Original Research Paper Medical Science

WHAT IF THE RESULTS OF A CLINICAL RESEARCH ARE 'STATISTICALLY INSIGNIFICANT'?

Vikram Singh Yadav	Assistant Professor,College of Physiotherapy, Pt. Bhagwat Dayal Sharma University of Health Sciences, Rohtak
Alka Pawalia	Ph.D. Scholar, Department of Physiotherapy, Guru Jambheshwar University of Science & Technology Hisar
Navinderpal Singh	Reader Jammu College of Physiotherapy, Jammu

ABSTRACT

Many researchers have faced a situation where after doing statistical analysis of data; results of research came out as insignificant, as the p value was different from what researcher expected it to be. Since the results were statistically insignificant majority of researchers feel disheartened and maximum of such research work is lying somewhere unattended or unpublished. Many feel they did something wrong since the null hypothesis could not be proved as false. Usually while planning for any research work there are certain steps such as setting the experimental & null hypothesis, data collection, data analysis & then results. Here if the null hypothesis is proved wrong the results are said to be significant and hence it is believed the research is useful or worth it. Ironically many researchers start the reverse process if the results on analysis are otherwise. All kinds of adjustments from data to hypothesis resetting are done to get a perfect statistically significant result. But does it really matter? Is getting a perfect statistically significant result the only aim for any researcher? What if the results are statistically insignificant but are clinically significant. The purpose of this commentary is to discuss whether proving a research's significance is more important statistically or clinically.

KEYWORDS Statistical Significance, Clinical Significance, 'p'-value

Introduction

These days because of evidence based practice the importance of clinical research work in various fields of health sciences is well known. Of the various steps involved in a research work, the analysis & interpretation of data is the most important. But this happens to be an uphill task for majority of the clinicians. Since results from clinical research work will directly affect the clinical decision making in terms of patient safety & deciding the effectiveness of a given treatment, it is important to properly analyse the data & even more important is to properly interpret them. Many researchers that have a research work with statistically insignificant results do not bother publishing their work, thinking it to be of low quality & of no worth publishing it. A study outcome can be statistically insignificant, but can be clinically significant, and vice-versa.

There could be various factors such as small sample size or measurement variability that could lead to non significant results but they could be clinically significant. In fact from a clinical perspective, the presence or absence of statistically significant differences is of limited value & clinical research is only of value if it is properly interpreted about its clinical significance. This basically means having a treatment protocol which benefits the patient most while giving him least inconvenience, is of low cost & does not cause any kind of harm to the patient. The purpose of this literature is to better understand the role of 'significance' from a statistical & a clinical point of view and understand the evidence based research in a simplified manner. This will be of great help to young researchers for better planning of their research.

Statistical Significance

Before starting any experimental research, researcher decides on what would be the expected outcomes of the study & it is assumed that whatever the researcher thinks is right. Researchers infer something about a population using representative sample & it is further checked with the help of the statistical hypothesis testing if proved otherwise. Statistics are used to answer questions of probability, using the scientific method. In order to determine if a hypothesis can be accepted or rejected, statistically significant differences are determined using a certain level of probability (the "p-value", or α). Statistical significance only addresses a hypothesis about whether or not differences exist, statistically, between groups. To understand them in a better way lets be clear with the following terms.

Hypothesis

A hypothesis is a speculation or theory based on insufficient evidence that lends itself to further testing and experimentation. With further testing, a hypothesis can usually be proven true or false.

Null Hypothesis

It is a type of hypothesis used in statistics that proposes that no statistical significance exists in a set of given observations. It attempts to show that no variation exists between variables, or that a single variable is no different than zero. The null hypothesis assumes that any kind of difference or significance in a set of data is due to chance. It is presumed to be true until statistical evidence nullifies it for an alternative hypothesis.

Experimental Hypothesis

It is also known as the alternate hypothesis, & simply is the inverse, or opposite, of the null hypothesis. It is the one, researcher would believe if the null hypothesis is concluded to be untrue. Experimental hypothesis is simply a prediction that the experimental manipulation will have some effect or that certain variables will relate to each other.

