The Design of Observational Longitudinal Studies

Xavier Basagaña[1,2,3,4], Donna Spiegelman[4,5]

[1]Centre for Research in Environmental Epidemiology (CREAL)

[2] Municipal Institute of Medical Research (IMIM-Hospital del Mar)

[3] CIBER Epidemiologia y Salud Publica (CIBERESP)

[4]Department of Biostatistics, Harvard School of Public Health

[5]Department of Epidemiology, Harvard School of Public Health

Summary

This paper considers the design of observational longitudinal studies with a continuous response and a binary time-invariant exposure, where, typically, the exposure is unbalanced, the mean response in the two groups differs at baseline and the measurement times might not be the same for all participants. We consider group differences that are constant and those that increase linearly with time. We study power, number of study participants (N) and number of repeated measures (r), and provide formulas for each quantity when the other two are fixed, for compound symmetry, damped exponential and random intercepts and slopes covariances. When both N and r can be chosen by the investigator, we study the optimal combination for maximizing power subject to a cost constraint and minimizing cost for fixed power. Intuitive parameterizations are used for all quantities. All calculations are implemented in freely available software.

1   Introduction

Sample size and power calculation in longitudinal studies with continuous outcomes that compare two groups have been considered previously (Yi and Panzarella, 2002; Schouten, 1999; Galbraith and Marschner, 2002; Frison and Pocock, 1992, 1997; Dawson and Lagakos, 1993; Raudenbush and Xiao-Feng, 2001; Overall and Doyle, 1994; Hedeker et al., 1999; Jung and Ahn, 2003; Schlesselman, 1973; Liu and Liang, 1997; Kirby et al., 1994; Rochon, 1998). These publications have based their formulas on several different test statistics, designed to maximize power over several typical hypotheses which arise in longitudinal studies. Most of this previous work was motivated by the design of clinical trials. In an observational setting, study design calculations must be based on tests which remain valid when baseline response levels of the exposed

and unexposed differ and when the exposure prevalence is not 0.5. Although most of the aforementioned formulas can be applied to a non-randomized setting, in-depth investigation of the formulas in settings relevant in observational research is lacking. For example, a study of the behavior of study power as the exposure prevalence deviates from 0.5 in longitudinal designs has not appeared previously. In clinical trials, the time scale of interest is usually time from randomization and the repeated measures are scheduled at a common set of times for all participants – therefore, all the sample size formulas were based on this case. Here, we consider situations where time in the study is not the time variable of interest, but rather age, time since exposure or other measures of time of importance. Design of longitudinal studies is complex, involve the a priori specification of up to ten parameters about which investigators may have little information unless pilot data are available. We therefore formulated intuitive parameterizations to our formulas, using percent changes for the specification of effects, and intuitive covariance parameters for three covariance structures in order to facilitate widespread use in applications.

In addition to exposure prevalence, we studied in detail the effect of the following factors on power: the number of repeated measures; the length of follow-up, the frequency of measurement; the use of age as the time metameter instead of time since randomization. The effect of these parameters on the required number of participants when the number of repeated measures is fixed was also studied. We studied the effect of the covariance parameters on the required number of repeated measures when the number of participants is fixed. Additionally, when the number of study participants (N) and the number of repeated measures (r) can be controlled by the investigator, their optimal combination for maximizing the power to detect a group difference subject to a cost constraint was derived. Cochran (1977) and Raudenbush (1997) examined this problem in the special cases of the alternative hypothesis of a group difference that is constant over time and under compound symmetry. Snijders and Bosker (1993) developed the methodology to obtain the optimal number of par-

ticipants and measurements, (N_{opt}, r_{opt}), subject to a cost constraint, under compound symmetry (CS) and random slopes (RS) covariance structures, for both a group difference that is constant over time (CMD) and for a group by time interaction (LDD). The model upon which they based these developments explicitly separates the between- from the within- subjects effects (B&W), and optimal designs are different than those given in this paper which follow the modeling approaches most commonly used in epidemiology (models (2.3)-(2.8)). In this paper, we will briefly address how design considerations differ for the B&W model from the models considered here, as relevant, and in addition, we extend results to the damped exponential (DEX) covariance and settings where subjects are observed at different times (e.g. when baseline ages vary). Finally, we study in detail for the first time the effect of all the parameters on the resulting optimal combination.

This paper is structured as follows. In section 2, we present intuitively parameterized models for the two alternative hypotheses commonly considered in longitudinal studies, and the test statistics that will serve as the basis for power and sample size calculation for each of them. We show that some of the test statistics that have been considered previously are biased or less efficient in observational (non-randomized) studies, when the expected value of the baseline measures is not equal in the two exposure groups. In section 3, we derive formulas for power and sample size for an arbitrary covariance matrix and study the effect of exposure prevalence. Using an intuitive formulation for the parameters of interest, as well as for the nuisance parameters, we provide the formulas under compound symmetry and study the effect of the covariance between repeated measurements on power. Then, we extend the formulas to other covariance structures by incorporating additional intuitively defined parameters, formulated in a manner that is intended to be accessible to non-statistical investigators and enables the use of existing pilot or published data and when unavailable, intuitive hunches. The effects of departure from compound symmetry on power, sample size and number of repeated measures are studied. In section 4, we

provide methods to find the optimal combination of number of participants and number of repeated measures for maximizing power under a fixed budgetary constraint. In section 5, we explore aspects of the design of an epidemiological study of the effects of cigarette smoking on lung function, based on publicly available data that we will treat as a pilot study. In section 6, we compare the functionality of currently available software for longitudinal study design, discuss their limitations, and introduce our comprehensive software for the observational longitudinal design setting. Finally, in section 7, we summarize our findings on the many different factors that need to be taken into account when planning a longitudinal observational study of a time-invariant exposure.

