Methodological and Statistical Issues in Research Proposals

How to design a pilot or feasibility study

Rich Jones (rich_jones@brown.edu)
NIDUS/CEDARTREE
6th Annual Delirium Boot Camp
October 5, 2018
The Nittany Lion Inn, State College PA

https://s0.geograph.org.uk/geophotos/05/52/49/5524950_0d016b97.jpg

https://s0.geograph.org.uk/geophotos/05/52/49/5524950_0d016b97.jpg

http://i1205.photobucket.com/albums/bb434/fromhollywood/Boiler%20HydroTherm%20model%20HC-145D-PV/Pilot_02.jpg

https://theaviationist.com/2013/07/03/top-guns-maverick-turns-51/

http://www.thefederationtimes.com/wp-content/uploads/2015/10/The-Cage-Poster_optimized.jpg

This talk is about designing pilot studies as a grant proposal, such as for

- K## project
- R03
- R21
- Foundation funding

Context

There are other contexts for conducting pilot studies. Most commonly, you are planning a big R01 or P project and want to convince reviewers your research plan is feasible, and this kind of work has not been done by you or someone on your investigator team previously.

This talk is not directly concerned with those situations. But it is still relevant to those situations.

Context

Think of the definitive study you would like to do to answer your important and clinically relevant question.

What are the things you need to know in order to design that study?

Finding out the things you need to know: that is the goal of a pilot study.

The big idea

What is a pilot study?

A small-scale study or experiment intended to inform the design of, or the decision as to whether to conduct, a larger study.

Typically focuses specific attention on aspects of research methodology for the subsequent (definitive) study: choice of measurements, suitability of research environment, participant availability, and resource allocations, etc.

What pilot studies are not

Not a little version of the definitive study

Not intended to develop target ‘effect size’ through an efficacy analysis*

Not a label applied post-hoc to a study designed to test a hypothesis, but the null hypothesis was not rejected

* But a pilot *may* be used to *inform* definition of minimal clinically important differences or some other method

Purpose of a pilot study (1)

To provide preliminary information about **feasibility** of doing a definitive study / trial

- Are participants truly available?
  - How many must be screened to enroll?
- Can measurements be done?
- Is the environment suitable?
- Is resourcing adequate (in particular: time)?

Purpose of a pilot study (2)

To provide preliminary information about **measurement variation**

- What is the degree of natural (biological) variation in endpoints?
- Are there stratifying factors that are critical to consider at design time? Can acknowledging these help us overcome variation?

Purpose of a pilot study (3)

To provide information about **outcomes performance characteristics**

- Are measurements affiliated with reference standards or reference standards?
- Do measurements display intra- and inter-rater reliability?

Purpose of a pilot study (4)

To provide information about **sample selection**

- Are there particular subpopulations ill-suited to enrollment? Should certain populations be over-represented?
- Is there evidence that efficacy / effectiveness or safety may vary across subpopulations?

Purpose of a pilot study (7)

To look for suggestive evidence of **efficacy/effectiveness**

- Is there evidence suggesting a favorable treatment effect?

https://s0.geograph.org.uk/geophotos/05/22/70/5227063_1a573d22.jpg

https://libreshot.com/wp-content/uploads/2015/01/freighter.jpg

Which boat do you want to use to bring your goods to market?

