Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.intl.elsevierhealth.com/journals/jden

Statistical methodology in oral and dental research: Pitfalls and recommendations

Ailish Hannigan [a,*], *Christopher D. Lynch* [b]

[a] *Biomedical Statistics, Graduate Entry Medical School, University of Limerick, Limerick, Ireland; Statistical Advisor, Journal of Dentistry*

[b] *Tissue Engineering & Reparative Dentistry, School of Dentistry, Cardiff University, Cardiff Wales, UK; Editor-in-Chief, Journal of Dentistry*

ARTICLE INFO

Article history:
Received 29 January 2013
Received in revised form
20 February 2013
Accepted 22 February 2013

Keywords:
Statistical methodology
Dental research
Errors
Misuse
Best practice

ABSTRACT

Objectives: This study describes the pitfalls for commonly used statistical techniques in dental research and gives some recommendations for avoiding them. It also explores the potential of some of the newer statistical techniques for dental research.

Methods: Each of the commonly used techniques e.g. descriptive statistics, correlation and regression, hypothesis tests (parametric and non-parametric) and survival analysis are explored with examples and recommendations for their use are provided. Common sources of error including those of study design, insufficient information, ignoring the impact of clustering and underuse of confidence intervals are outlined. The potential of statistical techniques such as multivariate survival models, generalized estimating equations and multilevel models are also explored.

Conclusions: Reviews of published dental research repeatedly identify statistical errors in the design, analysis and conclusions of the study. Educating researchers on common pitfalls and giving recommendations for avoiding them may help researchers to eliminate statistical errors. Developments in statistical methodology should be routinely monitored to ensure the most appropriate statistical methods are used in dental research.

1. Introduction

Evidence based dentistry has been defined as the integration of the best evidence with clinical expertise and patient preferences.[1] Establishing best evidence involves systematically collecting and analysing scientific evidence to answer a specific clinical question. Statistics has a central role in the production and analysis of scientific data and drawing inferences from that analysis, yet the misuse of statistics in published research is widespread. Reviews of statistical methodology in published papers in medicine estimated between 40 and 70% of the papers contained statistical errors, some of which were serious enough to have resulted in misleading conclusions.[2] Kim et al. assessed the level of misused statistics or statistical errors in dental research by reviewing 418 papers published between 1995 and 2009 in well- established dental journals.[2] For over a quarter of these papers, it could not be judged whether or not the use of statistics was appropriate because of insufficient information in the paper. For the remaining 307 papers, it was concluded that statistics was misused in 51.5% of them.

The most commonly used statistical techniques in the papers reviewed by Kim et al. were descriptive statistics, correlation and regression, hypothesis tests (parametric and non-parametric) and survival analysis.[2] The aim of this paper is to describe the pitfalls for each of these commonly used

* *Corresponding author*. Tel.: +353 61 202313; fax: +353 61 233778.
E-mail address: ailish.hannigan@ul.ie (A. Hannigan).

0300-5712/$ – see front matter
http://dx.doi.org/10.1016/j.jdent.2013.02.013

techniques in dental research and give some recommendations for avoiding them, along with a discussion of the potential of some of the newer statistical techniques for dental research.

2. Designing a study

Vähänikkilä et al. reviewed 928 published articles in four leading dental journals and categorised them by experimental study (clinical trials and studies on animals) and observational studies (cross-sectional studies, cohort studies and case–control studies).[3] Only a quarter of the studies were designed experimental studies. Observational studies can often 'happen' rather than being designed and use data which have often been collected for another purpose.[4] This can result in missing data particularly for subjects followed over time, and also variation in the measurement methods, number and quality of the measurements by different assessors. Other design issues can be the use of inappropriate controls, using a high risk sample to makes inferences about the general population and non-random sampling methods such as judgement, voluntary response or self-selected sampling. A justification for sample size is also important – is this a pilot or feasibility study or have power calculations been carried out? Lucena et al. reported that just 2 (1%) of the 226 research articles they reviewed in operative dentistry reported an a priori sample size calculation or effect size estimation.[5] 99% of the papers they reviewed gave no information on subjects who dropped out of the study and 87% did not report on the baseline characteristics and comparability of the study groups. Design flaws can be viewed as the most serious of all and cannot be fixed in the analysis phase of research, thus leading to inappropriate conclusions being drawn from a study.

> A justification for the sample size used in the study should be given. Baseline characteristics of the study groups should be compared and information given on non-response and dropouts.

3. Descriptive statistics

Graphical methods are a useful first step in exploring the data collected in a study and graphs are particularly useful for presenting data from individual subjects e.g. in a scatter plot or trends over time. Tufte describes the best principles for effective graphical presentation as showing all the data and not drawing the attention of viewers to irrelevant elements of the graph.[6] Three dimensional effects in a graph can be misleading as can presenting vertical axis without a true zero.[4] Two numbers are required to summarise a set of numeric data i.e. a measure of the centre of the data (mean or median) and a measure of the spread or variability of the data (standard deviation or quartiles). Graphs which present means or medians without an indication of variability are also misleading.

The standard deviation as a measure of variability is only appropriate to use with the mean not the median. Extreme outliers can distort the value of the mean and researchers should routinely compare the values of the mean and median for a set of numeric data. Large (clinically relevant) differences between the two would indicate the median is a more appropriate measure of the centre of the data or a transformation of the data (e.g. log transformation) should be investigated. Formal tests of normality can also be used to test the hypothesis that the data are normally distributed though these tests can be of limited used in very large samples where even small departures from normality will be statistically significant.

Even though commonly used, researchers should avoid the use of mean (±value) to summarise their results.[4] It is often not clear whether researchers are reporting the standard deviation or the standard error in the brackets. The standard deviation is a measure of the variability in the variable being investigated *between* individuals e.g. the variability in height in a sample of patients, and is thus a descriptive measure. The larger the standard deviation, relative to the mean, the more variation or dispersion is present in the data. By contrast, the standard error is a measure of the uncertainty in using a sample statistic to estimate a population parameter. For example, the standard error of the mean indicates the uncertainty of using the mean height for a sample of patients as an estimate of the mean value for the population of patients represented by the sample. The standard error of a mean is calculated by dividing the standard deviation by the square root of n where n is the sample size. Large sample sizes will therefore give rise to relatively small standard errors and it is important the reader knows which measure is being reported in a paper. There is also no special statistical importance given to the range (mean − one SD, mean + one SD) other than a property of the normal distribution tells us that this range typically includes 68% of observations. The clearest way of presenting numerical summaries is to use mean (SD = value) in a descriptive analysis. It is also suggested that even though computer packages produce results to four or more decimal places, care should be taken when choosing the appropriate level of numerical precision with which to present results so as not to detract from the readability of the article.[4] Similarly, when a p-value has been presented as 0.000 in the output from a statistical software package, this is usually as a result of rounding rather than a p-value of zero and it should be presented as $\leq$0.001 instead of 0.000.

> Both a measure of centrality and a measure of variability are required to describe a set of numeric data e.g. mean (SD=) or median (first quartile, third quartile). The standard deviation is only appropriate for use with the mean. The mean and the median should be routinely compared to investigate the impact of outliers.

4. Confidence intervals, hypothesis testing and p-values

Kim et al. reported that only 20 of the 307 papers reviewed in well-established dental journals reported using confidence