Type 2 Error Control

Type 2 error:

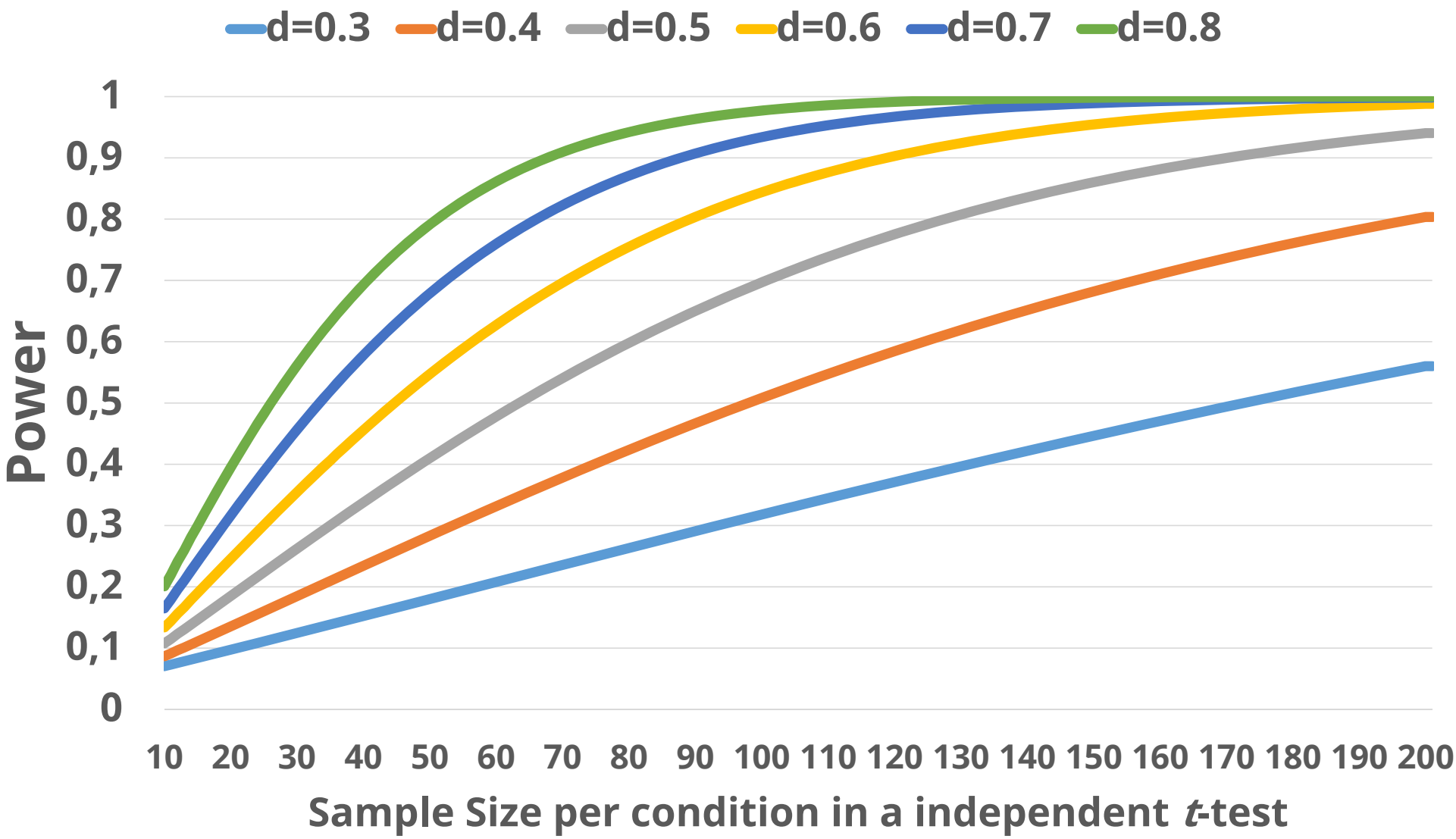
Saying there is nothing, when there is something.

1) It is “known” that an effect exists in the population.

2) A difference in a pilot between Group 1 (n = 22, $M = 5.68$, $SD = 0.98$) and Group 2 (n=23, $M = 6.28$, $SD = 1.11$), $p < .05$

You repeat the study

What is the chance you will observe a significant effect?