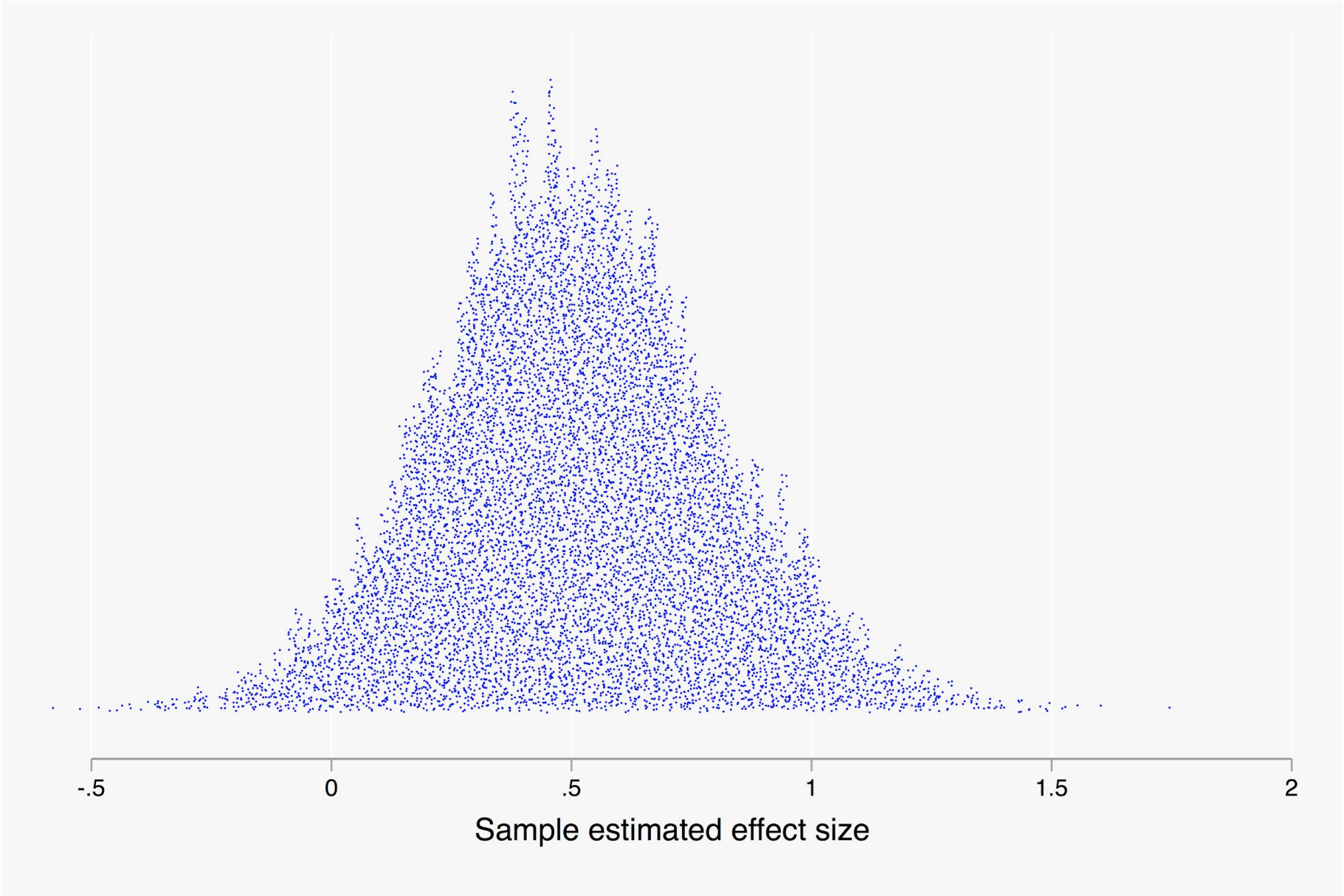
It is "wrong" to do so, but you will need to conduct an efficacy/effectiveness analysis with your pilot data [if you are able] and include that in your full, definitive study application.
Why?

Why is it "wrong"?