2 Notation and Preliminary Results

Consider the case where there are two groups, the exposed and the non-exposed. Let N be the total sample size and p_e be the prevalence of exposure. Let Y_{ij} be a normally distributed outcome of interest for the measurement taken at the jth time $(j = 1, \ldots, r)$ for the ith $(i = 1, \ldots, N)$ participant, and let k_i $(k_i = 0, 1)$ be the exposure group for subject i. We consider studies that obtain repeated measures every s time units, as is the usual design in epidemiologic studies. Thus, the total length of follow-up is $\tau = s\,r$. For example, a study that follows participants every 6 months $(s = 6)$ for five years $(\tau = 12*5 = 60$ months) would have 11 measures per participant, one at baseline plus $r = 10$ repeated measures. In epidemiology, there is often interest in the variation of the response by age and not by time in the study. Since participants enter the study at different ages, each participant has a different set of ages, $\mathbf{t}_i$, over which they are observed. Since each measurement is taken every s units, the vector of times is fully defined by the initial time (age at entry) t_{i0}, and then $\mathbf{t}'_i = (t_{i0}, t_{i0}+s, t_{i0}+2s, \ldots, t_{i0}+ r\,s)$. When $V(t_0) = 0$, where $V(t_0)$ is the variance of the primary time metameter of the analysis at baseline, as when using time since enrollment in the study as the time

variable of interest, all participants have the same time vector. We assume a linear form for the mean, $\mathbb{E}\left[\mathbf{Y}_i\right] = \mathbf{X}_i\mathrm{B}$ $(i = 1, \ldots, N)$, where $\mathbf{X}_i$ is the covariate matrix for participant i, and B is the vector of unknown regression coefficients relating the conditional mean of $\mathbf{Y}_i$ to its corresponding covariates; and $Var\left[\mathbf{Y}_i|\mathbf{X}_i\right] = \mathbf{\Sigma}_i$ $(i = 1, \ldots, N)$, where $\mathbf{\Sigma}_i$ is the $(r+1) \times (r+1)$ residual covariance matrix assumed equal for all participants. Note that $\mathbf{\Sigma}_i$ can be any valid covariance matrix, and can include terms associated with between-subjects variability as well as within-subjects variation. The generalized least squares (GLS) estimator of B is

$$\hat{\mathbf{B}} = \left(\sum_{i=1}^{N}\mathbf{X}'_i\mathbf{\Sigma}_i^{-1}\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{X}'_i\mathbf{\Sigma}_i^{-1}\mathbf{Y}_i\right)$$

One way to circumvent the problem of the design matrix not being known a priori in an observational study is to use the asymptotic variance of this estimator. Other possible approaches are discussed in section 7. The asymptotic variance of $\hat{\mathbf{B}}$ is $\frac{1}{N}\mathbf{\Sigma}_{\mathrm{B}}$, where

$$\mathbf{\Sigma}_{\mathrm{B}} = \left(\mathbb{E}_X\left[\mathbf{X}'_i\mathbf{\Sigma}_i^{-1}\mathbf{X}_i\right]\right)^{-1}. \tag{2.1}$$

and, provided $\mathbf{\Sigma}_i$ does not depend on the covariates, this covariance matrix can be fully specified by knowing the first and second order moments of the covariate distribution, and not their full distribution (Tu et al., 2004). We assume that the prevalence of exposure is p_e, the variance of the initial time is $V\left(t_0\right)$ and the correlation between exposure and initial time is ρ_{e,t_0}. We also assume that the variance of the initial time is the same in the two exposure groups.

Our sample size and power equations are based on the Wald test for the coefficient of interest. Thus, our test statistic has the canonical form

$$T = \frac{\sqrt{N}\,\mathbf{c}'\hat{\mathrm{B}}}{\sqrt{\mathbf{c}'\mathbf{\Sigma}_{\mathrm{B}}\mathbf{c}}}, \tag{2.2}$$

where $\mathbf{c}$ is a $(g+1) \times 1$ vector, where g is the number of explanatory variables in the model, with a one and g zeros isolating the particular component of B that is of interest. Our models will be written with the coefficient of interest always the last one, so $\mathbf{c}$ will be of the form $\mathbf{c} = (0, \ldots, 0, 1)$.
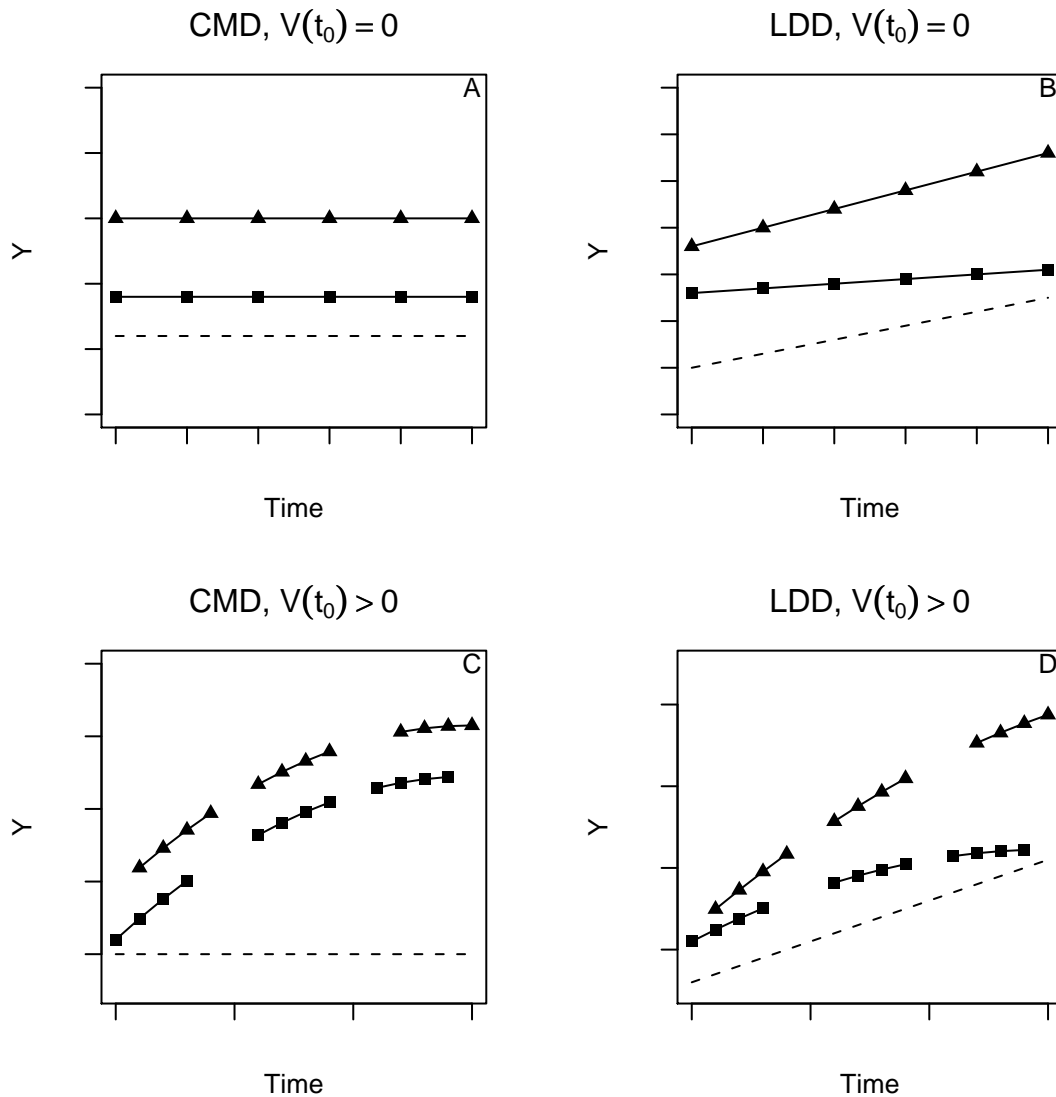
Figure 1: Possible patterns of response under the alternative hypotheses considered in this paper. In A and B, all participants have six measurements at the same time points. In C and D, the graph shows six participants (three exposed and three unexposed) with a total of four measurements each. The different lines represent unexposed (■), exposed (▲) and the difference between exposed and unexposed (- - -).

