Chapter 9: The two-sample t-test (3): False negatives, power, and necessary sample sizes

TXCL7565/PHSC7565

What This Chapter Covers

- False negatives (Type II errors)
- Power
- Calculating necessary sample size

FALSE NEGATIVES (TYPE II ERRORS)

P-values are related to type I error and represent the probability of making an error when there is no true treatment effect

False negatives (type II errors) are related to making an error when there IS a true treatment effect, i.e., failure to detect a difference that genuinely is present.

the beta value

- Alpha (α) is the risk of detecting a difference (significant result) when there is no true difference

- Beta (β) is the risk of failing to detect a true difference *of a stated size*

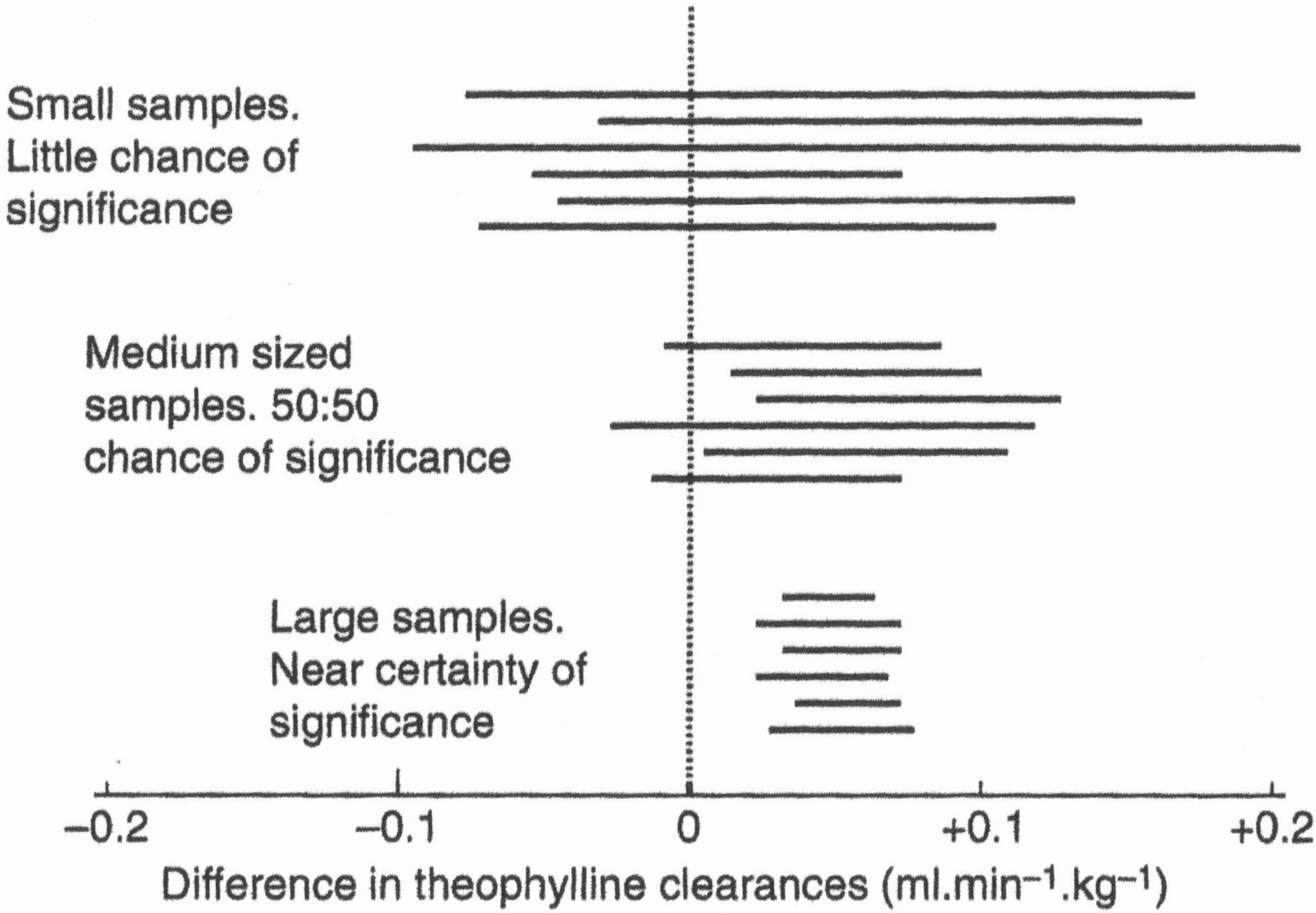
	Difference is not actually present	Difference is actually present
No difference detected	Correct	Type II Error False Negative (β)
Difference is detected	Type I Error False Positive (α)	Correct

