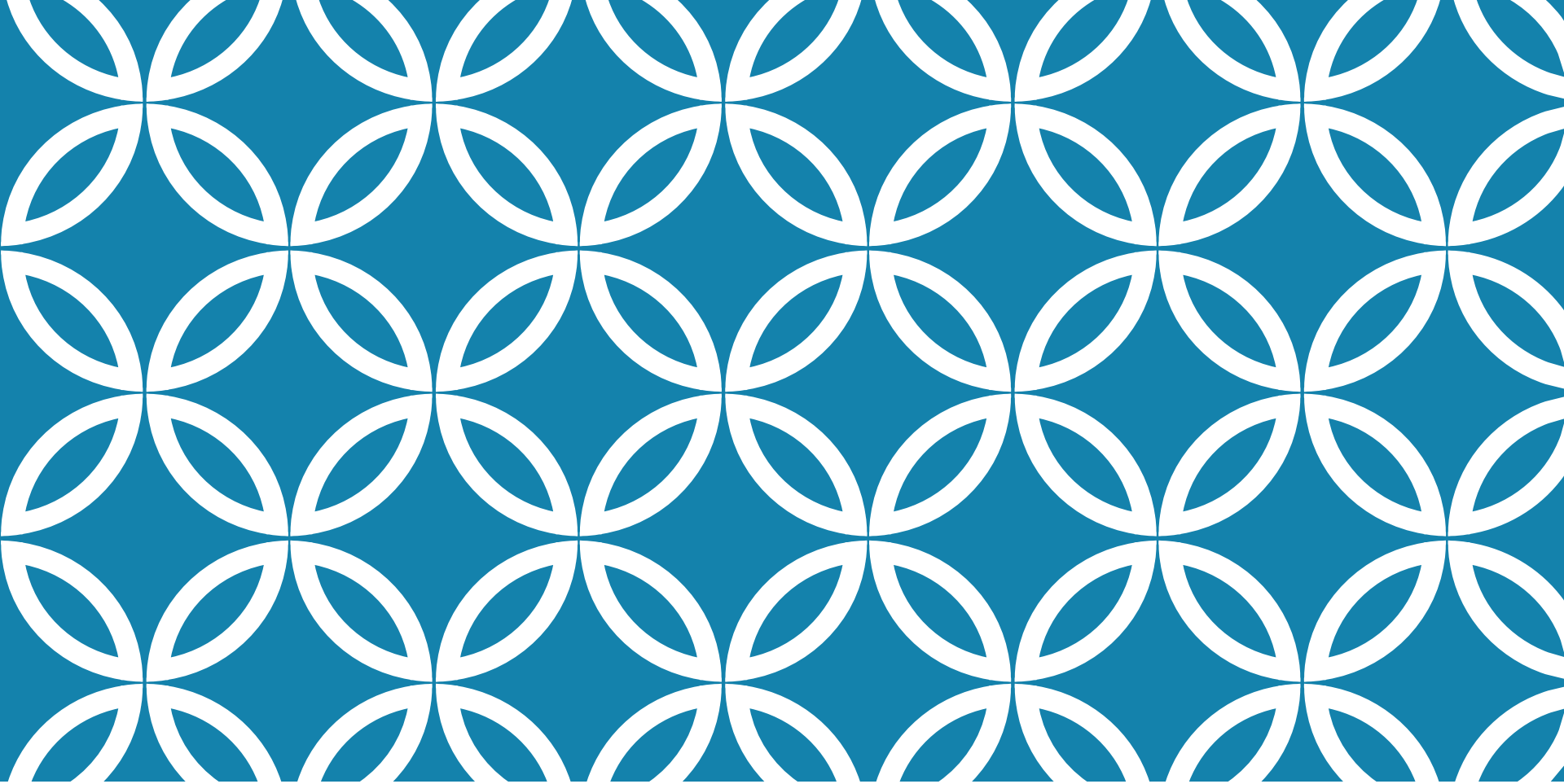


FLAWED INTUITIONS ABOUT POWER

Marjan Bakker, Jelte M. Wicherts, Chris H. J. Hartgerink, and Han L. J. van der Maas

WHAT IS POWER?

Which of the following statements best describes the statistical power of a significance test?