Power and sample size calculations depend on the alternative hypothesis under consideration. Two patterns have been commonly considered in past literature, and both are quite relevant for applications in epidemiology. We first consider a constant mean difference (CMD) (Frison and Pocock, 1992) between exposed and unexposed (figure 1A for $V(t_0) = 0$ and figure 1C for $V(t_0) > 0$). The CMD alternative hypothesis in the context of clinical trials assumes that the baseline means of the two groups are equal, and then the post-baseline means have a constant difference with respect to time. In observational studies, there is nothing special about baseline, and we need to allow the two groups to differ at baseline as well, as would usually be the case in observational studies. This situation can be modeled as

$$\mathbb{E}\left(Y_{ij}|X_{ij}\right) = \beta_0 + \beta_1 t_{ij} + \beta_2 k_i \tag{2.3}$$

if the effect of time can be considered linear. Our formulas will be based on model (2.3), however conclusions will extend to the more general model

$$\mathbb{E}\left(Y_{ij}|X_{ij}\right) = \beta_0 + f(t_{ij}; \beta_1) + \beta_2 k_i, \tag{2.4}$$

where $f(t_{ij}; \beta_1)$ is an arbitrary function of time, only if time and exposure can be considered independent. The resulting response profiles for the exposed and unexposed are parallel. The null hypothesis of interest is $H_0 : \beta_2 = 0$ vs. the alternative, $H_0 : \beta_2 \neq 0$, and serves as a basis for the test of whether the two response profiles coincide or not. When all participants are observed at the same set of time points, we can include indicators for the $(r + 1)$ time points and write model (2.4) as

$$\mathbb{E}\left(Y_{ij}|X_{ij}\right) = \mu_{0.0} + \mu_{0.1} + \cdots + \mu_{0.r} + \beta_2 k_i, \tag{2.5}$$

where $\mu_{0.0}, \mu_{0.1}, \cdots, \mu_{0.r}$ give the means at times $t_0, \ldots, t_r$ in the unexposed group, and β_2 is the constant difference in response between exposed and unexposed.

In the second pattern, linearly divergent differences (LDD), the effect of exposure varies linearly with time (figure 1B for $V(t_0) = 0$ and figure 1D for $V(t_0) > 0$) (Frison and Pocock, 1997). In clinical trials, the mean of the two groups at baseline are

assumed to be equal. Here, we allow for a baseline difference, as this would usually be the case in observational studies. In the simplest version of LDD, the effect of time is linear in both groups (figure 1B) and can be modeled as

$$\mathbb{E}\left(Y_{ij}|X_{ij}\right) = \gamma_0 + \gamma_1 t_{ij} + \gamma_2 k_i + \gamma_3\left(t_{ij} \times k_i\right), \qquad (2.6)$$

although we can allow the relationship between response and time in the unexposed to be more general, of form

$$\mathbb{E}\left(Y_{ij}|X_{ij}\right) = \gamma_0 + f\left(t_{ij}; \gamma_1\right) + \gamma_2 k_i + \gamma_3\left(t_{ij} \times k_i\right), \qquad (2.7)$$

where $f\left(t_{ij}; \gamma_1\right)$ is now a function of time that includes a linear term but is otherwise arbitrary. The formulas we derive, however, will be valid for model (2.7) only if exposure and time can be considered independent. When all participants are observed at the same set of time points, one can use the model

$$\mathbb{E}\left(Y_{ij}|X_{ij}\right) = \mu_{0.0} + \mu_{0.1} + \cdots + \mu_{0.r} + \gamma_2 k_i + \gamma_3(k_i \times t_j), \qquad (2.8)$$

which accommodates any shape over time. The test of interest is thus $H_0 : \gamma_3 = 0$ vs. the alternative $H_A : \gamma_3 \neq 0$, where γ_3 is the difference in the rates of change of the response over time between the exposed and unexposed, per a single unit of time. Note that if the null hypothesis is accepted there can still be a constant difference at baseline, which persists over time, between the exposed and unexposed, as in CMD. That is, we test whether the two response profiles are parallel or not, or, in other words, whether the effect of time is the same in exposed and unexposed.

Models (2.6)-(2.8) assume that the cross-sectional and longitudinal effects of time coincide. Models that separate the cross-sectional (between-subjects) and the longitudinal (within-subjects) effects have been developed (B&W) (Diggle et al., 2002; Fitzmaurice et al., 2004; Neuhaus and Kalbfleisch, 1998; Ware et al., 1990). Applying those models and allowing different effects for the exposed and the unexposed, one can fit the

following model to the data,

$$\mathbb{E}\left(Y_{ij}|X_{ij}\right) = \eta_0 + \eta_1 t_{i0} + \eta_2 \left(t_{ij} - t_{i0}\right) + \eta_3 k_i$$

$$+ \eta_4 \left(k_i \times t_{i0}\right) + \eta_5 \left(k_i \times \left(t_{ij} - t_{i0}\right)\right) \qquad (2.9)$$

$$= \eta_0 + \eta_1' t_{i0} + \eta_2 t_{ij} + \eta_3 k_i + \eta_4' \left(k_i \times t_{i0}\right) + \eta_5 \left(k_i \times t_{ij}\right),$$

where η_1 is the cross-sectional effect of time in the unexposed, η_4 is the difference in the cross-sectional effect of time between the exposed and the unexposed, η_2 is the longitudinal effect of time in the unexposed, and η_5 is the difference in the longitudinal effect of time between the exposed and the unexposed. The hypothesis to be tested is $H_0 : \eta_5 = 0$ vs. the alternative $H_A : \eta_5 \neq 0$. When there is no confounding, which in this context means that the exposed and unexposed do not differ with respect to the distribution of unmeasured risk factors, and when $V(t_0) = 0$, the distribution of time is the same among the exposed and unexposed, and $\eta_5 = \gamma_3$. Otherwise, they are different, and η_5 will be the parameter of interest in epidemiology. Another possibility is to fit a model for the differences from one visit to the next,

$$\mathbb{E}\left(Y_{i,j} - Y_{i,j-1}|X_{ij}\right) = \lambda_0 + \lambda_1 k_i. \qquad (2.10)$$

