Basics of interpreting results

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

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

Abstract
The correct interpretation of research results is of paramount importance to know the effectiveness of the study. Researchers should describe the result clearly, and in a way that other researchers can compare them with their own results. For correct interpretation of results, sound knowledge of research methodology and statistics is needed. Results should be analyzed using appropriate statistical methods to try to determine the probability that they may have been by chance, and may not be replicable in larger studies. Results need to be interpreted in an objective and critical way, before assessing their implications and drawing conclusions. The aim of the present review was to highlight the basic points to be kept in mind by the researcher while interpreting the results of a research paper.

Keywords
Bias, confidence interval, hypothesis testing, level of significance, research methodology, sample design

Correspondence
Dr. Sandhya Jain,
Department of Orthodontics,
Shantiniketan, Near Bombay Hospital,
Vijaynagar, Indore - 452 010, Madhya Pradesh, India. Phone: +91-9425045455,
Email: researchorthodontics@gmail.com

Received 27 May 2015;
Accepted 15 July 2015
doi: 10.15713/ins.idmjar.26

Introduction
Research requires a scientific and systematic approach involving logical thinking. Correct interpretations of results require a thorough knowledge of research methodology and statistics. Research is meaningful only when its methodology is sound. We must consider the logic behind the methods why we have used the particular technology to answer the research question. The first step in assessing a paper is its study design whether it is appropriate and valid. How well the study was conducted? Is the study design appropriate for the questions asked because different research questions require different study designs. There are different research designs ex-exploration (considers many different aspects of a problem), descriptive, diagnostic, and experimental designs. To evaluate the efficacy of any intervention randomized and non-randomized clinical trials (RCT) are required.

Observational studies (cohort) are required to look at the cause of the disease. Case-control studies identify the risk factors and are specially used for the identification of risk factors in rare diseases. Cross-sectional studies are planned to know the prevalence of diseases; they also examine the association of risk factor and the outcome. The diagnostic test should include the sensitivity and specificity. If the test is more sensitive, it means there would be less false negative results; if the test has more specificity there would be less false positive result. A test is said to be sensitive when it identifies the outcome measure correctly. A test is said to be specific when it identifies correctly those who do not have the disease.

Hypothesis testing involves estimating the likelihood of observed results occurring by chance. A result is said to be statistically significant when its P-value is equal to or lesser than the prespecified level of significance (0.01 or 0.05).

The level of significance is the criteria used for rejecting null hypothesis, i.e., to find out a significant difference between two variables; in simple words it is defined as the probability of making a decision to reject a null hypothesis.[1] The level of significance is usually set at 0.05. The smaller the P-value, the more significant the result will be. If the P-value of a statistical analysis is equal or <0.05 it indicates that the result is real and not by chance. There is a possibility that the difference is only one in 20 by chance, the rest are true. Hypothesis testing has certain limitations that it gives result in the dichotomy of significant and non-significant differences. Decision to reject null hypothesis on the basis of P-value depends on the type of analysis, sample size, etc. thus there may be false positive and false negative results. Even if the P-value is significant, the more important consideration is whether the
The interpretation of a result depends on the methodological soundness of a study. It depends on the methodological soundness that how far the study is closer to the actual value. There may be an inherent weakness in study design, conduct, and analysis. The treatment effect reported in the study is true, if the study is valid. With transparent reporting of the study, it is possible for the reader to judge how concrete the results are: Certain methodological characteristics may be associated with an effect size, that is, why it is necessary to describe the method.

One must consider, adequacy of research methodology, sample size and its design, validity, the reliability of the method, precision, power, CI, etc., while drawing a conclusion and interpreting results. All these things matter a lot to prevent false interpretation.

**Power**

If a study has less power, that means the results may have a false negative result. That means a difference is not detectable due to small sample size. Thus, to detect a difference (if present), we should calculate the sample size for a minimum of 80% power.

There are still 20% chances of missing the real difference. With 90% power, there is 10% probability of false negative result. Larger the sample size the study has greater power to detect the difference (if present).

An important issue is to consider the internal validity of the study. The study should be conducted in such a way that it should be free from any kind of bias, or systemic bias is minimized. Validity is the ability of the test to separate or distinguish those who have disease from those who do not. External validity refers to the fact whether results are generalized or restricted to a particular population. The degree to which results observed in a study are applicable to the outside world is external validity. Studies least susceptible to bias are placed at the top of the hierarchy, e.g., RCT is placed above observational study. RCT is the gold standard for deciding the efficacy of any intervention. However, a poor RCT is lesser reliable than a well-conducted cohort study.

Validity of method depends on,
1. Random allocation sequence
2. Concealment of allocation sequence
3. Blinding (single, double, triple)
4. Proportion of patients lost to follow-up
5. Stopping of trials early for benefit whether analysis followed the intention to test the principle.

Four main biases affecting the internal validity should be checked while interpreting results otherwise false interpretation may be induced. These are:

**Selection bias**

Procedure to select subjects for experimental and control groups is not proper. Selection bias refers to the selection of individuals, groups or data for analysis such that proper randomization is not achieved. The allocation should be unpredictable, and its sequences
should be concealed from the investigator. Random allocation by computer, dice throwing can be done. The allocation based on the date of birth, date of admission is open to manipulation.[5]

Recall bias (response bias, responder bias)

When cases and controls are asked to recall certain events, subjects in either group may recall the event more efficiently than the other group. Recall bias is a systematic error caused by differences in the accuracy or completeness of the recollections retrieved (“recalled”) by study participants regarding events or experiences from the past. Case-control and cross-sectional studies are susceptible to recall bias.

