Essentials of Statistics in Dental Research: Part 1

Ashwin C Shetty
Lecturer, Department of Restorative Dentistry, Riyadh Colleges of Dentistry and Pharmacy, Riyadh, Kingdom of Saudi Arabia

Dental professionals review articles in scientific dental journals to be informed about progress in the field and to remain up to date in the dental literature. As a general rule, knowing the basic statistical concepts such as study design, hypothesis testing, confidence interval, significance level, effect size, and sample size and power can help readers to better understand the whole concept of an article. Reviews of statistical methodology in published papers of well-established dental journals concluded that the majority of the papers contained statistical errors. Inappropriate use of statistical methods may lead to incorrect conclusions and a waste of valuable resources. The aim of this paper is to provide an overview on some implications of the statistical concepts and recommendations for avoiding statistical errors in dental research.

Keywords: Confidence interval, Dental research, Hypothesis, Significance level

INTRODUCTION

Over the past decade, the use of statistics in dental journals has greatly increased. Dental professionals often find it difficult to understand the statistical concepts. It is important to understand the basic statistical concepts for dentists involved in the research. A study on dentists in Kuwait reported a poor knowledge of statistical concepts such as confidence interval and \( P \) value.\(^1\) There has been a significant growth in dental literature in terms of the number of journals and frequency of publications. Statistical errors are common in scientific literature and about 50% of the published articles have at least one error.\(^2\) Statistical errors are often identified in the design and analysis of the study in the published dental research. These are understated and difficult to detect, but the damage they create can be appalling. When statistics is used appropriately, they shed further light and clarity upon subjects under study. However, if used wrongly, it can result in confusion and no validations. A manuscript should describe the statistical analysis in sufficient detail so a reader can reproduce calculations if data are available. Moreover, the clarification of the use of statistical methods has the potential of greatly enhancing the strength of a submitted manuscript. The aim of this paper was to provide an overview on some implications of the statistical concepts and recommendations for avoiding statistical errors in dental research.

STUDY DESIGN

It is essential to have a good study design as any errors in the analysis can be rectified if the design is sound. A study design with flaws can lead to data that are not retrievable. For a well-planned study, investigators usually first formulate a clearly defined research question and then carefully choose the design and analysis that are most suitable to answer their question. The research question drives the design of the study, the choice of methods and determines the appropriate data analysis strategy.\(^3\) The research questions identified are:

- Descriptive questions - description of phenomena
- Questions of relationships - relation of phenomena
  - Association: Refers to the extent to which two variables tend to occur together
  - Correlation: Refers to the extent to which two variables go together.
- Questions of comparison - determine cause and effect.

Different types of questions require different types of statistical analysis. The majority of research will start with descriptive statistics which form the basis of all research and new areas of investigation. Selecting an appropriate statistical test to address a specific research question is extremely important. The statistical analysis should be consistent with the research question and design scheme.
HYPOTHESIS TESTING

A hypothesis is a tentative explanation that accounts for a set of facts and can be tested by further investigations. The purpose of testing hypothesis is to assist researchers in making decisions. Quantitative research is well suited for the testing of these hypotheses, most common in experimental designs. In hypothesis testing, the researcher should define the population, state the hypotheses to be tested, specify the significance level, select a sample from the population, select a test statistics, perform the calculation for the statistical test, draw a conclusion, and develop appropriate interpretation of the conclusion. Traditionally, we look at two distinct types of hypothesis: The null hypothesis (H0) and the alternative hypothesis (H1) and we want the later to be true. In practice, we test a null hypothesis of no difference because standard statistical tests are usually designed to test the null hypothesis.

The null hypothesis is rejected, meaning the evidence from the sample indicates enough to say with a certain degree of confidence that the hypothesis is false. On the other hand, not rejected, meaning there is not enough evidence to reject the hypothesis. However, two types of incorrect decisions can be made: Rejecting null hypothesis when it is true (Type I error) and not rejecting null hypothesis when it is false (Type II error) (Figure 1). When the null hypothesis is not rejected, one should not say that the null hypothesis is accepted because we may have committed a type II error. The probability of committing a Type I error is denoted by α (alpha) and the probability of committing a Type II error is denoted by β (beta). The Type I error is generally considered the most serious and need to be minimized. However, having a small Type II error is also important. One way of doing this is by increasing the sample size as errors can occur in small samples due to the influence of a small number of extreme values (outliers). Therefore, a larger sample will decrease the chances of making both Type I and Type II errors.

CONFIDENCE INTERVALS AND SIGNIFICANCE LEVELS

Confidence intervals give a higher and lower bond between which values can fluctuate. A 95% confidence interval is the one in which there is 95% confidence in the value of the population mean being somewhere within that range. Thus, a 95% confidence interval means that there are 95 chances in 100 that the sample results represent the true condition of the population within a specified precision range against 5 chances in 100 that it does not. With a larger sample, the confidence interval would be narrower. The confidence interval gives an indication of how much uncertainty there is in the estimate of the true value. Therefore, the narrower the interval, the more precise is the estimate. Confidence intervals are underused in dental research. Only 20 of the 307 papers reviewed in established dental journals reported using confidence intervals.5

Significance level denote the chance of committing a Type I error, also known as coefficient alpha or the P value. P values are widely used in the dental literature, but many authors, reviewers, and readers are unfamiliar with a valid definition of a P value. The significance level can vary between 0 and 1, the smaller the significance level, the smaller the chance of making a Type I error and the more evidence there is against the null hypothesis (Figure 2). The most common is the 0.05 level. A significance level of <0.05 means that the probability to find the value we have in our sample if there was no relation in the population is <5%. In this case, we say that our finding is statistically significant. The word significant has a different meaning in statistics than in daily life, it does not mean important. It means generalizing the findings from the sample to the population. In larger samples, 0.01 or 0.001 cut-off points will be used. These cut-off points are arbitrary and need to be cautious not to rectify them.

