

Post-hoc power: Don't do it.

`www.biostat.ku.dk/~bxc/SDC-courses`

Bendix Carstensen

Steno Diabetes Center &

Biostatistisk afdeling, KU

`bxc@steno.dk` `www.biostat.ku.dk/~bxc`

December 2002

Power calculation

Power = $\mathbb{P} \{ \text{rejecting the null hypothesis} \mid \text{assumptions} \}$

Assumptions:

- Significance level, α
- Sample size, N
- True treatment (/exposure) effect, δ
- Variance of the difference σ_δ

Power calculation for $\delta = 0$ vs. $\delta \neq 0$

Observations x_1, \dots, x_N , variance σ . Test statistic:

$$z = \frac{\bar{x} - 0}{\sigma / \sqrt{N}}$$

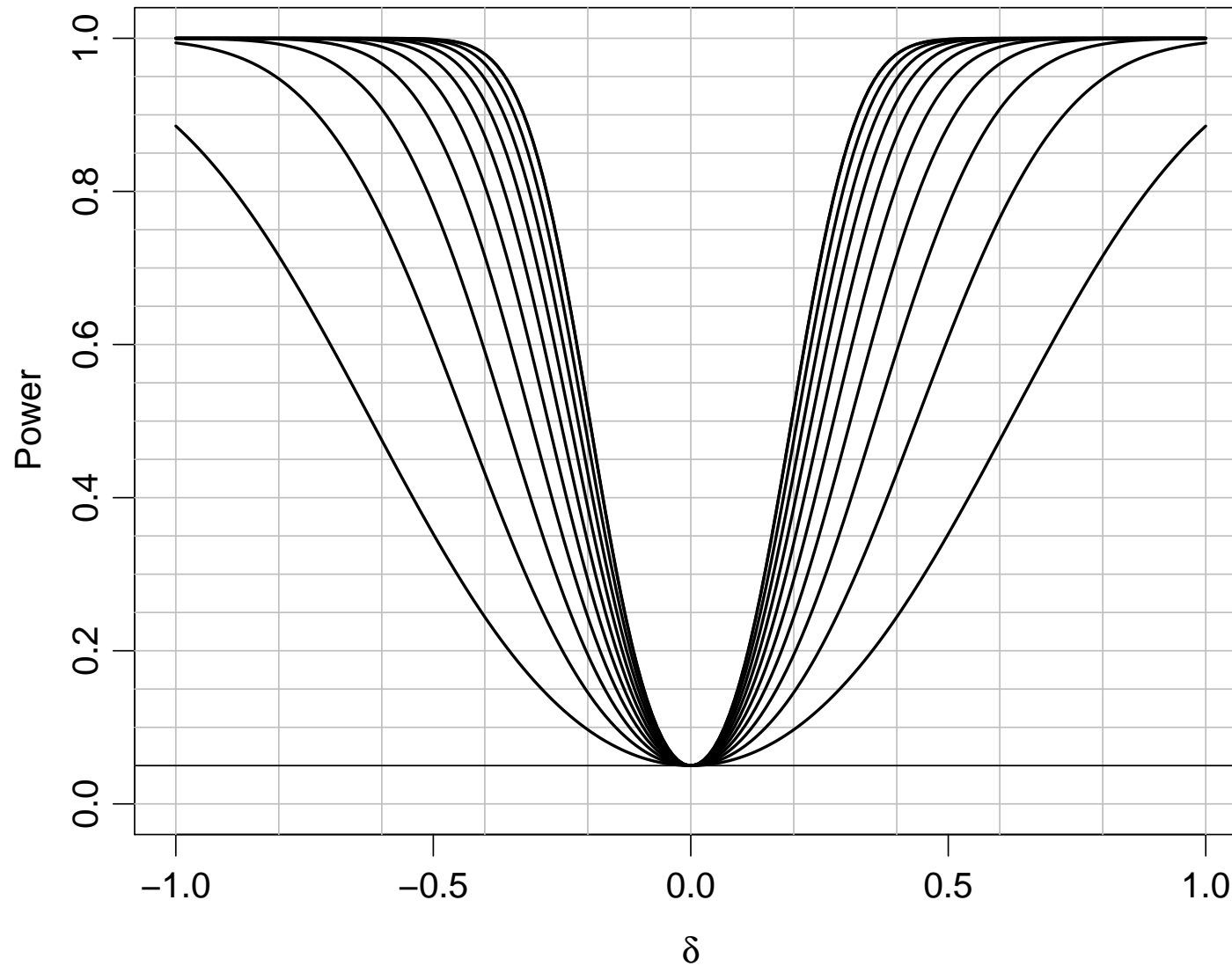
The null is rejected if $|z|$ is large;
two sided p-value:

$$p = 2 \times (1 - \Phi(|z|))$$

Power, if true effect is δ :

$$\begin{aligned} & \text{P}\{|z| > 1.96\} \\ &= 1 - \text{P}\left\{-1.96 < \frac{\bar{x} - 0}{\sigma/\sqrt{N}} < 1.96\right\} \\ &= 1 - \text{P}\left\{-1.96 - \frac{\delta}{\sigma/\sqrt{N}} < \frac{\bar{x} - \delta}{\sigma/\sqrt{N}} < 1.96 - \frac{\delta}{\sigma/\sqrt{N}}\right\} \\ &= 1 + \Phi\left(-1.96 - \frac{\delta}{\sigma/\sqrt{N}}\right) - \Phi\left(1.96 - \frac{\delta}{\sigma/\sqrt{N}}\right) \end{aligned}$$

Power functions for $N = 10, \dots, 100$



Post-hoc power calculations

Observed difference $\hat{\delta} = \bar{x}$.

