Sample size and Power calculations

The essential question in any trial/analysis:

"How many patients/persons/observations do I need?"
Sample size (an example)

“Twenty patients (10 in Arm A and 10 in Arm B) will be included initially as a "run in phase" of the study for the initial evaluation of feasibility and safety... The median survival in patients who are given (...) is taken to be 6.5 months. To detect an increase of at least 3 months in survival among patients given (...) (ie. to 9.5 months), the trial would need to recruit 50 patients; this with 70% power and a level of significance of 5% (two-sided)

Sample size and Power calculations

The essential question in any trial/analysis:
"How many patients/persons/observations do I need?"

a. Did we get an answer to our question?
b. Are we satisfied with the answer?
Sample size (an example)

“Twenty patients (10 in Arm A and 10 in Arm B) will be included initially as a “run in phase” of the study for the initial evaluation of feasibility and safety… The median survival in patients who are given (...) is taken to be 6.5 months. To detect an increase of at least 3 months in survival among patients given (...) (ie. to 9.5 months), the trial would need to recruit 50 patients; this with 70% power and a level of significance of 5% (two-sided)
Expect three questions in return (II)

- What is the size of the difference you would like to observe?
  - Which relates to the magnitude of the effect you aim to uncover, - from your clinical and biological point of view
  - In other words: What is the minimum difference that is of clinical importance (significance) to you?

Expect three questions in return (III)

- How sure do you want to be (ie. once you draw your conclusion)?
  - That your observed difference is a "true" one
  - That the difference you look for, is not overlooked
Sample size (an example)

"Twenty patients (10 in Arm A and 10 in Arm B) will be included initially as a "run in phase" of the study for the initial evaluation of feasibility and safety... The median survival in patients who are given (...) is taken to be 6.5 months. To detect an increase of at least 3 months in survival among patients given (...) (ie. to 9.5 months), the trial would need to recruit 50 patients; this with 70% power and a level of significance of 5% (two-sided)."

During this lecture we shall touch upon

- One sample test
- Two sample test
- One sided test
- Two sided test
- Continuous variables
- Proportions
Two quotations worth noticing:

"An hypothesis is a statement whose incorrect rejection one tries to avoid"

**#**

- "Hypotheses can only be tested, - but never proven"

Testing of hypothesis requires that you consider:

The Null vis á vis The Alternative hypothesis

**#**

\[ H_0: \mu = \mu_0 \text{ vs. } H_1: \mu = \mu_1 \]
For instance: A randomised controlled trial

Ie. A comparison between Trt. A vs Trt. B

\[ H_0: \mu_A = \mu_B \text{ vs. } H_1: \mu_A \neq \mu_B \]

Given that \( H_0 = \text{True} \)

| Our conclusion | Yes = True | No = False |
|----------------|------------|------------|

"The ABSOLUTE Truth" Yes = True ok Type I error (\( \alpha \))

No = False Type II error ok (\( \beta \))
Significance (statistical):

The probability to reject $H_0$ when $H_0$ is true.
Expressed as our choice of the level of $\alpha$
(= "Our willingness to commit a Type I error")

Power:

The ability of our study to reject $H_0$ när $H_0$ false.
Expressed as our choice of the level of $\beta$ (or really $1 - \beta$)
(= "Our willingness to commit a Type II error")

Power calculations - Practical

Suppose that a case-control study is planned for assessing the relationship between smoking in pregnancy and low birth weight in the offspring. Cases would be women giving birth to babies weighing 2500 grams or below, and controls would be women giving birth to babies over 2500 grams. Because only a small minority of the population falls into the case group, the overall prevalence of smoking in the general population of pregnant women serves quite well as an estimate of $p_0$, which in a case-control study denotes the proportion of controls who have the exposure (in this case 25%). Suppose that an odds ratio of 1.8 is regarded as important to detect, that 175 cases are available for study, and that a case-to-control ratio ($r$) is planned.