We prove in Appendix A.1.4 that $s\hat{\lambda}_1 = \hat{\eta}_5$ and $s^2 Var\left(\hat{\lambda}_1\right) = Var\left(\hat{\eta}_5\right)$, so inferences based on $\hat{\lambda}_1$ or $\hat{\eta}_5$ are equivalent. As relevant, we will discuss the impact on design when (2.9) or (2.10) is to be used instead of (2.6)-(2.8).

Other test statistics have been previously proposed for these settings. Frison and Pocock (1992, 1997) derived the optimal test statistics for both the CMD and the LDD hypothesis in the context of clinical trials, which they called ANCOVA and SLAIN respectively. They used the so called summary measure or two-stage approach to derive their test statistics, and in their derivation of these two tests some adjustment for the baseline response is involved. Since this baseline adjustment may be misleading, we derived in Appendix A.2 the form of the regression models that lead to the same ANCOVA and SLAIN test statistics when a Wald test of the parameter of interest is performed as in (2.2). Because differences at baseline among exposed and unexposed

commonly occur in observational studies, the properties of the test statistics are different in observational studies compared to randomized trials. We show in Appendix A.2 that, even though the ANCOVA test is still unbiased under CMD in an observational study, it is less efficient than the one we use; and that the SLAIN test is biased in general under LDD in an observational study. Therefore, these tests should not be used in observational studies. In the absence of additional model covariates, with all participants observed at the same time points, a two-stage estimator, where a regression of the response vs. time is performed for each participant, and in a second stage, these N independent estimates of the slopes are regressed on exposure, is algebraically identical to the estimating γ_3 from model (2.6) by OLS (Appendix A.3). It turns out that if we can assume a compound symmetry (CS) covariance structure or a random slopes (RS) covariance structure, where random effects are assumed both for the intercept and the time slope, the two stage estimator, and, the OLS estimator are equivalent to the GLS estimator of γ_3 (Appendix A.3). This result does not hold for damped exponential (DEX) correlation structure (Appendix A.3). The two-stage approach and GLS approach do not coincide when participants are observed at different times. Since $\hat{\beta}_2$ and $\hat{\gamma}_3$ in this paper are GLS and therefore are the best linear unbiased estimates for their respective models (2.3)-(2.5) and (2.6)-(2.8) (Searle, 1971), other valid options for the testing of the LDD hypothesis, such as comparing the maximum change over the exposed to the maximum change in the unexposed (Koh-Banerjee et al., 2003), are inefficient and will not be considered further.

The only parameters of the models given above that are needed for power calculations are β_2, γ_3 and η_5. Typically, it is difficult to provide a priori values for these parameters that are realistic and well justified and even more difficult to supply realistic and well justified values for their variance. Therefore, we reparameterized the key parameters of the models above in terms of quantities more likely to be known to the investigator, available from published papers, or easily calculated in pilot data. These parameters are:

1. the mean response at baseline (or at the mean initial time) in the unexposed group (μ_{00}), where $\mu_{00} = \mathbb{E}\left(Y_{i0}|k_i = 0\right)$, $i = 1, \cdots, N$.

2. the percent difference between exposed and unexposed groups (p_1) at baseline (or at the mean initial time), where

$$p_1 = \frac{\mathbb{E}\left(Y_{i0}|k_i = 1\right) - \mathbb{E}\left(Y_{i0}|k_i = 0\right)}{\mathbb{E}\left(Y_{i0}|k_i = 0\right)},$$

$i = 1, \cdots, N$.

3. the percent change from baseline (or from the mean initial time) to end of follow-up (or to the mean final time) in the unexposed group (p_2), where

$$p_2 = \frac{\mathbb{E}\left(Y_{i\tau}|k_i = 0\right) - \mathbb{E}\left(Y_{i0}|k_i = 0\right)}{\mathbb{E}\left(Y_{i0}|k_i = 0\right)},$$

$i = 1, \cdots, N$. For situations where τ is not fixed, p_2 is defined at time s instead of at time τ.

4. the percent difference between the change from baseline (or from the mean initial time) to end of follow-up (or mean final time) in the exposed group and the unexposed group (p_3), where

$$p_3 = \frac{\mathbb{E}\left(Y_{i\tau} - Y_{i0}|k_i = 1\right) - \mathbb{E}\left(Y_{i\tau} - Y_{i0}|k_i = 0\right)}{\mathbb{E}\left(Y_{i\tau} - Y_{i0}|k_i = 0\right)},$$

$i = 1, \cdots, N$. When $p_2 = 0$, p_3 will be defined as the percent change from baseline (or from the mean initial time) to the end of follow-up (or to the mean final time) in the exposed group,

$$p_3 = \frac{\mathbb{E}\left(Y_{i\tau}|k_i = 1\right) - \mathbb{E}\left(Y_{i0}|k_i = 1\right)}{\mathbb{E}\left(Y_{i0}|k_i = 1\right)},$$

$i = 1, \cdots, N$. For situations where τ is not fixed, p_3 is defined at time s instead of at time τ.

