INVITED ARTICLE

An Introduction to Statistics: Understanding Hypothesis Testing and Statistical Errors

Priya Ranganathan[1], Pramesh CS[2]

Abstract

The second article in this series on biostatistics covers the concepts of sample, population, research hypotheses and statistical errors.

Keywords: Biostatistics, Research design, Statistical bias

Indian Journal of Critical Care Medicine (2019): 10.5005/jp-journals-10071-23259

Two papers quoted in this issue of the Indian Journal of Critical Care Medicine report. The results of studies aim to prove that a new intervention is better than (superior to) an existing treatment. In the ABLE study, the investigators wanted to show that transfusion of fresh red blood cells would be superior to standard-issue red cells in reducing 90-day mortality in ICU patients.[1] The PROPPR study was designed to prove that transfusion of a lower ratio of plasma and platelets to red cells would be superior to a higher ratio in decreasing 24-hour and 30-day mortality in critically ill patients.[2] These studies are known as superiority studies (as opposed to noninferiority or equivalence studies which will be discussed in a subsequent article).

[1]Department of Anesthesiology, Critical Care and Pain, Tata Memorial Hospital, Mumbai, Maharashtra, India

[2]Department of Surgical Oncology, Tata Memorial Centre, Mumbai, Maharashtra, India

Corresponding Author: Priya Ranganathan, Department of Anesthesiology, Critical Care and Pain, Tata Memorial Hospital, Mumbai, Maharashtra, India, e-mail: drpriyaranganathan@gmail.com

How to cite this article: Ranganathan P, Pramesh CS. An Introduction to Statistics: Understanding Hypothesis Testing and Statistical Errors. Indian J Crit Care Med 2019;23(Suppl 3):S230–S231.

Source of support: Nil

Conflict of interest: None

Sample versus Population

A sample represents a group of participants selected from the entire population. Since studies cannot be carried out on entire populations, researchers choose samples, which are representative of the population. This is similar to walking into a grocery store and examining a few grains of rice or wheat before purchasing an entire bag; we assume that the few grains that we select (the sample) are representative of the entire sack of grains (the population).

The results of the study are then extrapolated to generate inferences about the population. We do this using a process known as hypothesis testing. This means that the results of the study may not always be identical to the results we would expect to find in the population; i.e., there is the possibility that the study results may be erroneous.

Hypothesis Testing

A clinical trial begins with an assumption or belief, and then proceeds to either prove or disprove this assumption. In statistical terms, this belief or assumption is known as a hypothesis. Counterintuitively, what the researcher believes in (or is trying to prove) is called the "alternate" hypothesis, and the opposite is called the "null" hypothesis; every study has a null hypothesis and an alternate hypothesis. For superiority studies, the alternate hypothesis states that one treatment (usually the new or experimental treatment) is superior to the other; the null hypothesis states that there is no difference between the treatments (the treatments are equal). For example, in the ABLE study, we start by stating the null hypothesis—there is *no* difference in mortality between groups receiving fresh RBCs and standard-issue RBCs. We then state the alternate hypothesis—There *is* a difference between groups receiving fresh RBCs and standard-issue RBCs. It is important to note that we have stated that the groups are different, without specifying which group will be better than the other. This is known as a two-tailed hypothesis and it allows us to test for superiority on either side (using a two-sided test). This is because, when we start a study, we are not 100% certain that the new treatment can only be better than the standard treatment—it could be worse, and if it is so, the study should pick it up as well. One tailed hypothesis and one-sided statistical testing is done for non-inferiority studies, which will be discussed in a subsequent paper in this series.

Statistical Errors

There are two possibilities to consider when interpreting the results of a superiority study. The first possibility is that there is truly no difference between the treatments but the study finds that they are different. This is called a **Type-1 error or false-positive error or alpha error.** This means falsely rejecting the null hypothesis.

The second possibility is that there is a difference between the treatments and the study does not pick up this difference. This is called a **Type 2 error or false-negative error or beta error.** This means falsely accepting the null hypothesis.

The power of the study is the ability to detect a difference between groups and is the converse of the beta error; i.e., power = 1-beta error. Alpha and beta errors are finalized when the protocol is written and form the basis for sample size calculation for the study. In an ideal world, we would not like any error in the results of our study; however, we would need to do the study in the entire

Table 1: Statistical errors

(a) Types of statistical errors

		Study findings: Null hypothesis is	
		True	False
Null hypothesis is actually	True	Correct results!	Falsely rejecting null hypothesis - Type I error
	False	Falsely accepting null hypothesis - Type II error	Correct results!

(b) Possible statistical errors in the ABLE trial

		Study findings	
		There is *no* difference in mortality between groups receiving fresh RBCs and standard-issue RBCs	There *is a* difference in mortality between groups receiving fresh RBCs and standard-issue RBCs
Truth	There is *no* difference in mortality between groups receiving fresh RBCs and standard-issue RBCs	Correct results!	Falsely rejecting null hypothesis - Type I error
	There *is a* difference in mortality between groups receiving fresh RBCs and standard-issue RBCs	Falsely accepting null hypothesis - Type II error	Correct results!

population (infinite sample size) to be able to get a 0% alpha and beta error. These two errors enable us to do studies with realistic sample sizes, with the compromise that there is a small possibility that the results may not always reflect the truth. The basis for this will be discussed in a subsequent paper in this series dealing with sample size calculation.

Conventionally, type 1 or alpha error is set at 5%. This means, that at the end of the study, if there is a difference between groups, we want to be 95% certain that this is a true difference and allow only a 5% probability that this difference has occurred by chance (false positive). Type 2 or beta error is usually set between 10% and 20%; therefore, the power of the study is 90% or 80%. This means that if there is a difference between groups, we want to be 80% (or 90%) certain that the study will detect that difference. For example, in the ABLE study, sample size was calculated with a type 1 error of 5% (two-sided) and power of 90% (type 2 error of 10%) (1).

Table 1 gives a summary of the two types of statistical errors with an example

In the next article in this series, we will look at the meaning and interpretation of '*p*' value and confidence intervals for hypothesis testing.

References

1. Lacroix J, Hébert PC, Fergusson DA, Tinmouth A, Cook DJ, Marshall JC, et al. ABLE Investigators; Canadian Critical Care Trials Group. Age of transfused blood in critically ill adults. N Engl J Med. 2015;372:1410–1418.
2. Holcomb JB, Tilley BC, Baraniuk S, Fox EE, Wade CE, Podbielski JM, et al. PROPPR Study Group. Transfusion of plasma, platelets, and red blood cells in a 1:1:1 vs a 1:1:2 ratio and mortality in patients with severe trauma: the PROPPR randomized clinical trial. JAMA. 2015;313:471–482.