

Power and Sample Size for Longitudinal Models in R – The longpower Package and Shiny App

by Samuel Iddi and Michael C. Donohue, for the Alzheimer's Disease Neuroimaging Initiative

Abstract 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 inferences. There are several methodologies for calculating sample sizes for longitudinal studies that depend on many considerations including the study design features, outcome type and distribution, and proposed analytical methods. We briefly review the literature and describe sample size formulas for continuous longitudinal data. We then apply the methods using example studies comparing treatment versus control groups in randomized trials assessing treatment effect on clinical outcomes. We also introduce a Shiny app that we developed to assist researchers with obtaining required sample sizes for longitudinal studies by allowing users to enter required pilot estimates. For Alzheimer's studies, the app can estimate required pilot parameters using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Illustrative examples are used to demonstrate how the package and app can be used to generate sample size and power curves. The package and app are designed to help researchers easily assess the operating characteristics of study designs for Alzheimer's clinical trials and other research studies with longitudinal continuous outcomes.

1 Introduction

Longitudinal designs are generally preferred over cross-sectional research designs as they provide richer data and greater statistical power. As such, many biomedical and medical studies employ longitudinal designs to study changes over time in outcomes 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 et al., 2009; Yan and Su, 2006). On the other hand, oversampling squanders resources and exposes participants to unnecessary risks associated with the intervention (Lu et al., 2009). Thus, optimal sample size and power analysis have become important prerequisites for any quantitative research design. Not only are these required during the design phase of research, but it has also become mandatory for ethical, scientific, or economic justification in submissions to institutional review boards and research funding agencies.

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 et al., 2013), cluster repeated measures (Liu et al., 2002), multivariate repeated measures (Vonesh and Schork, 1986; Guo and Johnson, 1996), longitudinal research designs (Lefante, 1990), the tasks of gathering pilot estimates of the necessary parameters and getting the right software to carry out the computation can be challenging. Researchers commonly rely on formulas for very basic cross-sectional studies and adjust for attrition and other longitudinal design effects. Although such approaches can yield appropriate approximations, the ideal approach is to use a formula derived directly from the longitudinal model that researchers plan to eventually use on the study data.

Guo et al. (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 covariate to control for differences in response at baseline. Sample size formulas are refined depending on the specific situation and design features. For example, Hedeker et al. (1999) considered a sample size for longitudinal designs comparing two groups that accounted for participant attrition or drop-out. Basagaña et al. (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 (Pan et al., 2018). However, these formulas usually require additional parameters such as exposure mean, variance, and intraclass correlations (Basagaña et al., 2011), mediation effect, number of repeated measures (Pan et al., 2018), covariance structures (Rochon, 1991), non-linear trends (Yan and Su, 2006), missing, attrition or dropout rates (Roy et al., 2007; Lu et al., 2008), among others. Advanced sample size methods simultaneously handle several practical issues associated with the research design and complications that may arise during data collection.

However, such methods are only available in commercial software (NCSS Statistical Software, 2021; nQuery, 2021).

Several R packages can be found on CRAN to compute sample size based on mixed-effect models and other specific designs depending on the area of applications. For example, Martin et al. (2011) proposed a simulation-based power calculation and an R package **pamm** for random regression models, a specific form of mixed-effect model that detects significant variation in individual or group slopes. In their approach, a power analysis was performed to detect a specified level of individual and environmental interactions within evolution and ecology applications. This is achieved by simulating power to detect a given covariance structure. Other simulation-based packages for power analysis are the **SIMR** by Green and MacLeod (2016) for linear and generalized linear mixed models and **clusterPower** by Kleinman (2021) for cluster-randomized and cross-over designs. Schoenfeld (2019) developed a power and sample size package called **LPower** (Diggle et al., 1994) to perform power analysis for longitudinal design accounting for attrition and different random effect specification. The approach requires the specification of a design matrix, and the variance-covariance matrix of the repeated measures (Yi and Panzarella, 2002). In pharmacokinetic study designs, Klopogge and Tarning (2015) developed the **PharmPow** power calculation package for mixed study designs including crossover and parallel designs. Quite recently, other R packages for performing power analysis exist for different designs; for example the **powerMediation** (Qiu, 2021) for mediation effect, mean change for longitudinal study with 2-time points, the slope for simple Poisson regression, etc.; **powerEQTL** (Dong et al., 2021) for unbalanced one-way ANOVA in a Bulk Tissue and Single-Cell eQTL Analysis; **WebPower** (Zhang et al., 2021) for basic and advanced power correlation, proportion, *t*-test, one-way ANOVA, two-way ANOVA, linear regression, logistic regression, Poisson regression, mediation analysis, longitudinal data analysis, structural equation modeling, and multilevel modeling; among others. These packages were tailored towards very specific areas of application although the methods can be adapted and utilized for other disciplines. In terms of software applications, attempts have been made in recent times to implement power and sample size calculations in Shiny applications to facilitate easy usage. For example, Hemming et al. (2020) developed an R Shiny app to conduct power analysis for several cluster designs including parallel, cross-over, and stepped-wedge designs. Schoemann et al. (2017) Shiny application conducts power analysis for mediation model, and Hu and Qu (2021) performs sample size and power calculations for a random coefficient regression model (RCRM) and a two-stage mixed-effects model.

As the analysis model and associated sample size formula become more sophisticated, estimating the parameters required by the formula becomes more challenging. A major hurdle to overcome is the availability of pilot data to inform these parameters. To assist Alzheimer's researchers with this challenge, we have developed a power and sample size Shiny app for Alzheimer's clinical trials. The app implements formulas for the linear mixed-effects model and mixed model for repeated measures (MMRM; Lu et al., 2008) allowing the user to input their pilot estimates, or allow the app to generate pilot estimates using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI; Weiner et al., 2015).

2 Review of methods

Continuous outcomes in clinical trial data collected longitudinally over time are commonly analyzed using linear mixed models (LMM; Laird and Ware, 1982) or MMRM (Mallinckrodt et al., 2001, 2003). Before such trials, it is necessary to estimate the required sample size for a given treatment effect with desired power and Type I error. 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 with continuous outcomes.

Sample size computation based on the linear mixed-effects model

Several sample size approaches have been developed by different authors. Diggle et al. (2002) proposed sample size formulas for two approaches to continuous longitudinal outcomes, one that assumes a constant mean over time and compares the average response over time between groups, and another that assumes a linear change over time and compares the mean rate of change or slope difference between groups. In either case, the null hypothesis is that there is no difference between groups. Suppose that Y_{ij} is the response for participant i taken at time j . To compare the average response between two groups (e.g. group A receiving active treatment and group B receiving control) across time, consider the model

$$Y_{ij} = \beta_0 + \beta_1 T_i + \epsilon_{ij}$$

where T_i is the treatment indicator which is coded 1 for participants in group A and 0 for participants in group B , and ϵ_{ij} the error term for subject, i at time j , which is normally distributed with common variance, σ^2 . The sample size required per group to detect an average response between the two groups, δ , is given by

$$m = \frac{2(Z_{1-\alpha/2} + Z_P)^2 \sigma^2 (1 + (J-1)\rho)}{J\delta^2}$$

where α is the type I error rate, P is the level of power, and $\text{corr}(Y_{ij}, Y_{ik}) = \rho$ for all $j \neq k$.