5. the residual variance of the response given the covariates, $\sigma^2 = Var\left(Y_{ij}|X_{ij}\right)$. Note importantly that this parameter is not equal to the marginal, cross-sectional

variance of $\mathbf{Y}$. It can be approximated by the variance of the response among the unexposed at baseline, a quantity that may be available from the literature or pilot data, or if age is the time variable of interest, within a reasonably narrow age group. If only a marginal response variance is available over a range of ages and exposure levels, as will often be the case in epidemiology, the investigator can approximate the residual variance by multiplying it by the quantity $1 - R^2$, where R^2 is the assumed proportion of the marginal variance of the response variable that is explained by the model to be fit when the study is conducted (Snijders and Bosker, 1994), here one of models (2.3)-(2.8). Typically, in epidemiology, R^2 ranges from 0.10 to 0.30 or so. Under CMD, the parameter of interest is $\beta_2 = p_1 \mu_{00}$, and under LDD, it is $\gamma_3 = \eta_5 = p_2 p_3 \mu_{00}/\tau$ or, when $p_2 = 0$, it is $\gamma_3 = \eta_5 = (1 + p_1) p_3 \mu_{00}/\tau$. Hence, when CMD is of interest, the investigator needs to specify the alternative through two parameters, p_1 and μ_{00}, and when LDD is of interest, four parameters are needed, p_2, p_3, μ_{00}, and τ. In our experience, investigators can readily provide values, or ranges of values, for these parameters, while it is difficult to directly obtain values for β_2, γ_3, or η_5. It is even more difficult, if not impossible, to obtain a priori values for $Var(\hat{\beta}_2)$, $Var(\hat{\gamma}_3)$ or $Var(\hat{\eta}_5)$, since these quantities depend on σ^2, along with other parameters.

3 Power and Sample Size when either N or r is Fixed

3.1 General case when $\Sigma_i = \Sigma \; \forall i$

We assume that the covariance matrix is the same for all participants, i.e. $\Sigma_i = \Sigma \; \forall i$. Section 3.4 will consider a particular where this is not true. The general power formula associated with the test statistic, T, is

$$\Phi \left[\frac{\sqrt{N} \; |(\mathbf{c}'\mathbf{B})_{H_A}|}{\sqrt{\mathbf{c}'\Sigma_{\mathrm{B}}\mathbf{c}}} \left(-z_{1-\alpha/2} \right] \right(\tag{3.1}$$

where $(\mathbf{c}'\mathrm{B})_{H_A}$ is the value of the regression parameter vector under the alternative (i.e. β_2 for CMD and γ_3 for LDD), α is the significance level, $\boldsymbol{\Sigma}_{\mathrm{B}}$ is defined in equation (2.1), and z_p and $\Phi\left(\cdot\right)$ are the pth quantile and the cumulative density of a standard normal, respectively. From (3.1), it is clear that power will increase as the number of participants, N, increases. From equation (2.1) we can see that the matrix $\boldsymbol{\Sigma}_{\mathrm{B}}$ depends both on the inverse of the residual covariance matrix, $\boldsymbol{\Sigma}^{-1}$, and the covariate matrices $\mathbf{X}_i$. Let $v_{jj'}$ be the (j, j')th element of $\boldsymbol{\Sigma}^{-1}$, and let

$$
\mathbf{A} = \begin{pmatrix} \sum_{j=0}^{r}\sum_{j'=0}^{r} v_{jj'} & \sum_{j=0}^{r}\sum_{j'=0}^{r} j' v_{jj'} \\ \sum_{j=0}^{r}\sum_{j'=0}^{r} j v_{jj'} & \sum_{j=0}^{r}\sum_{j'=0}^{r} j j' v_{jj'} \end{pmatrix}.
$$

Then, under CMD, we show that $\mathbf{c}'\boldsymbol{\Sigma}_{\mathrm{B}}\mathbf{c}$ is (Appendix A.1.1)

$$
\mathbf{c}'\boldsymbol{\Sigma}_{\mathrm{B}}\mathbf{c} = \Bigg(\Bigg(s^2 \det(\mathbf{A}) + \bigg(\sum_{j=0}^{r}\sum_{j'=0}^{r} v_{jj'} \bigg)^2 V\left(t_0\right) \Bigg)
$$
$$
\Bigg\{ p_e(1 - p_e) \sum_{j=0}^{r}\sum_{j'=0}^{r} v_{jj'} \bigg[s^2 \det(\mathbf{A}) \tag{3.2}
$$
$$
+ \bigg(\sum_{j=0}^{r}\sum_{j'=0}^{r} v_{jj'} \bigg)^2 \left(1 - \rho_{e,t_0}^2\right) V\left(t_0\right) \bigg] \Bigg\}^{-1}.
$$

If either $V\left(t_0\right) = 0$, i.e. all participants enter the study at the same time, or $\rho_{e,t_0} = 0$, i.e. exposure and initial time are uncorrelated, this formula reduces to

$$
\mathbf{c}'\boldsymbol{\Sigma}_{\mathrm{B}}\mathbf{c} = \Bigg[p_e(1 - p_e) \sum_{j=0}^{r}\sum_{j'=0}^{r} v_{jj'} \Bigg]^{-1} \tag{3.3}
$$

(Appendix A.1). Under LDD, we have

$$
\mathbf{c}'\boldsymbol{\Sigma}_{\mathrm{B}}\mathbf{c} = \sum_{j=0}^{r}\sum_{j'=0}^{r} v_{jj'} \left(p_e(1 - p_e) \right)^{-1}
$$
$$
\Bigg[s^2 \det(\mathbf{A}) + \left(1 - \rho_{e,t_0}^2\right) V\left(t_0\right) \sum_{j=0}^{r}\sum_{j'=0}^{r} v_{jj'} \bigg]^2 \Bigg]^{-1} \tag{3.4}
$$

14

(Appendix A.1.2). Under model (2.9), the B&W model, or model (2.10) we show that

$$\mathbf{c}'\Sigma_{\mathrm{B}}\mathbf{c} = \frac{\sum_{j=0}^{r}\sum_{j'=0}^{r} b_{jj'}}{p_e(1-p_e)s^2\det(\mathbf{A})}\left(\vphantom{\sum}\right.$$

(Appendix A.1.3), i.e. the formula for $\mathbf{c}'\Sigma_{\mathrm{B}}\mathbf{c}$ is the same as in the LDD case when $V(t_0) = 0$. Therefore, the results for LDD when $V(t_0) = 0$ apply to model (2.9) and model (2.10) and will not be presented in a separate section. When the follow-up period, τ, is fixed and the time points are equidistant, there are instances where $\mathbf{c}'\Sigma_{\mathrm{B}}\mathbf{c}$, under LDD and when $V(t_0) = 0$, is the same for $r = 1$ and $r = 2$. On pure efficiency grounds, in these situations, it is never cost-effective to add only one additional measure. In Appendix A.4, we derive a condition on the matrix Σ that needs to hold for this situation to occur, and we show that this will be the case for the three covariance structures considered in this paper. Of course, this result assumes that the interaction term is linear, and does not consider that the third measure is needed to assess the validity of this assumption. This counter-intuitive result is due to the fact that with equidistant points, the additional measure would be taken at half the follow up period, which is the mean of the time vector. This is similar to the fact that, in simple linear regression, adding an observation whose value for the explanatory variable is the explanatory variable mean produces no change on the variance of the slope.

