Biostatistics (2020), **0**, 0, *pp.* 1–11
doi:10.1093/biostatistics/Manuscript-Draft

Power and sample size for longitudinal data: the "longpower" R package and app

SAMUEL IDDI[1], MICHAEL C. DONOHUE[2*]

for the Alzheimer's Disease Neuroimaging Initiative[†]

[1] *Data Measurement and Evaluation Unit, African Population and Health Research Center (APHRC), Nairobi, Kenya*

[2] *Alzheimer's Therapeutic Research Institute, Keck School of Medicine, University of Southern California, San Diego, USA*

siddi@aphrc.org

Summary

Longitudinal studies are ubiquitous in medical and clinical research. Sample size computations are critical to ensure that these studies are sufficiently powered to provide reliable and valid inference. There are several methodologies for calculating sample sizes for longitudinal studies that depend on many considerations including the type of research design, outcome type, and proposed analytical methods. More advanced procedures take into account several factors leading to more complicated formulas for computing sample sizes. To provide easy access to common and widely used sample size and power calculation formulas, this tutorial briefly describes some methods for longitudinal data. We also enrich the discussion with real-life examples comparing treatment versus control groups in randomized trials assessing treatment effect on clinical out-

[*†] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`

comes. Accompanying this tutorial is also a web-based sample size Shiny app developed to help researchers to conduct different sample size and power calculations by allowing user-specified parameters or pilot parameter estimation using data from the Alzheimers Disease Neuroimaging Initiative (ADNI) study.

Key words: Mixed model for repeated measures, linear mixed model, Shiny app, power, sample size, longpower, longitudinal data

1. Introduction

Longitudinal designs are generally preferred over cross-sectional research design as they yield provide richer data and greater statistical power. As such, many biomedical and medical studies employ longitudinal design to study changes over time in outcome at the individual, group, or population level. Early in the design of a longitudinal experimental or natural history study, it is imperative to ensure that the study is adequately powered for its aims. Inadequate sample sizes leads to invalid or inconclusive inference and squandered resources (Lu *and others* (2009); Yan and Su (2006)). On the other hand, oversampling causing a waste of resources and exposure to many participants to harm that may be associated with the research (Lu *and others* (2009)). Thus, optimal sample size and power analysis have become an important prerequisite for any quantitative research design. Not only are these required during the design phase of research, but it has also become mandatory when preparing protocols for ethical review and research grant applications to guarantee both economic and ethical benefits. Determining the right sample size for a study is not a straightforward task. Despite the plethora of sample size formulas for repeated measures (Overall and Doyle (1994); Lui (1992); Rochon (1991); Guo *and others* (2013)), cluster repeated measures (Liu *and others* (2002)), multivariate repeated measures (Vonesh and Schork (1986); Guo (1996)), longitudinal research designs (Lefante (1990)), the tasks of gathering the

necessary inputs and getting the right software to carry out the computation are challenging to many. Investment in such efforts is not what many non-technical researchers desire to embark on. Commonly, the easiest route taken by most research is to use sample size formulas for very basic cross-sectional studies and adjust for design effect due to repeated measures. Although such approaches are valid, the best approach is the use of formulas derived directly from models for longitudinal or repeated measures to align with planned data analysis and yielding greater statistical power. Several factors need to be considered before choosing the right sample size formulas that increase the statistical power of a study. Guo *and others* (2013) describe practical methods for the selection of appropriate sample size for repeated measures addressing issues of missing data, and the inclusion of more than one covariates to control for differences in response at baseline.

Sample size formulas are refined depending on the specific situation and design features. For example, Hedeker *and others* (1999) considered a sample size for longitudinal design comparing two groups that accounted for subject attrition or drop-out. Basagaa *and others* (2011) derived sample size formulas for continuous longitudinal data with time-varying exposure variables typical of observational studies. Ignoring time-varying exposure was demonstrated to lead to substantial overestimation of the minimum required sample size which can be economically disadvantageous. In non-traditional longitudinal designs such as designs for mediation analysis of the longitudinal study, further refinements to sample size formulas are needed to ensure that sufficient sample sizes are obtained for conducting mediation analysis (Pan *and others* (2018)).

Extension to basic sample size formulae usually requires additional parameters such as exposure mean, variance, and intraclass correlations (Basagaa *and others* (2011)), mediation effect, number of repeated measures (Pan *and others* (2018)), covariance structures (Rochon (1991)), non-linear trends (Yan and Su (2006)), missing, attrition or dropout rates (Roy *and others* (2007); Lu *and others* (2008)), among others.

Advanced sample size methods simultaneously handle several practical issues associated with research design and complications that may arise during data collection. However, such methods are only available in commercial software.

Sample size calculation and power analysis are important components in designing a new trial or study. For many non-statisticians, the processes involved in performing these types of analyses can be daunting. A major hurdle to overcome is the availability of pilot parameters which are required inputs for generating sample size and power outputs. In this paper, we develop a power/sample size App for Alzheimer's Disease (AD) clinical trials. The web-based app implements the linear mixed model and basic mixed method repeated measures (MMRM) allowing user to input own pilot estimates, or use an ADNI-based pilot estimate generator to compute sample size and power.

