AJO-DO

Sample-size calculation for repeated-measures and longitudinal studies

Yi Guo[a] and Nikolaos Pandis[b]
Gainesville, Fla, Corfu, Greece, and Bern, Switzerland

In orthodontic research, investigators often design studies in which the main response variable is measured repeatedly over time. Compared with cross-sectional designs, repeated-measures designs allow a more definitive evaluation of within-person changes in the response variables across time. Moreover, collecting repeated measures can increase statistical power for detecting differences. Thus, fewer participants are required for conducting a study.

Despite the many advantages of repeated-measures designs, sample-size calculation for such designs is difficult because repeated measurements taken from the same participant are correlated. To compute a sample size, the researcher must specify the expected correlation pattern in the repeated measurements and find a software package that allows such a pattern.[1] In addition, many other inputs are also needed for sample-size calculation. In this article, we provide a step-by-step guide for how to calculate a sample size for repeated-measures designs. First, we discuss the greatest challenges of sample-size calculation for repeated-measures and longitudinal studies. Second, using an orthodontic example trial in which we compare the effectiveness of 2 appliances on overjet reduction, we guide the reader through the steps of sample-size calculation.

THE CHALLENGES

Choosing a primary hypothesis

Sample-size calculation is based on a particular hypothesis of the study. The researcher must choose a primary hypothesis and base the sample-size calculation on that hypothesis. With repeated measures, the researcher could possibly choose to test a several hypotheses. The treatment-by-time interaction hypothesis is usually of most interest. This hypothesis tests whether the trend of the response variable across time is the same between the treatment and the control groups. The Figure is an example of a treatment-by-time interaction. The change in overjet across time is assumed to be linear in both groups, and we could test whether the rate of overjet change differs between the treatment and control groups during the 18-month study period. Another possible hypothesis is the main-effect hypothesis, where we can test the effect of a particular predictor variable averaged across all other factors. For example, we could test whether the responses of participants in the treatment group differ from those of participants in the control group, averaged across all the repeated measurements.

Specifying the variances and correlations among the repeated measurements

A big challenge in sample-size calculation for repeated-measures designs is that the researcher must specify the variances (or standard deviations) and the correlations among the repeated measurements. With 4 repeated measurements, 4 variance or standard deviation values and 6 correlation values need to be specified. To compute an accurate sample size, the specified values should match as closely as possible the values expected to be observed in the data.[2]

To specify the variance values, it is often possible to estimate one variance value based on a previous study or to make an educated guess based on experience and then specify the other variance values based on the expected variance trend across time. To determine the variance trend across time, the scientific context often provides a reasonable model for variance change. For example, disease development often leads to monotonically increasing the variance among repeated measurements of responses. Often, we can assume that the variance stays constant across time.

To estimate the correlations, the same principles for specifying variances apply. We first estimate one

[a]Assistant professor, Department of Health Outcomes and Policy, College of Medicine, University of Florida, Gainesville, Fla.
[b]Private practice, Corfu, Greece; visiting assistant professor, Department of Orthodontics and Dentofacial Orthopedics, Faculty of Medicine, University of Bern, Bern, Switzerland.
Yi Guo's support included NIH/NIDCR 1R01DE020832 and NIH/NIDCR 1U54DE019261.
Address correspondence to: Yi Guo, Department of Health Outcomes and Policy, College of Medicine, University of Florida, PO Box 100177, Gainesville, FL 32610-0177; e-mail, yiguo@ufl.edu.
Am J Orthod Dentofacial Orthop 2015;147:146-9
0889-5406/$36.00