The power formula depends on N, r and s, producing a discrete three-dimensional surface of constant power. Fixing r and s, the formula for the required sample size in N to achieve a pre-specified power π is

$$N = \frac{(\mathbf{c}'\Sigma_{\mathrm{B}}\mathbf{c})\left(z_\pi + z_{1-\alpha/2}\right)^2}{(\mathbf{c}'\mathrm{B}_{H_A})^2}. \tag{3.5}$$

For some parameters, namely p_e, s, $V(t_0)$ and ρ_{e,t_0}, their effect on $\mathbf{c}'\Sigma_{\mathrm{B}}\mathbf{c}$, and by virtue of (3.1) and (3.5), on power and sample size with r fixed, can be derived for a general covariance structure. Unlike randomized clinical trials, in observational studies, p_e is rarely 0.5, which is the value of p_e that maximizes power. For other values of p_e
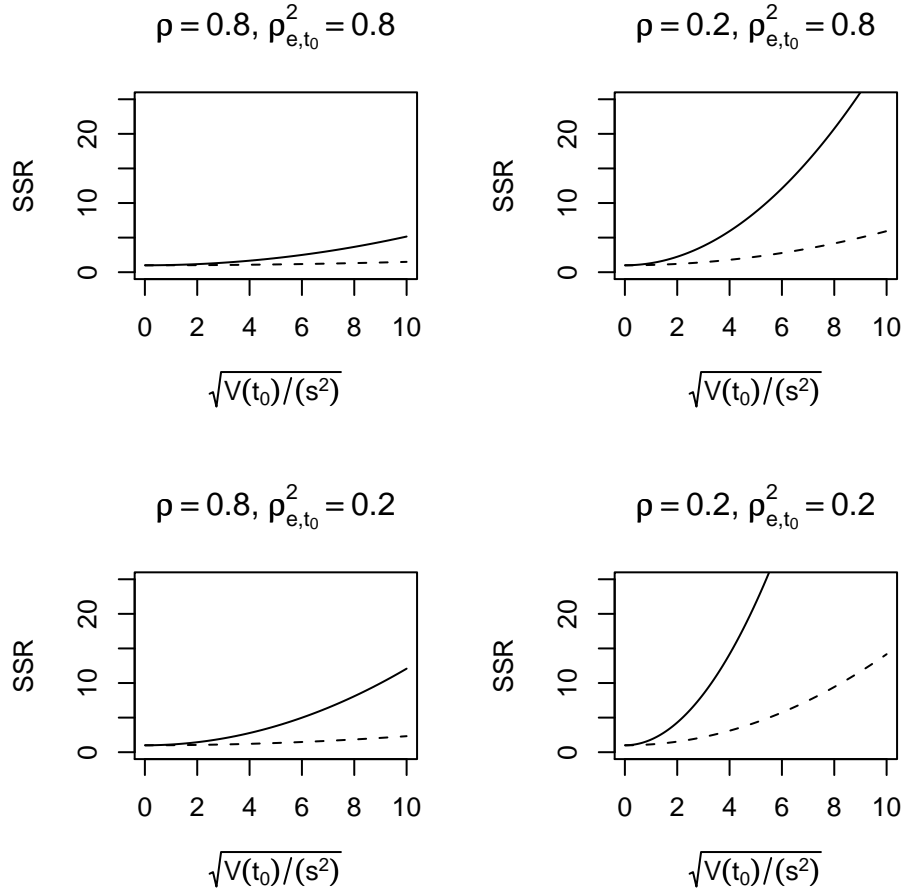
Figure 2: Ratio of required sample sizes (SSR) to achieve the same power comparing the $V(t_0) = 0$ case to $V(t_0) > 0$ under LDD and CS $\left(SSR = \frac{N_{V(t_0)=0}}{N_{V(t_0)}} \right)$. The groups are: $r = 2$ (——) and $r = 5$ (- - -).

the sample size obtained for $p_e = 0.5$ needs to be multiplied by $0.5^2 / (p_e(1 - p_e))$. For example, for $p_e = 0.2$ and r fixed we need 56% more participants to achieve the same power than for $p_e = 0.5$, and for $p_e = 0.1$, the sample size is multiplied by 2.7. It can also be derived from equations (3.2) and (3.4) that, unless Σ is an explicit function of s (as in the DEX covariance structure models considered in section 3.3), increasing s reduces $c' \Sigma_B c$ and therefore increases power and reduces the required sample size for both CMD and LDD. However, if either $V(t_0) = 0$ or $\rho_{e,t_0} = 0$, s does not play a role in the CMD case. The effect of $V(t_0)$, the baseline variance in the primary time scale, is different under CMD and LDD. Under CMD, as this parameter increases, keeping ρ_{e,t_0} and all other parameters fixed, so does $c' \Sigma_B c$ and therefore power decreases, since

$$\frac{\delta c' \Sigma_B c}{\delta V(t_0)} > 0.$$

The decrease in power will be larger as ρ_{e,t_0} departs from zero. When $\rho_{e,t_0} = 0$, inspection of (3.2) shows readily that the power is the same as that for the $V(t_0) = 0$ case (see (3.3)). Unlike the CMD case, under LDD, as $V(t_0)$ increases the power also increases, as is easily seen upon examination of (3.4), which depends on $V(t_0)$ in the denominator only. Increasing $V(t_0)$ increases the range of the regressor, t, which is known in simpler regression problems to increase the power of the study to detect a non-zero regression slope. Apparently, this result extends to interaction terms that are a function of a continuous variable, as well. The gain in power due to $V(t_0)$ is largest at $\rho_{e,t_0} = 0$, and it vanishes when $\rho_{e,t_0} = 1$ or $\rho_{e,t_0} = -1$, in which case power is equivalent to the $V(t_0) = 0$ case. It only makes sense to examine the effect of ρ_{e,t_0} when $V(t_0) > 0$, since if $V(t_0) = 0$ then t_{i0} is constant for all participants and ρ_{e,t_0} is zero. Figure 2 shows the gain in efficiency of having $V(t_0) > 0$ compared to $V(t_0) = 0$ for the CS case by showing the ratio of required sample size to achieve the same power in every case. The gains in efficiency can be very large for small r, large $\sqrt{V(t_0)}/s$, small ρ and small ρ_{e,t_0}.

