

Understanding statistics 5

Elliot Abt

Department of Dentistry, Illinois Masonic Medical Center, Chicago, Illinois, USA

The previous four articles in this series¹⁻⁴ have covered the basic principles of statistics which have included explanations of the null hypothesis, P-values and confidence intervals. Also discussed were the different types of data, non-parametric versus parametric tests and the different domains of statistical analysis: differences between groups, associations and time-to-event data.

In this article, several hypothetical examples of passages from dental papers will be given along with answers and explanations. All examples given are from the first domain of statistics: differences between groups.

Example 1-A	Answers to 1-A
<p>In a parallel group randomised controlled trial (RCT), investigators have compared deep scaling alone vs. deep scaling with antibiotic gel. They have found a reduction in pocket depth of 0.45mm (95%CI 0.23-0.62) in the test group.</p> <p>a. What type of data are these? b. Are the results statistically significant? c. What statistical test is likely to be used? Are the results clinically relevant? e. Are the results decisive?</p>	<p>a. These are continuous (quantitative) data as the outcome is pocket depth in millimetres. b. It is statistically significant. Although a P-value is not provided, the null value for a mean difference (0) is not included in the confidence interval (0.23-0.62) which corresponds to $P < 0.05$. c. An independent t-test would be the statistical test used, as we are comparing the means of two independent groups. d. To answer this question, a practitioner has to determine what is the minimal value for pocket depth reduction that is clinically relevant. This will vary from one clinician to the next, but a value of 1mm will be used in this example. As the value of 0.45mm is below the minimal value for clinical relevance, we would say that the results are statistically significant but not clinically relevant. e. Yes, as the upper level of the confidence interval (0.62) is below the minimal value for clinical relevance (1mm). This means we are confident that few patients will benefit from the intervention.</p>
<p>The same parallel group RCT as example 1-A, but now we obtain a mean difference of 1.25mm (95% CI 0.4-2.5).</p> <p>a. Are the results statistically significant? b. Are the results clinically relevant? c. Are the results decisive?</p>	<p>a. Yes, the results are statistically significant as the null value for a mean difference (0) is not included in the confidence interval (0.4-2.5), corresponding to $P < 0.05$. b. We would consider this result as clinically relevant as the mean difference of 1.25mm is greater than the minimal value for clinical relevance set at 1mm. c. No. As the lower level of the 95% confidence interval (0.4) is below the minimal value for clinical relevance (1mm), we would say the results are not decisive. This means that we are confident that some patients will not receive a clinical benefit from the intervention, and results should be interpreted with caution.</p>
<p>The same parallel group RCT as example 1-A, but now we obtain a mean difference of 0.45mm (95% CI -0.02 - 0.62).</p> <p>a. Are the results statistically significant? b. Are the results clinically relevant? c. Are the results decisive?</p>	<p>a. No. The null value (0) is included in the confidence interval (-0.02-0.62) corresponding to $P > 0.05$. b. No. 0.45mm is less than the minimal value of clinical relevance set at 1mm. c. Yes. These results are decisive as the upper level of the confidence interval (0.62) is less than the minimal value of clinical relevance of 1mm. This means the majority of patients will not receive a clinical benefit from this intervention.</p>

