Design Issues in Longitudinal Studies

Christopher H. Morrell[1], Veena Shetty[2], and Edward G. Lakatta[3]
[1]Mathematics and Statistics Department, Loyola University Maryland
[2]Medstar Health Research Institute, Hyattsville, MD
[3]Laboratory of Cardiovascular Sciences, Baltimore, MD

Abstract
In designing a longitudinal study one needs to decide on two critical components: duration of study and frequency of visits. In addition, other issues involving sample size, power, number of observations per subject must be addressed. If the study is meant to be completed within a certain time frame, would it better to have a fixed time between observations (which might allow the study to terminate early if its objectives are met) or to spread out the observations over the entire study period? At some point during the study, it may be of interest to see if additional data points would contribute substantially. Assume that the longitudinal data will be analyzed using a linear mixed-effects model. In this investigation we use the standard errors of estimates of model parameters as the criterion. We seek to address the issues using three approaches. First, subsets of a data set are constructed in a number of ways and the standard errors are examined. Second, using a variety of designs, the covariance matrix of the fixed-effects is computed and the standard errors are examined. Finally, a simulation study is conducted.

Key Words: Linear Mixed-Effects Model

1. Introduction

In designing a longitudinal study one needs to decide on two critical components: the duration of the study and the frequency of visits. In addition, other issues involving sample size, power, number of observations per subject must be addressed. When a study is being designed, the researchers might wish the study to be completed within a certain time frame. In this situation a choice could be made between having a fixed time between observations (which might allow the study to terminate early if its objectives are met) or to spread out the observations over the entire study period. In addition, as the study progresses, at some point it may be of interest to see if additional data points would contribute substantially.

A number of authors have considered issues related to planning longitudinal studies. In early work in a series of papers, Schlesselman (1973a, 1973b) addressed sample size determination and frequency of measurements and study duration. The first paper focuses on differences between the means of independent groups so does not address the repeated measures issue inherent to ongoing longitudinal studies. The second paper in the series assumes longitudinal trends can be modeled by a linear function that will be fit separately to the data from each individual. He shows how the standard error of the average of the slopes is related to the variability in slopes among subjects, the error variance, the design points and the number of subjects which, for design purposes, allows one to solve for the number of subjects given values of the other quantities. In a study of bone loss, Davis et al. (1991) use this approach to address the precision of long term bone loss.

Overall (1994) addresses issues in the design of clinical trials where the data will be analyzed using a repeated-measures ANOVA approach. He points out that "increasing the number of equally spaced repeated measurements spanning a specified total treatment period does not generally enhance the probability of detecting a true treatment effect."

Hedeker, Gibbons, and Waternaux (1999) discuss sample size estimation for comparing two groups when there is attrition and the data will be analyzed using a multilevel hierarchical model or mixed effects model. They provide tables of the number of subjects depending on the effect size, number of time points, and the correlation structure for the repeated measurements.

Tu et al. (2004) reviewed and extended power analysis approaches for clustered data analyzed by the generalized estimating equations or linear mixed-effects approaches. They base their formulae on the asymptotic distribution of the model estimates.

Recently Spiegelman and colleagues have published three papers on this topic (Basagaña and Spiegelman (2010); Basagaña, Liao, and Spiegelman (2011); and Barrera-Gomez, Spiegelman, and Basagaña (2013)). They provide a general framework for sample size calculations as well as an approach to obtain the optimal combination of number of repeated measurements and number of subjects. Software (OPTITXS) is available to apply their approach.

Since the study will consist of longitudinal observations on a number of subjects, we assume that the data will be analyzed using a linear mixed-effects model. One issue that must be addressed is how to assess the effect of the number of observations. In this study we choose to consider the standard errors of the longitudinal and cross-sectional effects in the model.

We seek to address the issues using two approaches.

1. Using a variety of designs, the covariance matrix of the fixed-effects is computed and the standard errors are examined.
2. Subsets of a data set are constructed in a number of ways and the standard errors are examined.