Sometimes, N is fixed and the interest is in finding the minimum r to achieve a certain power. This problem has not been examined in any detail in previous literature. It

may arise, for example, when an existing cross-sectional study is to form the basis of a new longitudinal one or when there are a fixed number of participants available, (e.g. nurses who returned a baseline questionnaire in 1989). For some covariance structures, an explicit formula for r as a function of N, π and s can also be obtained, and for other structures, r can be obtained only numerically. It is shown that the minimum r for fixed N is obtained when $p_e = 0.5$ (Appendix A.5). Since no other global results are available, we will therefore consider this problem in the next section.

Sections 3.2-3.4 provide power and sample size formulas for particular covariance structures. Then, for the two alternative hypotheses and for each covariance structure, we assess the effect of r and the covariance parameters on power. Two scenarios will be considered when studying the effect of r on power. First, the frequency of measurements, s, is fixed. For example, participants might visit the clinic every 6 months, and increasing the number of repeated measures increases duration of follow-up, τ. In the second situation, the length of follow-up, τ, is fixed, for example, to 5 years, and increasing the number of repeated measures involves increasing the frequency of measurement, s. Formulas for r as a function of N, π and s or τ will be given when closed-form solutions exist. The effect of the covariance parameters on r will also be studied.

3.2 Compound symmetry (CS)

The simplest residual covariance structure that can be assumed for longitudinal data is compound symmetry. The residual covariance matrix is fully defined using two parameters: σ^2 and ρ. The first one is the variance of the response given the covariates, which under CS is assumed to be constant over time, and was defined in section 2. The second parameter, ρ, is the correlation between two measurements from the same participant. Under CS, it is also the reliability coefficient, or intraclass correlation co-

efficient,

$$\rho = \frac{\sigma^2_{between}}{\sigma^2_w + \sigma^2_{between}},$$

where $\sigma^2_{between}$ and σ^2_w are the between- and within-subject variance, respectively, and $\sigma^2_w + \sigma^2_{between} = \sigma^2$. Typically, ρ is unavailable at the time a study is designed, and power or sample size will be calculated over a range. Under CS, then, the covariance matrix is

$$Var(\mathbf{Y}_i|\mathbf{X}_i) \quad = \quad \mathbf{\Sigma}_{(r+1)\times(r+1)} \quad = \quad \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}$$

3.2.1 CMD

Table 1 shows the necessary terms to plug in to equation (3.1) to obtain the power formula for the case $V(t_0) = 0$. Under this same scenario, the formula for power as a function of N and r was previously given by Bloch (1986). For the cases with $V(t_0) > 0$ the equations do not simplify a great deal and interested readers should use formulas (3.2) or (3.4) directly. For any $V(t_0)$, the power to detect a difference increases as either N or r increase, but while by increasing N power can get arbitrarily close to one, by increasing r the maximum power that can be reached (when $r \to \infty$) (Appendix A.6.1) is

$$\Phi \left[\frac{\sqrt{N p_e (1 - p_e)} \, |\beta_2|}{\sqrt{\sigma^2 \rho}} - z_{1-\alpha/2} \right]$$

It has been shown that, when $V(t_0) = 0$, as the correlation, ρ, increases, the power to detect a difference decreases (Hedeker et al., 1999). This is not necessarily the case when $V(t_0) > 0$. For example, when $N = 50$, $r = 2$, $\sigma^2 = 1$, $V(t_0) = 20$ and $\rho_{e,t_0} = 0.7$, the variance for $\rho = 0.8$ is 0.29 and for $\rho = 0.9$ it is 0.25. Plugging in the corresponding values in Table 1 to equation (3.5) one obtains an equation for sample size for $V(t_0) = 0$. The equation for the minimum value of r which achieves a specified

Table 1: Numerator and denominator of the test statistic T (equation (2.2)) to obtain power and sample size for several correlation structures. For LDD, it is assumed here that $V(t_0) = 0$.

Pattern, $(c'B)_{H_A}$	Σ	$c'\Sigma_B c$	
		Fixed s	Fixed τ
CMD, $\beta_2 = p_1\mu_{00}$	CS	$\dfrac{\sigma^2(1+r\rho)}{p_e(1-p_e)(r+1)}$ (1)	$\dfrac{\sigma^2(1+r\rho)}{p_e(1-p_e)(r+1)}$ (1)
	AR(1)	$\dfrac{\sigma^2(1+\rho^s)}{p_e(1-p_e)(1+r+\rho^s-r\rho^s)}$ (2)	$\dfrac{\sigma^2(1+\rho^{\tau/r})}{p_e(1-p_e)(1+r+\rho^{\tau/r}-r\rho^{\tau/r})}$ (2)
LDD, $\gamma_3 = \dfrac{p_2 p_3 \mu_{00}}{\tau}$	CS	$\dfrac{12\sigma^2(1-\rho)}{p_e(1-p_e)\,s^2\,r\,(r+1)(r+2)}$ (3)	$\dfrac{12\sigma^2(1-\rho)\,r}{p_e(1-p_e)\,\tau^2\,(r+1)(r+2)}$ (3)
	AR(1)	$\dfrac{12\sigma^2(1-\rho^{2s})\left[r\,s^2 p_e(1-p_e)\right]^{-1}}{(2+r(r+3)+8\rho^s-2r^2\rho^s+(r-2)(r-1)\rho^{2s})}$ (4)	$\dfrac{12\sigma^2(1-\rho^{2\tau/r})\,r\left[\tau^2 p_e(1-p_e)\right]^{-1}}{(2+r(r+3)+8\rho^{\tau/r}-2r^2\rho^{\tau/r}+(r-2)(r-1)\rho^{2\tau/r})}$ (4)
	RS	$\left(\dfrac{12\sigma^2(1-\rho_{t_0})}{s^2 p_e(1-p_e)}\right)\left(\dfrac{1}{r(r+1)(r+2)}+\left(\dfrac{\rho_{b_1,s,\tilde r}}{1-\rho_{b_1,s,\tilde r}}\right)\dfrac{1}{\tilde r(\tilde r+1)(\tilde r+2)}\right)$ (5)	$\left(\dfrac{12\sigma^2(1-\rho_{t_0})}{\tau^2 p_e(1-p_e)}\right)\left(\dfrac{r}{(r+1)(r+2)}+\left(\dfrac{\rho_{b_1,\tau,\tilde r}}{1-\rho_{b_1,\tau,\tilde r}}\right)\dfrac{\tilde r}{(\tilde r+1)(\tilde r+2)}\right)$ (5)

(1) (Bloch, 1986)

(2) Appendix A.6.3

(3) (Diggle et al., 2002; Dawson, 1998; Frison and Pocock, 1997; Hedeker et al., 1999; Kirby et al., 1994; Jung and Ahn, 2003; Yi and Panzarella, 2002)