'p' Value

To prove the null hypothesis as false i.e. to get the results as statistically significant every thing keeps revolving around the 'p' value. While performing a hypothesis test in statistics, a p-value

helps in determining the significance of the results. The p-value, or calculated probability, is the probability of finding the observed, or more extreme, results when the null hypothesis (H_0) of a study question is true – the definition of 'extreme' depends on how the hypothesis is being tested.

This value is chosen such that one does not reject the null hypothesis incorrectly even by chance, when in fact it is true (Type I error). The generally accepted p-level of $\alpha = 0.05$ suggests there is a 95% probability that the researchers correctly rejects the null hypothesis when there is no difference between groups. Determination of whether a statistically significant difference exists or does not exist is centered on accepting or rejecting a "null" or "alternate" hypothesis (Fig.1). Therefore, the p-value is only the chance that the researcher makes the correct "yes" or "no" decision regarding a hypothesis.

The p-value is a number between 0 and 1 & is interpreted in the following manner:

- A small p-value (≤ 0.05) indicates strong evidence against the null hypothesis, so as to reject the null hypothesis.
- A large p-value (>0.05) indicates weak evidence against the null hypothesis, so you fail to reject the null hypothesis.
- p-values very close to the cutoff (0.05) are considered to be marginal (could go either way). Always report the p-value so that readers can draw their own conclusions.

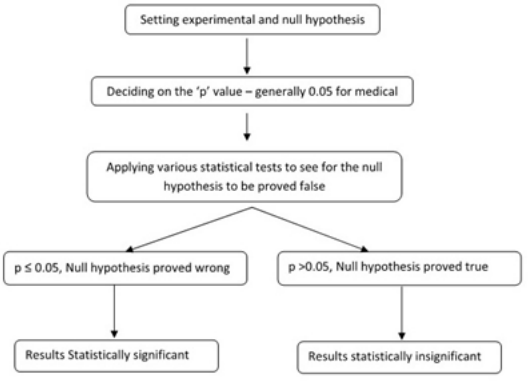

Fig.1.Steps generally followed during experimental research.

In health care research, it is generally agreed that researcher want there to be only a 5% or less probability that the treatment results, risk factor, or diagnostic results could be due to chance alone. When the p value is 0.05 or less, one says that the results are statistically significant. Results that do not meet this threshold are generally interpreted as negative. But, the p-value, gives limited information, essentially, significance versus non significance and it does not show how important the result of the statistical analysis are.

Testing statistical significance is all about the likelihood of a chance finding that will not hold up in future replications. Significance does not tell us directly how big the difference was. Various factors such as number of subjects, variability between them and the magnitude of effect determines the level of statistical significance difference. After completing the analysis and the results, the value of significance does not provide the clinical insight into various aspects such as treatment effect size, magnitude of change, or direction of the outcome. It means that p-values should be considered along with effect size, sample size, and study design. This simply means that in any study the research outcomes should also be seen for their clinical significance rather than just focusing on statistical significance.

Clinical Significance

The results of a study can be statistically significant but still be too small to be of any practical value. Clinical significance measures how large the differences in treatment effects are in clinical practice (Kazdin, 1999). This is of great importance to physicians when looking at research evidence. In medicine & physiotherapy, clinical significance is the practical importance of a treatment effect - whether it has a real genuine, palpable, noticeable effect on daily life. This concept of clinical significance is not clear to many researchers and hence also to their students. A supervisor generally negates a student's research work if results are negative & the concept of clinical significance is completely ignored.

Identical changes on a numerical scale may have different clinical importance in different patient populations (e.g. different ages, disease severity, & injury type). Furthermore, statistical significance is linked to the sample size. Many a times to bring statistically significant results researchers repeats the study with increased sample size, though the difference between the groups might be very small which are usually clinically meaningless. What is the use of such a study with large sample size if it has very little value or applicability when it comes to patient outcomes that are worth noting? Given a large enough sample, statistical significance between groups may occur with very small differences that are clinically meaningless.