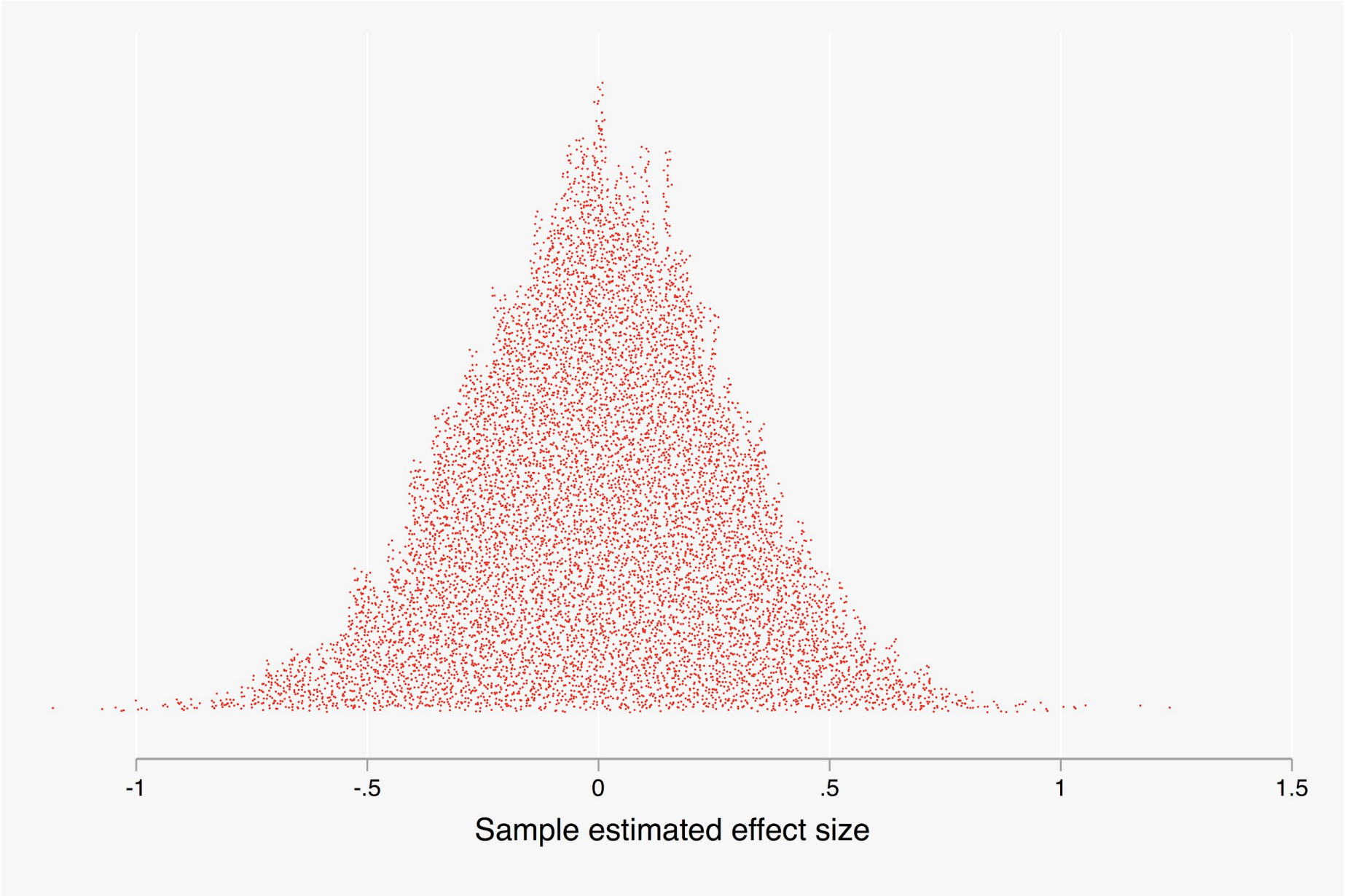
Your statistical power for finding an effect of minimal clinical importance is low (otherwise it would not be a pilot)

Chances are that even -- in truth -- your intervention is efficacious or effective, you might even find a result in a pilot study that suggests harm rather than benefit