What is the power in this study to detect a difference of $\hat{\delta}$?

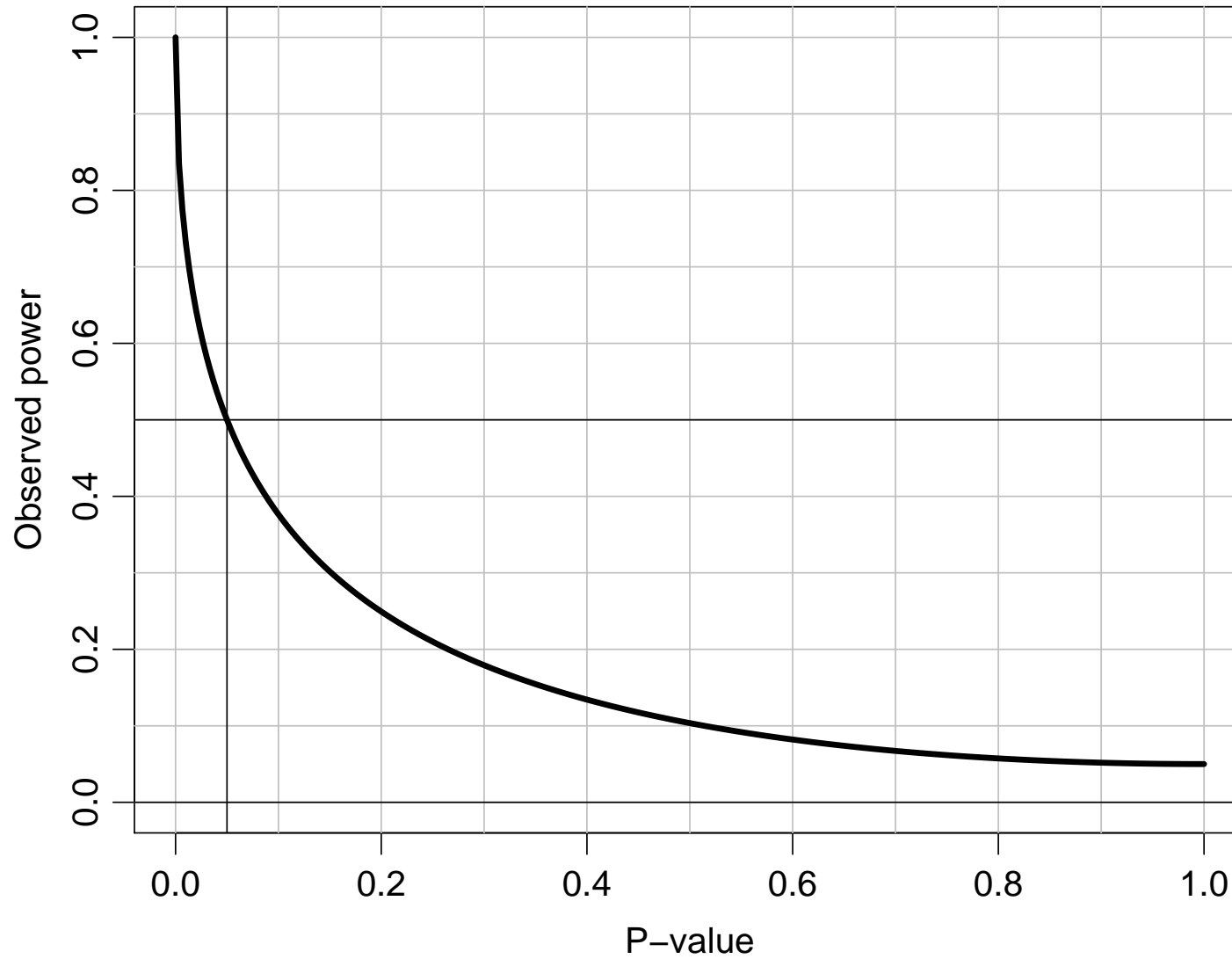
Answer:

$$1 + \Phi \left(-1.96 - \frac{\bar{x}}{\sigma/\sqrt{N}} \right) - \Phi \left(1.96 - \frac{\bar{x}}{\sigma/\sqrt{N}} \right)$$

But $\frac{\bar{x}}{\sigma/\sqrt{N}} = z$ is the test statistic, which is a 1 to 1 function of the p-value, $z = \Phi^{-1}(1 - p/2)$.