2. Methods

The aim of this study is to use longitudinal data sets of different sizes and to examine how the standard errors of the parameter estimates are affected by the particular study design. The linear mixed-effects model is used to model the longitudinal data sets (Verbeke and Molenberghs, 2000). The linear mixed-effects model for the vector of data for the i^{th} individual is

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i,\ i = 1,\ldots,N \tag{1}$$

where $\boldsymbol{y}_i$ is the $n_i{\times}1$ vector of observations for individual i, N is the number of individuals in the study, and $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are the design matrices for the fixed ($\boldsymbol{\beta}$) and random ($\boldsymbol{b}_i$) effects, respectively. It is assumed that the errors $\boldsymbol{e}_i \sim N(\boldsymbol{0},\ \sigma^2\boldsymbol{I})$ are independent of the random effects $\boldsymbol{b}_i \sim N(\boldsymbol{0},\ \boldsymbol{D})$, where $\boldsymbol{D} = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}$ is a $q \times q$ positive definite covariance matrix of the random effects where q is the number of random effects included in the mixed-effects model. The random components, σ^2 and $\boldsymbol{D}$, are estimated by maximum likelihood or restricted maximum likelihood. Once these are random components are estimated the fixed effects are obtained as the generalized least squares estimator,

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i^T V_i^{-1} X_i\right)^{-1} \sum_{i=1}^{N} X_i^T V_i^{-1} y_i$$

where $V_i = \sigma^2 I + Z_i D Z_i^T$ so that the covariance matrix of these estimates is given by

$$Cov(\hat{\beta}) = \left(\sum_{i=1}^{N} X_i^T V_i^{-1} X_i\right)^{-1} .$$

For the purposes of this study we consider a very simple mixed-effects model:
$y_i = \beta_0 + b_{i0} + \beta_1 FAge + (\beta_2 + b_{i1})Time + \varepsilon_i$.
For this model β_1 represents the cross-sectional effect and measures differences between subjects of different ages while β_2 represents the longitudinal effect which measures average changes within individuals over time. We will investigate how the standard errors of these parameters change as the design of the study changes.

2.1 Covariance Matrix Approach

In this study we vary: N (the number of subjects), n_i (the number of observations per subject), the design (the observations are added sequentially with a fixed between observation time or the observations are evenly spaced over the entire observations interval), the values of the variance components (σ^2, d_{11}, d_{22}, and $\rho_{12} = d_{12}\sqrt{d_{11} \times d_{22}}$, the correlation between the random effects).
The design uses the following values for each of these parameters:
N = 50, 100, 200, 500 n_i = 2, 3, 4, 5
σ^2 = 1, 5, 10, 25
d_{11} = 1, 5, 10 d_{22} = 0.5, 1, 2 ρ_{12} = -0.7, -0.4, 0, 0.4, 0.7
This leads to 2880 sets of standard errors which must be examined to determine the relationships of these input variables to the standard errors.

2.2 Data Analysis

A data set is constructed from the Baltimore Longitudinal Study on Aging (BLSA) (Shock et al., 1984). The data set consists of 1809 male BLSA participants with exactly five observations on systolic blood pressure (SBP) with a maximum of eleven years of follow-up from the first to the fifth visit. The simple model described above is fit to various subsets of this data set. One approach is to begin with the first two observations and then sequentially add the 3rd, 4th and 5th observations. The second approach uses data over the maximum study period for each participant. Here we attempt to pick increasing subsets of the data with visits that are approximately evenly spaced throughout the interval. For two observations, we use the first and last observations; for three the 1st, 3rd, and 5th; for four we pick either the 1st, 2nd, 3rd, and 5th or the 1st, 3rd, 4th and 5th picking the additional observation (2nd or 4th) in the larger gap between the 1st and 3rd or 3rd and 5th; for five we use all five observations. The standard errors of the fixed-effects parameter estimates are examined from the models fit to the various subsets of the data.