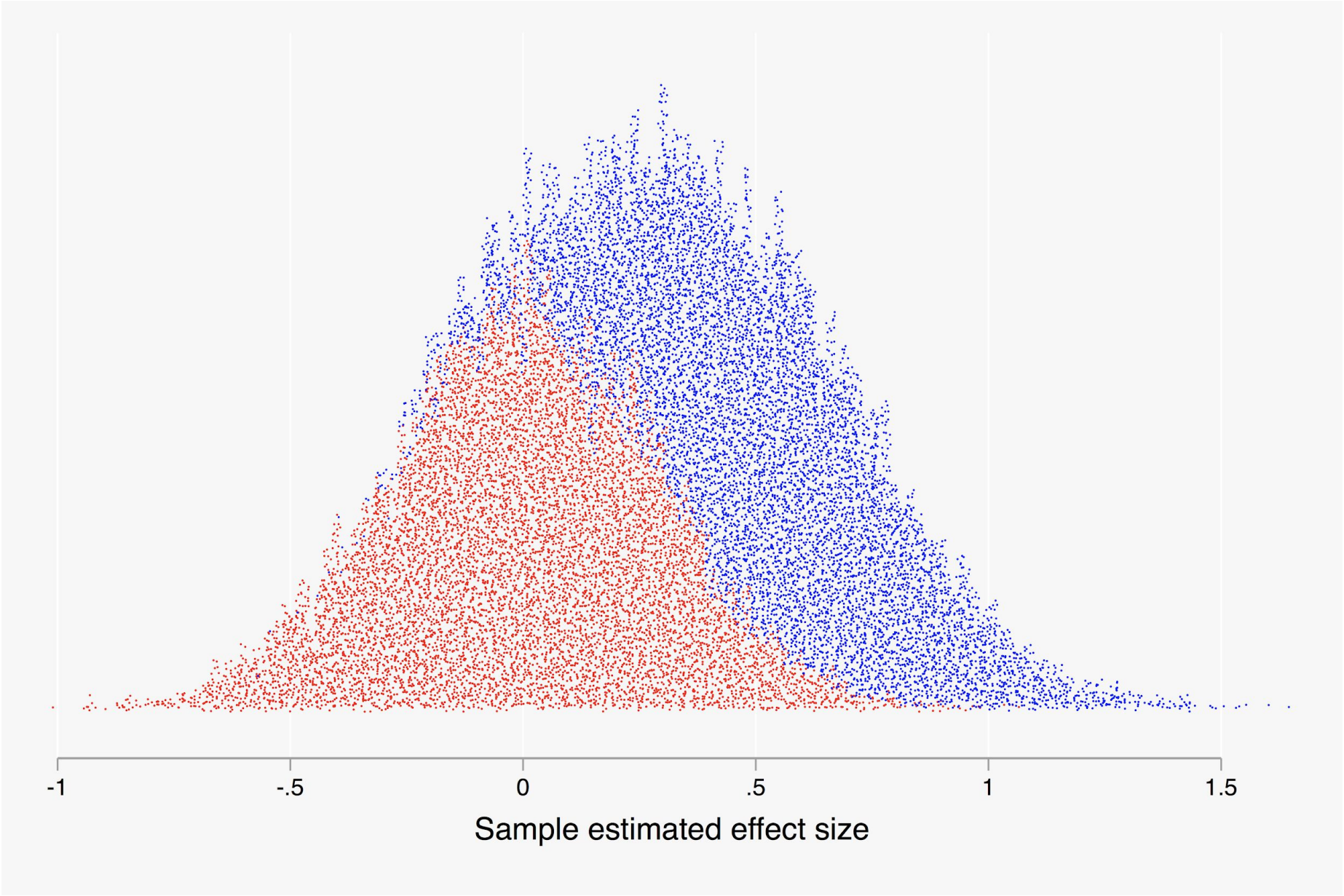
Just as bad is finding an inflated estimate of the efficacy or effectiveness of your intervention in a pilot study, if you use that effect estimate to power a full definitive study (your full study will be under-powered)



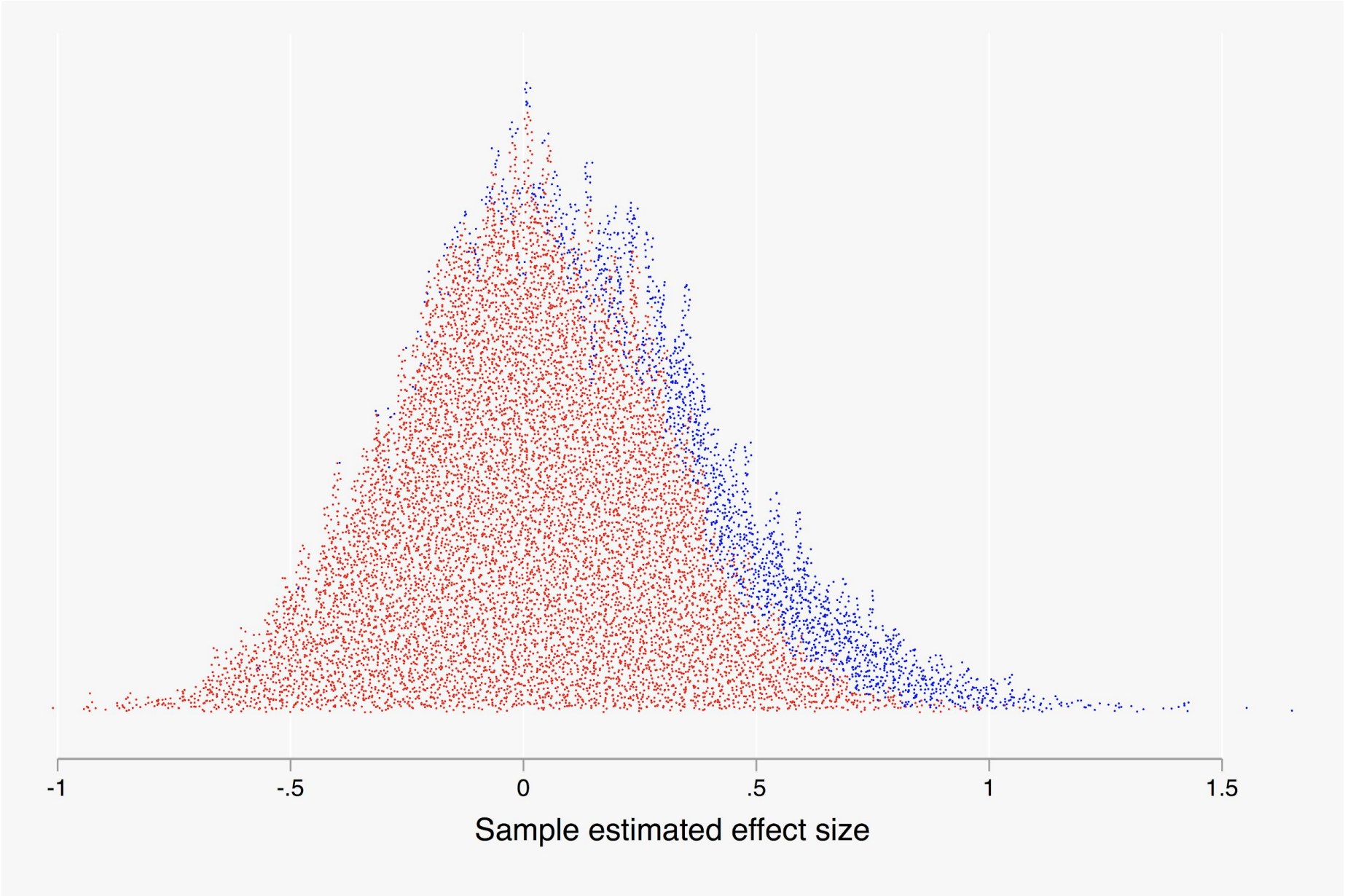
Distribution of 10,001 sample-based effect size statistics (Cohen's d) when n = 25 per group and population effect size is 0.5 (power is 41%)