To test difference in the slopes or average rate of change between the two groups, consider the parameterization:

$$Y_{ij} = \beta_{0A} + \beta_{1A}t_{ij} + \epsilon_{ij}, i = 1, 2, \dots, m; j = 1, 2, \dots, J.$$

A similar parameterization can be specified for group B with parameters β_{0B} , and β_{1B} representing the intercept and slope. The parameters β_{1A} and β_{1B} are the rate of change of the outcome for groups A and B , respectively. Assume that $\text{var}(\epsilon_{ij}) = \sigma^2$ and $\text{corr}(Y_{ij}, Y_{ik}) = \rho$ for all $j \neq k$. If the outcome of each participant is measured at the same time, t_j , then the sample size per group needed to detect a difference in the rate of change for each group, $\Delta = \beta_A - \beta_B$, is given by

$$m = \frac{2(Z_{1-\alpha/2} + Z_P)^2 \sigma^2 (1 - \rho)}{Js_t^2 \Delta^2}$$

where α is the type I error rate, P is the power, and $s_t^2 = \frac{\sum(t_j - \bar{t})^2}{J}$ is the within-participant variance of t_j .

Another sample size computation approach for correlated data is derived by [Liu and Liang \(1997\)](#) to detect differences in the average response between two groups. This approach derived sample size following the generalized estimating equation ([Liang and Zeger, 1986](#)) approach. Thus, different outcomes types can be handled. A special case is for a continuous response measured repeatedly over time and modeled using a linear model. Consider the model

$$Y_{ij} = \beta_0 + \beta_1 T_i + \epsilon_{ij}, i = 1, 2, \dots, N, j = 1, 2, \dots, J$$

where T_i is the treatment indicator, ϵ_{ij} is the error term assumed to follow a multivariate normal distribution with a mean vector of zeros, $\mathbf{0}$, and covariance matrix, $\sigma^2 \mathbf{R}$ and \mathbf{R} is a $J \times J$ working correlation matrix. Different covariance structures such as independent, exchangeable, auto-regressive, and unstructured can be assumed. To estimate the total sample size required, N for a given significance level (α) and pre-specified nominal power, P , the following formula is used:

$$N = \frac{(Z_{1-\alpha/2} + Z_P)^2 \sigma^2}{\pi_A \pi_B (\beta_1^2 \mathbf{1} \mathbf{R}^{-1} \mathbf{1})}$$

where π_A and π_B represent the proportion of observations in the control and treated groups, respectively. By assuming an exchangeable working correlation matrix, this formula reduces to

$$N = \frac{(Z_{1-\alpha/2} + Z_P)^2 \sigma^2 [1 + (J-1)\rho]}{\pi_A \pi_B \beta_1^2 J}.$$

It can be observed that for equal proportion of participants in each group i.e. $\pi_A = \pi_B = \frac{1}{2}$, the formula for the sample size per group is equivalent to the sample size formula for difference in average response between groups by [Diggle et al. \(2002\)](#) above. An appealing aspect of [Liu and Liang \(1997\)](#) approach is that it can allow for unequal allocation to the two groups, and different outcome types are considered.

Another approach is based on the LMM analysis ([Fitzmaurice et al., 2004](#)) comparing mean rate of change between groups ([Ard and Edland, 2011](#); [Zhao and Edland, 2020](#)). Consider the LMM

$$Y_{ij} = X'_{ij}\boldsymbol{\beta} + b_{0i} + b_{1i}t_{ij} + \epsilon_{ij}$$

where X_{ij} denotes a vector of indicator variables (treatment, time, and interaction terms), $\mathbf{b}_i = (b_{0i}, b_{1i})' \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_{b_0}^2 & \sigma_{b_0, b_1} \\ \sigma_{b_0, b_1} & \sigma_{b_1}^2 \end{bmatrix}\right)$ are random participant-specific effects, and $\epsilon_{ij} \sim N(0, \sigma_e^2)$ are the residual errors. The random terms, \mathbf{b}_i and ϵ_{ij} are assumed to be independent of each other. Following this model specification, [Ard and Edland \(2011\)](#) proposed a total sample size formula given by

$$N = \frac{4(Z_{1-\alpha/2} + Z_P)^2}{\Delta^2} \left[\sigma_{b_1}^2 + \frac{\sigma_e^2}{\sum(t_j - \bar{t})^2} \right]$$

where t_j indexes the times at which measures are made assuming all participants are observed at the same time points, \bar{t} is the mean of the times, $\sigma_{b_i}^2$ is the variance of the participant-specific random slope, and Δ is the difference in mean rate of change in treatment versus control group to be detected. For a random-intercept mixed-effect model, the total sample size formula reduces to

$$N = \frac{4(Z_{1-\alpha/2} + Z_P)^2}{\Delta^2} \left[\frac{\sigma_e^2}{\sum(t_j - \bar{t})^2} \right].$$

This sample size formula does not account for the correlation between the random intercept and slope, a limitation that can be mitigated by a recent approach by [Hu et al. \(2021\)](#).

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

An alternative approach is to treat time as a categorical variable. This approach, common in trials of treatments of Alzheimer's and other therapeutic areas, is often referred to as the MMRM ([Mallinckrodt et al., 2001, 2003; Lane, 2008](#)). The approach to fitting the MMRM is similar to that for other linear mixed-effect models for longitudinal or repeated measures except the unstructured modelling of time – treated as a categorical variable, and the specification of a within-participant correlation structure to account for association among the repeated measurements. The MMRM provides an estimate of the mean response for each time point category, for each group, and the resulting mean trajectory over time is unconstrained. The primary test statistic is usually the estimated group difference at the final time point. The null hypothesis is again that there is no difference between groups.

Suppose that Y_{ij} is the j th measurement taken of participant i ; $i = 1, 2, \dots, N$ at time, t_{ij} and T_i represents treatment or group indicator. Assuming we have two groups, $a; = (A, B)$, let n_{aj} be the number of participants at time $j = 1, 2, \dots, J$ for group a . Thus, at time point 1, the total number of participants is $n_{A1} + n_{B1}$. The response variable is assumed to follow a multivariate normal distribution with mean, variance-covariance and correlation matrix given respectively as

$$\begin{aligned} E(Y_{ij}|T_i = a) &= \mu_{aj} \\ \text{cov}(Y_{ij}, Y_{ik}|T_i = a) &= \sigma_{ajk} = \Sigma_a \\ \text{corr}(Y_{ij}, Y_{ik}|T_i = a) &= \frac{\sigma_{jk}}{\sigma_j \sigma_k} = R \end{aligned}$$

where $\sigma_j = \sigma_{jj}^{\frac{1}{2}}$. It is common to assume that the variance-covariance matrix is the same for each group in which case we can have $\Sigma_A = \Sigma_B = \Sigma$. Different correlation structures, R , can be assumed (eg., compound symmetry, autoregressive and unstructured). The sample size formula for MMRM ([Lu et al., 2008](#)) can account for the attrition rate at each time point. Define the attrition rate $\zeta_{aj} = 1 - r_{aj}$, where $r_{aj} = \frac{n_{aj}}{n_{a1}}$ is the retention rate at time point, j . The sample size for group A needed to achieve a power, P , at a significance level, α is given by

$$n_{A1} = (\psi_A + \lambda \psi_B)(Z_{1-\alpha/2} + Z_P)^2 \frac{\sigma_J^2}{\delta_J^2}$$

where $\lambda = \frac{n_{A1}}{n_{B1}}$ is called the allocation ratio, δ_J is the effect size (difference in mean response between the two groups) at the last time point, J , and ψ_a is the (J, J) th component of the variance inflation factor, I_a^{*-1} defined with respect to an estimation of the μ_{aj} for treatment group, a . Specifically, $\psi_a = [I_a^{*-1}]_{JJ}$ where

$$I_a^* = \sum_{j=1}^J (r_{a,j} - r_{a,j+1}) \begin{bmatrix} R_j^{-1} & 0_{j \times (J-j)} \\ 0_{(J-j) \times j} & 0_{(J-j) \times (J-j)} \end{bmatrix}$$

