Power Analysis & Sample Size Calculation:
Why is it important?

F. Farrokhyyar, MPhil, PhD
Department of Surgery
Department of Clinical Epidemiology & Biostatistics
Casual relationship

Any association shown in a study:

- Errors and biases
- Confounding
- Chance
- Causation/association
Errors and biases

- Error is inaccuracy which is similar between different groups of study. “non-differential error”

- Bias is inaccuracy which is different in size and direction between different groups of study; such as selection bias, recall bias, observation bias, measurement bias.
Confounding bias – a distortion of an exposure–outcome relationship brought about by the association of another factor with both outcome and exposure.

- **Low exercise**
  - Positive association
  - Overall OR=4.2
  - True OR=3.0

- **MI**
  - True OR=2.0

**Obesity**
The third non-causal explanation for an association is due to chance variation. Is the difference in outcome between the two groups larger than we would expect to occur purely by chance?

There are two general approaches to assess the role of chance in a clinical observation:

- Confidence intervals – concept of precision
- Hypothesis testing – P value
Stating Hypotheses

✓ **Clinical hypothesis** – Does laparoscopic surgery reduces post-op pain in patients with hernia repair?

✓ **Null Hypothesis** – Mean post-op pain with laparoscopic surgery is no different from mean post-op pain with open surgery.

\[ H_0: \mu_L = \mu_O \]

✓ **Alternate Hypothesis** - Mean post-op pain with laparoscopic surgery is different from mean post-op pain with open surgery.

\[ H_a: \mu_L \neq \mu_O \quad \text{or} \quad H_a: \mu_L < \mu_O \]
Assumptions in hypotheses testing

- The reason for having a null hypothesis is to be able to calculate errors in decision making.

- Most common assumptions about hypothesis testing are:
  - Errors are drawn from normal distribution except in non-parametric tests.
  - Errors are independent.
<table>
<thead>
<tr>
<th>Truth</th>
<th>Decision</th>
<th>H₀ true</th>
<th>H₀ false</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₀ true</td>
<td>Correct decision</td>
<td>True negative (Probability 1-α)</td>
<td>False positive (probability α)</td>
</tr>
<tr>
<td>H₀ false</td>
<td>Type II error</td>
<td>False negative (probability β)</td>
<td>Correct decision</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>True positive (probability 1- β)</td>
</tr>
</tbody>
</table>
Errors probabilities

✓ Type I ($\alpha$) error – the null hypothesis may be rejected when it should have been accepted. It’s probability of occurring by chance is denoted as $\alpha$.

$$\alpha = \text{prob}[\text{rejecting } H_0|H_0 \text{ true}]$$

✓ Type II ($\beta$) error – the null hypothesis may be accepted when it should have been rejected. It’s probability of occurring by chance is denoted $\beta$.

$$\beta = \text{prob}[\text{accepting } H_0|H_0 \text{ false}]$$

$$1 - \beta = \text{power of the test.}$$
Why do we need to calculate sample size?

- To make sure our study sample is large enough to detect the minimum clinically important differences under the fixed chances of type I and II errors.

- Larger samples provide better estimates, narrower confidence intervals and smaller test errors.

- Note: The calculated sample size is the minimum number required.
Sample size calculation or “Power Analysis”

Key concepts:

- Type I error or $\alpha$ error
- Type II error or $\beta$ error
- Outcome of interest; whether the outcome of interest is a categorical (proportions), or a continuous (means) variable.
- Effect size: the magnitude of the difference to be detected.
- One-sided or two-sided test
Steps in sample size calculation

To calculate sample size:

- we fix type I error at $\alpha=5\%$
- try to make type II error ($\beta$) small - usually 20% (power=$1-\beta=80\%$).
- we find the information on the proportion or mean and standard deviation using the data from our study or from literature.
- we decide on the minimum clinically important difference we are willing to detect ($d$).
Sample Size Formula: Continuous Variable

ONE MEAN:

\[ n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2}{(m - \mu)^2} \]

\[ \sigma = \text{population standard deviation}, \]

\[ \mu = \text{population mean}, \quad m = \text{sample mean} \]

\[ Z_{1-\alpha/2} = 1.96 \text{ when } \alpha = 5\% \text{ for two-sided test} \]

\[ Z_{1-\alpha} = 1.645 \text{ when } \alpha = 5\% \text{ for one-sided test} \]

\[ Z_{1-\beta} = 0.842 \text{ when } \beta = 20\% \text{ (80\% power)} \]
Example for one mean

An emergency physician wants to test the effectiveness of a "GI cocktail" to treat emergency dyspeptic symptoms on a scale of 1-10 pain score. Standard deviation of pain score without treatment is $\sigma = 1.73$. He considers a reduction in pain score of $d = 1.5$ points as clinically relevant. He believes that the treatment cannot increase pain. How many patients will he need with $\alpha = 5\%$ and power $= 80\%$?
Example for one mean

An emergency physician wants to test the effectiveness of a “GI cocktail” to treat emergency dyspeptic symptoms on a scale of 1-10 pain score. $\sigma$ of pain score without treatment is 1.73. He considers a reduction in pain scale of $d=1.5$ points as clinically relevant. He believes that the treatment cannot increase pain. How many patients will he need using $\alpha=5\%$ and power=80%?

