0363-5465/96/2424-0564$02.00/0
THE AMERICAN JOURNAL OF SPORTS MEDICINE, Vol. 24, No. 4
© 1996 American Orthopaedic Society for Sports Medicine

Current Concepts

A Statistics Primer

Statistical Terminology—Part 2

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

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

In our never-ending struggle to evaluate scientific data, an important study component is the critique of the steps used to collect or to interpret data. If the steps or tests used to differentiate data points are appropriate, then the differences or similarities between data become apparent. If tests are not chosen or designed appropriately, subtle differences become indistinct or blurred. The more efficiently that a test can separate similar-but-different data points, the better the test. Obviously, the more a test lumps dissimilar data points, the less helpful the test.

The *accuracy* of a test addresses this issue of differentiation. Accuracy reflects the number of data points or variates that are correctly identified by a test. An accurate test is not "fooled" by similar-but-different variates; it can differentiate positives and negatives. The Lachman test for ACL insufficiency can be used to examine the issue of test accuracy. If, for instance, the Lachman test can be performed in such a way that those knees with excessive secondary restraint laxity, hypermobility, or the upper ranges of normal anterior tibial translation can be differentiated from those with a torn ACL, then the Lachman test would be described as an accurate test for a torn ACL. If, however, the test cannot separate those knees with a torn ACL and excessive muscle tightness, intact secondary restraints, and minimal increases in anterior tibial translation from normal, loose knees, the test would not be considered accurate for this lesion.

Accuracy should be differentiated from *precision,* which deals with repeated measures. For instance, if a knee has 3 mm of anterior tibial translation but three arthrometer measurements yield 3.1, 3.2, and 3.3 mm, the precision of that measurement is within 0.3 mm. Accuracy, on the other hand, only deals with how close each individual measurement comes to the true laxity.

Ideal tests are not only accurate, they are both highly sensitive and specific (Table 1). *Sensitivity* refers to the ability of the test to detect true positives. *Specificity* refers to the ability of the test to detect true negatives. More often than not, there is a tradeoff: as the sensitivity of a test rises, the specificity falls, and vice versa, because in most clinical tests there are three possible results: positives, negatives, and a gray zone. How this gray zone of a data set is handled is determined by the clinical situation. Some clinical situations will dictate the implementation of highly sensitive tests and others will stress specificity. For example, in screening for acquired immunodeficiency syndrome (AIDS), the ramifications of false-negative tests could be catastrophic. So a highly sensitive test that detects the largest number of true positives is preferred—even at the risk of falsely identifying some cases as positives when they are not. As long as repeat testing or another test can separate these false-positive and true-positive results, a highly sensitive test would be the best clinical choice. Although a false-positive test could cause a lot of anxiety, if it will eventually be identified correctly in secondary testing, it would be the lesser of two evils. That is, it would be better to suffer the temporary anxiety of thinking you might have AIDS than to be HIV-positive and have that totally missed in the diagnostic workup. In this scenario, a highly sensitive test, even though not as specific, would be the preferred choice if high sensitivity and specificity could not be achieved simultaneously.

On the other hand, if the clinical consequences of a false-positive test are unidirectional and cannot be adequately confirmed or repeated by other testing, then a highly specific test may be the best clinical choice. For instance, if an amputation will be recommended for a positive biopsy of a lytic bone lesion, then the methods used to detect cancer in the biopsy must be extremely specific. In fact, no false-positive results would be toler-

* 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 study.

TABLE 1
Knee Injuries: Clinical Scenarios[a]

Lachman test	Torn ACL	Normal ACL
Positive	A	C
Negative	B	D

$$\text{Prevalence} = \left[\frac{A+B}{A+B+C+D} \times 100\right]$$

$$\text{Sensitivity} = \left[\frac{A}{A+B} \times 100\right]$$

$$\text{Specificity} = \left[\frac{D}{C+D} \times 100\right]$$

$$\text{PPV} = \left[\frac{A}{A+C} \times 100\right]$$

$$\text{NPV} = \left[\frac{D}{B+D} \times 100\right]$$

Incidence: Number of new cases diagnosed during a specific time period.

[a]The amount of laxity detected on the Lachman test depends on several factors, especially the integrity of the active and passive restraints of the knee. Therefore, the results of the Lachman test must be interpreted with caution. PPV, positive predictive value; NPV, negative predictive value.

ated. These examples represent extreme circumstances, but they emphasize the need to determine the worst- and best-case scenarios before clinical tests or studies are performed. In other words, the implications of the possible test results need to be considered before the test is ordered.

Another helpful concept in evaluating tests in clinical decision-making is predictive value, which measures whether an individual has a certain condition given the results of the screening test. *Positive predictive value* is the probability that an individual actually has the condition if the test is positive, and *negative predictive value* is the probability that the individual does not have the lesion or condition if the test is negative. If a clinical condition is quite rare (such as rabies), then the negative predictive value should be very high (99%) for a screening test. However, if a condition is quite common, a high positive predictive value would be desirable. For this reason, both aspects of predictive values depend on the prevalence of a condition in the population along with the sensitivity and specificity of the test itself. The more sensitive a test becomes, the less likely that a patient with a negative test has the problem and, therefore, the higher the negative predictive value. Conversely, the more specific a test is, the less likely that a person with a positive test will not have the condition, yielding a greater positive predictive value.

The population screened plays a large role in predictive value determinations. Even with the ideal test (highly sensitive and specific), if the number of patients with the problem is low, the positive predictive value of the test will be low and many patients will have false-positive results. For example, the positive predictive value of screening an Afro-American population for sickle cell disease would be much higher than that for the general population. Similarly, targeting smokers when looking for lung cancer will be more productive than targeting the entire population, thus improving the positive predictive value of whatever screening method is chosen.

Two other terms that are helpful in discussing the frequency of a clinical condition are prevalence and incidence. *Prevalence* refers to the number of cases that exist within a population at any one time, whether diagnosed or not. *Incidence,* on the other hand, refers to the number of newly diagnosed cases during a specified time period. Regarding torn ACLs, many of these go undiagnosed, so that the prevalence may be much higher than the diagnosed incidence. Frequently, knowledge of the natural history of a problem will help put these concepts in order. For instance, if a clinician knows that many people can live normal lives with torn, untreated ACLs if their lifestyles are modified, the prevalence of the problem is less likely to be underestimated by focusing only on those who seek medical treatment.

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.

SUGGESTED READINGS

Hennekens CH, Buring JE: *Epidemiology in Medicine.* Boston, Little, Brown and Company, 1987

Lieber RL: Experimental design and statistical analysis, in Simon SR (ed): *Orthopaedic Basic Science.* Chicago, AAOS, 1994, pp 623–665

Stratford PW: Letters to the editor. *Am J Sports Med 22:* 567–568, 1994