0363-5465/96/2424-0702$02.00/0
THE AMERICAN JOURNAL OF SPORTS MEDICINE, Vol. 24, No. 5

A Statistics Primer

Hypothesis Testing

John E. Kuhn, MS, MD, Mary Lou V. H. Greenfield,* MPH, and Edward M. Wojtys, MD

From MedSport and the Section of Orthopaedic Surgery, University of Michigan, Ann Arbor, Michigan

In a truly philosophical sense, the scientific method establishes rules to determine the validity of an observation. Every scientific endeavor should follow these rules that guide the researcher in the design of an experiment. After observing an event or phenomenon, the researcher develops an hypothesis. An *hypothesis* is defined as a supposition or an unproved theory.[1] A *theory* has evidence to support it, and is defined as a formulation of apparent relationships or underlying principles of certain observed phenomena that has been verified to some degree.[1] A collection of theories, if repeatedly substantiated, leads to a *natural law,* which is a sequence of events in nature or in human activities that has been observed to occur with unvarying uniformity under the same conditions.[1] All of these gradations, each supported with more evidence, lead toward the truth.

Consider the observation that certain ice hockey officials do a poor job of officiating and are frequently missing injury-causing penalties. You might hypothesize that these officials are responsible for an increased risk of injury. You now want to test this hypothesis. This hypothesis is the substantive or *research hypothesis.* The logic used in hypothesis testing bears striking resemblance to the procedure used in the American system of justice, where the accused is presumed innocent and the prosecution must provide enough data to demonstrate guilt by disproving innocence. In testing this research hypothesis, you are like a prosecuting attorney and must produce enough evidence to convince a jury of your peers that your hypothesis is correct. To work within the rules of the system, you must first reconstruct your hypothesis, presuming innocence, and state that these officials have no effect on injury rates in ice hockey. This becomes the *null hypothesis* and states there is no (null) effect.

In the American system of justice, the prosecuting attorney attempts to demonstrate guilt by providing enough evidence to convince the jury that the accused is not innocent. In the same way, the scientist using statistics attempts to use collected evidence to convince his or her peers to reject the null hypothesis (and hence accept the research hypothesis). In this example, you are trying to disprove the null hypothesis and to demonstrate that these officials *do* miss infractions and cause more injuries. Accepting or rejecting the null hypothesis forms the basis for all scientific studies.

In statistics, evaluating a null hypothesis gives a number of choices and outcomes. Although you think the null hypothesis is probably false, it may in reality be true. In addition, the data from your experiment (which are a sample and are greatly influenced by chance) will lead you to either reject or accept the null hypothesis (Table 1). For example, suppose your observations are accurate and that the null hypothesis is, in reality, false. The hockey officials in question may, in truth, have an effect on the injury rate. Suppose you now have observed one hockey game where the officials did a reasonable job and no injuries occurred. Keep in mind that observations are influenced by probability and chance. These data may lead you to erroneously accept a false null hypothesis as truth. (Based on your observations, you would wrongly conclude that these officials have no effect on the injury rate.) By definition in statistics, accepting a false null hypothesis is committing a Type II error and is denoted by the Greek symbol β, and means you have a false-negative result (Table 2). Clearly the way to avoid this error would be to attend more hockey games and make more observations. Increasing the number of observations or size of the data set improves the *statistical power* of the experiment and lends validity to the study by reducing the influence of probability and chance on conclusions.

* Address correspondence and reprint requests to Mary Lou V. H. Greenfield, MPH, University of Michigan, Orthopaedic Surgery, TC2914G-0328, 1500 East Medical Center Drive, Ann Arbor, MI 48109.

No author or related institution has received any financial benefit from research in this work.

TABLE 1
Statistical Errors and the Null Hypothesis

Decision[a]	Null Hypothesis		
	In reality:	True	False
Rejected		Type I error	Correct decision
Accepted		Correct decision	Type II error

[a] Decision to accept or reject the null hypothesis is based on the data from the experiment, which is influenced by probability and chance.

TABLE 2
Interpretation and Control of Statistical Error

Condition	Greek symbol	Meaning	Controlled using
Type I error	α	False-positive	Significance level
Type II error	β	False-negative	Statistical power

On the other hand, suppose that, in reality, the null hypothesis is false. Now you have attended a number of hockey games and made the observation that certain officials were present when the majority of injuries occurred. Again this finding may be due to probability and chance. If your data lead you to reject the null hypothesis—that is, you reject the hypothesis that the officials have no effect on the injury rate—then you have made the correct decision.

If, however, in reality, the truth is that these officials are not so bad and have no effect on the injury rate (i.e., the null hypothesis is true), and by chance you happened to catch those games in which injuries occurred, then you have committed a Type I error by erroneously rejecting a true null hypothesis. This is denoted by the Greek symbol α and means that you have a false-positive result. We would like to avoid convicting innocent men, and likewise avoid stating that our experiment demonstrates an effect, when in reality there is no effect. Statistical significance or α helps us in this regard. The statistical significance gives us the probability of committing a Type I error in a given experiment. The probability of committing a Type I error is the P value.

A few important points about research studies must be emphasized. First, if authors state in their conclusions that two groups were found to have no statistically significant differences, it is important to look at the numbers of observations and perform a power analysis on the data. A *power analysis* is a method of reviewing the data to determine if the sample size is adequate to demonstrate statistical significance. It may be that the numbers of observations are too small to demonstrate statistical differences. You should not accept that the two groups are the same (accepting the null hypothesis, and committing a Type II error) until you know the groups are large enough to demonstrate a statistical difference, if one existed. Typically, the power analysis should be described in the "Materials and Methods" section of the paper or presentation.

With regard to statistical significance, there is nothing special about the expression $P < 0.05$. What this means is that the probability of committing a Type I error (rejecting a true null hypothesis) is 5%. This might be reasonable for some experiments; however, if the penalty for poor hockey officiating was execution, you might want to decrease your chance of committing an error and set your statistical significance level to $P < 0.01$. The significance level must be considered within the context of the experiment!

When evaluating manuscripts in the literature, or designing an experiment, three initial steps are helpful to understand the experiment and to determine if it has been designed well.[2] 1) Identify and write down the research hypothesis and the null hypothesis. 2) Identify and write down the meaning of the Type I error in the experiment, then determine if the significance level is appropriate in the context of the experiment. 3) Identify and write down the meaning of the Type II error in the experiment. After performing this task, you will have a better appreciation for the study you are evaluating.

In future editions, the meaning of statistical significance and statistical power, and tests for these parameters will be explored.

ACKNOWLEDGMENT

The authors thank Dr. M. A. Schork from the University of Michigan, School of Public Health, Department of Biostatistics, for his review of this manuscript.

REFERENCES

1. Guralnik DB (ed): *Webster's New World Dictionary of the American Language.* Second College Edition. New York, William Collins & World Publishing, 1978
2. Lieber RL: Statistical significance and statistical power in hypothesis testing. *J Orthop Res 8:* 304–309, 1990