The total sample size at the first time point is therefore given by

$$n_{A1} + n_{B1} = (1 + \lambda^{-1})(\psi_A + \lambda \psi_B)(Z_{1-\alpha/2} + Z_P)^2 \frac{\sigma_J^2}{\delta_J^2}.$$

The **longpower** package

The aforementioned sample size approaches have been implemented in an R package, **longpower** ([Donohue and Edland, 2020](#)), that can be found on CRAN via the URL <https://cran.r-project.org/web/packages/longpower/index.html>. The package also contains functions which translate pilot mixed effect model parameters (e.g. random intercept and/or slope) into marginal model parameters

so that the formulas of Diggle et al. (2002) or Liu and Liang (1997) or Lu et al. (2008) formula can be applied to produce sample size calculations for two sample longitudinal designs assuming known variance.

A web-based app

The interactive Shiny (Chang et al., 2021) application available from the URL <https://atrichub.shinyapps.io/power/> is an interface to the **longpower** package 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 similarly implements the sample size formula of Liu and Liang (1997) and Diggle et al. (1994, 2002) using functions in the **longpower** package. A novel feature of the app is that it can generate required pilot estimates by sourcing data from the Alzheimer's Disease Neuroimaging Initiative (ADNI).

3 Illustrations

Illustration of the longpower package

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 uses of the data for advancing knowledge for AD trials (Weiner et al., 2015), this big data resource can be used to improve study design. Specifically, in this paper, the data is used to generate pilot estimates for the computation of sample size and power.

We consider an Alzheimer's disease example using ADAS-Cog (Rosen et al., 1984) pilot estimates from the ADNI database. Suppose that we want to compute the sample size required to detect an effect of 1.5 at the 5% level of significance and 80% power. The LMM fit to ADNI data has an estimated variance of random slope for group A (placebo or control group) of 74, and a residual variance of 10. Assuming study visits at (0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50) years, the sample size using the 'edland' approach can be obtained by using the `edland.linear.power()` functions in the **longpower** R package as follows.

```
> t = seq(0,1.5,0.25)
> edland.linear.power(delta=1.5, t=t, sig2.s = 24, sig2.e = 10, sig.level=0.05,
power = 0.80)
```

Zhao and Edland, in process

```
N = 414.6202
n = 207.3101, 207.3101
delta = 1.5
t = 0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50
p = 0, 0, 0, 0, 0, 0, 1
p_2 = 0, 0, 0, 0, 0, 0, 1
sig2.int = 0
sig.b0b1 = 0
sig2.s = 24
sig2.e = 10
sig2.int_2 = 0
sig.b0b1_2 = 0
sig2.s_2 = 24
sig2.e_2 = 10
sig.level = 0.05
power = 0.8
alternative = two.sided
```

NOTE: N is *total* sample size and n is sample size in *each* group

An alternative approach is to use `lmpower()` and specify the argument `method="edland"` in the following way.

```
> lmpower(delta=1.5, t=t, sig2.s = 24, sig2.e = 10, sig.level=0.05,
power = 0.80,method="edland")
```

Zhao and Edland, in process

```
N = 414.6202
n = 207.3101, 207.3101
delta = 1.5
t = 0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50
p = 0, 0, 0, 0, 0, 0, 1
p_2 = 0, 0, 0, 0, 0, 0, 1
sig2.int = 0
sig.b0b1 = 0
sig2.s = 24
sig2.e = 10
sig2.int_2 = 0
sig.b0b1_2 = 0
sig2.s_2 = 24
sig2.e_2 = 10
sig.level = 0.05
power = 0.8
alternative = two.sided
```

NOTE: N is *total* sample size and n is sample size in *each* group

The Diggle and Liu & Liang approaches can be applied with the `diggle.linear.power()` and `lui.liang.linear.power()` functions, respectively. The `lmpower()` functions can be used for either approach with the appropriate specification of the `'method'` argument.

The second illustrative example is the hypothetical clinical trial discussed in Diggle et al. (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 participants between a control and active treatment group to have equal size. Three visits are envisaged with assessments planned at years 0, 2, and 5. Thus, $n = 3$ and $s_x^2 = 4.22$. Assuming a type I error rate of 5%, power of 80% and smallest meaningful difference of 0.5 mmHg/year, we want to determine the number of participants needed for both treated and control groups for varying correlations (0.2, 0.5 and 0.8) and response variance (100, 200, 300).

```
> 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[1]))))
> colnames(tab) = paste("sigma2 =", sigma2)
> rownames(tab) = paste("rho =", rho)
> tab
sigma2 = 100 sigma2 = 200 sigma2 = 300
rho = 0.2      313      625      938
rho = 0.5      196      391      586
rho = 0.8       79      157      235
```

The above code reproduces the table on page 29 of Diggle et al. (2002). We also reproduce the table on page 30 of Diggle et al. (2002) for detecting a difference in average response between two groups through the method by Liu and Liang (1997) as follows:


```

> u = list(u1 = rep(1,n), u2 = rep(0,n)) #a list of covariate vectors or
  matrices associated with the parameter of interest
> v = list(v1 = rep(1,n), v2 = rep(1,n)) #a respective list of covariate
  vectors or matrices associated with the nuisance parameter
> rho = c(0.2, 0.5, 0.8) #correlations
> delta = c(20, 30, 40, 50)/100 #effect size
> tab = outer(rho, delta, Vectorize(function(rho, delta){
+   ceiling(liu.liang.linear.power(
+     delta=delta, u=u, v=v,
+     sigma2=1,
+     R=rho, alternative="one.sided",
+     power=0.80)$n[1]))))
> colnames(tab) = paste("delta =", delta)
> rownames(tab) = paste("rho =", rho)
> tab
delta = 0.2 delta = 0.3 delta = 0.4 delta = 0.5
rho = 0.2      145      65      37      24
rho = 0.5      207      92      52      33
rho = 0.8      268     120      67      43

```

The sample size formula for the MMRM approach is also implemented in the **longpower** package's `power.mmrn()` function. To illustrate how this approach is implemented, consider a hypothetical example with a correlation matrix having 0.25 as off-diagonal entries (exchangeable), a retention vector (1, 0.90, 0.80, 0.70) and standard deviation of 1, for group A. Assuming these values to be the same for group B, then the sample size required to detect an effect size of 0.5 at 5% level of significance and 80% power is computed as follows:

```

> Ra <- matrix(0.25, nrow = 4, ncol = 4)
> diag(Ra) <- 1 #exchangeable correlation matrix for group A
> ra <- c(1, 0.90, 0.80, 0.70)#retention in group A
> sigmaa <- 1 #standard deviation for group A
> power.mmrn(Ra = Ra, ra = ra, sigmaa = sigmaa, delta = 0.5, power = 0.80)

```

Power for Mixed Model of Repeated Measures (Lu, Luo, & Chen, 2008)

```

n1 = 86.99175
n2 = 86.99175
retention1 = 1.0, 0.9, 0.8, 0.7
retention2 = 1.0, 0.9, 0.8, 0.7
delta = 0.5
sig.level = 0.05
power = 0.8
alternative = two.sided

```

Suppose the allocation ratio is 2, then the function argument `lambda=2` can be added.

