

# MSMS 304 - Biostatistics

Ananda Biswas

31.07.2025

• **Superiority, Equivalency, Non-Inferiority Trials** : Suppose the control has a cure rate of  $x\%$ . The new intervention is said to be superior than the current control if the cure rate of the new intervention is greater than or equal to  $(x + \delta)\%$  where  $\delta > 0$  is a predefined margin called superiority margin. Many clinical trials are designed to demonstrate that a new intervention is superior to some established intervention. Such trials are called Superiority trials.

However, not all trials share this goal. New interventions may have no superiority to existing therapy, but as long as they are not materially worse, maybe of interest because they are less toxic, bear less cost, require fewer doses, or have some other value to patients. In trials where we test whether the new intervention is no worse than or at least as good as, some established intervention, are called Non-inferiority trials.

Moreover, in some trials the goal is not to determine whether the new intervention is better or worse, rather we want to test whether it is similar enough in effectiveness to be used interchangeably with the current intervention. Such trials are called Equivalency trials.

• **Sample Size Determination** : Clinical trials should have sufficient statistical power to detect differences between groups considered to be of clinical importance. Therefore calculation of sample size with provision for adequate levels of significance and power is an essential part of planning.

If the sample size is drastically large, then we may put human lives at risk unnecessarily. On the other hand, if the sample size is too small, there is a good chance that the trial will fall short in demonstrating any difference between study groups.

To detect a large difference between two study groups as statistically significant, small sample size is enough. But to detect a small difference between the study groups as significant, we require big enough sample size.

Note : Prevalence of a condition in a region = 
$$\frac{\text{total number of cases of the health condition}}{\text{total population in the region}}$$

📖 **A Case Scenario** : A researcher is interested in estimating the prevalence of Type II diabetes amongst the population in Jaipur. It was assumed that the anticipated prevalence of Type II diabetes would not be more than 15% in the population. What is the minimum sample size required to estimate this prevalence at 95% confidence level and 5% absolute precision ? (Absolute prevalence is the maximum allowable difference between the estimated prevalence and the true prevalence.)


$$p = 0.15, 1 - \alpha = 0.95, d = 0.05$$

$$\begin{aligned}
n &\geq \frac{Z_{1-\frac{\alpha}{2}}^2 \cdot p \cdot (1-p)}{d^2} \\
&= \frac{(1.96)^2 \cdot 0.15 \cdot 0.85}{(0.05)^2} \\
&= 196
\end{aligned}$$

From the formula used above, sample size  $n$  increases as  $p$  increases up to  $p = 0.5$  *i.e.* 50%.

Absolute Precision	Anticipated Prevalence		
	0.10	0.15	0.20
0.15	16	22	28
0.20	9	13	16
0.25	6	8	10

As you might notice, for a fixed absolute precision, as anticipated prevalence increases required sample size also increases. This directly follows from the above formula. The sample size  $n$  and  $p$  are directly proportional.

 A Case Scenario : A researcher is interested in estimating the prevalence of Type II diabetes amongst the population in Jaipur. It was assumed that the anticipated prevalence of Type II diabetes would not be more than 15% in the population. What is the minimum sample size required to estimate this prevalence at 95% confidence level and 20% relative precision ? (Relative prevalence is the maximum allowable error in the estimated prevalence expressed as a proportion of the true value of prevalence.)

$$p = 0.15, 1 - \alpha = 0.95, d = 15 \times \frac{20}{100} = 3$$

$$\begin{aligned}
n &\geq \frac{Z_{1-\frac{\alpha}{2}}^2 \cdot p \cdot (1-p)}{(pd)^2} \\
&= \frac{(1.96)^2 \cdot 0.15 \cdot 0.85}{(0.15 \times 3)^2} \\
&= 545
\end{aligned}$$

Relative Precision	Anticipated Prevalence		
	0.10	0.15	0.20
0.15	1537	968	683
0.20	865	545	385
0.25	534	349	246

Here, for a fixed relative precision, as anticipated prevalence increases required sample size decreases. This directly follows from the formula (Notice an additional  $p^2$  in the denominator). So as  $p$  increases sample size  $n$  decreases.



A few advantages of relative precision over absolute precision are noteworthy.

For rare diseases, prevalence is too small, not more than 3% to 5%. Then the choice of an absolute precision 10% will result the tolerance boundary of estimated prevalence to go beyond 0 in the negative region. To prevent such scenarios, relative precision is preferred.

Another edge case might be for very frequent conditions the prevalence is high, for example 95% or 97%. Then a choice of absolute precision 10% will push the tolerance boundary of estimated prevalence beyond 100%, which is fallacious. This adds to demerits of absolute precision.