Observed power is a simple function of the p -value.

P-value and observed power for ANY study



The meaning of power (post hoc)

Power is the probability of the event:

“Rejecting H_0 ” in a (future) study.

Once the study is conducted, this probability is either

1 — if H_0 were rejected

0 — if it was not.

Power is meaningless post hoc.

Dealing with reviewers

Refuse to do post-hoc power calculations.

Use a reference to Hoening & Heisey.

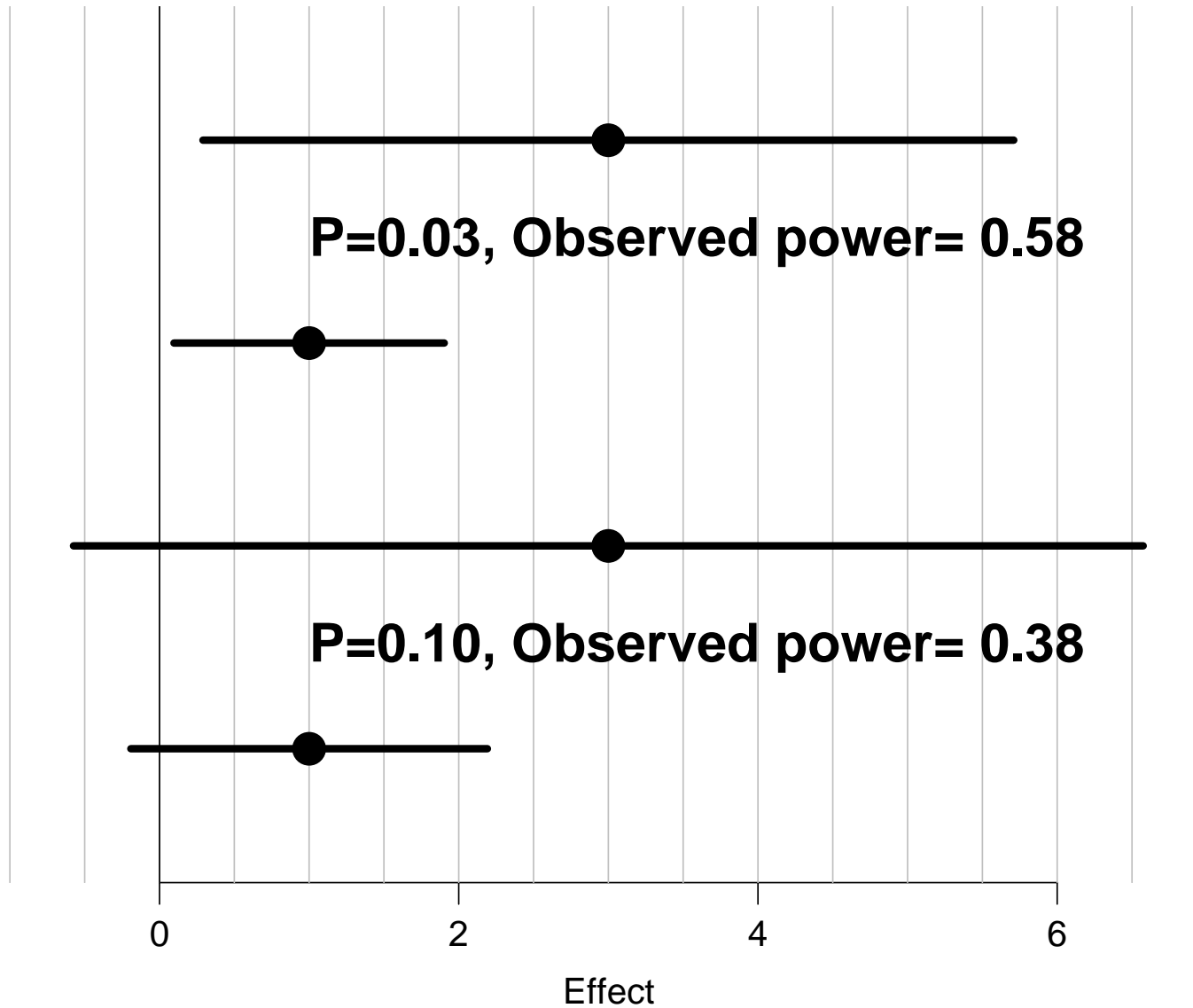
A copy can be obtained from BxC.

Reviewers do not always have sound statistical advice to offer. A good discussion is given in Bacchetti's paper.

See also the discussion in BMJ following it.

Some discussants argue that mention of the initial power calculation (pre-study) is useful — it shows that thought has been given to the design.

Report and use the confidence interval



References

- [1] Peter Bacchetti. Peer review of statistics in medical research: the other problem. *British Medical Journal*, 324:1271–1273, 2002.
- [2] JM Hoening and DM Heisey. The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Statistician*, 55(1):19–24, 2001.