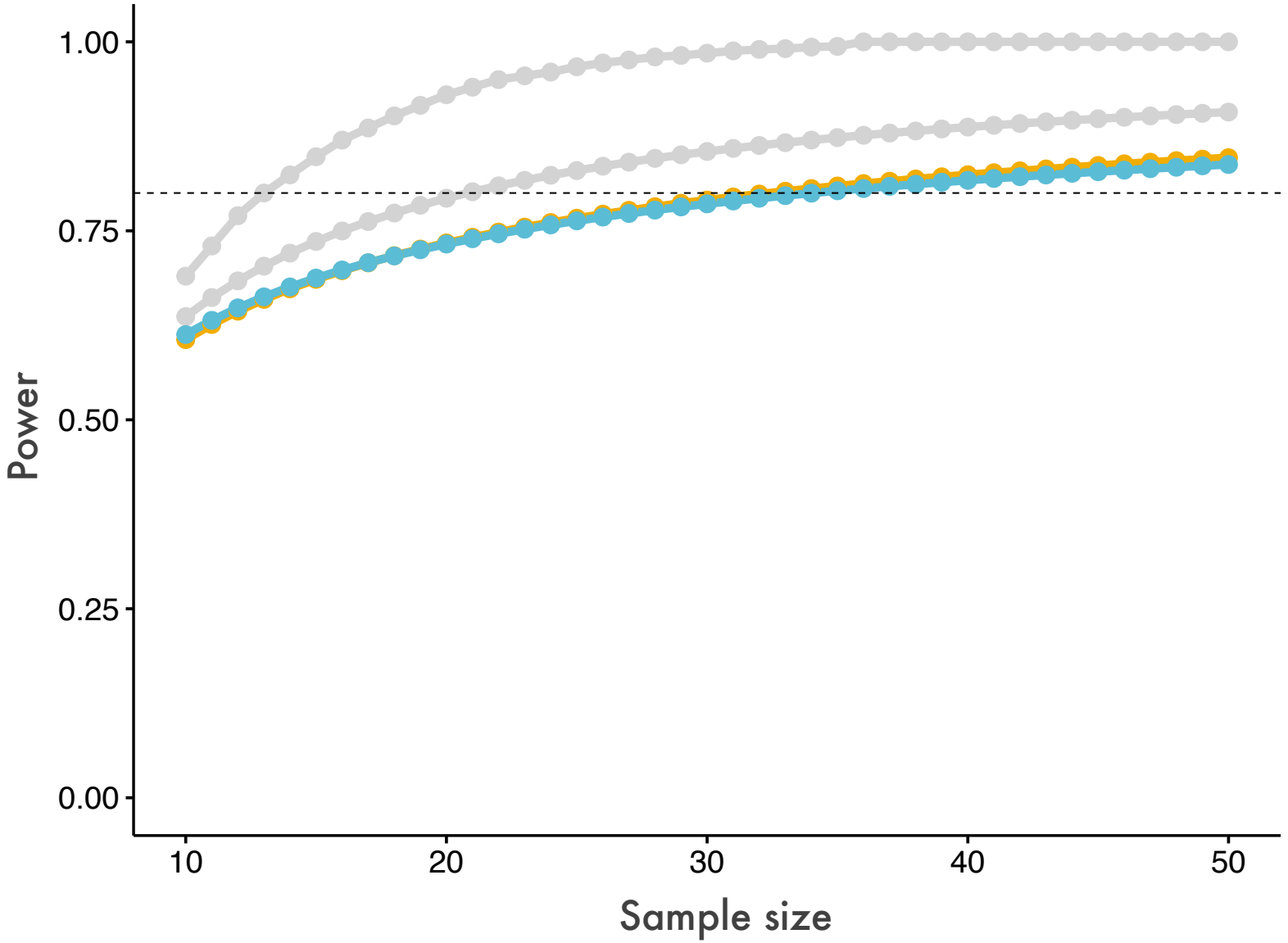
Publication bias necessitates larger samples