In clinical research it is not only important to assess the significance of the differences between the evaluated groups but it is also recommended, if possible, to measure how meaningful the outcome is (for instance, to evaluate the effectiveness & efficacy of an intervention). Only statistical significance does not provide information about the clinical relevance, effect size or confidence intervals (CI).

Outcomes with small p-values are often misunderstood as having strong effect sizes. On one hand, a large sample size study may have a statistically significant result but a small effect size because of that, researchers often misinterpret statistically significance as clinical one. On the other hand, another misinterpretation is present when non statistical significant difference could lead to a large effect size but a small sample may not have enough power to reveal that effect.

What can be inferred from this is, whether the study results have any practical application or usefulness for the researcher keeping in mind, the cost, inconvenience & side effects of the therapy. Besides from significance; confidence intervals & measures of effect sizes (i.e., the magnitude of the change) should also be included in the research findings, as they can provide more information regarding the magnitude of the relationship of the studied variables (e.g., changes after an intervention, differences between groups,). For instance, CI's facilitate the range of values within the true difference value of the studied parameter lies. There are various terms which are used interchangeably to understand this concept such as "clinically meaningful difference" (CMD), "minimal clinically important differences" (MCID) and "minimally important changes" (MIC), "minimally important difference" (MID). It is basically defining what really matters to the patients (Page et al, 2014). In general, all these terms refer to the smallest change in an outcome score that is considered "important" or "worthwhile" by the practitioner or the patient and/or would result in a change in patient management after considering the side effects, costs and inconvenience of the therapy given. These are crucial in both planning of clinical trials and the interpretation of their results.

Minimal Clinically Important Difference (MCID)

The smallest treatment efficacy that would lead to a change in a patient's management or the smallest benefit of value to patients is called the minimal clinically important difference (MCID) (Anna E. et al, 2014). It is a patient-centered concept, capturing the magnitude of the improvement. MCID can help determine the effect of a given therapy on a patient and adds meaning to statistical inferences made in clinical research. Of the various

available methods to measure/determine the MCID, they can be divided into well-defined categories: distribution-based & anchor-based approaches. Distribution-based approaches are based on statistical characteristics of the obtained samples. There are various methods within the distribution-based approach, including the standard error of measurement, the standard deviation, the effect size, the minimal detectable change, the reliable change index, and the standardized response mean.

The effect size is one of the most important indicators of clinical significance, is evaluated by comparing the size of the effect of intervention to some measure of variability, such as the between-person variability in outcomes or the variability associated with repeated measures of the outcome. Effect size reflects the magnitude of the difference in outcomes between groups; a greater effect size indicates a larger difference between experimental and control groups. Some researchers claim, that if standardised effects are small then the difference between groups is clinically unimportant, whereas large effects are clinically important.

Cohen J (1988) established traditional calculation of standardised effect size values based on group differences (change in experimental vs. control group score), divided by the pooled standard deviation (SD of both groups). Cohen J (1988) quantified effect sizes that have been operationally described in ranges: <0.2= trivial effect; 0.2-0.5 = small effect; 0.5-0.8 = moderate effect; > 0.8= large effect. Cohen's effect sizes may be positive or negative, indicating the direction of the effect.

All of these indices have potential uses. Standardized effect sizes (or standardized mean differences) are important when comparing treatment effects between different studies, as can be used for meta-analyses. And the standard error of the mean and the minimum detectable change are useful clinimetric indices. Clinical researchers should include standardized effect sizes in their results.

In Anchor-based approaches when measuring the effects of intervention, compares the change in a patient-reported outcome (pain) to a second, external measure of change (one that is more clearly understood, such as the global rating of perceived effect), which serves as the anchor. Though Anchor based methods can help researcher's associate descriptors with outcomes but they do not assist researchers to decide if an effect of intervention is clinically important.

Limitation of both methods is that neither attempts to evaluate whether patients feel that the effect is large enough to make the costs, inconvenience, and harms associated with intervention worthwhile. A second issue is that these approaches evaluate properties of outcome measures, not of interventions. Another limitation of methods used to evaluate the clinical significance of intervention is that such estimates are almost always based on within-group changes (i.e. changes in outcome from baseline) rather than between-group changes (i.e. the difference in outcome between the intervention and control groups). Within-group changes may be due to the intervention, but they may also be due to natural recovery, statistical regression, and placebo (Herbert et al, 2005).