1. The probability that an effect is real
2. The probability you can find an effect, given that it exists
3. The probability that the effect will replicated in future studies of this kind
4. The accuracy of the estimated effect size
5. One minus the probability that the null hypothesis is true

WHAT IS POWER?

Which of the following statements best describes the statistical power of a significance test?

1. The probability that an effect is real
2. **The probability you can find an effect, given that it exists**
3. The probability that the effect will replicated in future studies of this kind
4. The accuracy of the estimated effect size
5. One minus the probability that the null hypothesis is true

HIGHER POWER

Larger sample sizes

Larger effect sizes

Observation

Experimental studies is Psychology have often small sample sizes…

- On average 20 to 24 participants per group (Marszalek et al., 2011; Wetzels et al., 2011)

But effects are generally subtle and small…

- The average effect size found in meta-analyses in psychology is $d = 0.50$ (Anderson et al., 1999; Hall, 1998; Lipsey & Wilson, 1993; Meyer et al., 2001; Richard et al., 2003)
- This is probably an overestimation!

Thus power is very low!

- Power is .35 for a independent samples t-test with 20 subjects per group and $d = 0.50$ (Bakker & Wicherts, 2012)

Why would a researcher do such a study?

- Publishing a non-significant effect is very difficult…

Intuitions about Power: Study 1

Sample:

- Emailed 1304 corresponding authors who published in 2012 in one of 10 different psychology journals.
- 291 finished the survey

Questions:

- How do you generally determine sample size?
- Typical alpha, power, ES (in Cohen's *d*) and *N* (cell size)

Sample size decisions

Power analysis	47%
Practical constraints	20%
Some rule of thumb ▪ e.g., 20 subjects per condition	23%
Common practice in their field of research	21%
As many as possible, to have the highest power	9%

"I usually aim for 20-25 participants per cell of the experimental design, which is typically what it takes to detect a medium effect size with .80 probability"

Typical alpha, ES, N, and Power

	Literature	Researchers
Alpha	.05	.05
Effect size (d)	0.50	0.39
N (cell size)	20	34.6
Reported power		0.80
Calculated power	0.35	0.40

This table contains 20% trimmed means

Study 2

Are the intuitions really flawed?

- Most participants indicated a power of .80 (normative), but they might know that it is less in their research

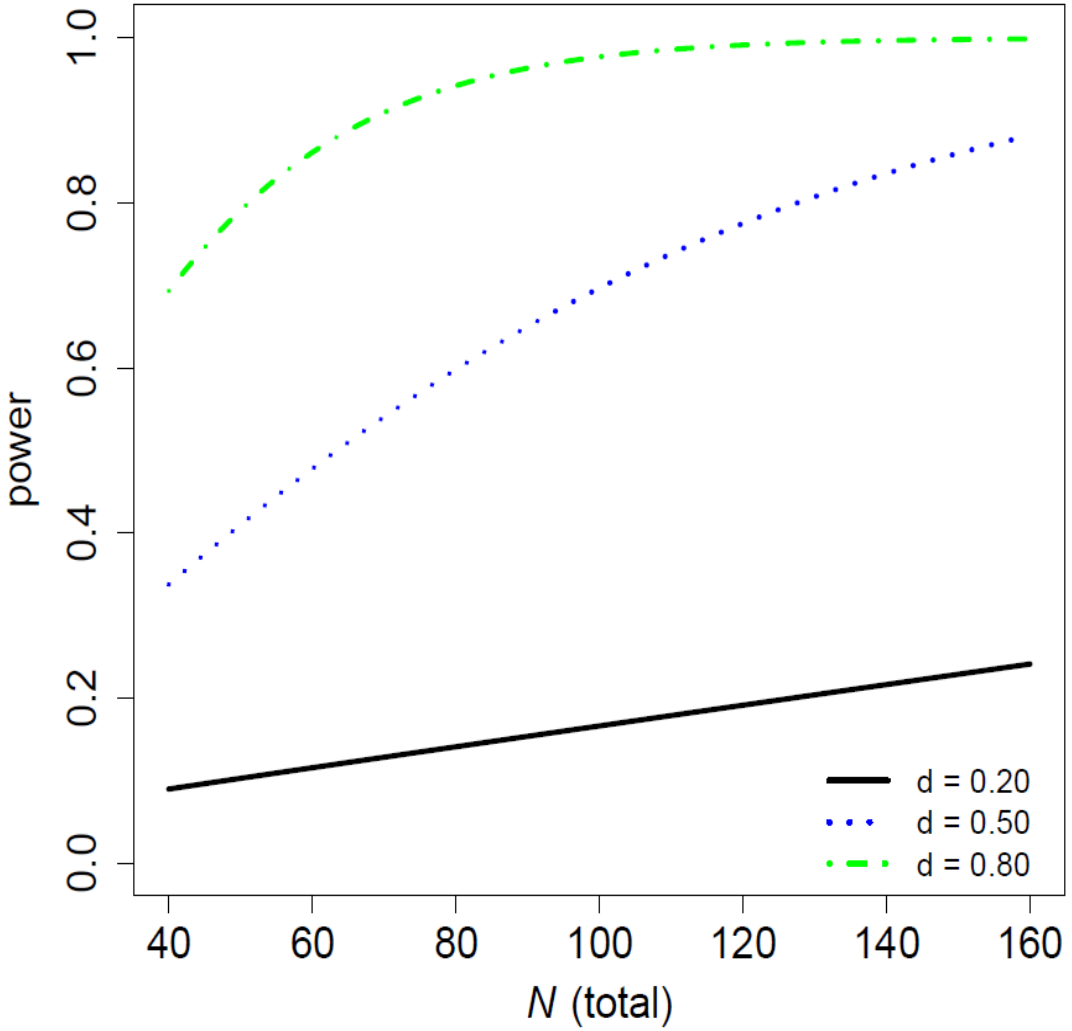
Present a research situation with alpha, ES, and N, and ask to estimate the power