Observed estimate: N=13
Mean power: N=21
Low bias: N=32

janssen | PHARMACEUTICAL COMPANIES OF Johnson&Johnson

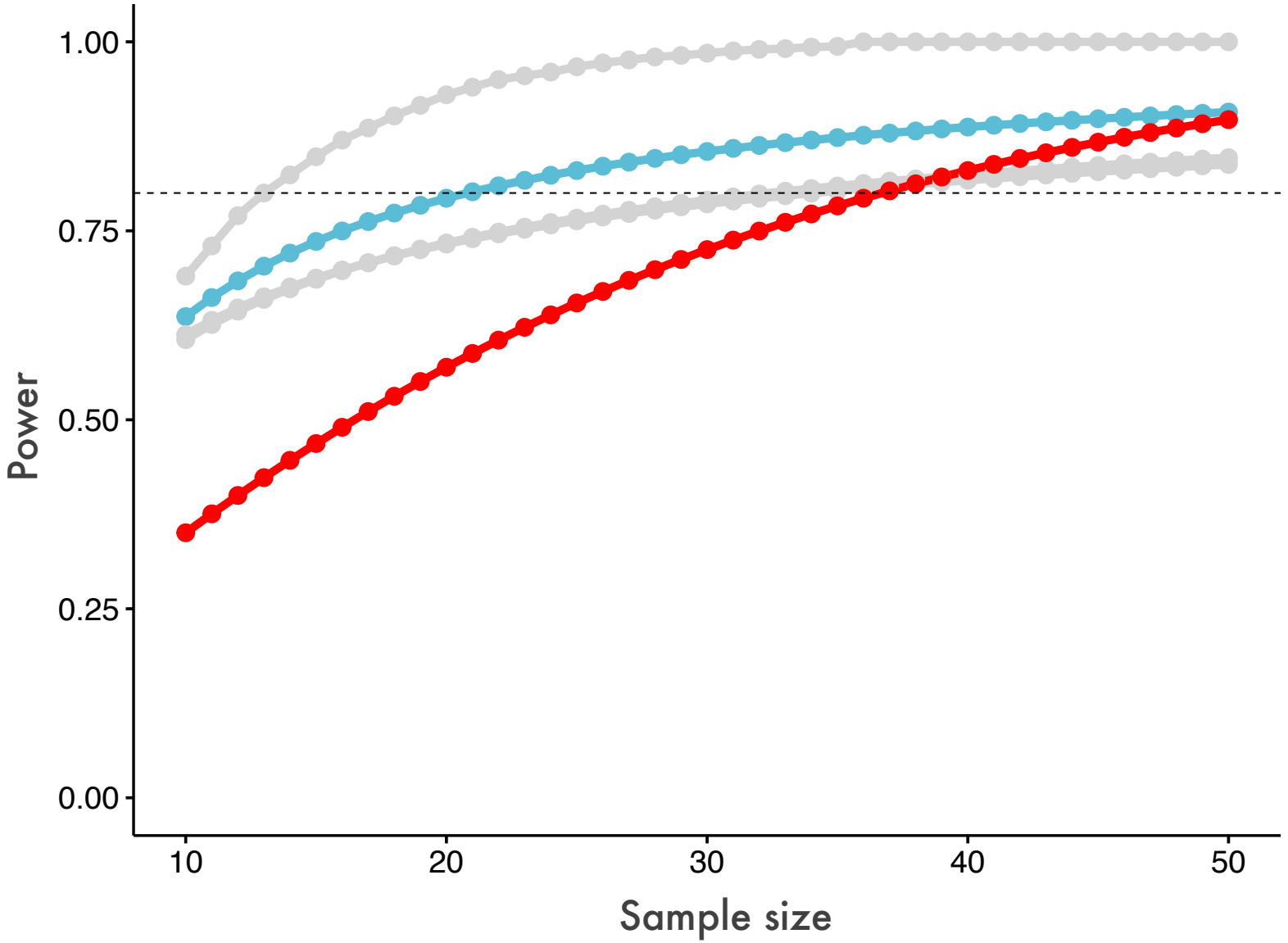
Publication bias necessitates larger samples