3. Results

3.1 Covariance Matrix Approach

The standard errors have been computed for all 2880 combinations of the parameters given above and for the two designs. Various plots and descriptive statistics were constructed and computed. For illustration, the Figures 1 to 3, shows the relationships among:

1. the cross sectional standard errors and the correlation between the random effects, the error variance, and the variance of the intercept random effect variance for 50 subjects and two observations within subject and for the two designs.
2. the longitudinal standard errors and the correlation between the random effects, the error variance, and the variance of the longitudinal random effect variance for 50 subjects and two observations within subject and for the two designs.
3. the standard errors for the two designs.

Figure 1 shows that the cross sectional standard error: increases with the error variance, increase with the variance of the intercept random effect, and has a curved association with the correlation among the random effects, particularly for smaller error variances. Figure 2 shows that the longitudinal standard error: increases with the error variance, increase with the variance of the longitudinal random effect, but is unrelated to the correlation among the random effects. Figure 3 shows that the intercept and cross sectional standard errors are highly related for the two designs. However, the longitudinal standard errors are not as closely related but increase with the error variance (from bottom to top) and with the variance of the longitudinal random effect (from left to right).

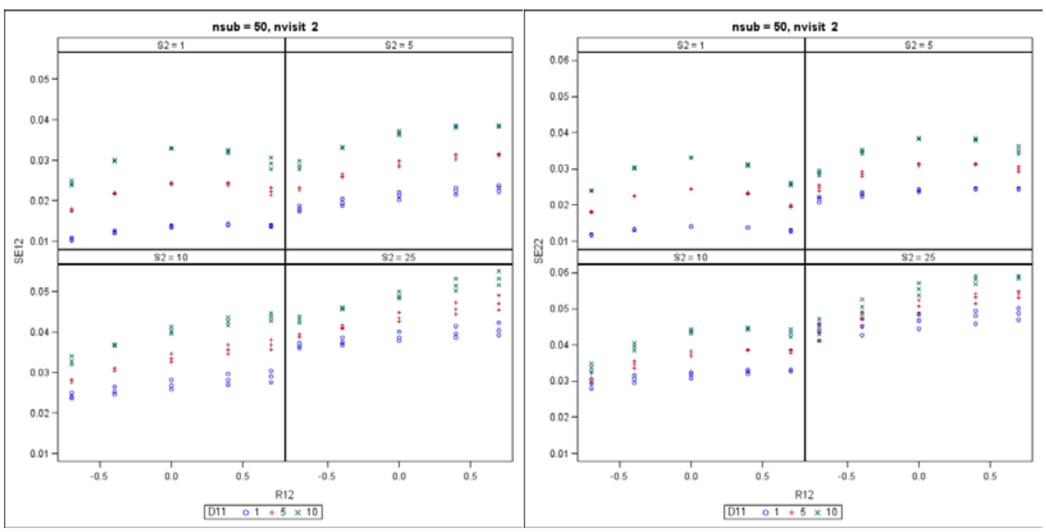

Figure 1. Illustrative plot showing the association between the cross-sectional standard errors (SE12 and SE22) and the correlation between the random effects (R12), the error variance (S2) and the variance of the intercept random effect variance (D11) for 50 subjects and two observations within subject and for the two designs (sequential on the left and spread out on the right).