What power has the study to detect an odds ratio = 1.8?
Table 10–11. Definitions of symbols used in equations for calculating power and required sample size

| Symbol | Definition |
|--------|------------|
| $d^*$  | Non-null value of the difference in proportions or means (i.e., the magnitude of difference one wishes to detect) |
| $n$    | In a cohort or cross-sectional study, the number of exposed individuals studied; in a case-control study, the number of cases |
| $r$    | In a cohort or cross-sectional study, the ratio of the number of unexposed individuals studied to the number of exposed individuals studied; in a case-control study, the ratio of the number of controls studied to the number of cases studied |
| $\sigma$ | Standard deviation in the population for a continuously distributed variable |
| $p_e$  | In a cohort study (or a cross-sectional study), the proportion of exposed individuals who develop (or have) the disease; in a case-control study, the proportion of cases who are exposed |
| $p_c$  | In a cohort study (or a cross-sectional study), the proportion of unexposed individuals who develop (or have) the disease; in a case-control study, the proportion of controls who are exposed |
| $\bar{p}$ | $\frac{p_e + p_c}{1 + r}$ – weighted average of $p_e$ and $p_c$ |

Table 10–12. Formulas for use in calculating the power of a study to detect an association

$Z_d$ for difference in means: $Z_d = \frac{d^*}{\sigma} \sqrt{\frac{nr}{r + 1}} - Z_{\alpha/2}$

$Z_p$ for difference in proportions: $Z_p = \frac{n(d^*)}{\sqrt{r(1 + r)[p(1 - p)]}} - Z_{\alpha/2}$

Value of $p_e$ in terms of $p_c$ and a specified odds ratio (OR): $p_e = \frac{p_c \cdot OR}{1 + p_c \cdot (OR - 1)}$

Value of $p_c$ in terms of $p_e$ and a specified risk ratio (RR) (for use in cohort or cross-sectional studies only): $p_c = p_e \cdot RR$
Table 10-15. Formulas for use in calculations of required sample size

Difference in means:

\[ n = \left( \frac{Z_{\alpha/2} + Z_{\beta}}{\sigma} \right)^2 \left( \frac{1}{\pi} + 1 \right) \]

Difference in proportions:

\[ n = \left( \frac{Z_{\alpha/2} + Z_{\beta}}{d} \right)^2 \left( \frac{\pi(1-\pi)}{\pi^2} + 1 \right) \]

---

Table 10-14. Conversion of \( z \) to the percentage corresponding to the power for detecting an association

| \( z \)   | 0.000 | 0.010 | 0.020 | 0.030 | 0.040 | 0.050 | 0.060 | 0.070 | 0.080 | 0.090 | 0.100 | 0.110 | 0.120 | 0.130 | 0.140 | 0.150 | 0.160 | 0.170 | 0.180 | 0.190 | 0.200 | 0.210 | 0.220 | 0.230 | 0.240 | 0.250 | 0.260 | 0.270 | 0.280 | 0.290 | 0.300 | 0.310 | 0.320 | 0.330 | 0.340 | 0.350 | 0.360 | 0.370 | 0.380 | 0.390 | 0.400 | 0.410 | 0.420 | 0.430 | 0.440 | 0.450 | 0.460 | 0.470 | 0.480 | 0.490 | 0.500 | 0.510 | 0.520 |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1.3       | 9.7   | 9.5   | 9.5   | 9.7   | 9.9   | 10.1  | 10.4  | 10.6  | 10.4  | 10.2  | 10.0  | 9.9   | 9.6   | 9.3   | 8.9   | 8.5   | 8.1   | 7.6   | 7.1   | 6.6   | 6.1   | 5.6   | 5.1   | 4.6   | 4.1   | 3.6   | 3.1   | 2.6   | 2.1   | 1.6   | 1.1   | 0.6   | 0.1   | 0.6   | 1.1   | 1.6   | 2.1   | 2.6   | 3.1   | 3.6   | 4.1   | 4.6   | 5.1   | 5.6   | 6.1   | 6.6   | 7.1   | 7.6   | 8.1   | 8.6   |
| 0.9       | 7.5   | 7.3   | 7.3   | 7.5   | 7.8   | 8.1   | 8.4   | 8.6   | 8.4   | 8.2   | 8.0   | 7.8   | 7.6   | 7.4   | 7.2   | 7.0   | 6.8   | 6.6   | 6.4   | 6.2   | 6.0   | 5.8   | 5.6   | 5.4   | 5.2   | 5.0   | 4.8   | 4.6   | 4.4   | 4.2   | 4.0   | 3.8   | 3.6   | 3.4   | 3.2   | 3.0   | 2.8   | 2.6   | 2.4   | 2.2   | 2.0   | 1.8   | 1.6   | 1.4   | 1.2   | 1.0   | 0.8   | 0.6   |
| 0.6       | 5.5   | 5.3   | 5.3   | 5.6   | 5.9   | 6.2   | 6.5   | 6.8   | 6.5   | 6.3   | 6.1   | 5.9   | 5.7   | 5.5   | 5.3   | 5.1   | 4.9   | 4.7   | 4.5   | 4.3   | 4.1   | 3.9   | 3.7   | 3.5   | 3.3   | 3.1   | 2.9   | 2.7   | 2.5   | 2.3   | 2.1   | 1.9   | 1.7   | 1.5   | 1.3   | 1.1   | 0.9   | 0.7   | 0.5   | 0.3   | 0.1   | 0.9   | 0.7   | 0.5   | 0.3   | 0.1   | 0.9   |
| 0.4       | 3.5   | 3.3   | 3.3   | 3.6   | 3.9   | 4.2   | 4.5   | 4.8   | 4.6   | 4.4   | 4.2   | 4.0   | 3.8   | 3.6   | 3.4   | 3.2   | 3.0   | 2.8   | 2.6   | 2.4   | 2.2   | 2.0   | 1.8   | 1.6   | 1.4   | 1.2   | 1.0   | 0.8   | 0.6   | 0.4   | 0.2   | 0.0   | 0.8   | 0.6   | 0.4   | 0.2   | 0.0   | 0.8   |
| 0.2       | 1.5   | 1.3   | 1.3   | 1.6   | 1.9   | 2.2   | 2.5   | 2.8   | 2.6   | 2.4   | 2.2   | 2.0   | 1.8   | 1.6   | 1.4   | 1.2   | 1.0   | 0.8   | 0.6   | 0.4   | 0.2   | 0.0   | 0.8   | 0.6   | 0.4   | 0.2   | 0.0   | 0.8   |
| 0.1       | 0.5   | 0.4   | 0.4   | 0.6   | 0.9   | 1.2   | 1.5   | 1.8   | 1.6   | 1.4   | 1.2   | 1.0   | 0.8   | 0.6   | 0.4   | 0.2   | 0.0   | 0.8   | 0.6   |