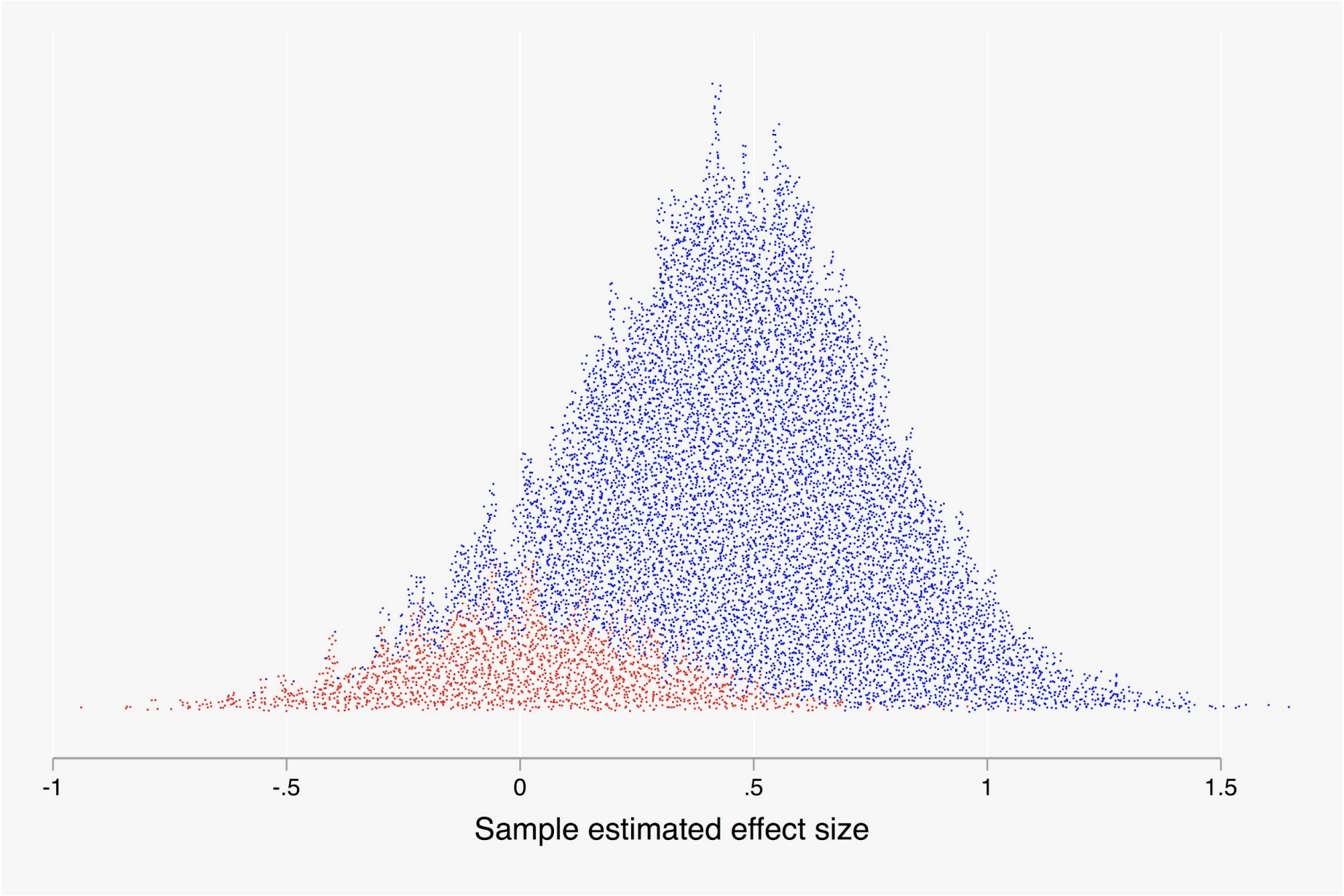
Distribution of 10,001 sample-based effect size statistics (Cohen's d) when n = 25 per group and population effect size is 0 (type I error rate = 5%)