Illustration of the Shiny app

We demonstrate how the app is used to perform power analysis by inputting user-specified values and a specific sample size approach. We make use of the values shown in Table 1. For the MMRM method, we specify the options assuming an exchangeable correlation structure.

For the ADNI-based pilot estimate generator, we select the full range of values for some selected variables (See Figure 3).

App menus and contents

The app has three main menus on the sidebar. The first menu has two sub-menus which let the user perform power and sample size calculations based on longitudinal designs when necessary pilot estimates are known. The first submenu, the default, accepts user-specified inputs and generates a graphical output and summary of the inputs. In the 'input' box, as shown in Figure 1a, users are aided with the selection of their inputs through widgets such as select input, numerical input, slider input, and radio button to facilitate the user selection. The select input for 'Analysis Type' allows the user to select whether the user desires to determine power or sample size. Input parameters that are not

Table 1: Parameter values considered in Shiny app demonstrations. Each row represents a different parameter and the values considered for the linear mixed model (LMM, left column) and the mixed model for repeated measures (MMRM, right column) demonstrations. The "-" indicates the parameter is "not applicable" for the given model.

Parameter	LMM	MMRM
Start time	0	1
End time	1.5	4
Timestep	0.5	1
Type I error rate	0.05	0.05
Effect size	1.5	-
Estimate of variance of random intercept	55	-
Estimate of variance of random slope	22	-
Estimate of covariance of random intercept and slope	29	-
Estimate of the error variance	10	-
Standard deviation of observation in group A	-	1
Standard deviation of observation in group B	-	1
Exchangeable correlation	-	0.25
Allocation ratio	1	2

applicable for a given selection or are not applicable for the selected sample size method are grayed out to enhance the user experience. For example, the allocation ratio input widget is grayed out when the 'diggle' and 'liuliang' procedures are selected. The 'output' box, shown in Figure 1b, displays a graph of power versus sample size, a note on the method for the sample size computation, and a summary of the selected inputs by the user. The second submenu enables the user to conduct power analysis based on the MMRM methodology. Widgets such as select inputs, slider inputs, numerical inputs, and radio buttons display the current values of the model parameter in the 'input' box (See Figure 2a). As the MMRM requires the specification of an association structure and a vector of the retention rates, the app display vectors, and matrix inputs widgets. The size of these widgets depends on users' choice of the number of time points for the study. However, these widgets are not reactive, and therefore the user must use the "Update/Enter" action button when changes are made to the number of time points to update the vector and matrix input widgets. The corresponding graphical and summary outputs are reactively displayed within the 'output' box, immediately below the 'input' box (See Figure 2b).

The second menu item also has two sub-menus for the LMM and the MMRM methods, respectively. In this menu, the app enables the user to generate pilot estimates from the Alzheimer's Disease Neuroimaging Initiative (ADNI) data that is fed into the application to perform the power calculations.

In the box for 'baseline selection', the user can select variables that define their study population. By default, all variables are selected, utilizing all the data in the ADNI. A variable can be deselected from the left side of the box with a click. After the user selection of the variable, the user can submit by clicking on the 'submit selected criteria' action button to activate the corresponding widgets in the next box. The 'Inclusion/exclusion criteria' box is made of a slider and select input widgets which allow the user to select a range of values for the selected population characteristics (See Figure 3). Baseline summaries are produced by the app according to the selections of the user. Summaries by gender, education, ethnicity, and race are provided for the selected population. For continuous variables, the number of observations, number of missing observations, mean, median, lower and upper quantiles are displayed, while for categorical variables, the total and level-specific number of observations, and percentages are displayed. These bivariate summaries give the user a clear presentation of the data per the selected inclusion and exclusion criteria.

Next, the user can choose a primary outcome from among some that are commonly used for Alzheimer's disease studies. Pilot estimates are obtained from fitting a LMM for the selected outcome for the power analysis. The estimates of the model can be adjusted with some user-selected covariate options. Data on the selected outcome from baseline to the selected number of years of the study are presented by individual-level and smooth mean profile plots. The app allows the selection of options for selecting the sample size method, type of test, type I error rate, percentage change, and allocation ratio as inputs for the power analysis (see Figure 4). Finally, the graphical output for the power analysis and summary of the inputs used are displayed in the 'output' box. The summary of the LMM is also shown (see Figure 5).

The interface of the second sub-menu is very similar to the first except that the sample size methodology is based on the MMRM. Additionally, the model fitting options include specifying an

association structure, allocation ratio, and percentage retention for the two groups.

The final menu on the sidebar is the 'About' menu, which provides brief information of the dashboard, data description, and acknowledgment of the ADNI resources, packages used for developing the dashboard, and contact information of the developers of the dashboard.

4 Discussion

In this manuscript, we have presented the **longpower** R package, and a Shiny app dashboard that facilitates sample size and power analysis for a longitudinal study design with two-group comparisons of a continuous outcome. The app implements the sample size formulas of [Liu and Liang \(1997\)](#), [Diggle et al. \(1994, 2002\)](#), [Lu et al. \(2008\)](#), and [Ard and Edland \(2011\)](#) using functions in the **longpower** package. The package also handles models in which time is treated either as continuous (e.g. with random intercepts and slopes) or categorical (MMRM).

The **longpower** package was created to allow R users easy access to sample size formulas for longitudinal data that were already available in the literature. Many of the earlier papers on the topic provided no software, and so considerable effort was required by each reader to program implementations of the formulas. Collecting these formulas into an R package makes the methods more accessible and easy to compare. The package includes unit tests to ensure the software can adequately reproduce published results, and alternative approaches for the same study design are validated against each other.

A novel feature of the app is the ability to source pilot data for Alzheimer's disease trials to generate required parameter estimates. We focus on Alzheimer's data as our primary area of interest, but future work could bring in data from other disease areas. Other future directions include accommodating other outcome types, and keeping up with the evolving landscape of model parameterizations.

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 Research 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.