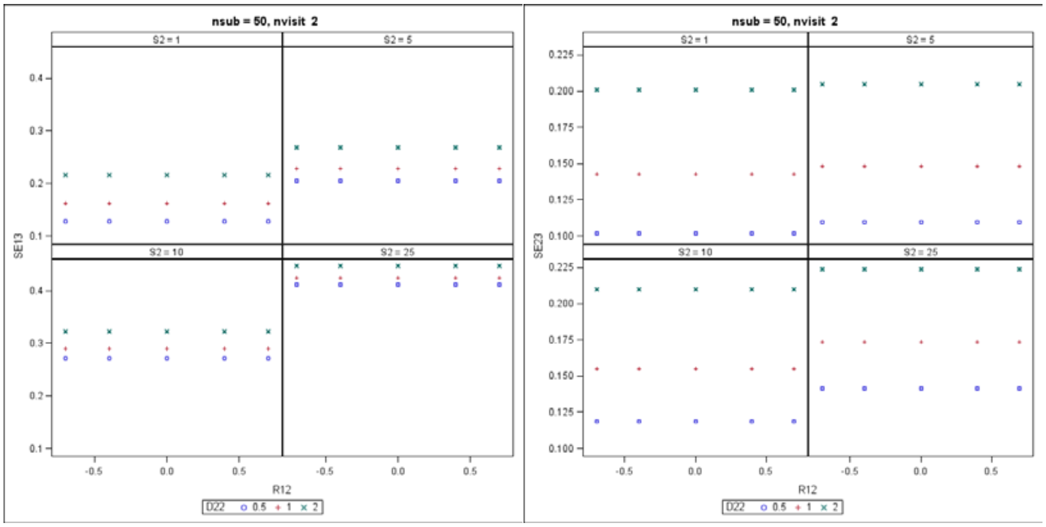

Figure 2. Illustrative plot showing the association between the longitudinal standard errors (SE13 and SE23) and the correlation between the random effects (R12), the error variance (S2) and the variance of the longitudinal random effect variance (D22) for 50 subjects and two observations within subject and for the two designs (sequential on the left and spread out on the right).

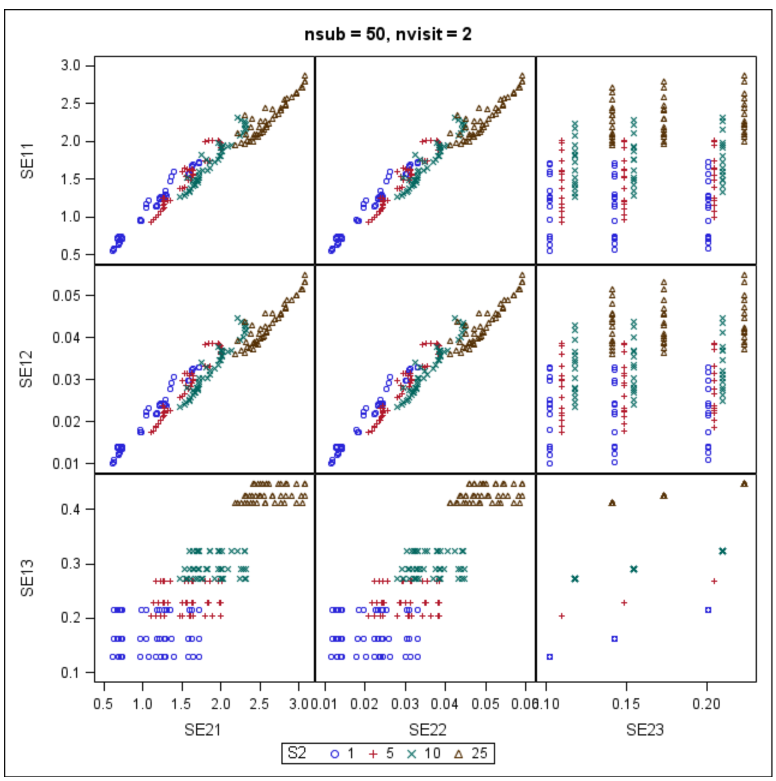

Figure 3. Matrix plot showing associations among standard errors for the two designs: sequential and spread out.

Standard errors are frequently related to the square root of the sample size. Consequently we computed the correlation of the standard errors with a number of factors: the number of subjects (N), the number of observation within subject (n_i), the total number of observations (M = N×n_i), the square roots of N and M, and the variance components, σ^2, d_{11}, d_{22}, and ρ_{12}, the correlation between the two random effects. Table 1 provides the results. Most of the standard errors are most highly correlated with $\sqrt{N}$ except for the standard error for the longitudinal change for the sequential design which is most highly correlated with $\sqrt{M}$. Among the variance components, the error variance has the highest correlation. Interestingly, neither the random intercept variance nor the correlation among the random effects is associated with the longitudinal standard error for either design. This is, perhaps, not surprising as the variability in the intercept should not provide any information about the longitudinal effect. In addition, the variance of the longitudinal random effect is not significantly related to the standard errors for the intercept nor the cross-sectional effect for either design. Again the might be expected as the variability on the longitudinal trajectories should not have an impact on these two terms.