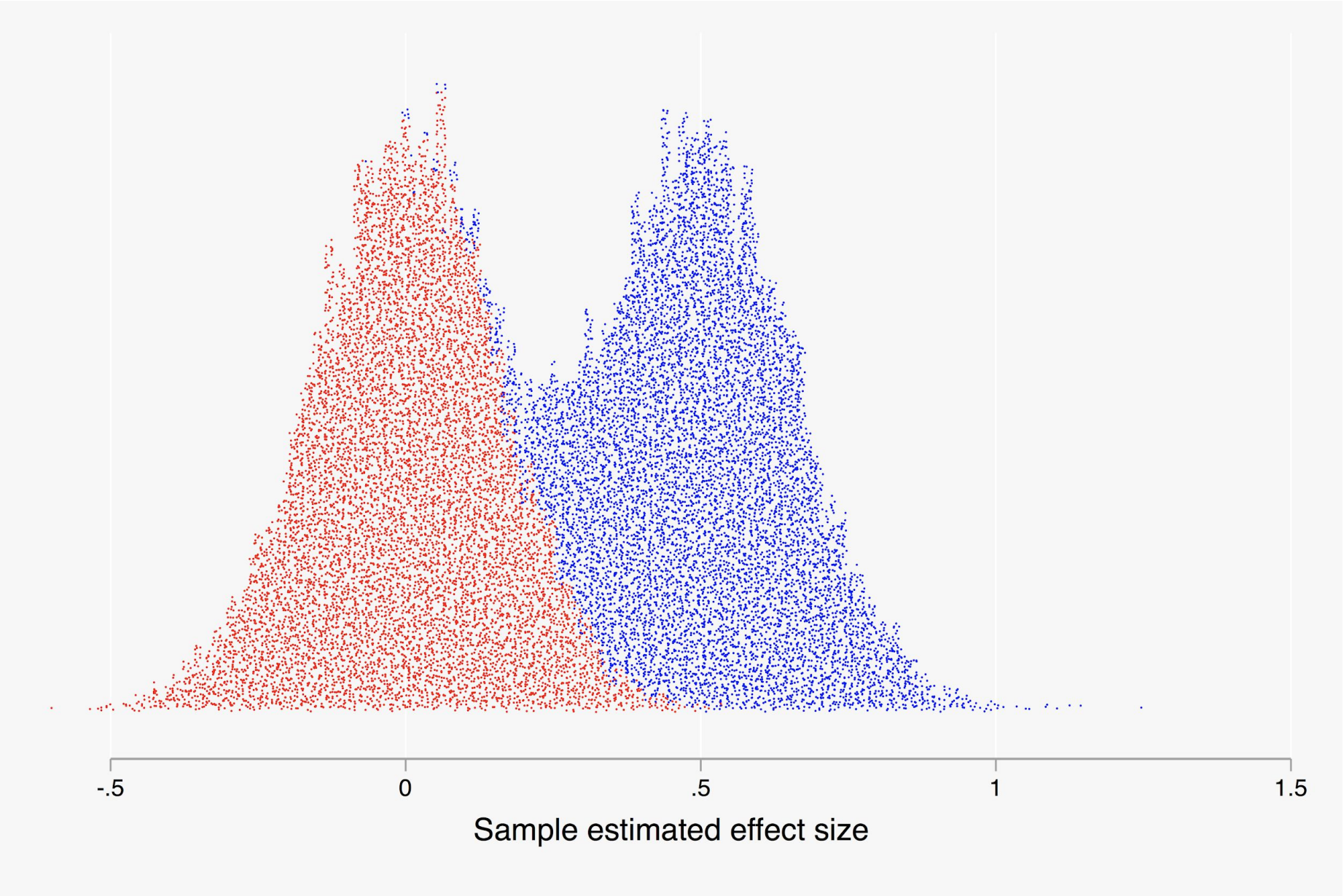
Distribution of 20,002 sample-based effect size statistics (Cohen's d) when n = 25 per group and population effect size is 50% d=0, 50% d=0.5