2. Methods

Clinical trial data collected longitudinally over time are commonly analyzed using the linear mixed model and mixed model for repeated measures methods for continuous outcomes. Prior to such trials, sample size and power analysis are conducted to obtain the required sample size needed to assess treatment effect with optimal power. Various sample size approaches for longitudinal data have been proposed. We review a few of the most commonly used methods applied in Alzheimer's disease trials focusing on continuous outcome type.

2.1 *Sample size computation based on the linear mixed model (LMM)*

One of the most common sample size computation approaches for correlated data is derived by Liu and Liang (1997). This approach derived sample size from a generalized estimating equation (Liang and Zeger (1986)). Thus, different outcomes types can be handled. A special case is for continuous response measured repeatedly over time and modeled using a linear mixed model

(LMM). The model specification involves fixed covariates and random-effects to the outcome for between-subject variability. The error component is assumed to follow a multivariate with a mean vector of zeros and covariance matrix. Different covariance structures such as independence, exchangeable, auto-regressive, and unstructured to estimate the minimum sample size for a given significant level and pre-specified nominal power. A reduced form of the linear mixed model with only the treatment effect and a random-intercept component, the sample size formula generated by this approach is equivalent to that of Diggle *and others* (1994).

2.2 *Sample size computation based on the mixed model for repeated measures (MMRM)*

An alternative sample size computation method used in many areas of application including AD trials is that based on the mixed model for repeated measures (MMRM; Mallinckrodt *and others* (2001, 2003); Lane (2008). The MMRM method of model fitting is similar to the mixed-effect model for longitudinal or repeated measures except for the unstructured modelling of time treated as a categorical variable, and the specification of a within-subject error structure. Additionally, the MMRM considers the interaction between time and treatment.

2.3 *The 'longpower' package*

2.4 *A Web-based App*

This Shiny app dashboard is developed to easily generate sample size and conduct power analysis for a longitudinal study design with two-group comparisons for a continuous outcome. The App implements the sample size formula of Liu and Liang (1997) and Diggle *and others* (1994, 2002) using functions developed for the R 'longpower' package. The 'longpower' package handles cases where time is treated either as continuous or categorical. The former approach uses the linear mixed model with random intercept and slope while the later leads to the well-known Mixed Model of Repeated Measures (MMRM) used in many clinical trial applications for conducting

statistical analysis.

The dashboard is in two parts. The first part accepts user inputs to generate sample size when time is treated as both categorical and continuous. Thus, this part assumes that the user already has pilot parameter estimates including effect size and known variance. Users can generate sample sizes and perform power analysis using different sample size methods (`diggle, liuliang`, and `edland`). The basic inputs of the App include an option for whether sample size or power is desired, a slider for sample size and power enabled when appropriate and difference in means (treatment effect) estimate. Additional options include the type I error rate, type of test (one/two sided), estimates of random effect variances, sample size computation method, allocation ratio and time intervals. For the MMRM method, options for the association structure and retention in each group is enable depending on the association structure assumed by the user.

The second part of the dashboard uses the methodology that is similar to the first part except that Alzheimer's Disease Neuroimaging Initiative (ADNI)-based pilot parameters are generated based on user-selected inclusion and exclusion criteria, primary outcome, duration of the study, and covariate options to be included in the linear mixed and MMRM models.

The App interface display plots of power against sample size over range of values, and text summary of imputed and selected estimates. For the ADNI-based generator, additional summary estimates from fitted model and descriptive summary of baseline participants characteristics are also dynamically generated.

3. Illustrative examples

To illustrate how sample size/power can be obtained using the 'longpower' package and the web app, we discuss two illustrative examples.

3.1 *Hypothetical clinical trial*

We consider the hypothetical clinical trial example discussed in Diggle *and others* (2002). Suppose that we are interested in testing the effect of a new treatment in reducing blood pressure through a clinical trial. The investigator is interested in randomizing subjects between a control and active treatment group to have equal size. Three visits are envisaged with assessments planned at years 0, 2, and 5. Suppose that the within-subject variance is given by and assuming an error rate of 0.05, power of 80% and smallest meaningful difference of 0.5 mmHg/year, we want to determine the number of subjects needed for both treated and control groups for varying correlation (0.2, 0.5 and 0.8) and response variance (100, 200, 300).