Note: For values of \( z \) less than 1.99, the power is less than 0.2%; for values of \( z \) greater than 2.45, the power is greater than 99.8%.
Suppose that a case-control study is planned for assessing the relationship between smoking in pregnancy and low birth weight in the offspring. Cases would be women giving birth to babies weighing 2500 grams or below, and controls would be women giving birth to babies over 2500 grams. Because only a small minority of the population falls into the case group, the overall prevalence of smoking in the general population of pregnant women serves quite well as an estimate of $p_0$, which in a case-control study denotes the proportion of controls who have the exposure (in this case 25%). Suppose that an odds ratio of 1.8 is regarded as important to detect, that 175 cases are available for study, and that a case-to-control ratio ($r$) is planned.

What power has the study to detect an odds ratio = 1.8?

**Power Calculation: Practical**

- $p_0$: Proportion of exposed controls
- OR of importance to identify = 1.8
- Number ♂ available for study: $n = 175$
- Control - to - case ratio = 2

$$p_1 = \frac{(0.25) \times (1.8)}{[1+(0.25) \times (1.8 - 1)]} = 0.375$$

$$d^* = p_1 - p_0 = 0.375 - 0.250 = 0.125$$

$$\bar{p} = \frac{[(0.375) + (2) \times (0.25)]}{(1 + 2)} = 0.292$$

$$Z_\theta = \sqrt{\frac{(175) \times (0.125)^2 \times (2)}{(2+1) \times 0.292 \times 0.708}}^{1/2} = 1.96 = 1.01$$
Factors affecting the power ($1 - \beta$)

- If the significance level is made smaller ($\alpha$ decreases), $z_{\alpha}$ decreases and hence the power decreases
- If the alternative mean is shifted further away from the null mean ($|\mu_1 - \mu_0|$ increases), then the power increases
- If the standard deviation of the distribution of individual observations increases ($\sigma$ increases), then the power decreases
- If the sample size increases ($n$ increases), then the power increases

Factors affecting the sample size ($n$)

- The sample size increases as $\sigma^2$
- The sample size increases as the significance level is made smaller ($\alpha$ decreases)
- The sample size increases as the required power increases ($1 - \beta$ increases)
- The sample size decreases as the absolute value of the distance between the null and alternative means ($|\mu_1 - \mu_0|$) increases
Medical statistics, Part 1
Faculty of medicine, NTNU 2009

What you need to know and decide
1. $\alpha$ Level and $Z_{\alpha/2}$
2. $\beta$ level and $Z_{\beta}$
3. $\delta$ level = difference (in prevalence, incidence, or any outcome variable) between the groups you want to observe

Then – and only then – can you calculate the number needed in your study (ie. $n$ in each arm)