Present a research situation with alpha, ES, and power, and ask to estimate the N

Sample:

- Emailed 1625 corresponding authors who published in 2014 in one of 10 different psychology journals. 214 finished the survey

What is the power?



What is the power?

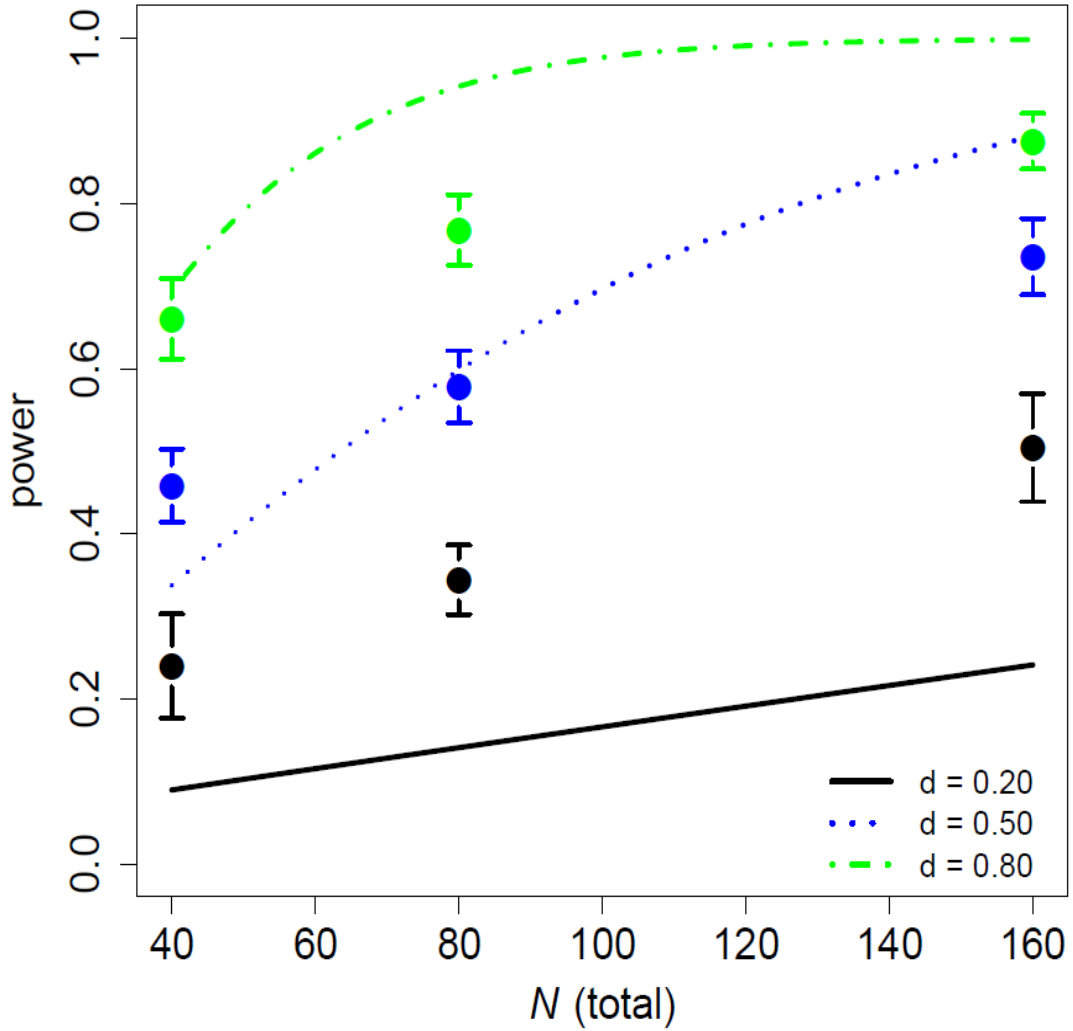
True power

	d = 0.20 (small)	d = 0.50 (medium)	d = 0.80 (large)
N = 40	0.09	0.34	0.69
N = 80	0.14	0.60	0.94
N = 160	0.24	0.88	>.99

Estimated power: 20% trimmed mean

	d = 0.20 (small)	d = 0.50 (medium)	d = 0.80 (large)
N = 40	0.240	0.459	0.660
N = 80	0.344	0.578	0.768