Distribution of ~12,000 sample-based effect size statistics (Cohen's d) when n = 25 per group and population effect size is 80% d=0, 20% d=0.5



Distribution of ~12,000 sample-based effect size statistics (Cohen's d) when n = 25 per group and population effect size is 20% d=0, 80% d=0.5



Distribution of 20,002 sample-based effect size statistics (Cohen's d) when n = 86 per group and population effect size is 50% d=0, 50% d=0.5 (Power = 90%)

It is “wrong” to do so, but you will need to conduct an efficacy/effectiveness analysis with your pilot data and include that in your full, definitive study application. Why?

Why do you have to do it even tho’ it’s wrong?

Your reviewers might not be aware of the problems of doing efficacy/effectiveness studies in small pilot samples

You don’t want to appear as if you are suppressing evidence

Do this: In full/definitive trial, provide the pilot effect estimate and it’s confidence interval. Remind reviewers that pilot effect sizes are poor predictors of actual effect sizes in fully powered study (Kraemer et al., 2006). Hope and pray that the full study hypothesized effect size lies in the confidence region for the pilot effect size. Wave hands.

Kraemer, H. C., Mintz, J., Noda, A., Tinklenberg, J., & Yesavage, J. A. (2006). Caution regarding the use of pilot studies to guide power calculations for study proposals. Archives of General Psychiatry, 63(5), 484-489.

So if the pilot cannot be used to estimate the [target] effect size, how is this to be determined?

A big question

The critical effect size is the minimum value of the population effect size that would be clinically or practically significant

Kraemer and Blasey (2016) How Many Subjects? Statistical Power Analysis in Research (second edition). SAGE, Los Angeles

How to come up with an effect size

Nowadays this concept is called, by some

Minimal Clinically Important Difference (MCID), or
Minimal Clinically Important Change (MCIC), or
sometimes “Clinically” is dropped, so:
MID, MIC

How to come up with an effect size

But there are analogous ideas and terms used in various fields....

How to come up with an effect size

Health Services Research and Patient Reported Outcomes	**Psychology**	**Public Health, Epidemiology, Economics**
Minimally clinical important difference (MCID) and related concepts	Reliable change index	Prevented fraction, Numbers needed to treat (NNT), Cost [benefit, effective-ness, utility]
How much difference must we observe to conclude the difference is equal to or greater than some threshold that is considered to be clinically or practically important	How much difference must we observe to conclude the difference is equal to or greater to a difference we might expect given measurement error or other artifacts	How much difference must we observe to conclude there is a tangible public health benefit (c.f. Raferty (2000))

Health Services Research and Patient Reported Outcomes	Psychology	Public Health, Epidemiology, Economics
Minimally clinical important difference (MCID) and related concepts	Reliable change index	Prevented fraction, Numbers needed to treat (NNT), Cost [benefit, effectiveness, utility]
Engel L, Beaton DE, Touma Z. Minimal Clinically Important Difference: A Review of Outcome Measure Score Interpretation. Rheum Dis Clin N Am. 2018.	Schennach R, et al. Challenging the understanding of significant improvement and outcome in schizophrenia–the concept of reliable and clinically significant change methods. International journal of methods in psychiatric research. 2016 Mar;25(1):3-11.	Brownson RC, Petitti DB, editors. Applied epidemiology: theory to practice. Oxford University Press on Demand; 1998. (or any epi textbook) Raferty (2000) BMJ 321(7262): 697.