Table 1. Correlations of the standard errors for the two designs (sequential (1) and spread out (2)) with a number of factors. SE11 and SE21 are the standard errors for the intercept, SE12 and SE22 are the standard errors for the cross sectional effect, and SE13 and SE23 are the standard errors for the longitudinal effect.

Pearson Correlation Coefficients, N = 2880
Prob > |r| under H0: Rho=0

	M	N	n_i	$\sqrt{M}$	$\sqrt{N}$	S2	D11	D22	R12
Sequential Design:									
SE11	-0.6107	-0.6671	-0.0604	-0.6714	-0.7087	0.4670	0.3208	0.0282	0.1395
	<.0001	<.0001	0.0012	<.0001	<.0001	<.0001	<.0001	0.1302	<.0001
SE12	-0.6084	-0.6657	-0.0559	-0.6687	-0.7072	0.4605	0.3213	0.0313	0.1546
	<.0001	<.0001	0.0027	<.0001	<.0001	<.0001	<.0001	0.0931	<.0001
SE13	-0.6379	-0.6107	-0.3762	-0.7134	-0.6488	0.3374	0.0000	0.2803	0.0000
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	1.0000	<.0001	1.0000
Spread Out Design:									
SE21	-0.6106	-0.6536	-0.1089	-0.6727	-0.6943	0.5035	0.2831	0.0249	0.1130
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.1808	<.0001
SE22	-0.6089	-0.6515	-0.1096	-0.6709	-0.6921	0.5015	0.2800	0.0275	0.1243
	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.1403	<.0001
SE23	-0.6636	-0.7394	-0.0129	-0.7278	-0.7855	0.1399	0.0000	0.4904	0.0000
	<.0001	<.0001	0.4882	<.0001	<.0001	<.0001	1.0000	<.0001	1.0000

To understand the association between the 2880 standard errors computed for each of the parameters with the various input factors used, the standard errors are modeled as a function of these factors using multiple regression. Since standard errors are frequently related to the square root of the sample size, we include various sample size-related variables in the regressions as follows: $\sqrt{N}$ (or $\sqrt{M}$ for $\hat{\beta}_2$) and n_i. Since the plots of some standard errors vs. the correlation in the random effects covariance matrix exhibited some curvature, ρ_{12}^2 is entertained as a covariate in the regression models. Backward elimination is used to remove statistically nonsignificant factors from the models. At this time, only main effects have been entertained in the multiple regression models – there are no interaction terms.

Table 2 presents the results of the final models. The intercept and cross sectional effect (Table 2a and 2b) exhibit similar results: the standard errors decrease with $\sqrt{N}$ and n_i but increase with the three variance components (σ^2, d_{11}, and d_{22}) and have negative curvature with the correlation (ρ_{12}^2). The magnitudes of the various terms indicate the differences in the rates of change in the standard errors between the two designs. Interestingly, while d_{22} did not have a significant bivariate correlation, after accounting for the other terms in the model it is highly statistically significant.

Table 2c and 2d give the results for the longitudinal standard error. Table 2c includes $\sqrt{N}$ as a factor while Table 2d uses $\sqrt{M}$. For both designs, d_{11} and ρ_{12}, fall out of the models. The longitudinal standard errors decrease with $\sqrt{N}$ and n_i but increase with two of the variance components (σ^2 and d_{22}). When $\sqrt{M}$ is used in place of N, the sequential design leads to similar results. However, for the spread out design, n_i now has a positive regression coefficient. But one cannot interpret this as leading to an increase in the standard errors as one cannot hold $\sqrt{M}$ constant while changing n_i.

