MakroCare

New Possibilities, Everyday

WHITE PAPER

Statistical Significance versus Clinical Significance in Clinical Trials

Dr. A.K. Mathai

www.makrocare.com

Introduction

Tests of statistical significance are invariably applied nowadays by research scientists not only in clinical trials but also in various other domains such as sociology, psychology etc. In addition, most of the medical journals refuse to accept papers for publication if the authors have not used significance testing for evaluating their results, wherever appropriate. Almost all scientific reports carry statement like, `the success rates in the test drug and reference drug groups were 84% and 81% respectively. However, the difference is not statistically significant (P > 0.05)'. It is not adequate to mechanically undertake significance tests. The research scientist must fully understand the basic concepts underlying in a significance test, the assumptions involved and the limitations, for making proper interpretation. At the same time the researcher should be aware about the distinction between the statistical significance and clinical significance in research. The objective of this paper is to highlight the difference between statistical significance versus clinical significance in medical research.

Statistical Significance vs. Clinical Significance

In order to understand the concept of statistical significance versus clinical significance, the researcher needs to understand the idea behind **sampling variation**. As we know, the research studies are undertaken to investigate certain hypotheses. Based on the data collected from the study, the investigator has to draw an inference, which could be that the hypothesis is true or that it is false. But we all know that if we repeat the same study, even under exactly similar conditions, we will not necessarily get identical results. For example, let us assume that in a clinical trial of 500 patients, we find that the efficacy of a particular drug is 80%. If we repeat the study using the same drug in another group of similar 500 patients, we will not necessarily get the same efficacy of 80%. It could be 78% or 81%. Thus, we will get different results from different trials though all of them were conducted under the same conditions. Variation of this type is known as sampling variation. Therefore, when taking decisions based on experimental data, we must give some allowance for sampling variation. For example, if we are testing the claim of a pharmaceutical company that the efficacy of a particular drug is 70%, we have to allow in practice for some difference from 70%. Intuitively, we may accept the company's claim if we observe the efficacy in the trial to be 68%, 71%, 73%, or 67%. But if the efficacy in the trial, based on sufficiently large number of patients, happens to be 40% we would have good reason to feel that the true efficacy cannot be 70%. This is because we intuitively feel that if the real efficacy is 70% we are very unlikely to get an efficacy of 40% in the trial, i.e., the chance of such a happening must be very low. We then tend to dismiss the claim that the efficacy of the drug is 70%. From the above example, it is clear that while taking decisions, we have to allow for some differences due to sampling variation.

Again, the researcher should be well aware about the **decision rule** in statistics. A decision rule based on the probability of getting a difference of the magnitude found in a trial can be formulated. For this, we start with the hypothesis that there is no difference in efficacy between the two drugs. This is known as the null hypothesis (denoted by Ho). Assuming the null hypothesis is true, and using statistical methods we compute the probability of getting the observed difference in the trial. The decision rule is that if the probability is low (say, $P \leq 0.05$) the null hypothesis of no difference should be rejected. On the other hand, if the probability is high (say, $P > 0.05$) the null hypothesis is to be accepted.

However, every decision making process, including that based on significance testing, carries two types of errors. The first one is that even when there is no difference in reality, we may get, purely due to sampling fluctuations, a difference which is statistically significant (this can occur in 5% of the occasions) and thereby we reject the null hypothesis which is true. This is known as **Type I** error (i.e., rejecting the null hypothesis when it is true). The second type of error is that

even though there is a real difference (i.e., the null hypothesis is not true), but due to inadequate sample size and / or vagaries of sampling fluctuations, we may have obtained a `non-significant' result, and consequently accepted the null hypothesis of no difference. This is called **Type II** error (i.e., accepting the null hypothesis when it is false). It is impossible to eliminate both the errors but their magnitudes can be controlled by appropriate increase in sample size. The relative importance of the two types of errors vary from situation to situation - e.g., in drug trials for diabetes where many effective drugs are available Type I error is more important, whereas in drug trials for AIDS where very few effective drugs are known, Type II error is more important.

Finally, how the researchers distinguish between statistical significance and clinical significance? It should be emphasized that a difference being statistically significant does not necessarily mean that it is large or clinically important. Statistical significant only means that it is unlikely (≤5%) that the difference is due to chance. In other words, it means that the observed difference is unlikely to be due to sampling variation. By `unlikely' we usually mean a probability of less than or equal to 0.05. Sometimes trivial differences can be statistically significant if they are based on large sample size. For example, in a clinical trial on diabetes patients, the mean reduction of fasting blood sugar from baseline is 50±7 in the test group (n=223) and the mean reduction of the same in the control group (n=225) is 45±5. However, a comparison of these mean values shows that a difference of 5 units between test and control groups is statistically significant [P<0.0001]. But this difference of 5 units may not be considered as clinically significant.

Again, a non-significant result (P>0.05) does not necessarily mean that there is no real difference. It means only that the observed difference could easily be due to chance. There could be a real or important difference but due to inadequate sample size we might have obtained a non-significant result. For example, in a clinical trial on diabetes patients, the mean reduction of fasting blood sugar from baseline is 50±47 in the test group (n=25) and the mean reduction of the same in the control group (n=23) is 30±24. However, a comparison of these mean values between test and control groups shows that a difference of 20 units is not statistically significant [P=0.0735], but the researcher feels that this difference is clinically significant. In this scenario, the researcher needs to look into the inadequacy of sample size and needs to be increased to get a statistical significance along with clinical significance. Hence the researcher should be cautious about the proper interpretations of his results by taking into account of both statistical significance and clinical significance. To regard a statistically significant difference as proof of the existence of an important difference, or to equate a non-significant difference to proof of no difference is very naive and such thinking deserves to be strongly discouraged.

Conclusion and recommendation

It is important to understand the concept of statistical significance versus clinical significance in medical research. It would be more helpful, if statistician try to understand the importance of clinical significance and a clinician do the same about statistical significance. Then the communication between two professionals would be much easy and the researcher can take the most appropriate decision at the end of the research.