http://dx.doi.org/10.1016/j.ajodo.2014.10.009

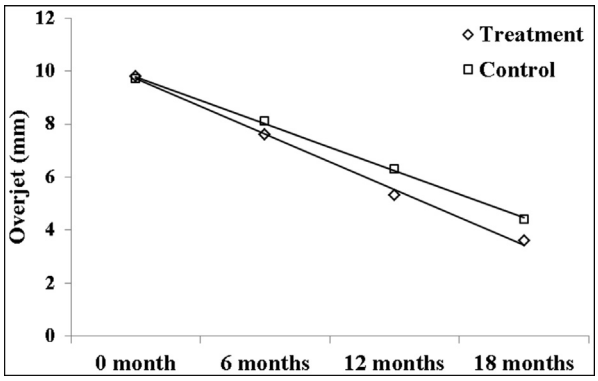

Fig. Illustration of an interaction between the treatment and the trend of overjet. The graph shows that the 2 lines are not parallel, indicating a treatment-by-time interaction. On the y-axis, on average the control and the treatment groups start with similar average overjets; however, with time, the reduction in overjet is greater in the treatment group compared with the control group.

correlation value based on a previous study or an educated guess, and then we specify the other correlation values based on the expected correlation patterns. Three types of correlation patterns can be considered, in increasing complexity: equal correlations, rule-based patterns, and "unstructured" correlations (no specific pattern). The simplest pattern of correlations assumes a constant correlation among the repeated measurements. However, this pattern is often not valid in longitudinal studies because measurements taken farther apart in time are usually less correlated than measurements taken closer in time. For instance, in our example study of overjet reduction, we would expect the correlation between the baseline measurement and the 6-month measurement to be larger than the correlation between the baseline measurement and the 12-month measurement. A more realistic hypothesis is that correlations among repeated measurements decline exponentially with time. In this case, we could consider modeling the decline in correlation with a rule-based pattern. A commonly used pattern is the first-order autoregressive, which belongs to the linear exponent first-order autoregressive family.[3] Using the more general linear exponent first-order autoregressive model requires providing 2 correlation parameters: the base correlation and the decay rate. The base correlation is the correlation between 2 measurements taken next to each other in time. The decay rate is the rate of decline in base correlation as time between repeated measurements increases. Our experience with biologic and behavioral data leads us to suggest specifying a decay rate between 0.05 and 0.5. Finally, we could assume that there are no particular correlation patterns and that each correlation between any 2 repeated measurements is unique and needs to be specified. The number of correlations that need to be specified is $p \times (p - 1)/2$, where p is the number of repeated measurements. Assuming "unstructured" correlations requires estimating the most correlations. Moreover, the number of correlations that needs to be specified increases dramatically as the number of repeated measurements increases.

Choosing an appropriate software

Currently, there are only a few software packages or programs that compute sample sizes for a limited range of repeated-measures and longitudinal designs. Therefore, it is crucial that the researcher carefully chooses a program that can support the assumptions of the study design. For example, if a program operates under the assumption of equal correlation among repeated measurements, this program is most likely unsuitable for computing sample sizes for longitudinal studies, for which the assumption of equal correlations is rarely true.

Another challenge is that powerful sample-size programs may require strong knowledge in statistical theories and programming skills.[4] Few graphical user interface sample-size programs have the ability to support longitudinal designs and various variance and correlation patterns. PASS (NCSS) and GLIMMPSE (http://glimmpse.samplesizeshop.org/) are graphical user interface programs designed for applied researchers.[5,6] PASS is a commercial product that must be purchased and installed on a computer. GLIMMPSE is a free, Internet-based program. Both programs support repeated-measures and longitudinal designs and require no programming experience.

THE EXAMPLE

In this section, we provide a step-by-step tutorial on how to calculate a sample size for repeated-measures and longitudinal studies using the "guided study design" mode of GLIMMPSE. We show what information is needed and how to gather the information. Readers are reminded that although GLIMMPSE is the choice of sample-size program in this example, the same information is required even if another software program is preferred.

Overview of the orthodontic study

In the study being planned, the researcher is interested in comparing the effect of the Twin-block appliance with the headgear appliance on overjet reduction. Patients eligible for the study are those between the ages of 10 and 13 years, and with overjet greater than