Table 2. Multiple regression modeling of the standard errors as a function of input variables

a) se($\hat{\beta}_0$)

Factor	**SEQUENTIAL VISITS**		**VISITS SPREAD OVER STUDY DURATION**	
	Regression Parameter Estimate	p-value	Regression Parameter Estimate	p-value
Intercept	1.477	<0.0001	1.623	<0.0001
n_i	-0.02895	<0.0001	-0.05516	<0.0001
$\sqrt{N}$	-0.06601	<0.0001	-0.06832	<0.0001
σ^2	0.02753	<0.0001	0.03135	<0.0001
d_{11}	0.04671	<0.0001	0.04354	<0.0001
d_{22}	0.02425	<0.0001	0.02265	0.0006
ρ_{12}	0.14662	<0.0001	0.12554	<0.0001
ρ_{12}^2	-0.19234	<0.0001	-0.22905	<0.0001
Model R^2	85.2%		84.7%	

b) se($\hat{\beta}_1$)

Factor	**SEQUENTIAL VISITS**		**VISITS SPREAD OVER STUDY DURATION**	
	Regression Parameter Estimate	p-value	Regression Parameter Estimate	p-value
Intercept	0.0280	<0.0001	0.0311	<0.0001
n_i	-0.00051026	<0.0001	-0.00106	<0.0001
$\sqrt{N}$	-0.00125	<0.0001	-0.00130	<0.0001
σ^2	0.00051703	<0.0001	0.00059763	<0.0001
d_{11}	0.00089096	<0.0001	0.00082418	<0.0001
d_{22}	0.00051326	<0.0001	0.00047762	0.0002
ρ_{12}	0.00310	<0.0001	0.00264	<0.0001
ρ_{12}^2	-0.00408	<0.0001	-0.00483	<0.0001
Model R^2	85.0%		84.5%	

c) se($\hat{\beta}_2$)

Factor	SEQUENTIAL VISITS		VISITS SPREAD OVER STUDY DURATION	
	Regression Parameter Estimate	p-value	Regression Parameter Estimate	p-value
Intercept	0.2581	<0.0001	0.1395	<0.0001
n_i	-0.02486	<0.0001	-0.00056325	0.0482
$\sqrt{N}$	-0.00833	<0.0001	-0.00665	<0.0001
σ^2	0.00274	<0.0001	0.00074983	<0.0001
d_{11}				
d_{22}	0.03321	<0.0001	0.03832	<0.0001
ρ_{12}				
ρ_{12}^2				
Model R^2	75.5%		87.7%	

d) se($\hat{\beta}_2$) using $\sqrt{M}$ in place of $\sqrt{N}$.

Factor	SEQUENTIAL VISITS		VISITS SPREAD OVER STUDY DURATION	
	Regression Parameter Estimate	p-value	Regression Parameter Estimate	p-value
Intercept	0.1971	<0.0001	0.0921	<0.0001
n_i	-0.00929	<0.0001	0.01226	<0.0001
$\sqrt{M}$	-0.00426	<0.0001	-0.00350	<0.0001
σ^2	0.00274	<0.0001	0.00074983	<0.0001
d_{11}				
d_{22}	0.03321	<0.0001	0.03832	<0.0001
ρ_{12}				
ρ_{12}^2				
Model R^2	71.8%		85.9%	

3.2 Data Analysis

The analyses of the SBP measurements from the male BLSA subjects yield the results in Table 3. It is clear that the standard errors decline with increasing n_i. Not surprisingly, the longitudinal standard errors (for $\hat{\beta}_2$) are smaller when the observations are more spread out, though the standard errors for the intercept and cross sectional terms are smaller for sequential visits. The intercept standard errors tend to be slightly smaller for sequential visits though do not change substantially as n_i increases. This is not surprising as with a fixed number of subjects (N), additional longitudinal data does not have much effect on the intercept. Similarly, the standard errors for the cross sectional effects also declines with increasing n_i and are slightly smaller for the sequential visits. As before, once we know the ages of the subjects in the study, additional longitudinal data does not add much to estimating the cross sectional effect which measures differences among subjects of different ages. Finally, the spread out design has much smaller standard errors for the longitudinal effect for few observations though both decline with increasing n_i. The sequential design's longitudinal standard errors decrease more dramatically than the spread out design's longitudinal standard errors which only decrease modestly. Once data has been placed at the extremes, one does not estimate the longitudinal effect much better by including additional intermediate observations. Of course, these additional points will be beneficial for checking whether the model is correctly specified via residual plots.