POWER

power

If Beta is the chance that we will fail to detect a difference, the ***power*** is simply the opposite; it is the chance that we will successfully detect the difference

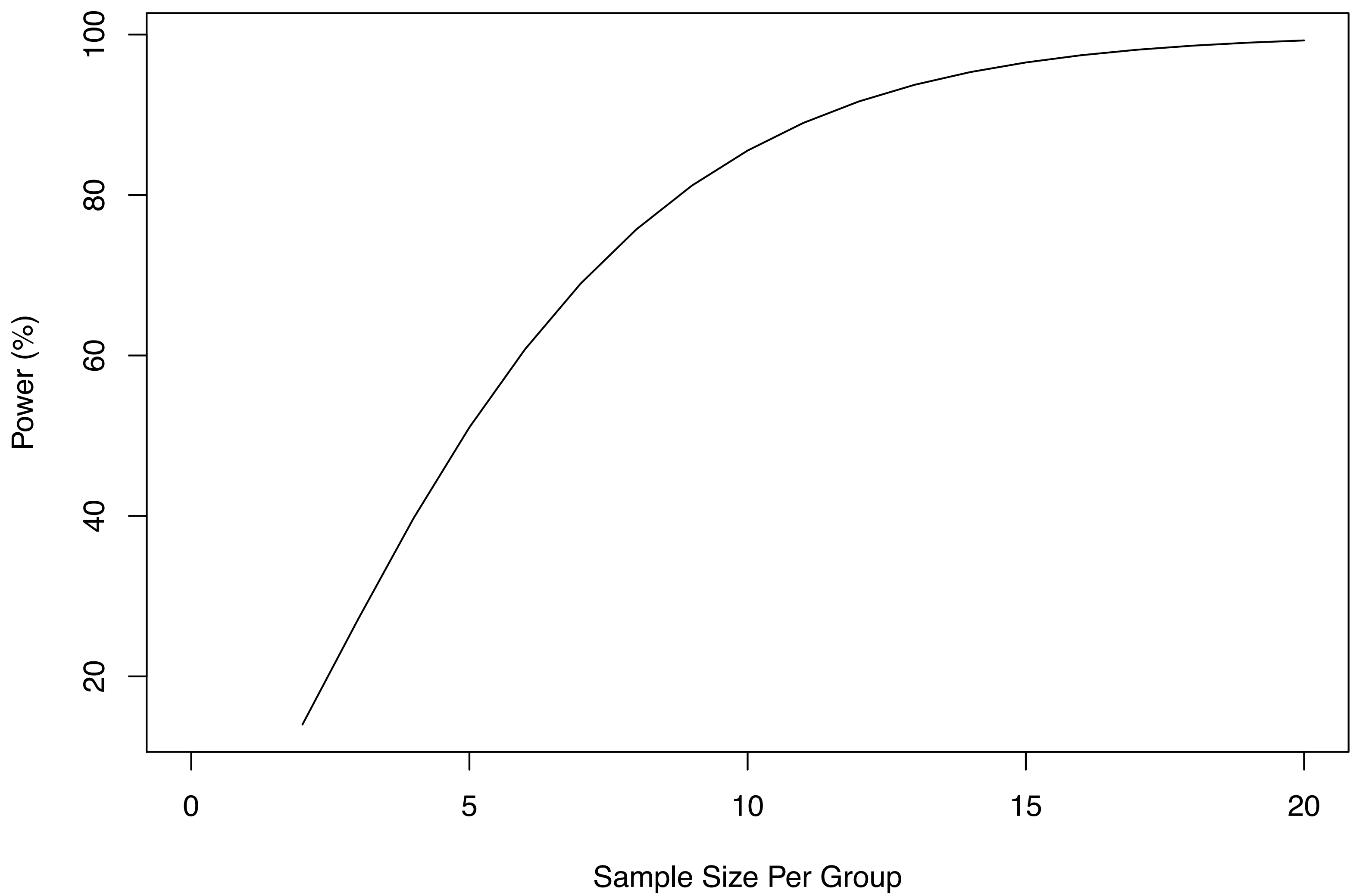
$$\text{Power} = 100\% - \beta$$


Small samples. Little chance of significance
Medium sized samples. 50:50 chance of significance
Large samples. Near certainty of significance
–0.2
–0.1
0
+0.1
+0.2
Difference in theophylline clearances (ml.min⁻¹.kg⁻¹)

