Statistical seminar at MedViz
2015

MedViz
Haukeland University Hospital
2015-09-04

Power and precision in clinical studies

by

Geir Egil Eide

Centre for Clinical Research
Helse Vest RHF
&
Lifestyle Epidemiology Research Group
Department of Global Public Health and Primary Care
University of Bergen

Topics

Introduction

- Comparing two groups
- Precision
- Software

Summary

What determines the size of your study?

- This depends on your research question:
  - Compare groups (hypothesis test)?
    - Statistical power:
      - For a specified true difference between the two groups, how large must your study be to have a reasonable chance to give a statistically significant result ($p \leq 0.05$)?
  - Estimation?
    - Precision:
      - To obtain a required precision of your estimate specified as a standard error (or 95% confidence interval), how large must your study be?

1. Comparing two groups

- *Hypothesis test.*

The result is decided by an hypothesis test. Starting point is the formulation of two competing hypotheses in the research protocol:

- *The null hypothesis* (H_0) *.*

Postulates normally that there is no difference between the two groups.

- *The alternative hypothesis* (H_A) *.*

Postulates normally that there is a difference between the two groups.

- *One- or two-sided test?*

If the alternative hypothesis postulates a difference in only one direction a one-sided test can be performed.

Possible conclusions from a clinical study:

- either the null hypothesis is ‘rejected’: There is a difference!

- or the null hypothesis is ‘accepted’: There is no difference!

Wrong conclusions:

Rejection error:

The null hypothesis is rejected although it is actually correct, i.e. we conclude there is a difference when there in fact is none.

Acceptance error:

The null hypothesis is accepted although the alternative hypothesis is the correct one, i.e. we conclude that there is no difference when there in fact is one.

The risks for wrong conclusions

In an empirical study there is always a positive risk of drawing the wrong conclusion. Therefore we wish to design our study so that these risks are minimized.

Definition of risk of error.

The significance level: The risk of rejection error. It is denoted by α. The study is usually designed so that α is 0.05 (or 5 %).

The risk of acceptance error given a difference of Δ is denoted as β(Δ), where Δ is a measure of the difference, e.g. the difference in mean blood pressure or in cured proportion of patients. One would like to design the study so that β(Δ) is at most 0.10 (10%) if the true difference is the smallest difference considered to be of clinical importance.

Which error is most important to avoid?

Rejection error brings you one step in the wrong direction!

Acceptance error delay the development of new knowledge.

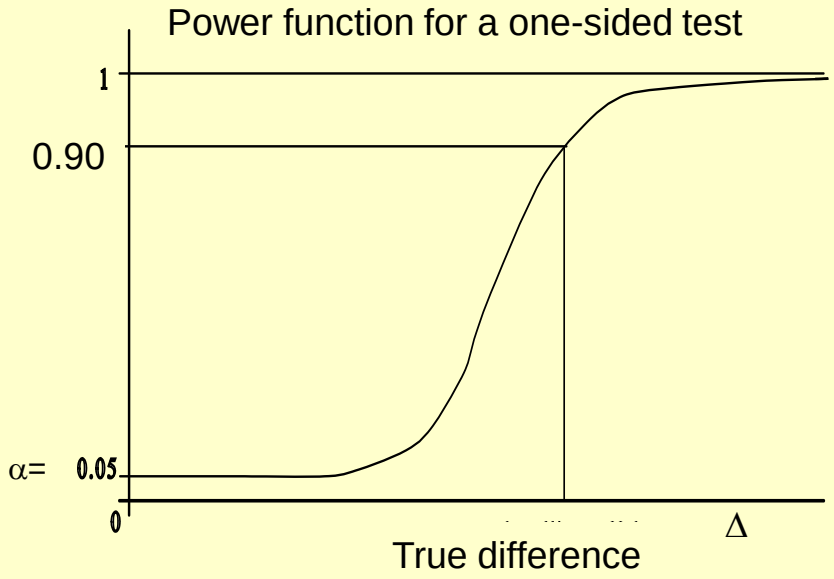
The chance of drawing the correct conclusion.

Statistical power.

The study's statistical power given a specified difference Δ is the chance that the hypothesis test will reject the null hypothesis given that the true difference is as specified (Δ).

The power is denoted by Π(Δ) and is related to the risk of acceptance error as follows:

$$\Pi(\Delta) = 1 - \beta(\Delta).$$



I recommend that you design your study so that Π(Δ) is at least 0.90 for the least difference you judge to be of clinical importance.

Controlling the chance of a correct conclusion.

