What is the Proper Sample Size for Studies of Periodontal Treatment?

Jihnhee Yu, PhD*, Luge Yang, MS, Alan D. Hutson, PhD

KEYWORDS

• Rules of thumb for the sufficient sample • Exact nonparametric tests
• Bootstrap method • Periodontal treatment • Small sample size

KEY POINTS

- The authors provide some rules of thumb regarding sufficient sample sizes for a few statistical methods.
- The authors examine distributional characteristics of data from periodontal studies and relevant sampling distributions.
- The authors provide some strategies to perform statistically sound data analysis with a small sample size.

INTRODUCTION

A key feature of scientific research activities is the design of an experiment, the collection of data from the realization of the given experiment, and testing of statistical hypotheses based on the accumulated data. The data itself may take a variety of forms. Thus, a broad knowledge of various statistical methods is crucial for performing scientifically sound research. Basic statistics texts well explain appropriate statistical techniques for different characteristics of the data corresponding to the various hypotheses of interests. Less emphasized is whether those statistical methods are suitable regardless of the sample sizes. What needs to be considered are the 2 types of errors involved in statistical testing termed type I and type II errors, as well as the statistical power (1 minus the probability of a type II error). Type I error is the probability of rejecting the statistical hypothesis of interest (termed the null hypothesis) when in fact it is true. Type II error rate is the probability of failing to reject the statistical null

Department of Biostatistics, University at Buffalo, State University of New York, 706 Kimball Tower, Buffalo, NY 14214-3000, USA
* Corresponding author.
E-mail address: jinheeyu@buffalo.edu

Dent Clin N Am ■ (2015) ■–■
http://dx.doi.org/10.1016/j.cden.2015.06.006 **dental.theclinics.com**

hypothesis when in fact it is false. The statistical power is the probability of rejecting the statistical null hypothesis of interest when in fact it is false.

Many scientific studies suffer from a small sample size owing to limitations on available resources or poor planning. Periodontal research outcomes may include outcomes such as bacteria colonization, measure of plaque, pocket depth, and clinical attachment loss.[1] With these types of data, one may ask whether it is appropriate that someone tests the difference between 2 treatments based on only 10 subjects per group. One may also ask whether several predictors can be included in a regression analysis when the total sample size is only, for example, 30. It is possible through trial and error to have a model that incorrectly overfits the data given a large number of covariates or a small sample size. This is due to an inflation of the type I error across multiple tests of multiple models. That is, if we test enough models, by chance alone we are likely to find one that fits the data well even, although there is no relationship. These issues are to be considered before carrying out the data analysis. Many statistical textbooks extensively address the power of hypothesis testing; however, not many statistical textbooks sufficiently address the validity (especially in the context of the type I error) of statistical inference with a small sample size, even though hypothesis testing with a small sample size may be of great interest to many researchers. The concept of the type I error is equivalent to the confidence level in a confidence interval, thus the validity of tests in terms of the type I error is directly addressing the validity of a confidence interval. This article provides some reasonable guidelines for researchers who want to draw statistically sound results and relevant interpretation from studies with small sample sizes.

It is well understood that studies with a small sample size have a low statistical power to detect a true effect. The statistical literature offers a rule of thumb for sample size in the context that the minimum sample size requirement often assures a certain statistical power for detecting a moderate effect size (eg, Refs.[2,3]). The suggested sample sizes based on the statistical power are usually large enough that the accompanying statistical methods are justified in terms of maintaining the desired type I error rate. In practice, the available samples are often limited because of time, budget, or ethical reasons. For researchers with limited resources, the statistical analysis needs to be performed based on the available sample size rather than on the sample size for a certain study power. As a result, the right question to be asked regards the suitability of the statistical methods with a limited sample size. Because many statistical methods rely on so-called large sample properties, a liberal use of statistical tests, regardless of inadequate sample sizes, may lead to inflated statistical type I errors (more detail for this point is given in the section Caution on Using the Bootstrap Method for a Small Sample Size), meaning that nonexistent study effects may too often be declared to be significant. Statistical software packages print out results but in general do not warn users that the results may be inaccurate because of small sample sizes. Inflated type I errors in turn give rise to low reproducibility of the results by future research. The problem of inflated type I error is less commonplace in practice by the incorporation of computationally intensive exact statistical methods (details given in section Alternative Methods for Small Sample Sizes), although the exact methods have a problem to adapt a variety of complex modeling schemes.

This article therefore discusses the sample sizes adequate for several popular statistical methods. It discusses distributional characteristics of the data based on a study to reduce the oral colonization of pathogens in the oral cavity[4] and other periodontal studies, followed by the bootstrap method that can be used to cope with small sample size problems. Alternative statistical methods that can be used with the small sample sizes are also described.