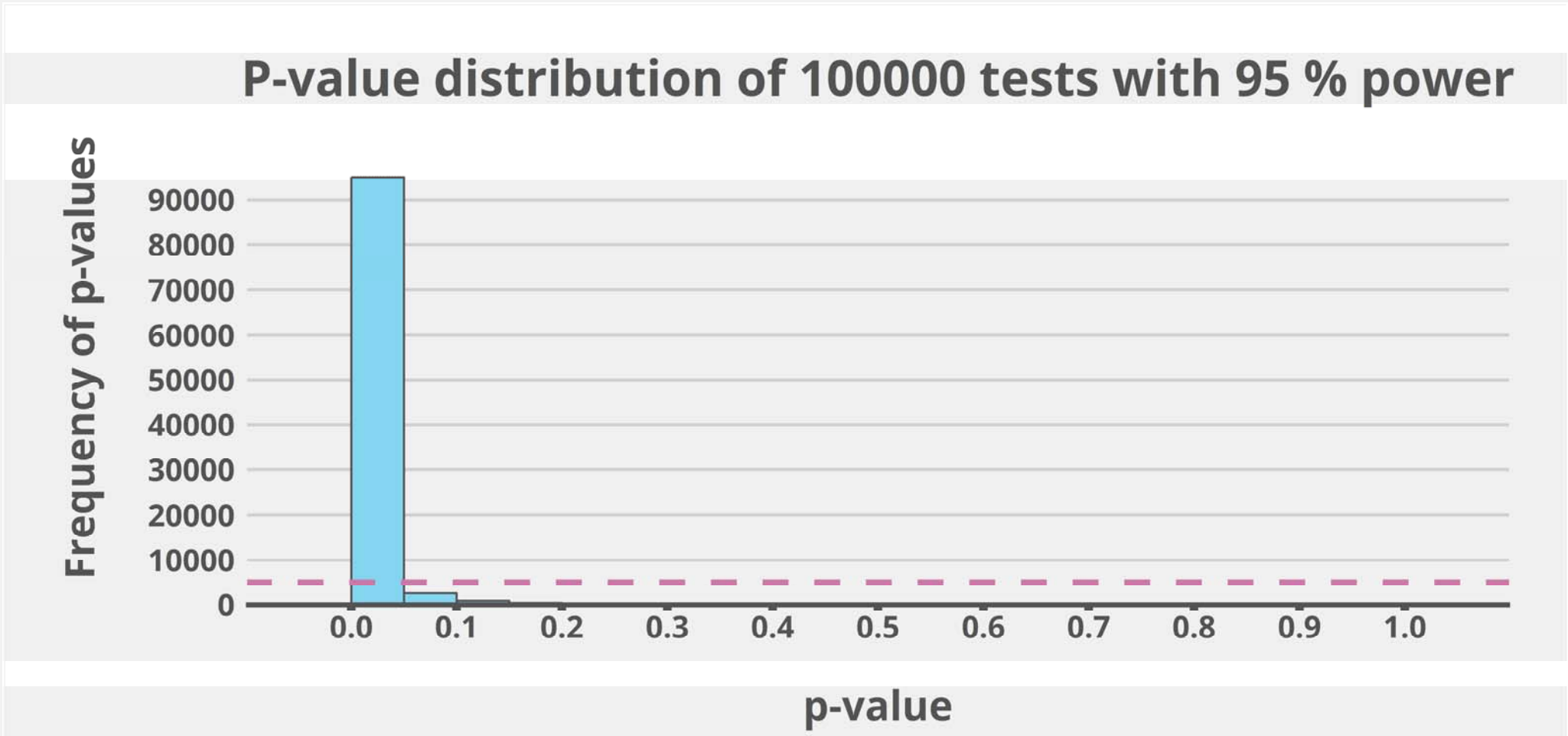

d=0.3
d=0.4
d=0.5
d=0.6
d=0.7
d=0.8
Power
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200
Sample Size per condition in a independent t-test

With $n = 100$, you had 95% power to observe a $d = 0.5$

Underpowered studies have low informational value

Studies with high power (low Type 2 error) are ‘severe’ tests.