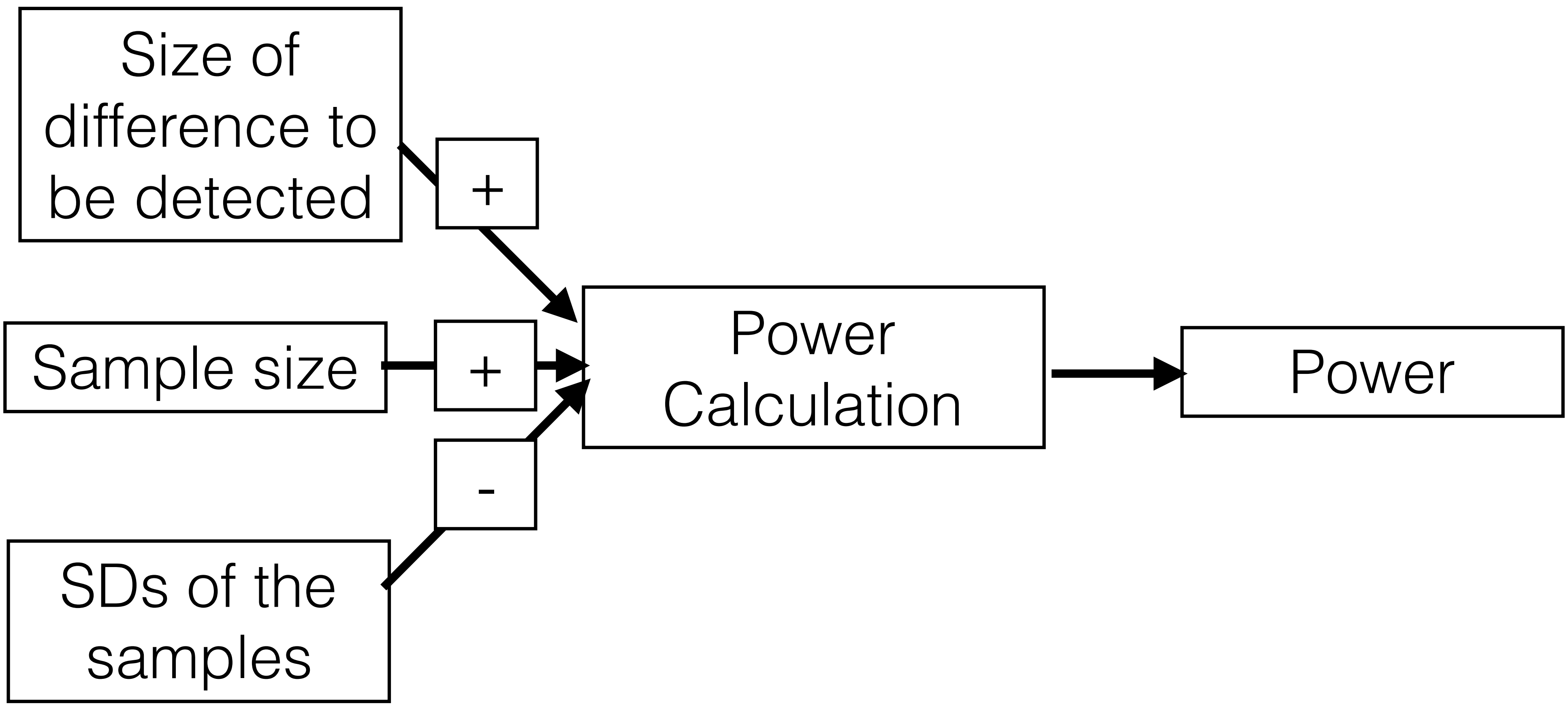
aspects of the data that determine the power of an experiment

- Size of difference that we are trying to detect
- Sample size
- Variability of the data
- Standard of proof demanded (α)