A new method for assessment of clinically important effects of intervention was developed by Barret & colleagues in 2005 namely 'benefit-harm trade-off method'. The method involves presenting patients with estimates of the benefits, risks, costs & inconveniences associated with the intervention and then asking them whether, they would choose to have the intervention. If the patients say they would have the intervention. Then the process is repeated, holding the costs, inconveniences, and harms constant & the patient is asked to imagine that the benefit of the intervention is larger (or smaller). This is repeated until it is possible to establish, with sufficient precision, the smallest worthwhile effect of intervention (threshold benefit) for which the patient would choose to have the intervention. This threshold is called the 'sufficiently important difference'. The benefit-harm trade-off method overcomes all of the main shortcomings of anchor-based & distribution-based evaluations of clinical importance: instead of researchers or clinicians, patients estimate directly how large the effects of intervention must be to make the intervention worthwhile, the estimate is intervention-specific & the method focuses on the effects of intervention (between-group differences) rather than on change over time (within-group differences).(Manuela L Ferreira et al, 2008)

Method of Interpreting Results from the Perspective of Clinical Significance

A valuable strategy to be utilized for clinical interpretation is combining the MCID with the CI, especially when hypothesis testing reveals statistically non significant differences. Clinical importance can take different forms, depending on the relationship of the MCID of the intervention to the point estimate (the best single value of the efficacy of the intervention that has been derived from the study results) and the 95% CI surrounding it (Man-Son-Hing M et al, 2002):

1. Definite - When the MCID is less than the lower limit of 95% CI.

2. Probable - When the MCID is more than the lower limit of 95% CI, but less than the point estimate of the efficacy of the intervention.

3. Possible - When the MCID is smaller than upper limit of 95% CI, but greater than the point estimate of the efficacy of the intervention.

4. Definitely not - When the MCID is greater than upper limit of 95% CI.

Conclusion

Any research should not just focus on the 'p' value or the statistical significance that is calculated but should also focus on the clinical relevance of the study too. Blindly increasing sample size or doing other adjustments to get the results significant is of no use if the patient's perceptions of the treatment effects are not worth clinically. So clinical researchers should present clinically meaningful results, & clinicians should also know how to interpret & implement those results in their evidence based practice to clinical decision making. Clinician's interpretation of clinical research outcomes should be based on clinically-relevant measures such as effect size, clinically meaningful differences, confidence intervals, and magnitude-based inferences.

References:

1. Alan E. Kazdin. The Meanings and Measurement of Clinical Significance. Journal of Consulting and Clinical Psychology, June 1999 Vol. 67, No. 3, 332-339
2. Anna E. McGlothlin, Roger J. Lewis. Minimal Clinically Important Difference: Defining What Really Matters to Patients. JAMA Guide to Statistics and Methods Clinical Review & Education. JAMA October 1, 2014 Volume 312, Number 13
3. Barrett B et al. (2005a) Med Decis Making 25: 250–261.
4. Barrett B et al. (2005b) Med Decis Making 25: 47–55.
5. Cohen J. Statistical power analysis for the behavioural sciences. 2nd ed. New Jersey: Lawrence Eribaum; 1988.
6. Herbert RD et al. (2005) Aust J Physiother 51: 3–4.
7. Man-Son-Hing, M., Laupacis, A., O'Rourke, K., Molnar, F. J., Mahon, J., Chan, K. B. Y., & Wells, G. (2002). Determination of the Clinical Importance of Study Results: A Review. Journal of General Internal Medicine, 17(6), 469–476.
8. Manuela L Ferreira, Robert D Herbert. What does 'clinically important' really mean? Australian Journal of Physiotherapy 2008 Vol. 54
9. Phil Page. Beyond statistical significance: Clinical Interpretation of rehabilitation research Literature. The international journal of sports physical therapy. Volume 9, number 5, October 2014: 726-736