SHORT COMMUNICATION

Relevance of statistical and clinical significance in dental research

Sandhya Jain¹, Prateek R. Punyani¹, Deshraj Jain², Sandesh Jain³

¹Department of Orthodontics, Government College of Dentistry, Indore, Madhya Pradesh, India, ²Department of Prosthodontics, Government College of Dentistry, Indore, Madhya Pradesh, India, ³Sri Aurobindo College of Dentistry, Indore, Madhya Pradesh, India

Abstract

Tests of statistical significance do not provide information about the clinical significance of research results. It is of utmost importance for a researcher to not only know the statistical significance of the difference, but also the magnitude of the difference and whether it is clinically significant or not. Results might be statistically significant yet it may not have some clinical relevance and vice versa. The concept of clinical significance has received little discussion or debate in the dental literature. The purpose of this paper was to compare statistical and clinical significance, to provide an overview of the two, and to discuss some implications of these concepts for dental research with the help of some suitable examples.

Keywords

Clinical significance, effect size, statistical significance

Introduction

Statistical significance tests were first studied long back 1700s. Statistical testing did not become fully understood until the early 1900s. In 1900, Karl Pearson developed the chi-square goodness-of-fit test. In 1908, William S. Gossett published his t-test under the pseudonym "student."¹

Statistical significance is when a finding allows the researcher to reject the null hypothesis at a pre-specified level of confidence or probability, while this statistically significant finding may not always be clinically or practically significant in the outcome, and sometimes the reverse may be true although there are very meager chances of the results being clinically significant and not showing a statistical difference (false negative result) may be due to small sample size.²

Purpose of the Present Article

The purpose of the present article, in general, is to review about statistical significance and clinical significance and compare the two with examples so as to help the clinician interpret results accordingly. Having a difference between two readings has no value unless it is of some clinical relevance.

Before going into details about the statistical and clinical significance and their relevance in dental research, it is of utmost importance for us to know what effect size is (ES).

In statistics, the difference between the value of the variable in the control group and that in the test group is known as ES. This difference can be expressed as the absolute or relative difference, for e.g. if we are calculating root resorption in two groups of orthodontic patients, i.e., control and test group, the root resorption in the control group is 2 mm and in the test group it is 4 mm, the absolute ES is 2 mm and the relative reduction with the test intervention is 2/4 or 50%.

If the ES between two groups or entities is smaller than the standard error (SE) of mean it is likely that the difference is due to method error, although this difference might be of statistical significance but since it is less than the SE it has little or no practical significance, for example a difference of 2 IQ is practically minimal or very small in relation to SE of most IQ tests e.g., standard error of the mean (SEM) 4.2 or more. On the contrary, if the difference or the ES is more than SEM it implies
that it has a practical significance and that this difference is not merely due to method or measurement error [Table 1].

Why ES interpretation should be required

ES interpretation is required because just knowing that two values A and B differ is not enough, we should know the exact difference that by how much the difference exists between the two values.

Statistical significance

A finding is statistically significant when statistical test allows the researcher to reject the null hypothesis at a pre-specified level of confidence or probability. This level is called the alpha level (α) and for educational purposes it is usually set at 0.05 which means there are 5% chances of false positive results. For critical results we generally set a level at 0.01 i.e., 1 out of 100 may be false positive. Thus, we can be more confident that the result is real rather than random. Findings are said to be statistically significant, when the P value associated with the test statistic is smaller than the predetermined alpha level (usually 0.05). A P < 0.05 would result in rejecting the null hypothesis. The problem is that statistically significant differences can be found even in very small differences, if the sample size is large enough.

That means the difference between sample means will be significant if the sample is large enough.

Clinical significance

It reflects the magnitude or size of the difference of means that the difference is not just by chance. Arguably it is much more important than statistical significance, especially for clinical questions. Measures of ES are used to quantify the practical significance of a finding. When determining practical significance the researcher must consider the following questions: The quality of the research questions, the relative size of the effect, the size of the sample, the importance of the finding, confidence intervals (CIs), and the link to previous research and the strength of correlation.

Clinical significance asks the larger question about differences “Are the differences between samples big enough to have real meaning.” i.e., just knowing that two entities are different is not enough. The researcher needs to know the size of the difference and the error associated with the estimate. For example: There is a wide range of insulin-like growth factor (IGF-1) level for each cervical vertebral maturation stage, the normal range during a particular stage varies in different growers and among genders. Therefore, skeletal growth based solely on statistics of serum IGF-1 level at a particular point of time in an individual cannot be assessed. A statistically significant difference of IGF levels between subjects may show little or no clinical relevance, with regards to growth assessment unless some other method of growth assessment is combined with.

Using CI and minimal clinical importance difference (MCID), we can know the clinical significance of the study. MCID is a statistical method that defines the smallest change in a treatment outcome that a patient would identify as important. Study is clinically significant when it is 95% CI is higher than MCID. Value of statistical significance cannot convey the effectiveness of the intervention. Both clinical and statistical are important for clinical research. Statistically significant result may not have clinical effect in clinical setting [Table 2].

In statistics, a CI is a type of interval estimate of a population parameter. It is an observed interval (i.e., is calculated from the observations), in principle different from sample to sample, that frequently includes the parameter of interest if the experiment is repeated. CI provides a range of a parameter within which the true value lies. How frequently the observed interval contains, the parameter is determined by the confidence level or confidence coefficient.

Clinical significance is generally assessed with some measure of ES.

Sometimes studies demonstrate the statistically significant difference that they have no clinical significance. For example: studies of the size of the mandible with or without treatment aimed at stimulating its growth almost always show small differences in the ultimate size of the jaws. In some publications, the differences are reported as statistically significant and in others as not significant. At that level, the argument is over whether average differences of 1-2 mm in the size of 120 mm mandible are treatment related. The more important consideration is whether it would make any real difference clinically.

According to the fifth edition of the American Psychological Association 2001 Publication manual:

“it is almost always necessary to include some index of ES or strength of relationship in results section. The general principle to be followed is to provide the reader not only with information about statistical significance but also with enough information to assess the magnitude of the observed effect or relationship.”

Conclusion

It is of utmost importance to the researcher to not only know that there exists a statistical difference between two entities but the magnitude of the difference i.e., the ES, and whether the difference actually has a real meaning when considered in the

<table>
<thead>
<tr>
<th>Table 1: Relationship between ES and SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES&lt;SEM</td>
</tr>
<tr>
<td>ES&gt;SEM</td>
</tr>
</tbody>
</table>

ES: Effect size, SEM: Standard error of the mean

<table>
<thead>
<tr>
<th>Table 2: Relationship between ES and MCID</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES&lt;MCID</td>
</tr>
<tr>
<td>ES&gt;MCID</td>
</tr>
</tbody>
</table>

ES: Effect size, MCID: Minimal clinical importance difference
Clinical scenario, i.e., whether it is clinically significant or not. Thus, it is necessary to include the magnitude of ES or strength of relationship in the results section of any research paper.

References