Power vs. Sample Size



calculating power



CALCULATING NECESSARY SAMPLE SIZE

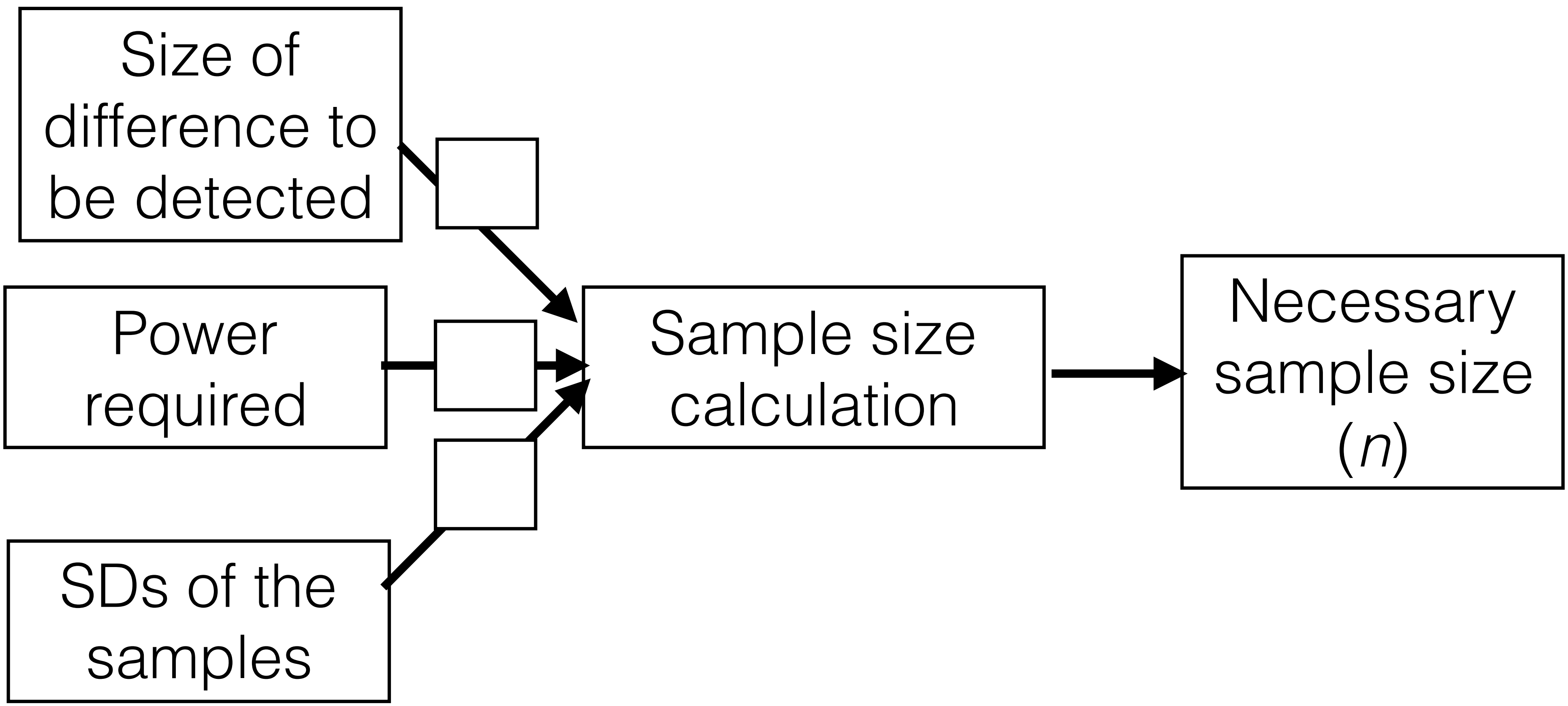
It is unethical to perform an experiment that is either too small or too large

- too large - waste of resources
- too small - useless conclusions and a waste of resources

factors that influence necessary sample size

- Size of experimental effect to be detectable
- Standard deviation
- Power required ($1 - \beta$)
- Standard of proof demanded (α)

calculating necessary sample size



size of effect to be detected

The number we choose for the smallest difference we want to be able to detect, should match the smallest effect that would be of real practical significance.