Health Services Research and Patient Reported Outcomes	Psychology	Public Health, Epidemiology, Economics
Minimally clinical important difference (MCID) and related concepts	Reliable change index	Prevented fraction, Numbers needed to treat (NNT), Cost [benefit, effectiveness, utility]
How is the outcome distributed in the target population? And/or how is the outcome related to an external standard for important difference (clinician, patient)	How is the outcome distributed and what are the psychometric properties in a normative sample? Are they different in the target population? How large are practice/retest or other measurement artifacts in a normative sample? How does the outcome change in a normative sample?	What NNT is of practical importance? What are the costs and/or utilities associated with the condition and treatment?

MEDICAL CARE
Volume 41, Number 5, pp 582–592

When all else fails...

Point/Counterpoint

Interpretation of Changes in Health-related Quality of Life

The Remarkable Universality of Half a Standard Deviation

GEOFFREY R. NORMAN, PHD,* JEFF A. SLOAN, PHD,† AND KATHLEEN W. WYRWICH, PHD‡

BACKGROUND. A number of studies have computed the minimally important difference (MID) for health-related quality of life instruments.

CONCLUSION. In most circumstances, the threshold of discrimination for changes in health-related quality of life for chronic diseases appears to be approximately half a SD.

Key words: Quality of life; threshold; interpretation; MID; effect size. (Med Care 2003; 41:582–592)

Archetypal aims for a pilot RCT

Aim 1: Treatment development activities

Aim 2: Feasibility study to assess acceptability design issues

Aim 3: Estimate key statistics of major outcomes in order to design a larger, definitive trial. Key statistics include, for example:

(a) standard deviation of the outcome variable, and/or
(b) reliability of the outcome variable, and/or
(b) correlation of outcome variable from baseline to post-test, and/or
(c) expected distribution of outcome variable in the absence of treatment or in normative sample

Feasibility and acceptability issues

1. How many people can be approached (e.g., what is the size of the patient pool)?
2. What proportion of those people approached will be eligible?
3. What proportion of those people eligible will be willing to participate in research?
4. What proportion of those people randomized to treatment will comply with treatment?
5. What proportion of those people enrolled will complete the follow-up?

How large should your pilot sample be?

Right answer:

- Be clear about the questions
- The answer is based on the desired *precision of parameter estimation* rather than hypothesis testing
- Consider working with an expert statistician or methodologist, the answer is nuanced

Plan on this answer while waiting for the right answer:

- about 25 per treatment arm

How large of a sample do I need to estimate a percent within ±5%?

What do you expect the proportion to be?

75%

How confident do you want to be in obtaining margin of error of ±5%?

??? 

???

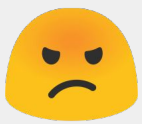
Sometimes people say 95% when I ask them that...

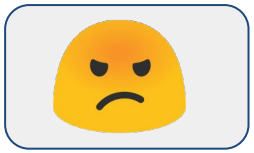
OK, 95% then

...but I think people really want 80% confidence...

OK, 80% then

...or even 1-sided 80%, which is 2-sided 60%...





1-sided 80% means 80% confidence that proportion is at least 70% (i.e.,75-5)

OK, 80% 1-sided

For that you'd need a denominator of 54 people

What about 95%, 2-sided?

289

mmmkay

Do you have a citation for this calculation?

It's pretty basic, it's just the equation for a confidence interval on a proportion...

$p \pm z^*\sqrt{pq/n}$

rearranged to solve for smallest n where $(z^*\sqrt{pq/n}) <= .05$

...please?

…please?

I also checked my calculation with a Monte Carlo simulation, 54 is right

That’s wonderful! Have a nice trip to Monaco! How can I cite your calculation?

Daniel WW. Biostatistics: A Foundation for Analysis in the Health Sciences. 5 ed. New York: John Wiley & Sons; 1991.

Suggested reading

Arain, M., Campbell, M. J., Cooper, C. L., & Lancaster, G. A. (2010). What is a pilot or feasibility study? A review of current practice and editorial policy. BMC medical research methodology, 10(1), 67.

Kraemer, H. C., Mintz, J., Noda, A., Tinklenberg, J., & Yesavage, J. A. (2006). Caution regarding the use of pilot studies to guide power calculations for study proposals. Archives of General Psychiatry, 63(5), 484-489.

Lancaster, G. A., Dodd, S., & Williamson, P. R. (2004). Design and analysis of pilot studies: recommendations for good practice. Journal of Evaluation in Clinical Practice, 10(2), 307-312.

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. Journal of Psychiatric Research, 45(5), 626-629.

Orsmond, G. I., & Cohn, E. S. (2015). The distinctive features of a feasibility study: Objectives and guiding questions. OTJR: occupation, participation and health, 35(3), 169-177.

Questions