Bibliography

- M. C. Ard and S. D. Edland. Power calculations for clinical trials in alzheimer's disease. *J Alzheimers Dis.*, 26(Suppl 3):369–377, 2011. doi: [doi:10.3233/JAD-2011-0062](https://doi.org/10.3233/JAD-2011-0062). [p266, 272]
- X. Basagaña, X. Liao, and D. Spiegelman. Power and sample size calculations for longitudinal studies estimating a main effect of a time-varying exposure. *Statistical methods in medical research*, 20:5, 2011. doi: <https://doi.org/10.1177/0962280210371563>. [p264]

- W. Chang, J. Cheng, J. J. Allaire, C. Sievert, B. Schloerke, Y. Xie, J. Allen, J. McPherson, J. Dipert, and B. Borges. *shiny: Web Application Framework for R*, 2021. URL <https://CRAN.R-project.org/package=shiny>. R package version 1.6.0. [p268]
- P. Diggle, K.-Y. Liang, and S. L. Zeger. *Analysis of Longitudinal Data*. Oxford University Press, New York, 1994. [p265, 268, 272]
- P. Diggle, P. Heagerty, K. Liang, and S. Zeger. *Analysis of longitudinal data. Second Edition*. Oxford Statistical Science Series, 2002. [p265, 266, 268, 269, 272]
- X. Dong, X. Li, T.-W. Chang, S. T. Weiss, and W. Qiu. '*powerEQTL*': Power and Sample Size Calculation for Bulk Tissue and Single-Cell eQTL Analysis, 2021. URL <https://cran.r-project.org/web/packages/powerEQTL/powerEQTL.pdf>. R package version 0.3.4. [p265]
- M. C. Donohue and S. D. Edland. '*longpower*': Sample Size Calculations for Longitudinal Data, 2020. URL <https://cran.r-project.org/web/packages/longpower/longpower.pdf>. R package version 1.0-21. [p267]
- G. M. Fitzmaurice, N. M. Laird, and J. Ware. *Applied Longitudinal Analysis*. John Wiley & Sons, Hoboken, N.J., 2004. [p266]
- P. Green and C. J. MacLeod. Simr: an r package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7:493–498, 2016. doi: doi:10.1111/2041-210X.12504. [p265]
- X. Guo and W. D. Johnson. Sample size for experiments with multi-variate repeated measures. *J. Biopharmaceutical Statistics*, 6:155–176, 1996. [p264]
- Y. Guo, H. Logan, and D. e. a. Glueck. Selecting a sample size for studies with repeated measures. *BMC Medical Research Methodology*, 13:100, 2013. doi: <https://doi.org/10.1186/1471-2288-13-100>. [p264]
- D. Hedeker, R. D. Gibbons, and C. Waternaux. Sample size estimation for longitudinal designs with attrition: comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics*, 24:70–93, 1999. [p264]
- K. Hemming, J. Kasza, R. Hooper, and M. Forbes, A. Taljaard. A tutorial on sample size calculation for multiple-period cluster-randomized parallel, cross-over, and stepped-wedge trials using the shiny crt calculator. *International Journal of Epidemiology*, 49(3):979–995, 2020. [p265]
- N. Hu and Z. Qu. *Mixed Model Power & Size*, 2021. URL https://rcrm-power-size.shinyapps.io/rcrm_power_size/. a Shiny application accessed on 18/11/2021. [p265]
- N. Hu, H. Mackey, and R. Thomas. Power and sample size for random coefficient regression models in randomized experiments with monotone missing data. *Biometrical Journal*, 2021:1–9, 2021. doi: DOI:10.1002/bimj.202000184. [p267]
- K. Kleinman. '*clusterPower*': Power Calculations for Cluster-Randomized and Cluster-Randomized Crossover Trials, 2021. URL <https://cran.r-project.org/web/packages/clusterPower/clusterPower.pdf>. R package version 0.7.0. [p265]
- F. Klopogge and J. Tarning. '*PharmPow*': Pharmacometric Power calculations for mixed study designs, 2015. URL <https://cran.r-project.org/web/packages/PharmPow/PharmPow.pdf>. R package version 1.0. [p265]
- N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982. [p265]
- P. Lane. Handling drop-out in longitudinal clinical trials: a comparison of the locf and mmrm approaches. *Pharmaceutical Statistics*, 7:93–106, 2008. [p267]
- J. J. Lefante. The power to detect differences in average rates of change in longitudinal studies. *Statistics in Medicine*, 9:437–446, 1990. [p264]
- K.-Y. Liang and S. L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73: 13–22, 1986. [p266]
- A. Liu, W. J. Shih, and E. Gehan. Sample size and power determination for clustered repeated measurements. *Statistics in Medicine*, 21:1787–1801, 2002. doi: <https://doi.org/10.1002/sim.1154>. [p264]

- G. Liu and K. Y. Liang. Sample size calculations for studies with correlated observations. *Biometrics*, 53(3):937–47, 1997. [p266, 268, 269, 272]
- K. Lu, X. Luo, and P. Chen. Sample size estimation for repeated measures analysis in randomized clinical trials with missing data. *The International Journal of Biostatistics*, 4(1):Article 9, 2008. [p264, 265, 267, 268, 272]
- K. Lu, D. V. Mehrotra, and G. Liu. Sample size determination for constrained longitudinal data analysis. *Statistics in Medicine*, 28:679–699, 2009. [p264]
- K. J. Lui. Sample size requirement for repeated measurements in continuous data. *Statistics in Medicine*, 11:633–641, 1992. [p264]
- C. Mallinckrodt, T. Sanger, S. Dubé, D. DeBrot, G. Molenberghs, R. Carroll, W. Potter, and G. Tollefson. Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. *Biological Psychiatry*, 53:754–760, 2003. [p265, 267]
- C. H. Mallinckrodt, W. S. Clark, and S. R. David. Accounting for dropout bias using mixed-effects models. *Journal of Biopharmaceutical Statistics*, 11:9–21, 2001. [p265, 267]
- G. A. J. Martin, H. D. Nussey, J. A. Wilson, and D. Reale. Measuring individual differences in reaction norms in field and experimental studies: a power analysis of random regression models. *Methods in Ecology and Evolution*, 2:362–374, 2011. [p265]
- NCSS Statistical Software. *Power Analysis & Sample Size*, 2021. URL <https://www.ncss.com/software/pass/pass-documentation/>. NCSS, LLC, Kaysville, Utah, USA. [p265]
- nQuery. *Sample Size and Power Calculation*, 2021. URL <https://www.statsols.com/nquery/sample-size-software-for-teams>. ‘Statsols’ (Statistical Solutions Ltd), Cork, Ireland. [p265]
- J. E. Overall and S. R. Doyle. Estimating sample size for repeated measurement designs. *Controlled Clinical Trials*, 15:100–123, 1994. [p264]
- H. Pan, S. Liu, D. Miao, and et al. Sample size determination for mediation analysis of longitudinal data. *BMC Medical Research Methodology*, 18:32, 2018. doi: <https://doi.org/10.1186/s12874-018-0473-2>. [p264]
- W. Qiu. ‘powerMediation’: Power/Sample Size Calculation for Mediation Analysis, 2021. URL <https://cran.r-project.org/web/packages/powerMediation/powerMediation.pdf>. R package version 0.3.4. [p265]
- J. Rochon. Sample size calculations for two-group repeated-measures experiments. *Biometrics*, 47:1383–1398, 1991. [p264]
- W. Rosen, R. Mohs, and K. Davis. Alzheimer’s Disease Assessment Scale—Cognitive and Non-Cognitive sections (ADAS-Cog, ADAS Non-Cog). *Journal of Psychiatry*, 141:1356–1364, 1984. [p268]
- A. Roy, D. K. Bhaumik, S. Aryal, and G. R. D. Sample size determination for hierarchical longitudinal designs with differential attrition rates. *Biometrics*, 63:699–707, 2007. doi: <https://doi.org/10.1111/j.1541-0420.2007.00769.x>. [p264]
- A. M. Schoemann, A. J. Boulton, and S. D. Short. Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science*, 8:379–386, 2017. [p265]
- D. A. Schoenfeld. ‘LPower’: Calculates Power, Sample Size, or Detectable Effect for Longitudinal Analyses, 2019. URL <https://cran.r-project.org/web/packages/LPower/LPower.pdf>. R package version 0.1.1. [p265]
- E. F. Vonesh and M. A. Schork. Sample sizes in multivariate analysis of repeated measurements. *Biometrics*, 42:601–610, 1986. [p264]
- M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, J. Cedarbaum, M. C. Donohue, R. C. Green, D. Harvey, C. R. Jack Jr, and others. Impact of the alzheimer’s disease neuroimaging initiative, 2004 to 2014. *Alzheimer’s & Dementia*, 11(7):865–884, 2015. [p265, 268]
- X. Yan and X. Su. Sample size determination for clinical trials in patients with nonlinear disease progression. *Journal of Biopharmaceutical Statistics*, 16(1):91–105, 2006. doi: 10.1080/10543400500406579. [p264]