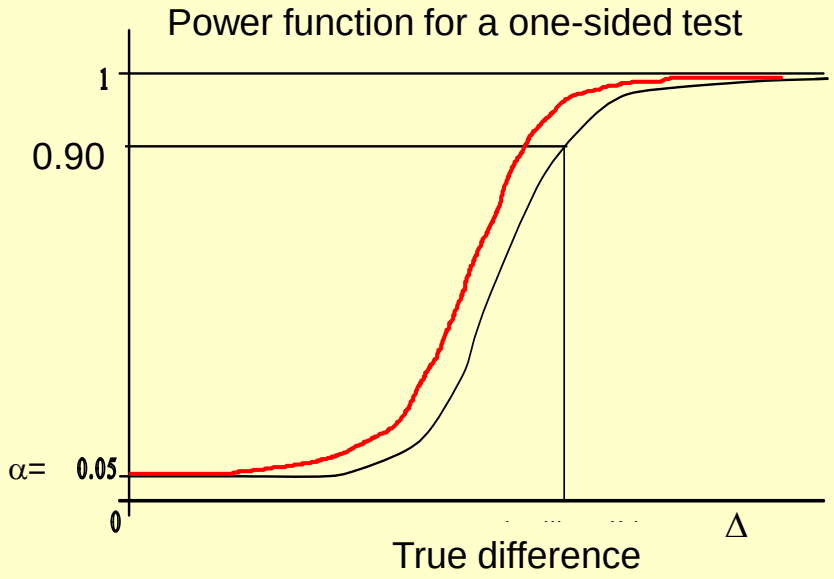
What determines the statistical power?

I) The chosen statistical test:
Optimal choice gives largest power.

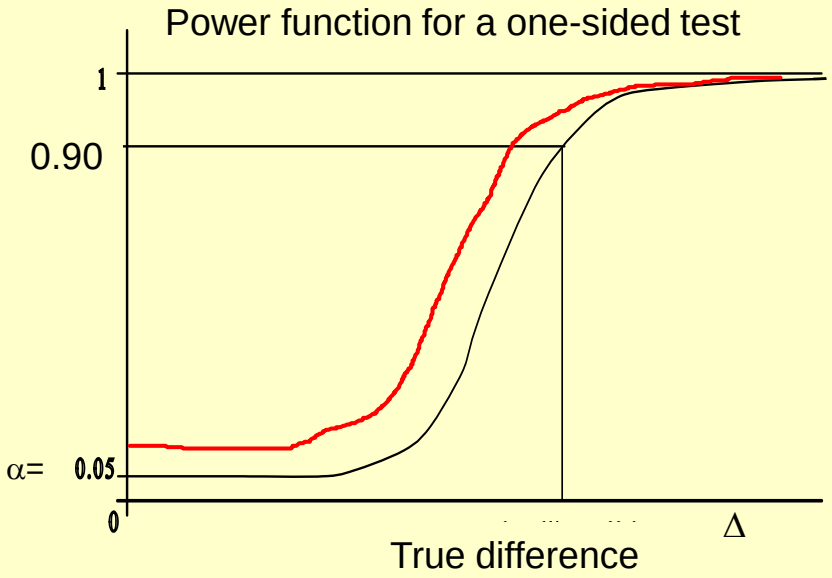
II) The significance level (α):
Increasing the significance level increases the power.

III) The difference (Δ):
Larger difference gives better power.

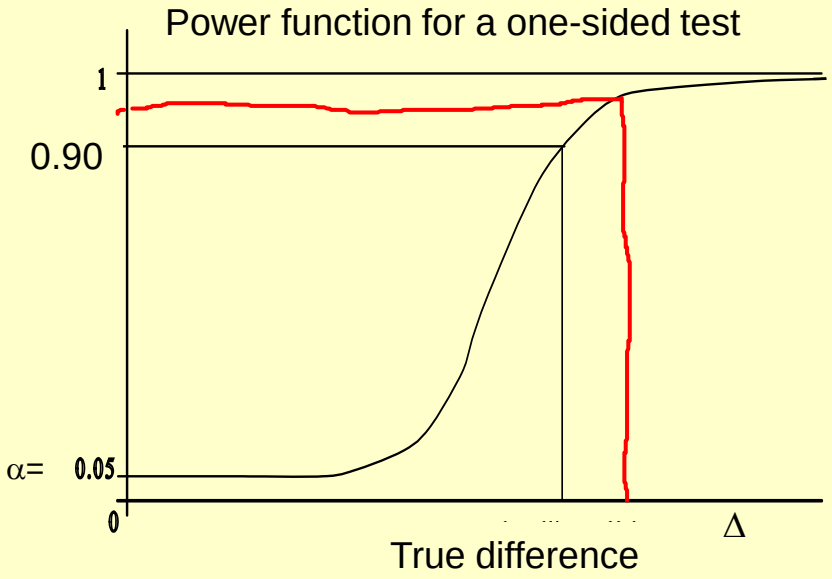
IV) The size of the study groups(n):
Larger groups give larger power.