$z_{1-\alpha}=1.645$, for 1- sided test when $\alpha=5\%$

$$n= \frac{(1.645+ 0.842)^2 (1.73)^2}{(1.5)^2}$$

$n= 8.21$
Sample Size Formula: Continuous variable

TWO MEANS:

\[
n_1 = n_2 = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 (\sigma_1^2 + \sigma_2^2)}{(m_1 - m_2)^2}
\]

- \( m_1, m_2 \) = mean value of each group
- \( \sigma_1, \sigma_2 \) = standard deviation for each group
Example for two means

From literature, mean LV mass indexed regression in patients who had a mechanical valve replacement is $115 \pm 20 \text{ g/m}^2$ and those who had a tissue valve is $131 \pm 35 \text{ g/m}^2$. We want to detect a $15 \text{ g/m}^2$ difference in mean LV mass indexed regression between two valve types with alpha of 0.05 and 90% power. How many patients do we need?
From literature, mean LV mass indexed regression in patients who had a mechanical valve replacement is $115 \pm 20 \, \text{g/m}^2$ and those who had a tissue valve is $131 \pm 35 \, \text{g/m}^2$. We want to detect $15 \, \text{g/m}^2$ difference in mean LV mass indexed regression between two valve types with alpha of 0.05 and 90% power. How many patients do we need?

$Z_{1-\beta} = 1.282$, is the statistic for $1-\beta = 10%$

$$n_1 = n_2 = \frac{(1.96 + 1.282)^2 \times (35^2 + 20^2)}{(15)^2}$$

$n_1 = n_2 = 76$
Sample Size Formula: Categorical variable

ONE PROPORTION:

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

\( \pi \) = population proportion, \( p \) = sample proportion

\( Z_{1-\alpha/2} = 1.96 \) when \( \alpha = 0.05 \) for two-sided test
\( Z_{1-\alpha} = 1.645 \) when \( \alpha = 0.05 \) for one-sided test
\( Z_{1-\beta} = 0.842 \) when \( \beta = 0.2 \) (80% power)
\( Z_{1-\beta} = 1.282 \) when \( \beta = 0.1 \) (90% power)
Example for one proportion

The prevalence of IBS in general population is 20%. We think the prevalence might be higher in IBD patients. How many patients with IBD do we need to detect a difference of 10% with 80% power and 5% significance level.
The prevalence of IBS in general population is 20%. We think the prevalence might be higher in IBD patients. How many patients with IBD do we need to detect a difference of 10% with 80% power and 5% significance level.

\[
n = \frac{[(1.96\sqrt{0.2(1-0.2)} + 0.84\sqrt{0.3(1-0.3)}]^2}{(0.3-0.2)^2}
\]

\[
\pi = 0.2, \quad p = 0.3, \quad z_{1-\alpha/2} = 1.96, \quad z_{1-\beta} = 0.842
\]

N = 117
Sample Size Formula: Categorical variable

TWO PROPORTIONS: when $p_m$ is not close to 0.

$$n_1 = n_2 = \frac{[(z_{1-\alpha/2}\sqrt{2p_m(1-p_m)} + z_{1-\beta}\sqrt{p_1(1-p_1) + p_2(1-p_2)}]^2}{(p_1 - p_2)^2}$$

$p_1, p_2 = \text{sample proportions from our data}$

$P_m = p_1 + p_2 / 2$
Example for two proportions

From literature the prevalence of wound infection in patients after colon resection is 18%. How many patients do we need to detect a difference of 6% wound infection between new antibiotic and standard treatment when $\alpha=5\%$ and $\beta=20\%$. 
Example for two proportions

From literature the prevalence of wound infection in patients after colon resection is 18%. How many patients do we need to detect a difference of 6% wound infection between new antibiotic and standard treatment when \( \alpha = 5\% \) and \( \beta = 20\% \).

\[
\begin{align*}
\frac{1}{n_1} &= \frac{1}{n_2} = \frac{[(1.96\sqrt{2\times0.15\times0.85}) + 0.84\sqrt{0.12\times0.88+0.18\times0.82})]^2}{(0.12-0.18)^2} \\
p_1 &= 0.18, \quad p_2 = 0.12, \quad p_m = 0.15, \\
n_1 = n_2 = 555
\end{align*}
\]
Sample Size Formula: Categorical variable

TWO PROPORTIONS: when $p_m$ is close to 0.

\[
n_1 = n_2 = \frac{[(z_{1-\alpha/2} + z_{1-\beta}) \sqrt{p_1 + p_2}]^2}{(p_1 - p_2)^2}
\]

$p_1$, $p_2$ = sample proportions from our data
Example for two proportions

From our pilot data, the wound infection from open hernia repair is 6% and laparoscopic repair is 2%. How many patients do we need to detect a difference of 4% in wound infection between laparoscopic and open technique when $\alpha = 5\%$ and $\beta = 20\%$. 
Example for two proportions

From our pilot data, the wound infection from open hernia repair is 6% and laparoscopic repair is 2%. How many patients do we need to detect a difference of 4% in wound infection between laparoscopic and open technique when $\alpha = 5\%$ and $\beta = 20\%$.

\[
\begin{align*}
n_1 &= n_2 \frac{(1.96 + 0.842) \sqrt{0.06 + 0.02}}{(0.06 - 0.02)^2} \\
p_1 &= 0.06, \quad p_2 = 0.02, \quad n_1 = n_2 = 393
\end{align*}
\]
Sample size calculated vs. sample size needed

The calculated sample size should be increased by a ‘safety factor’ to account for the possible sources of error in the calculation of sample size that could affect “study power” such as wrong data source, randomness, lost to follow up, mortality, low compliance, allocation ratio and so on.