No bigger, no smaller.

how powerful we want our experiment to be

- Traditionally, power = 80%
- Can range from 80% to 95%
- if resources are plentiful and the experiment is critical, 95% power is reasonable (e.g., clinical trials)

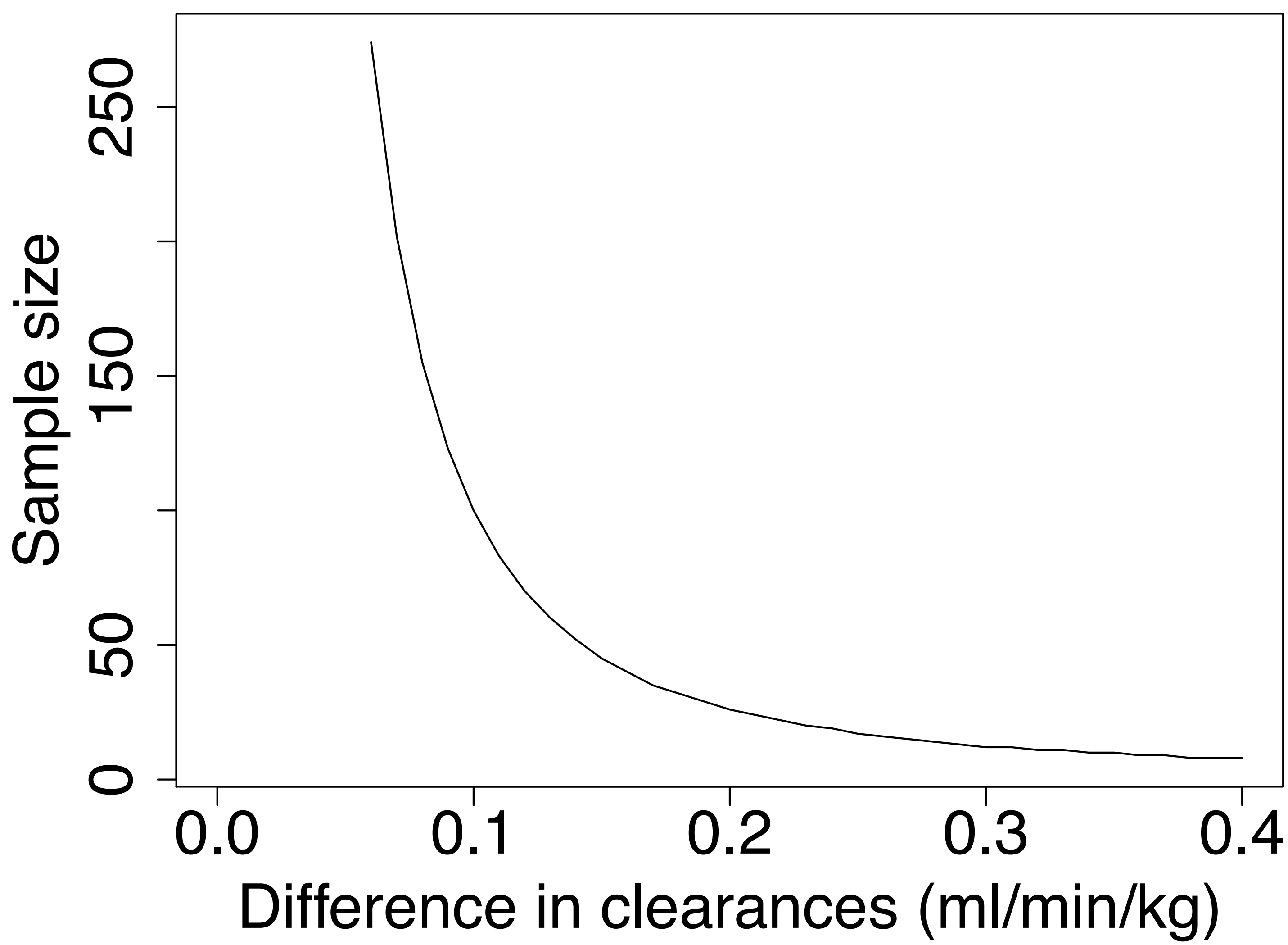
where do I find SD estimates

- look in literature (can even estimate from a range of values)
- conduct a pilot experiment
- take a peek at the first few samples