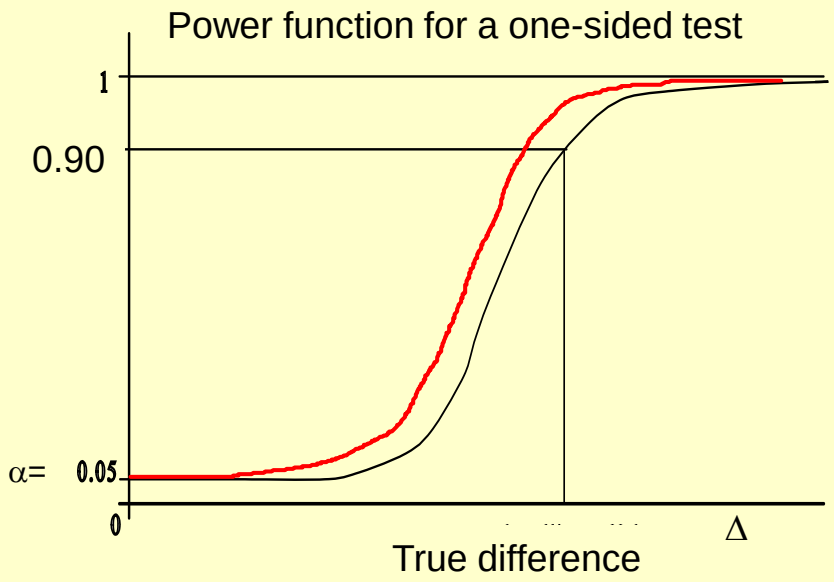
Optimal statistical test gives better power



Increased risk of rejection error increases the power



Increased true difference increases the power



Increased sample size (n) increases the power

2. Precision

- An estimate (of a population mean or a proportion) is usually reported with a measure of uncertainty like a standard error (*se*) or 95 % confidence interval.
- The ideal would be to have no uncertainty (*se* = 0), but *se* is never 0 unless all observations are equal or you have infinitely many observations. However, *se* is a decreasing function of the sample size (*n*).

The standard error

- For a sample of size n with a standard deviation of *s* the standard error is

  – $se = s/\sqrt{n}$

  – The standard deviation is a ‘fixed’ population quantity
  – The standard error is a decreasing function of the sample size

Calculating the sample size

- If you require an estimate with (at most) standard error *se* you will need a sample from the population of (at least) size

$$n = (s/se)^2$$

- Example: If you wish to estimate mean blood pressure in a population where the standard deviation is $s = 5$ with a standard error of $se = 1$ you will need $n = (5/1)^2 = 25$ persons.

3. Software

- Sample Power 3.0 is a program that comes free with SPSS, but is installed separately. Stata, StatXact, R and other have options for sample size and power.
- If you Google ‘sample power’ you’ll find other (also free of charge) programs.
- You can even find apps for your smart phone for the simplest scenario.

Example

A randomized blinded smoking cessation trial

Group	Proportion Positive	N Per Group	Standard Error	95% Lower	95% Upper
Population 1	0,025	906			
Population 2	0,050	906			
Rate Difference	-0,025	1 812	0,009	-0,042	-0,008

Alpha= 0,050, Tails= 2, Power = 0,800

Power computation: Normal approximation (unweighted mean p)
Precision computation: Log method

Example: How to formulate

- To detect a true difference of 2.5 % points in the proportions of smokers with a statistical power of 80% you would need 1872 persons randomized to smoking cessation advice and not smoking cessation advice in a two-sided hypothesis test at significance level 5 %.

Summary

- Calculating the minimum number of patients necessary to get the required power or precision is important.
- A study with too small power or bad precision will be unethical to perform (waisted resources, uneccessary patient inclusion)
- Such calculation must be included in applications and protocol.
- The calculation of necessary patients is an academic exercise that can be based on previous empirical knowledge, qualified guessing or a pilot study.
- Such calculation(s) should at least be done for the study’s primary research goal(s).