Observed estimate: N=13
Mean power: N=21
Low bias: N=32
High bias: N=34

janssen | PHARMACEUTICAL COMPANIES OF Johnson&Johnson

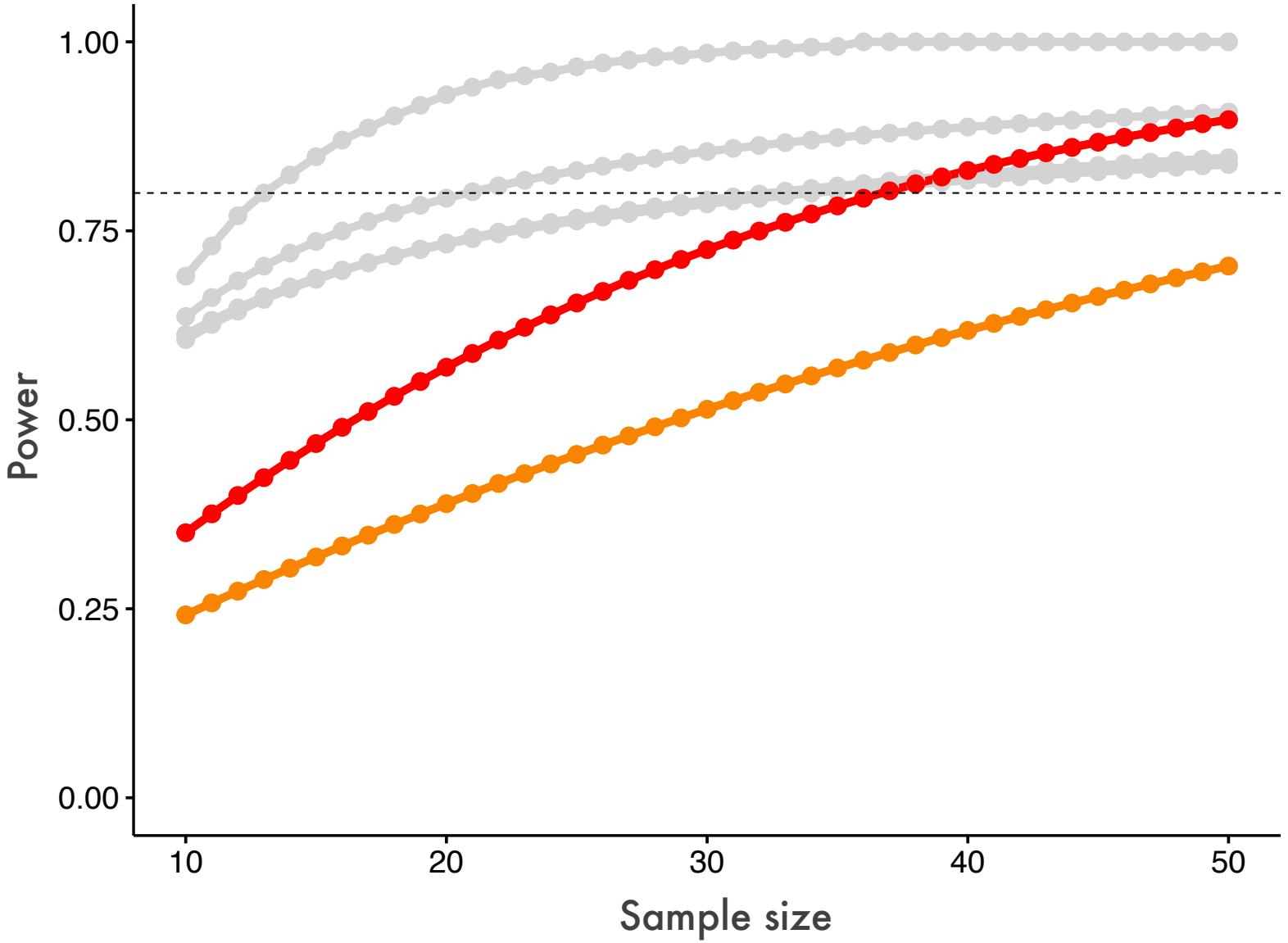
Publication bias necessitates larger samples



Observed estimate: N=13
Mean power: N=21
Low bias: N=32
High bias: N=50
20th percentile: N=37