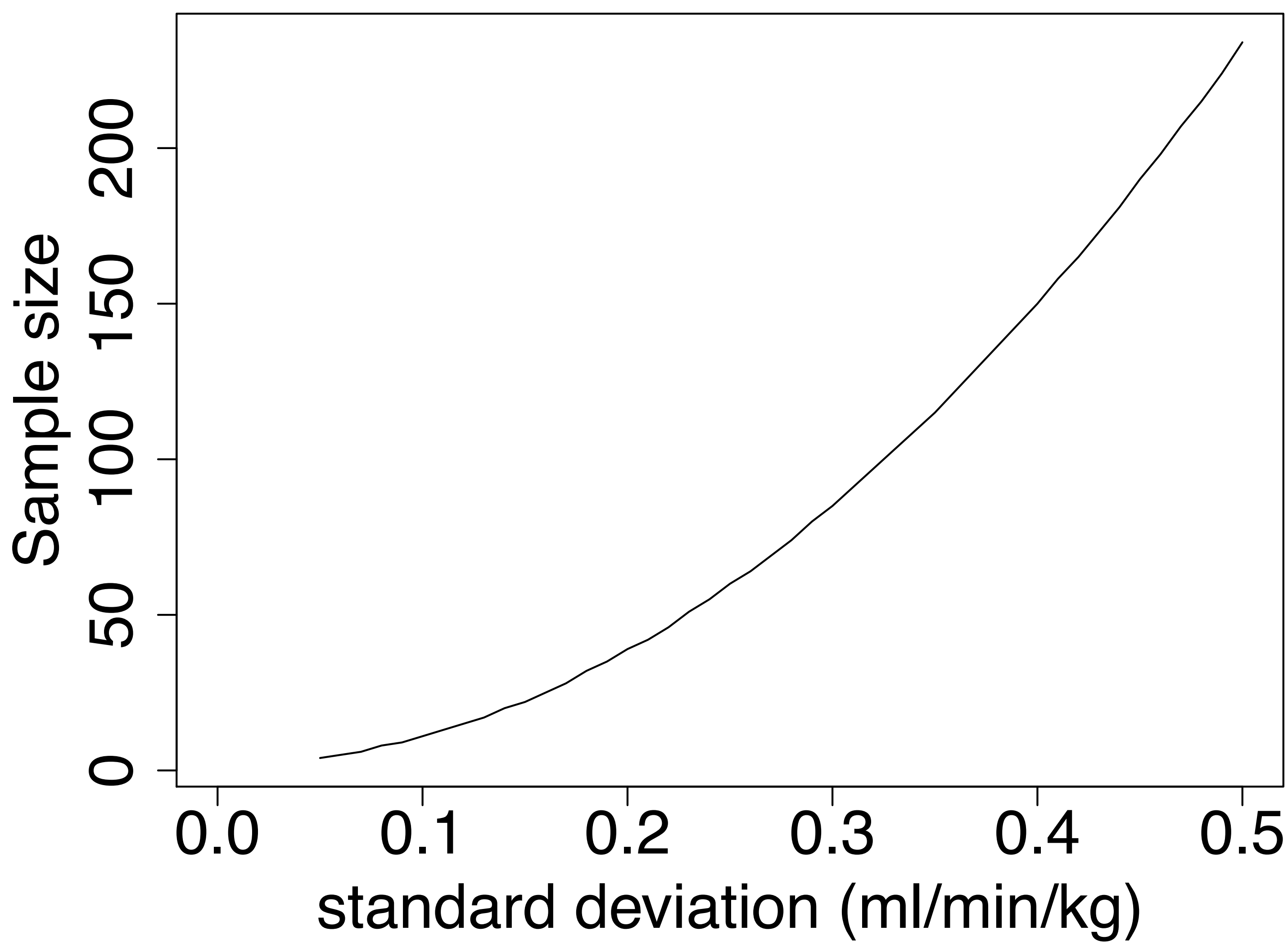
theophylline clearance experiment

- minimum effect to detect
- SD among theophylline clearances
- power to be achieved