- O. Yi and T. Panzarella. Estimating sample size for tests on trends across repeated measurements with missing data based on the interaction term in a mixed model. *Controlled Clinical Trials*, 23:481–496, 2002. [p265]
- Z. Zhang, Y. Mai, and M. Yang. 'WebPower': *Basic and Advanced Statistical Power Analysis*, 2021. URL <https://webpower.psychstat.org>. R package version 0.6. [p265]
- Y. Zhao and S. D. Edland. Power formulas for mixed effects models with random slope and intercept comparing rate of change across groups. *International Journal of Biostatistics*, 2020:20200107, 2020. doi: <https://doi.org/10.1515/ijb-2020-0107>. [p266]

Samuel Iddi
Department of Statistics and Actuarial Science
University of Ghana
Legon-Accra, Ghana
<https://orcid.org/0000-0002-2366-2774>
siddi@ug.edu.gh; isamuel.gh@gmail.com

Data Measurement and Evaluation Unit
African Population and Health Research Center (APHRC)
Nairobi, Kenya
siddi@aphrc.org

Michael C Donohue
Alzheimer's Therapeutic Research Institute
Keck School of Medicine
University of Southern California
San Diego, USA
<https://orcid.org/0000-0001-6026-2238>
mdonohue@usc.edu

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

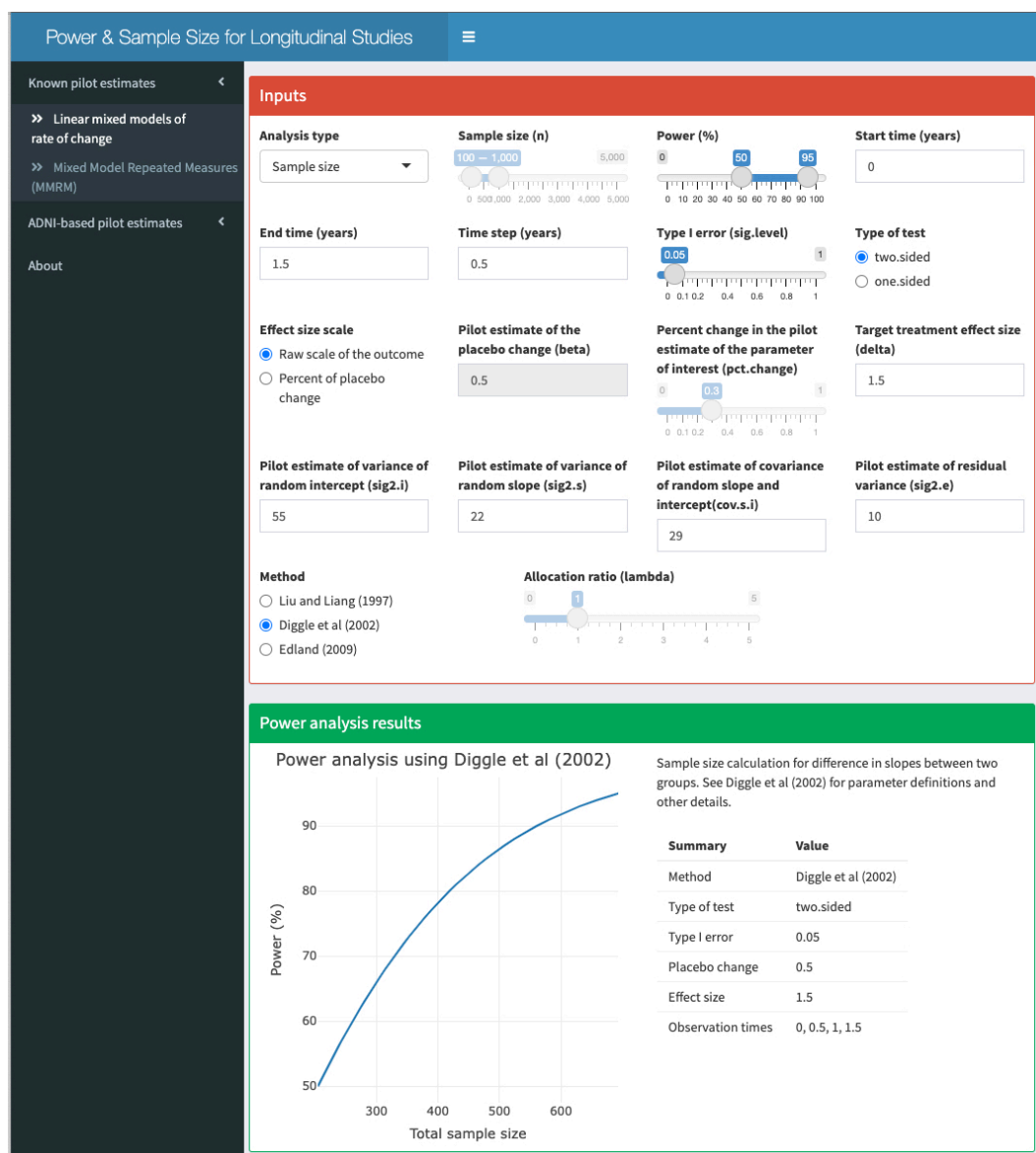


Figure 1: Screenshot of Shiny app menu for linear mixed models (LMM) of rates of change. The red box (top) denotes an area where user input of pilot estimates and study design features is required. Output, including a power curve and a summary table of study design characteristics, is displayed in the green box (bottom).