N = 160	0.504	0.736	0.876

What is the power?



What should N be?

True needed N

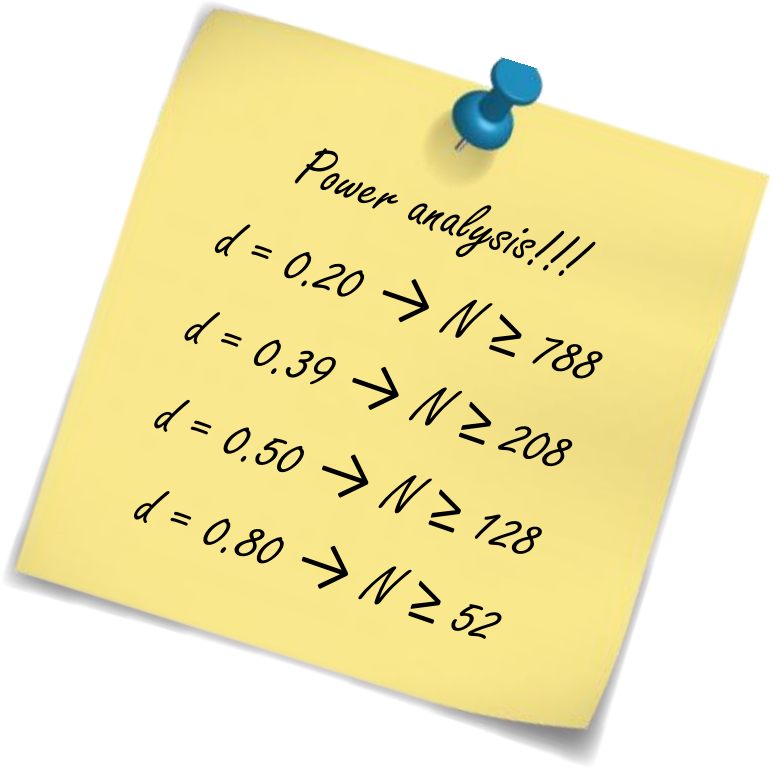
d = 0.20 (small)	d = 0.50 (medium)	d = 0.80 (large)
788	128	52

Estimated N: 20% trimmed mean

d = 0.20 (small)	d = 0.50 (medium)	d = 0.80 (large)
216	124	77

N was underestimated by 96% when d = 0.20

Conclusions



Power intuitions are flawed

- Especially when effects are small!
- N should be 3 times larger for small effects!

Do a power analysis!

- Don't trust your intuitions
- Use a range of proper effect size estimates