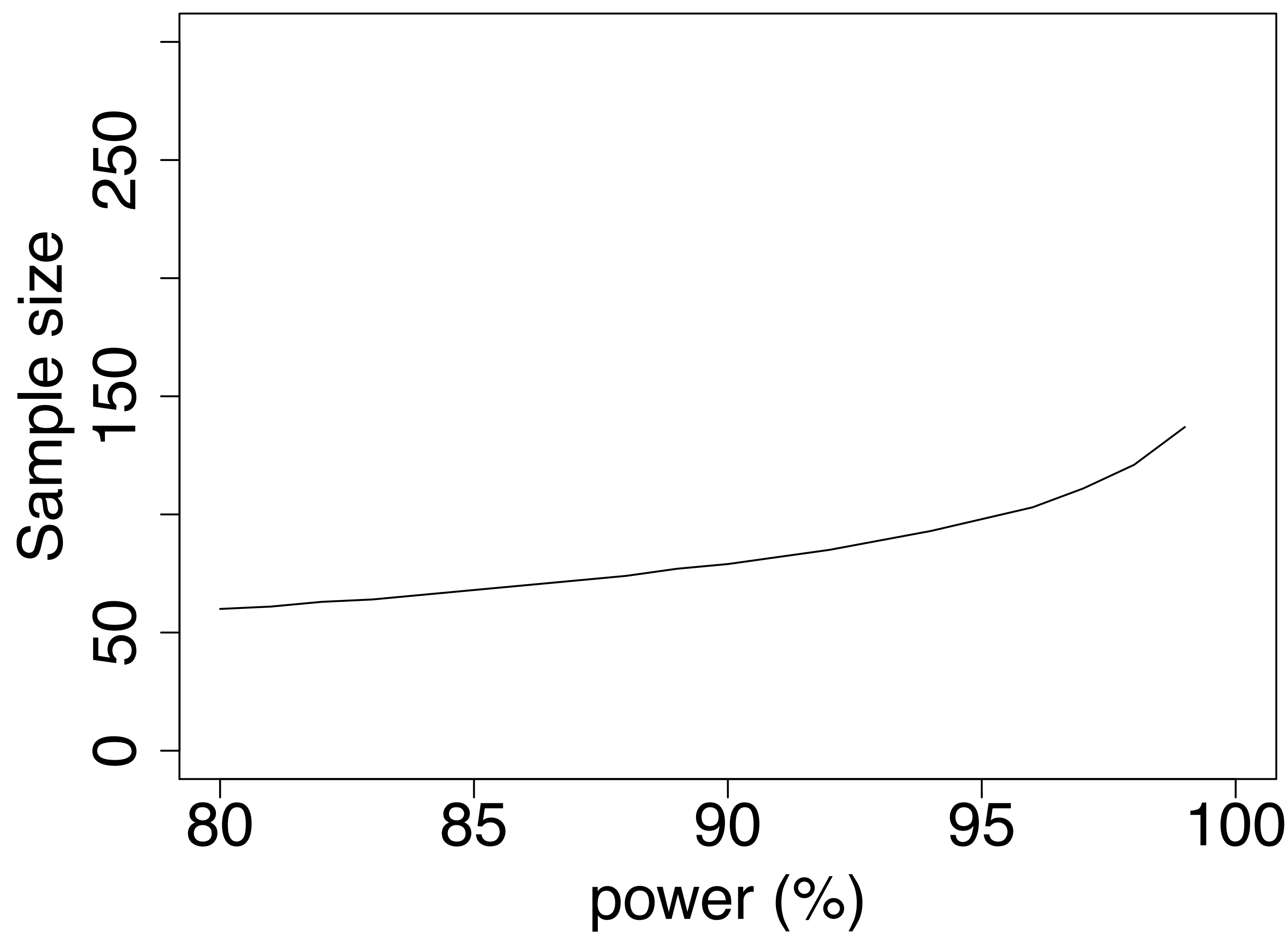
sensitivity of necessary sample size to minimum effect to be detectable



sensitivity of necessary sample size to standard deviation



sensitivity of necessary sample size to planned power of study



factors influencing necessary sample size

- Highly sensitive to minimum effect to be detected
- Highly sensitive to the standard deviation within the samples
- Modestly sensitive to the planned power of the experiment

what if you only have a specific sample size available?

- Can calculate the detectable difference
- Decide if the detectable difference is worthy of a study

What did we learn?

- False negative/type II error is when we get a ‘non-significant’ results when there is a true difference
- Power (probability of a ‘significant’ results if a difference of a stated size is present) is influenced by sample size, size of difference, variability
- Can rationally calculate the sample size needed to detect a specific effect with estimates of SD and a targeted power
- The minimum effect to be detectable should be practically significant, i.e., clinically relevant