Table 3. Standard errors for the intercept, cross-sectional effect, and longitudinal effect from the BLSA data as a function of number of repeated measurements.

Number of visits, n_i	**SEQUENTIAL VISITS**			**VISITS SPREAD OVER STUDY DURATION**		
	se($\hat{\beta}_0$)	se($\hat{\beta}_1$)	se($\hat{\beta}_2$)	se($\hat{\beta}_0$)	se($\hat{\beta}_1$)	se($\hat{\beta}_2$)
2	1.56	0.0283	0.264	1.64	0.0297	0.0677
3	1.51	0.0275	0.125	1.57	0.0283	0.0672
4	1.48	0.0268	0.079	1.54	0.0278	0.0654
5	1.46	0.0266	0.062	1.46	0.0266	0.0624

4. Conclusions

As expected, the standard errors of the intercept, cross-sectional, and longitudinal effects all decline with increasing n_i. The longitudinal standard errors are smaller when the observations are more spread out, though the standard errors for the intercept and cross sectional terms are smaller for sequential visits.

When modeling the standard errors as a function of the sample sizes and variance components, all standard errors decline with increasing number of subjects ($\sqrt{N}$) and increasing number of longitudinal observations, and increase with increasing error variance. The intercept and cross sectional standard errors increase with all increasing variance components while the longitudinal standard error increases with increasing variance of the longitudinal random effect. For the intercept and cross sectional standard errors there is a curved association with the correlation between the random effects – the standard errors are largest for zero or low correlations and decrease as the correlation deviates from zero.

Acknowledgement:

This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging. A portion of that support was through a R&D contract with MedStar Research Institute.

References

Barrera-Gomez J, Spiegelman D, and Basagaña X (2013). “Optimal combination of number of participants and number of repeated measurements in longitudinal studies with time-varying exposure” In Press, *Statistics in Medicine*.

Basagaña 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)**, 471-487.

Basagaña X and Spiegelman D (2010). “Power and sample size calculations for longitudinal studies comparing rates of change with a time-varying exposure”. *Statistics in Medicine,* **29**, 181-192.

Davis JW, Ross PD, Wasnich RD, MacLean CJ, and Vogel JM. (1991). "Long-term precision of bone loss rate measurements among postmenopausal women." *Calcif Tissue Int.*, **48(5)**, 311-8.

Hedeker, D., Gibbons, R. D., and Waternaux, C. (1999). "Sample size estimation for longitudinal designs with attrition," *Journal of Educational and Behavioral Statistics*, **24**, 70-93.

Overall JE. (1994). "Issues in the design and analysis of controlled clinical trials." *J Clin Psychol.*, **50(1)**, 95-102.

Schlesselman JJ. (1973a). "Planning a longitudinal study: I. Sample Size Determination." *J Chronic Dis*, **26**, 553–560.

Schlesselman JJ. (1973b). "Planning a longitudinal study: II. Frequency of measurement and study duration." *J Chronic Dis,* **26**, 561–570.

N.W. Shock, R.C. Greulich, R. Andres, E.G. Lakatta, D. Arenberg, and J.D. Tobin (1984). *Normal Human Aging: The Baltimore Longitudinal Study of Aging*. NIH Publication No. 84-2450. Washington, D.C.: U.S. Government Printing Office.

Tu XM, Kowalski J, Zhang J, Lynch KG, and Crits-Christoph P. (2004). "Power analyses for longitudinal trials and other clustered designs," *Statistics in Medicine,* **23(18)**, 2799-2815.

Verbeke, G. and Molenberghs, G., (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.