Performance bias

A bias in outcomes found in clinical trials without blinding and attributable to behavioral responses by subjects or researchers. E.g., for comparison of manual versus power toothbrushes to evaluate oral hygiene, if the investigator is biased, may give oral hygiene instructions to the group he favors. Blinding of study participants and investigators to the type of treatment helps in minimizing performance bias.

Attrition bias

To prevent attrition bias, all participants should be counted for the analysis because excluding patients might underestimate or overestimate the result.

If systemic bias has been avoided then, result is considered if they fulfill the aim of the study. In order to avoid or eliminate their biases following three methods i.e. randomization, blinding and matching may be used.

Randomization is a statistical procedure by which participants are allocated to the study groups in a manner eliminating bias and thus allowing study groups for comparability. Confounding factors, which are important yet unrecognized will be distributed equally in both groups by lottery method, or by a table of random numbers, or by a computer program. The degree of comparability is high with randomization; if the degree of comparability is low, the chances of spurious results will be high.

Binding is a statistical procedure to prevent consciously, as well as subconscious bias, from being included in the study. Subjects, investigators, analyzers observers are not aware of the type of treatment subject are exposed to, thus subject bias, investigator bias, observer bias are minimized.

Matching is a process of selecting controls in such a way that they are similar to cases except the factor, which is to be evaluated ex-match the age, gender, no history of preceding orthodontic treatment. Matching distributes confounding factor equally between the groups. Matching reduces known confounding factor while randomization prevents known and unknown confounding factor, as well as selection bias.[6]

Reliability (repeatability, reproducibility, or precision) is the level of agreement between repeated measurement of the same variable, highly reliable test give consistent results e.g. The index should give the same score if patient is examined by two different examiners (interexaminer reliability) or by same examiner at two different occasions (intraexaminer reliability).

The use of the appropriate statistical test is of utmost importance. Without proper statistical analysis even a well-conducted study is not able to conclude with concrete results, thus selecting appropriate statistical tests involves: Data type (qualitative/quantitative, paired/unpaired), type of distribution (normal or not) etc.[7]

There are two major types of statistical tests to measure hypothesis [Table 1].

Assessing Sample Design

Consideration needs to be given to sample size of the study. Prior calculation of sample size with a minimum of 80% power is required so that it produces the more precise estimate. Smaller studies are underpowered and may be unable to detect the statistically significant results. Results cannot be held conclusive if there is a larger standard error of mean than the effect size, for e.g., if standard error of mean is 4 and the effect size is 3, or even 4; the difference might be because of method error, even if the difference is statistically significant. The result of the study showing large standard deviation indicates that the sample size might be small. If there is a large variation in the population, we must include a large number of the subject for more precise results.

The way of selecting sample size can be either probability or non-probability, in probability sample each element has known the probability of being included in the sample, whereas non-probability sample do not allow the researcher to determine probability. Whenever, possible one must opt for random sampling, to eliminate bias and to estimate sampling error. Purposive sampling is considered when the population is small, and a known characteristic of it is to be studied.[5]

The sample design depends on the nature of inquiry and other related factors such as the type of data collected.

If data collection is by mailing the questionnaire then before applying this method usually a pilot study for testing the

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Parametric tests (distribution dependent test)</th>
<th>Non-parametric tests (distribution independent test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Quantitative e.g.: Weight, B.P.</td>
<td>Qualitative e.g.: Gender, blood pressure, etc.</td>
</tr>
<tr>
<td>Observation form</td>
<td>Mean, SD, range, variance</td>
<td>Proportion percentage fraction</td>
</tr>
<tr>
<td>Scales</td>
<td>Interval or ratio</td>
<td>Nominal or ordinal</td>
</tr>
<tr>
<td>Examples</td>
<td>Student's t-test a. Paired t-test b. Unpaired t-test ANOVA</td>
<td>Chi-square test Fisher exact test Sign test Wilcoxon sign rank sum test Wilcoxon signed rank test</td>
</tr>
</tbody>
</table>

Table 1: Comparison between parametric and non parametric tests
A questionnaire is conducted to reveal if there is any weakness in the questionnaire. If yes then it is reformatted, the questions must be prepared very carefully so that it may prove effective in collecting information. The questions should be easily understood by the subjects. In the structured questionnaire, the questions and possible answers must be coded. If data are to be collected by an interviewer, there should be proper training of interviewer with an instruction manual. Summary must be made as much realistic as possible. The collected information should be in accordance with the predetermined standard of accuracy. The efforts should be made from securing a response from non-responders. The procedural design should be carefully planned, so as to yield results that are as objective as possible.

Validity and reliability of data should be checked carefully, the conclusion should be confined to those justified by the data of the research. Conclusion based on guessing and intuition should not be entertained. A good research is systematically structured with specific sequential steps in accordance with a well-defined set of rules. Logical reasoning makes research more meaningful if the research is verified by replicating the study, if builds a sound basis for decisions.

Conclusion

The knowledge of research methodology provides a tool to look at things objectively. The interpretation of the result is based on objective, logical, and systematic method. The method should be free from any personal bias or prejudice. The investigation proceeds in an orderly manner and research is guided by rules of logical reasoning, only then the method would be valid (internal consistency); thus, the results could be evaluated and concluded with reasonable confidence. The knowledge of research methodology (adequacy of research methods) helps us to evaluate research results and enable us to take a rational decision. Conclusions drawn from the poor quality research are not reliable.

References
