hypothesis is false or the degree to which the phenomenon is present in the population. In the Black and White example, the degree to which the null hypothesis is false (the effect size) is identical in both studies. Yet one trial (White) reported statistically significant results and the second (replication) trial, by Black, reported statistically nonsignificant results. Why is there statistical confusion?

The apparent contradiction between the two statistical results is because of the difference in sample size and the impact of sample size on statistical power. The White and Black trials were identical in terms of research design, independent variable, dependent variable and type of analysis. However, there was one important difference, ie, Black's investigation had a smaller sample size. The size of the sample for White's trial was 80 and for Black's investigation the sample size was 20. The difference in sample size is reflected in the different degrees of freedom (df) values contained in the statistical information about the two clinical trials presented above (df = N − 2). The df for White's trial was 78 and for Black's trial the df was 18. Sample size can have a dramatic impact on the statistical power of clinical trials.

STATISTICAL POWER AND EXPERIMENTAL ERRORS IN CLINICAL RESEARCH

Statistical power is defined as the ability of a statistical test to find a significant difference that really does exist, or the probability that a test will lead to the rejection of the null hypothesis when the null hypothesis is false and should be rejected.[21] Any time a statistical test is conducted the results will lead the researcher to either fail to reject the null hypothesis (ie, decide on the basis of the data collected from the sample that the null hypothesis is probably true), or to reject the null hypothesis (ie, decide, based on the sample data, that the null hypothesis is probably false). The possibility exists that the researcher will make a mistake in the decision to reject or fail to reject the null hypothesis. If the researcher rejects the null hypothesis, when it was, in fact, true and should have not been rejected, then a type 1 error has occurred. The probability of making a type 1 error is equal to the level of significance used in the investigation. For example, if the significance level is set at $p < .05$ then there is a 5% chance that the sample data will mislead the researcher into rejecting the null hypothesis when there really was no difference between the two groups and the null hypothesis should have not been rejected. A type 1 error can only occur when the researcher rejects the null hypothesis, ie, reports a statistically significant result.

The opposite of a type 1 error is a type 2 error. A type 2 error occurs when the researcher fails to reject the null hypothesis when it is really false and should have been rejected. The mistake that occurs when a type 2 error is made is that a difference between the groups is present, but the analysis is not sensitive enough to detect the difference as statistically significant. A type 2 error can only take place when the researcher reports statistically nonsignificant results. The relationship between hypothesis testing, and type 1 and type 2 errors is presented in table 1.

Unlike the type 1 error, no simple, direct relationship exists between the level of significance and the probability of committing a type 2 error. Factors that influence the probability of committing a type 2 error include; the level of significance selected for testing the hypothesis (usually $p < .05$), the size of the treatment effect (effect size), and the size of the sample. Type 2 errors and statistical power are closely related. Power is defined in relation to the probability of a type 2 error. Specifically, power is equal to 1 − type 2 error rate. A power of 0.80 is often recommended for clinical research. This level of power would be associated with a type 2 error rate of 0.20, ie, 1 − 0.20 = 0.80.[18,21]

Table 1: Relationship of Statistical Hypothesis Testing Results With Type 1 and Type 2 Errors. The Researcher's Decision is Based on the Results of the Statistical Significance Tests

	State of Null Hypothesis	
Researcher's Decision	H_0 True	H_0 False
Reject null hypothesis	Type 1 error	Correct decision
Not reject null hypothesis	Correct decision	Type 2 error

A reciprocal relationship exists among the four factors influencing error rates ie, power (type 2 error), significance level (type 1 error), sample size, and effect size. For instance, if the significance level and sample size remain constant, then the larger the effect size, the more powerful the statistical test. In contrast, greater statistical power is required to show a small effect size as statistically significant when the sample size and significance level are held constant. In clinical investigations the type 1 error rate is usually fixed at $p < .05$ or smaller. The other factors that are free to vary are power (type 2 error), sample size, and effect size.

If we apply this information on type 1 and type 2 errors to the White and Black investigations we can begin to understand the apparent contradiction between the results of the statistical analysis and the reported effect size. White reported a statistically significant result in her trial. The statistically significant result means that White rejected the null hypothesis. There are two possible outcomes associated with rejecting the null hypothesis: (1) the researcher made the correct decision or (2) the researcher made a type 1 error (table 1). If the researcher rejects the null hypothesis, there is no chance of making a type 2 error. In the White trial, the probability of committing a type 1 error is 0.05 or 5%. The type 1 error probability is equal to the significance level set before conducting the statistical test to evaluate the null hypothesis.

In the second clinical trial by Black, the quantitative result was not statistically significant. This means that Black's decision was to not reject the null hypothesis. This decision could either be correct, or a type 2 error. Because Black failed to reject the null hypothesis, the only type of error he could have made was a type 2 error (table 1). A key question in interpreting the result of the Black trial is: What is the probability of Black committing a type 2 error? To answer this question we need the following information; (1) the type of statistical test, (2) the significance level for the test, (3) whether the test is one-tailed or two-tailed, (4) the effect size, and (5) the sample size. Using this information we can find the statistical power by consulting tables in Cohen's text.[21] For a two-tailed t test, using the $p < .05$ significance