Mayo & Spanos, 2006


P-value distribution of 100000 tests with 95 % power
Frequency of p-values
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
p-value

Increase power by:

1) Decreasing measurement error

Increase power by:

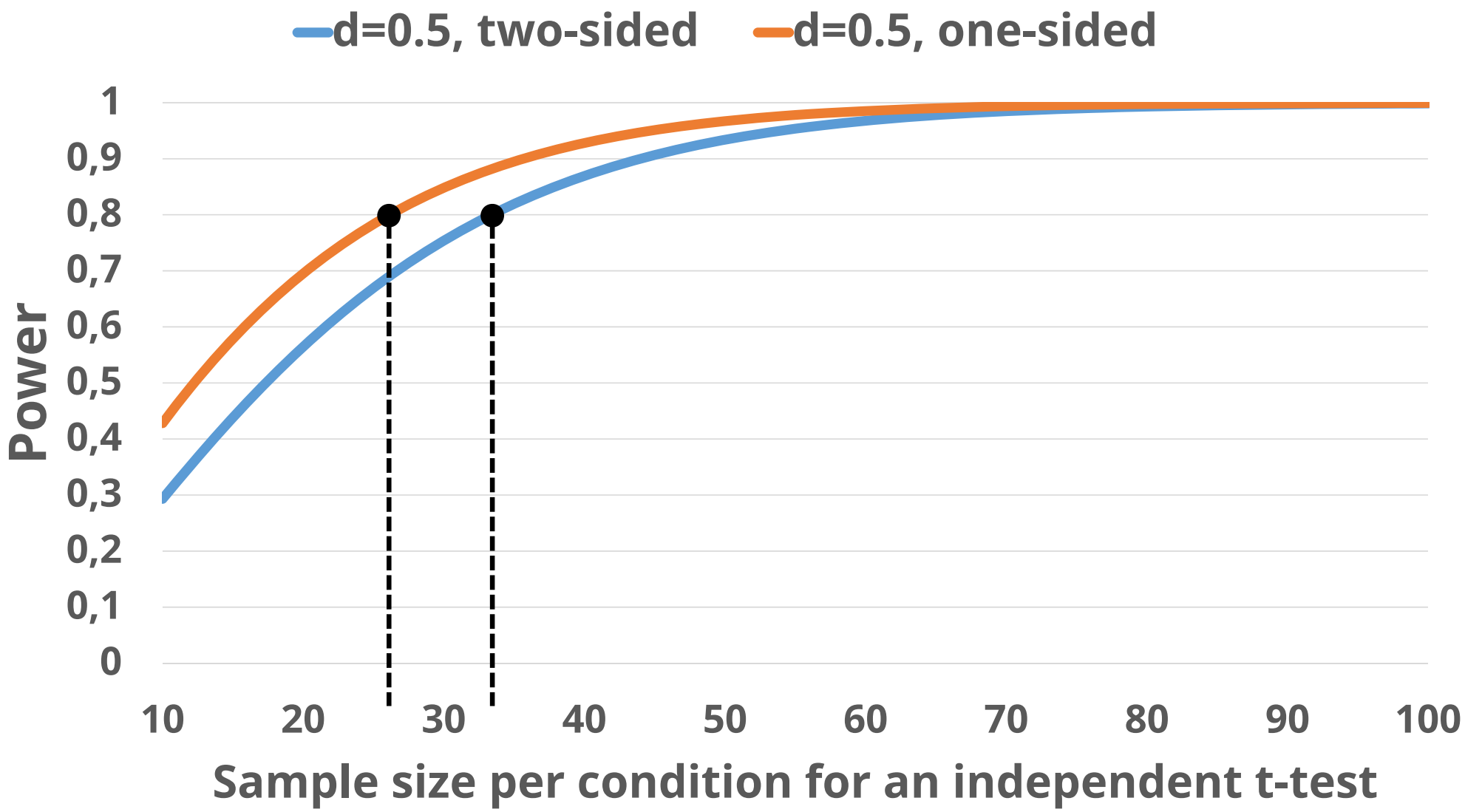
2) Using within designs (when within-measure correlation > 0.5)

Increase power by:

3) Increasing variability (e.g., use 7 or 9 instead of 3 or 5 item scales)

Increase power by:

4) Use one-sided tests (if you have a directional prediction)


d=0.5, two-sided
d=0.5, one-sided
Power
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
10
20
30
40
50
60
70
80
90
100
Sample size per condition for an independent t-test

As long as replications are performed, Type 1 errors will be corrected.

Type 2 errors might be more severe.

Type 2 errors are more difficult to control than Type 1 errors, but just as important.