TOOLBOX

Example 1-D	Answers to 1-D
<p>The same parallel group RCT as example 1-A, but now we obtain a mean difference of 1.25mm (95% CI 1.13-1.37).</p> <p>a. Are the results statistically significant?</p> <p>b. Are the results clinically relevant?</p> <p>c. Are the results decisive?</p>	<p>a. Yes, as the null value (0) is not included in the confidence interval.</p> <p>b. Yes, as 1.25mm is above the minimal value of clinical relevance set at 1mm.</p> <p>c. Yes, as the lower level of the confidence interval (1.13) is above the minimal value of clinical relevance set at 1mm. The majority of patients will receive a clinical benefit from the intervention.</p>
Example 2	Answers to Example 2
<p>A systematic review has examined the effects of smoking on implant failure. The overall OR for implant failure was reported to be 2.09 (95% CI 1.67-2.83).</p> <p>a. How might we verbally interpret these results?</p> <p>b. What type of data are these?</p> <p>c. Are the results statistically significant?</p> <p>d. What statistical test would we expect?</p> <p>e. Are the results clinically relevant?</p>	<p>a. As odds ratios are calculated as the odds exposed/odds control group, we would say that the odds of implant failure in smokers are slightly more than twice that of non-smokers.</p> <p>b. As the results are expressed as an odds ratio, we are dealing with binary data (such as implant survival/failure).</p> <p>3. Yes, as the null value for an odds ratio (1) is not contained within the confidence interval (1.67-2.83). This corresponds to $P < 0.05$.</p> <p>4. A chi-square test would be needed to obtain a P-value.</p> <p>5. To evaluate clinical relevance, the first step is to look at the control, or baseline failure rate. If we accept 5% as a baseline failure rate for dental implants, we are now looking at failure jumping from 5% to roughly 10%. Although a doubling of failure is undesirable, the baseline failure rate is rather low, and a 10% failure rate may be acceptable to many practitioners. Thus, we might state that an odds ratio of approximately two is clinically irrelevant. However, if baseline failure rates were in the order of 15%, a doubling of this value may be looked at as clinically relevant.</p>
Example 3	Answers to Example 3
<p>A parallel group RCT examined amalgam vs. composite failure due to recurrent caries. RR were reported to be 3.4 (95% CI 2.6-5.3).</p> <p>a. How might we verbally interpret these results?</p> <p>b. What type of data are these?</p> <p>c. Is this statistically significant?</p> <p>d. What statistical test would be needed to get a P-value?</p> <p>e. Is this clinically relevant?</p>	<p>a. We would say that the risk of failure due to recurrent caries is roughly 3½ times higher with composite as compared with amalgam.</p> <p>b. As risk ratios are reported, we are dealing with binary data (such as recurrent caries/no recurrent caries).</p> <p>c. Because the null value for a risk ratio (1) is not contained in the confidence interval, this is statistically significant.</p> <p>d. A chi-square test would be needed to obtain a P-value.</p> <p>e. As baseline failure rates for amalgam have generally been reported at approximately 5-6%, we are now looking at failure rates for composite at around 20%, a value that may be considered clinically relevant.</p>
Example 4	Answers to Example 4
<p>A parallel group RCT was done comparing manual vs. electric toothbrushes on oral hygiene (expressed as excellent, good, fair or poor). The authors ran an independent t-test and obtained a P-value of 0.04.</p> <p>a. What type of data are these?</p> <p>b. Is this statistically significant?</p> <p>c. Did the authors run the proper statistical test?</p> <p>d. What statistical test should they have done and why was it not employed?</p>	<p>a. These are ordinal data, as one category is better than the next, but the differences are not quantifiable.</p> <p>b. Yes, as $P < 0.05$.</p> <p>c. No, an independent t-test would be appropriate for two independent groups with continuous data.</p> <p>d. The Mann-Whitney U test is the appropriate test for ordinal data on two independent groups. The non-parametric Mann-Whitney U tests differences in rank order, whereas the t-test examines differences in means. The Mann-Whitney U test was not done as, in general, it is easier to obtain statistical significance with a parametric test, such as a t-test, as compared with a non-parametric test.</p>

1. Abt E. Understanding statistics 1. *Evid Based Dent* 2010; 11: 60-61.
2. Abt E. Understanding statistics 2. *Evid Based Dent* 2010; 11: 93-94.
3. Abt E. Understanding statistics 3. *Evid Based Dent* 2010; 11: 118.
4. Abt E. Understanding statistics 4. *Evid Based Dent* 2011; 12: 25-27.