janssen | PHARMACEUTICAL COMPANIES OF Johnson&Johnson

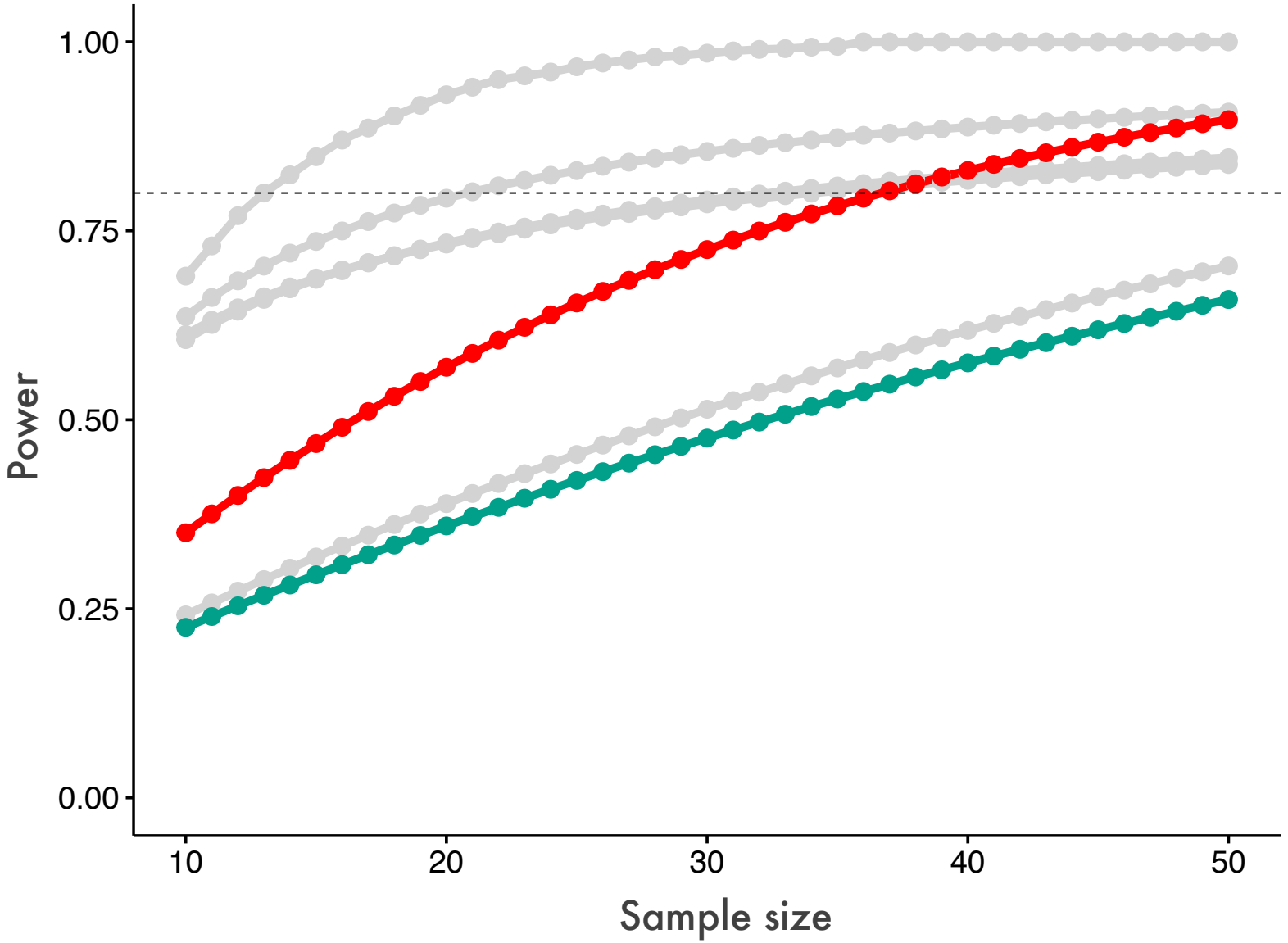
Publication bias necessitates larger samples



Observed estimate: N=13
Mean power: N=21
Low bias: N=32
High bias: N=34
20th percentile: N=37
Low bias: N=65

janssen | PHARMACEUTICAL COMPANIES OF Johnson&Johnson

Publication bias necessitates larger samples



Observed estimate: N=13
Mean power: N=21
Low bias: N=32
High bias: N=34
20th percentile: N=37
Low bias: N=65
High bias: N=73

Replications require large sample sizes

- Assuming observed parameters reflect reality leads to underpowered studies
- P-values are not the only problem; estimates can vary in magnitude, direction
- The proposed method can be tailored to a given replication study
    - Uses all available information, can extend to unequal variance
    - Quantifies uncertainty in power calculation
    - Accommodates publication bias and hidden multiplicity
    - Can be computationally intensive