Two things determine the size of significance level which is important to remember, size of the relationship or difference found in the sample and the sample size. The significance level does not tell us the strength of the relation, it only tells us the probability that the relationship in the sample would exist if there are no relationship in the population. Therefore, a smaller P value does not mean a stronger relationship. It may be purely from an increase in the sample size. Confidence level indicates the likelihood that the answer will fall within that range and the significance level indicates the likelihood that the answer will fall outside that range. If the confidence level is 95%, then the significance level will be 5% (100-95); if the confidence level is 99%, the significance level is 1% (100-99); and so on. Confidence interval should be used along with P value as they convey more information than P value.6

Confidence intervals give an indication of how much uncertainty there is in the estimate of the true value. Therefore, the narrower the interval, the more precise is the estimate. Confidence intervals are underused in dental research. Only 20 of the 307 papers reviewed in established dental journals reported using confidence intervals.5

### Table: P-value

<table>
<thead>
<tr>
<th></th>
<th>Strong</th>
<th>Moderate</th>
<th>Weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.05</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 1: Type 1 and Type 2 errors

Figure 2: Strength of evidence against null hypothesis and interpretation of P value
EFFECT SIZE

The traditional interpretation of $P$ value that the null hypothesis is rejected if $P < 0.05$ and accepted if $P > 0.05$ does not make sense.\(^7\) The use of arbitrary cut-off point like $P < 0.05$ can be argued that in many cases the difference between a significance level of 0.05 (not significant) and 0.049 (significant) is literally a couple of respondents. Another criticism is based on the fact that null hypothesis is almost always interpreted to mean a difference of literally zero or no relation in the population. Very few relationships are exactly zero as there is usually some element of relationship or difference present. In larger samples, most relationships or differences will be statistically significant. The problem of always testing a hypothesis of no (zero) difference might be absurd. However, most statistical tests available in the statistical packages only test the zero difference (null hypothesis).

Few researchers believe in not using any significance test at all. However, most of the researchers admit that the significance test has problems associated with it and should not be used as a sole measure. Two measures have been proposed: one is replacing significance test by a confidence interval. However, currently, confidence intervals are not as a rule produced in the output of most procedures in statistical software packages and are not used very often. Another measure that is being increasingly used is the effect size. It solves the problem by giving us a measure of the strength of the difference or relationship so that we can compare to results from other studies. Use of effect sizes provides us with very important information when using statistical analysis and some journals currently are requesting these measures from authors.

SAMPLE SIZE AND POWER

During study design, an appropriate sample size should be calculated based on a sample-size calculation as small samples often fails to yield useful findings and thus a poor use of resources. On the other hand, recruiting more participants than needed can waste valuable resources. The method of calculating sample size will differ with each study design. Sample size calculation generally involves quite complicated algebra. The sample size required for a study depends on significance level, power, one or two-tails, ratio of the group sizes, and standardized difference (used with means). In general, descriptive studies need hundreds of subjects to give an acceptable confidence interval for small effects and experimental studies need a lesser sample.

Power calculations are performed to determine the number of participants needed for the study or to work backward to determine the power for a given study size. A larger sample means higher power and traditionally this has been taken as 0.8 or occasionally 0.9. A power of 0.5 implies that there is only a 50% chance that a true alternative hypothesis will be detected; this is an unacceptable risk. The precision with which sample statistics estimate population parameters is strongly influenced by the sample size. Statistical power analysis provides a method of determining the sample sizes needed to control for errors in hypothesis testing. Power is largest when the effects being studied are large and the sample is large. A justification for the sample size and power calculations should be reported when publishing results.

A review of 226 articles from 22 journals of operative dentistry indexed in ISI Web of Science, only two articles reported a priori sample size calculation.\(^8\) Another review of 125 research articles published in the journal of conservative dentistry, only three articles reported sample size calculation and sampling method.\(^9\) Using an adequate sample size along with high quality data collection efforts will result in more reliable, valid, and generalizable results. A study that is sufficiently powered has a statistical rescannable chance of answering the questions put forth at the beginning of research study.\(^10\) There are many different software packages that can be used for sample size estimation. G*Power software package is freely downloadable and capable of estimating sample size for many of the different statistical tests.

CONCLUSION

The concept of significance testing has come under increasing criticism recently. The cut-off points are seen as arbitrary, the reliance on a hypothesis of no difference in the population is seen as unrealistic, and the lack of information on the strength of the effect is seen as unhelpful. For these reasons, many researchers have suggested replacing or supplementing significance level estimates with confidence intervals and effect size measures. An appropriate sample size should be calculated before commencing a study as there are important statistical and ethical implications in the choice of sample size. Finally, the statistician should participate from the beginning to the end of the study.

REFERENCES


How to cite this article: Shetty AC. Essentials of Statistics in Dental Research: Part 1. IJSS Case Reports & Reviews 2015;2(1):35-38.

Source of Support: Nil, Conflict of Interest: None declared.