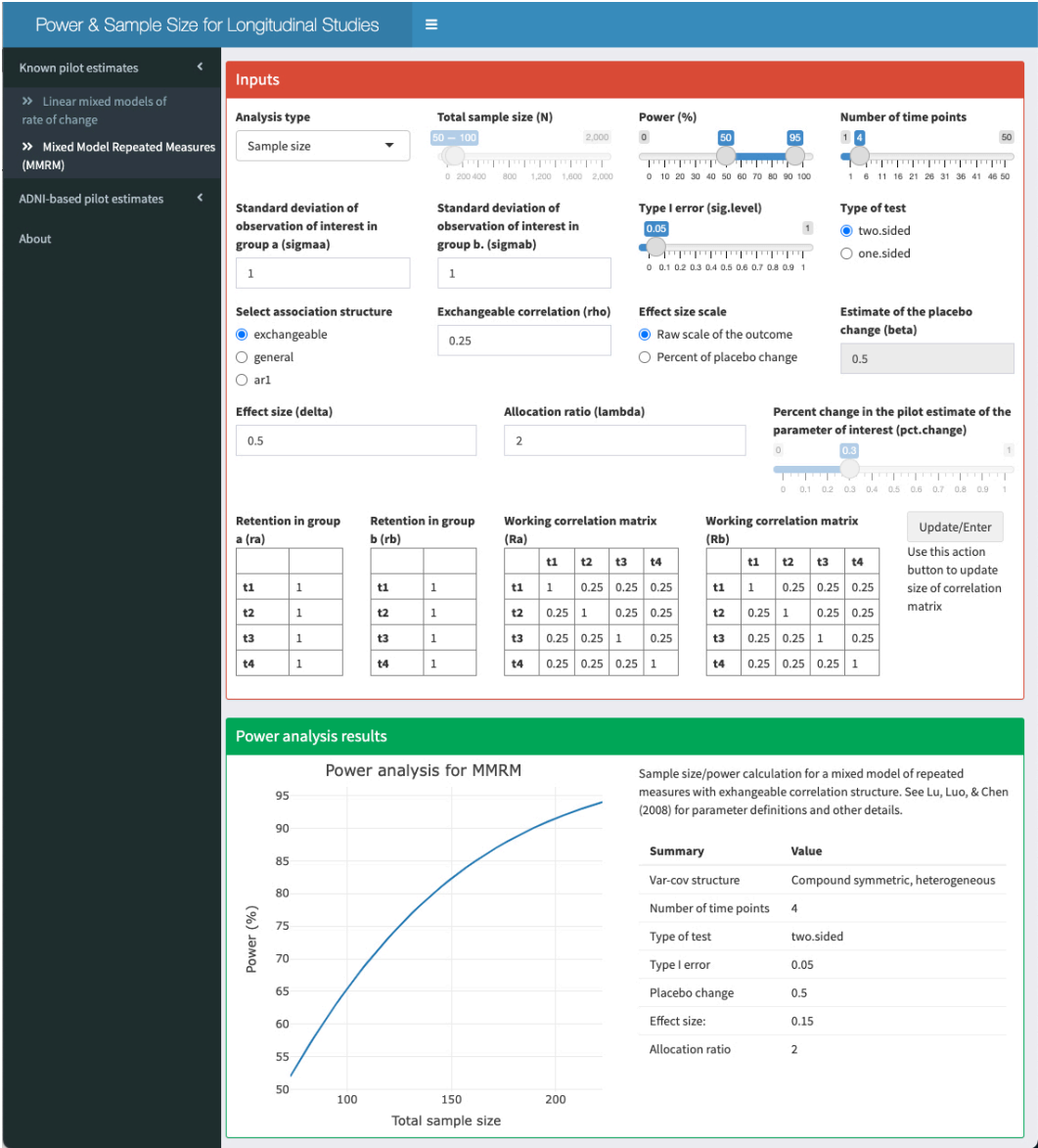


Figure 2: Screenshot of Shiny app menu for mixed models for repeated measures (MMRM). The red box (top) denotes an area where user input of pilot estimates and study design features is required. Output, including a power curve and a summary table of study design characteristics, is displayed in the green box (bottom).

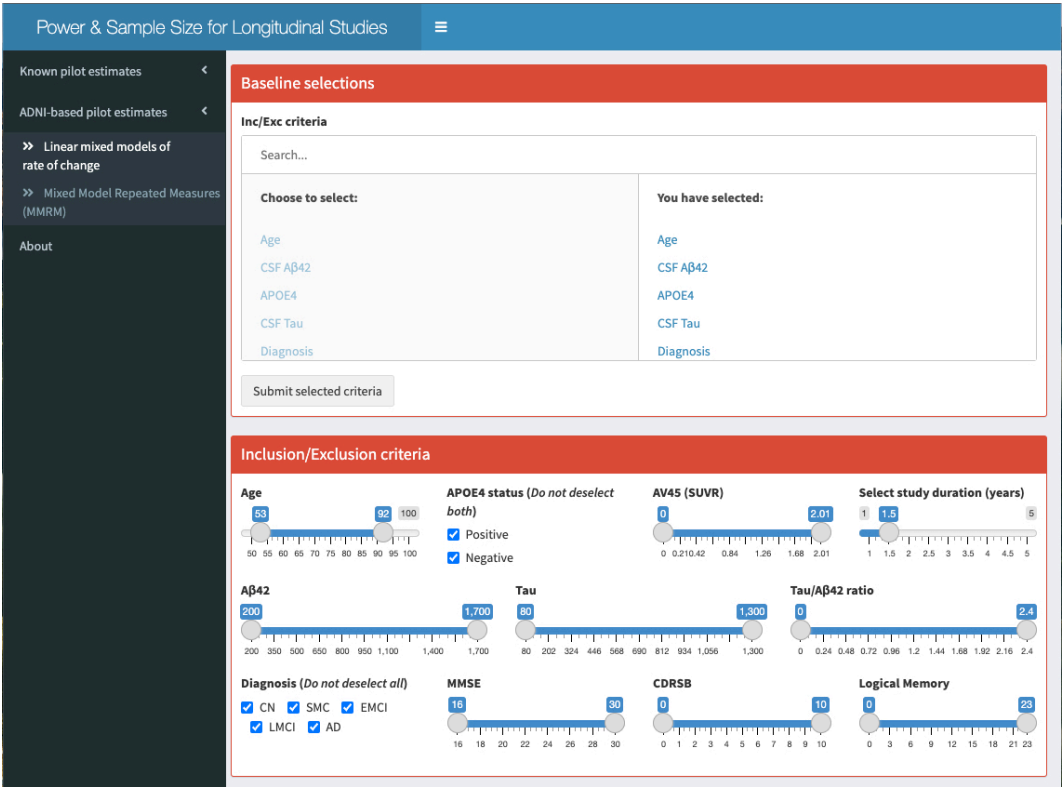


Figure 3: Screenshot of Shiny app menu for ADNI-based pilot estimate generator. The user can select variables and associated ranges for desired inclusion/exclusion criteria. The app then filters the ADNI data accordingly, summarizes the characteristics of the filtered data, fits a linear mixed model (LMM), and extracts pilot estimates required for power calculations.