3.2 *Alzheimer's Disease Neuroimaging Initiative (ADNI)*

ADNI is a population-based longitudinal cohort study that follows study participants to collect data on their clinical, cognitive, imaging including MRI and PET images, genetic, and biochemical biomarkers. The study was designed to discover, optimize, standardize, and validate clinical trial measures and biomarkers that are used in AD clinical research. This multi-site longitudinal study, runs at about 63 sites in the US and Canada and began in 2004. All the data generated from the ADNI study are entered into a data repository hosted at the Laboratory of Neuroimaging (LONI) at the University of Southern California, the LONI Image & Data Archive (IDA). The data can be freely accessed upon request. Apart from the many utility of the data for advancing knowledge for AD trial, this big data resource can be used to improve study design. In particular, the data is used to generate pilot estimates for the computation of sample size and power analysis.

```
n = 3
t = c(0,2,5)
rho = c(0.2, 0.5, 0.8)
sigma2 = c(100, 200, 300)
tab = outer(rho, sigma2,
Vectorize(function(rho, sigma2){
```

```
ceiling(diggle.linear.power(
delta=0.5,
t=t,
sigma2=sigma2,
R=rho,
alternative="one.sided",
power = 0.80)$n)}))
colnames(tab) = paste("sigma2 =", sigma2)
rownames(tab) = paste("rho =", rho)
tab
```

4. Discussion

Acknowledgments

We are grateful to the ADNI study volunteers and their families. This work was supported by Biomarkers Across Neurodegenerative Disease (BAND-14-338179) grant from the Alzheimer's Association, Michael J. Fox Foundation, and Weston Brain Institute; and National Institute on Aging grant R01-AG049750. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Re-

search is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. *Conflict of Interest*: None declared.

References

Basagaa, X., Liao, X. and Spiegelman, D. (2011). Power and sample size calculations for longitudinal studies estimating a main effect of a time-varying exposure. *Statistical methods in medical research* **20**, 5.

Diggle, P.J., Heagerty, P.J., Liang, K. and Zeger, S.L. (2002). *Analysis of longitudinal data. Second Edition*. Oxford Statistical Science Series.

Diggle, P., Liang, K.-Y., and Zeger, S. L. (1994). *Analysis of Longitudinal Data*. New York: Oxford University Press.

Guo, Johnson W. D. X. (1996). Sample size for experiments with multi-variate repeated measures. *J. Biopharmaceutical Statistics* **6**, 155176.

Guo, Y., Logan, H.L. and Glueck, D.H. et al. (2013). Selecting a sample size for studies with repeated measures. *BMC Med Res Methodol* **13**, 100.

Hedeker, D., Gibbons, R. D. and Waternaux, C. (1999). Sample size estimation for longitudinal designs with attrition: comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics* **24**, 7093.

Lane, P. (2008). Handling drop-out in longitudinal clinical trials: a comparison of the locf and mmrm approaches. *Pharmaceut. Statist.* **7**, 93106.

Lefante, J. J. (1990). The power to detect differences in average rates of change in longitudinal studies. *Statistics in Medicine* **9**, 437446.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Liu, A., Shih, W. J. and Gehan, E. (2002). Sample size and power determination for clustered repeated measurements. *Statistics in Medicine* **21**, 17871801.

Liu, G. and Liang, K. Y. (1997). Sample size calculations for studies with correlated observations. *Biometrics* **53**(3), 937–47.

Lu, K., Luo, X. and Chen, P. (2008). Sample size estimation for repeated measures analysis in randomized clinical trials with missing data. *The International Journal of Biostatistics* **4**(1), Article 9.

Lu, K., Mehrotra, D. V. and Liu, G. (2009). Sample size determination for constrained longitudinal data analysis. *Statist. Med.* **28**, 679699.

Lui, K. J. (1992). Sample size requirement for repeated measurements in continuous data. *Statistics in Medicine* **11**, 633641.

Mallinckrodt, C.H., Sanger, T.M., Dub, S., DeBrota, D.J., Molenberghs, G., Carroll, R.J., Potter, W.Z. and Tollefson, G.D. (2003). Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. *Biological Psychiatry* **53**, 754–760.

Mallinckrodt, C. H., Clark, W. S. and David, S. R. (2001). Accounting for dropout bias using mixed-effects models. *Journal of Biopharmaceutical Statistics* **11**, 9–21.

Overall, J. E. and Doyle, S. R. (1994). Estimating sample size for repeated measurement designs. *Controlled Clinical Trials* **15**, 100123.

Pan, H., Liu, S., Miao, D. and et al. (2018). Sample size determination for mediation analysis of longitudinal data. *BMC Med Res Methodol* **18**, 32.

Rochon, J. (1991). Sample size calculations for two-group repeated-measures experiments. *Biometrics* **47**, 13831398.

Roy, A., Bhaumik, D. K., Aryal, S. and D., Gibbons R. (2007). Sample size determination for hierarchical longitudinal designs with differential attrition rates. *Biometrics* **63**, 699707.

Vonesh, E. F. and Schork, M. A. (1986). Sample sizes in multivariate analysis of repeated measurements. *Biometrics* **42**, 601610.

Yan, X. and Su, X. (2006). Sample size determination for clinical trials in patients with nonlinear disease progression. *Journal of Biopharmaceutical Statistics* **16**(1), 91–105.

[*Received August 1, 2010; revised October 1, 2010; accepted for publication November 1, 2010*]