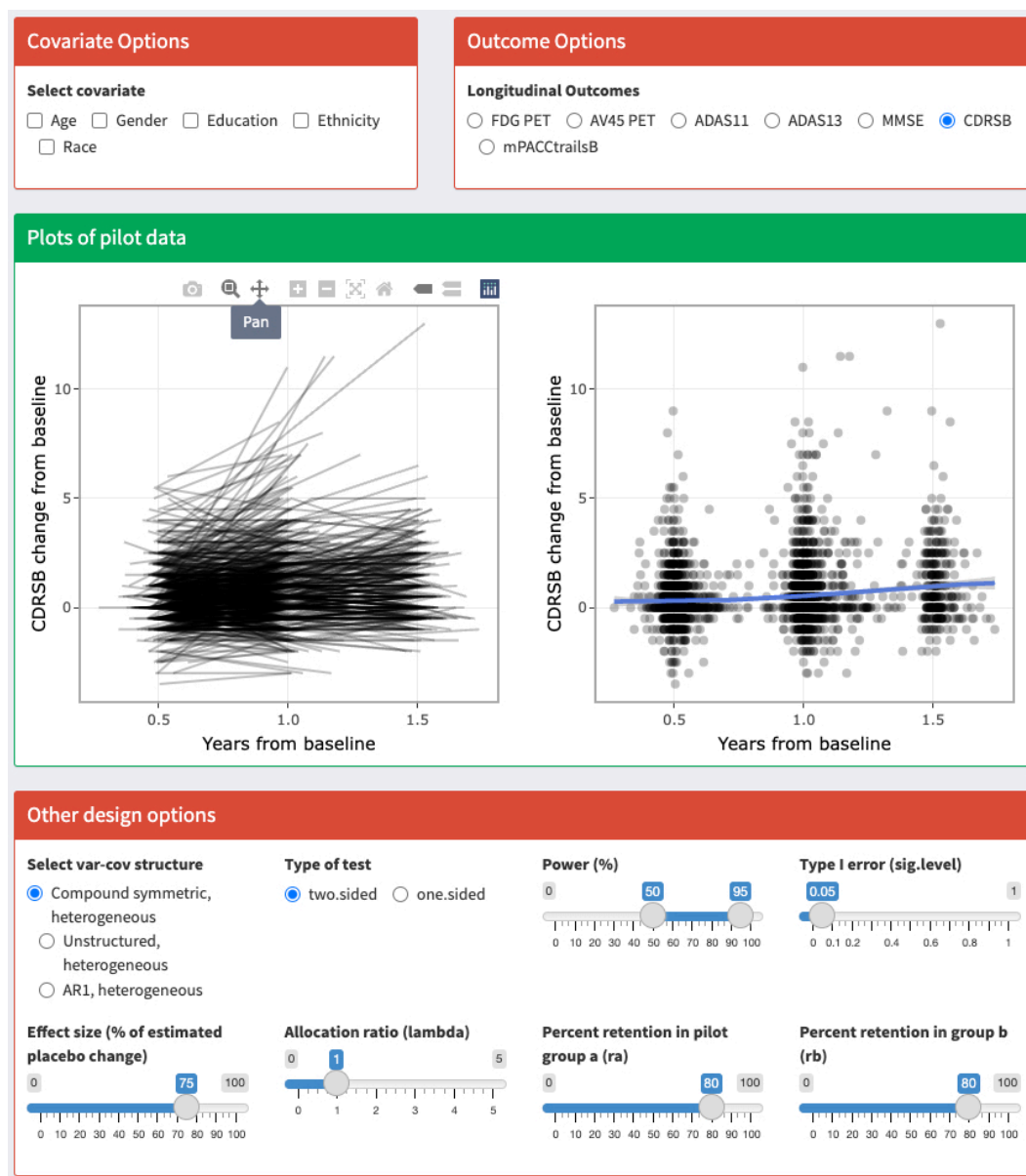


Figure 4: Screenshot of Shiny app input and output for ADNI-based pilot estimate generator. Once the user has selected variables and associated ranges for desired inclusion/exclusion criteria (Figure 3), the app summarizes the characteristics of the filtered data, including these spaghetti and dots plots (middle green box). The user must select the desired analysis model covariates and outcome (top red boxes); and other analysis model and study design features (bottom red box).

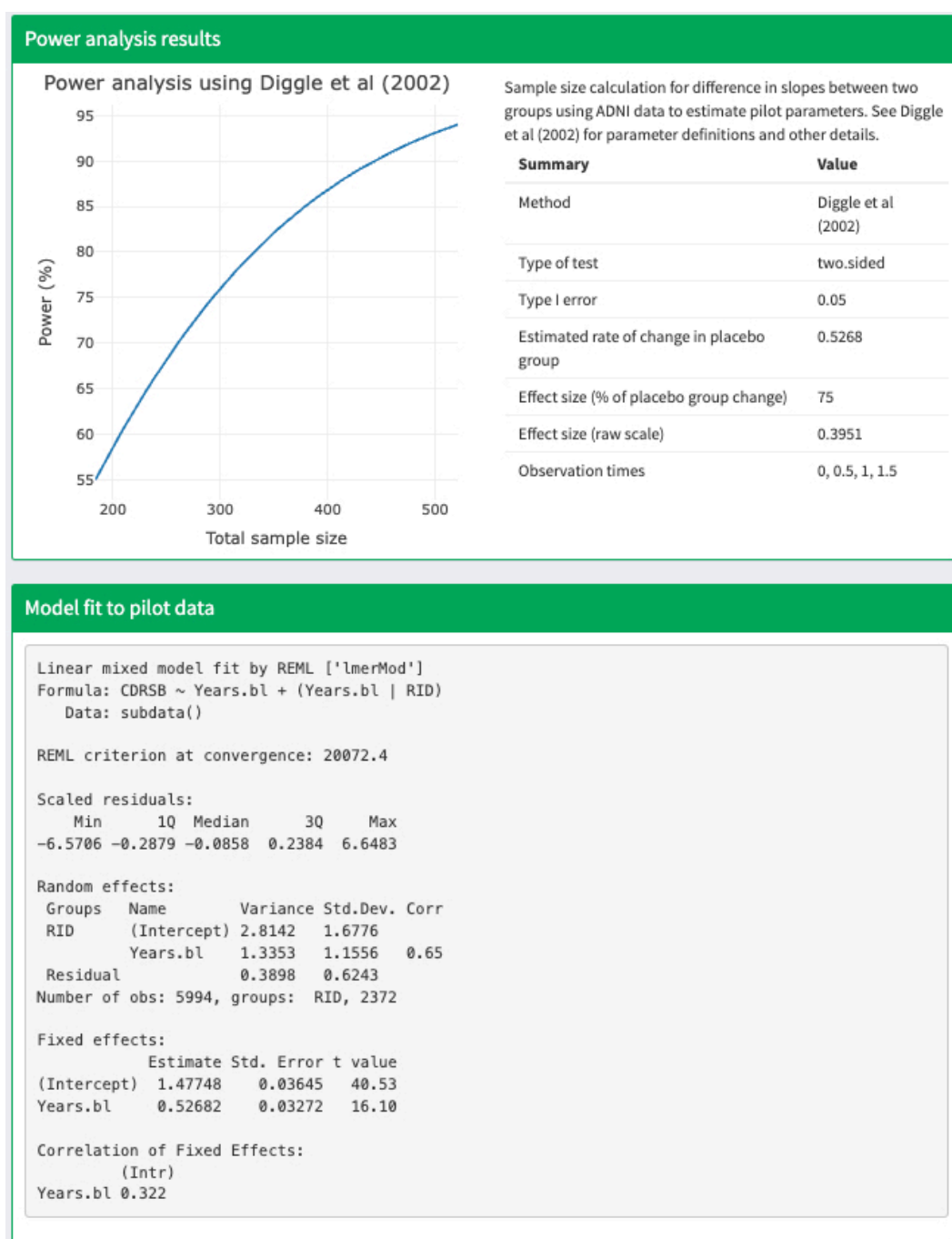


Figure 5: Screenshot of Shiny app output for ADNI-based pilot estimate generator. The bottom box shows R output summarizing a linear mixed model fit to a subset of ADNI data with desired inclusion/exclusion criteria applied. The top box shows study characteristics and a power curve derived using parameters from the fitted model.