(4) Appendix A.6.4

(5) (Fitzmaurice et al., 2004; Galbraith and Marschner, 2002; Raudenbush and Xiao-Feng 2001; Schlesselman, 1973; Yi and Panzarella, 2002)

particular power π, with N fixed under CS, CMD and $V(t_0) = 0$ is

$$r = \frac{\beta_2^2 N\, p_e(1 - p_e) - \left(z_\pi + z_{1-\alpha/2}\right)^2 \sigma^2}{\left(z_\pi + z_{1-\alpha/2}\right)^2 \sigma^2 \rho - \beta_2^2 N\, p_e(1 - p_e)}. \tag{3.6}$$

As noted before, the desired power cannot always be reached by increasing the number of repeated measures, so equation (3.6) will not always have a positive solution. The effect of the intraclass correlation, ρ, on r depends on the additional parameters. If $\left(z_\pi + z_{1-\alpha/2}\right)^2 \sigma^2 > \beta_2^2 N\, p_e(1 - p_e)$, then r increases as ρ increases, otherwise r decreases as ρ increases (Appendix A.7.1). Our program also computes the required r for the case where $V(t_0) > 0$.

3.2.2 LDD

Here,

$$\mathbf{c}' \boldsymbol{\Sigma}_{\mathrm{B}} \mathbf{c} = 12\sigma^2 (1 - \rho)(1 + r\rho) \left[N p_e (1 - p_e)(r + 1)\right]^{-1}$$
$$\left\{r(r + 2)(1 + r\rho)s^2 + 12(1 - \rho)(1 - \rho_{\exp,t_0}^2) V(t_0)\right\}^{-1}. \tag{3.7}$$

For $V(t_0) = 0$, this variance was given previously (Diggle et al., 2002; Dawson, 1998; Frison and Pocock, 1997; Hedeker et al., 1999; Kirby et al., 1994; Jung and Ahn, 2003; Yi and Panzarella, 2002) (Table 1). The power formula can be obtained plugging in (3.7) into (3.1).

Power is a monotone function of r, and the limit of power when r goes to infinity is one, both when s is fixed and when τ is fixed (Appendix A.6.2). Thus, any pre-specified power can be achieved by increasing the number of repeated measures. The effect of ρ on power depends on a complicated fashion on r, s, $V(t_0)$ and ρ_{e,t_0}^2. However, when all participants are observed at the same time points ($V(t_0) = 0$), then increasing ρ will always increase power, the opposite effect that it has under CMD (Hedeker et al., 1999).

Fixing r and s, an expression for the required number of participants, N, is readily obtained using (3.5) and (3.7). With N and s fixed, a closed form solution for r is not

available. Our program (see Section 6) can be used to calculate the required r for this case. As noted before, any pre-specified power can be reached by increasing r. If $V(t_0) = 0$, then as the intraclass correlation, ρ, increases, the required r decreases, both when s is fixed and when τ is fixed (in the latter case, provided $r > 1$) (appendices A.5.2-A.5.3). When $V(t_0) > 0$, the effect of ρ on r is not necessarily monotone.

3.3   Damped exponential

In this section, we consider a covariance structure that generalizes CS as a particular case. Following Munoz et al. (1992), the damped exponential (DEX) covariance matrix can be expressed as

$$\Sigma = \sigma^2 \left(\begin{pmatrix} 1 & \rho^{s^\theta} & \rho^{(2s)^\theta} & \cdots & \rho^{(rs)^\theta} \\ \rho^{s^\theta} & 1 & \rho^{s^\theta} & \ddots & \vdots \\ \rho^{(2s)^\theta} & \rho^{s^\theta} & 1 & \ddots & \rho^{(2s)^\theta} \\ \vdots & \ddots & \ddots & \ddots & \rho^{s^\theta} \\ \rho^{(rs)^\theta} & \cdots & \rho^{(2s)^\theta} & \rho^{s^\theta} & 1 \end{pmatrix} \right) . \tag{3.8}$$

Under this covariance model, ρ is now the correlation between two measures from the same participant separated by one time unit. The correlation between two consecutive measures is ρ^{s^θ} and the correlation between two measurements of the same participant decreases as their separation in time increases. If the correlation between two measurements separated by s units, ρ_s, is known, the correlation per one unit is $\rho = \rho_s^{s^{-\theta}}$. The parameter $\theta \in [0, 1]$ controls the degree of attenuation of the correlation over time. This covariance structure includes compound symmetry when $\theta = 0$ and AR(1) covariance structure when $\theta = 1$. Thus, an investigator can vary the value of θ to determine the sensitivity of sample size and power calculations to departures of this sort from compound symmetry. For example, in a study on pulmonary function loss, the damping coefficient, θ, was 0.48, and in a study on CD4 cell count in HIV infected subjects, it was 0.35 (Munoz et al., 1992). Under DEX, $\mathbf{c}'\Sigma_B\mathbf{c}$ cannot be simplified to a simpler expression since the inverse of a DEX matrix is a complicated expression, and there is no general expression for all values of r. Design calculations

can be performed using our program. However, when $\theta = 1$, that is, AR(1), a simple expression is obtained (Table 1).

3.3.1   CMD

As noted before, with DEX, formulas do not have closed form, but the computations can be performed with our program. The formula for the variance of $\hat{\beta}_2$ under AR(1) and $V(t_0) = 0$ is given in Table 1. With N fixed, under DEX and CMD, power increases as r increases. For small values of θ, we observed in some cases that by increasing r to large values, the limit of the power when r tends to infinity was not one, as in the CS case. However, we observed in some cases that when the frequency of measurements, s, is fixed, as θ gets large the limit of power as r goes to infinity gets closer to one. With AR(1) covariance, we proved that the limit of power is one, so any pre-specified power can be reached by increasing the number of repeated measures (Appendix A.6.3). When the follow-up period, τ, is fixed, this limit of power is not one. For example, when $V(t_0) = 0$ this limit is

$$\Phi\left[\frac{\sqrt{N\, p_e\, (1 - p_e)(2 - \tau \log \rho)}\ |\beta_1|}{\sqrt{2\sigma^2}} - z_{1-\alpha/2} \right] ($$

(Appendix A.6.3). Regarding the influence of the covariance parameters, the power to detect an exposure effect under CMD decreases as the correlation ρ increases, as in the CS case, if $V(t_0) = 0$; otherwise the relationship with ρ is not always monotone. When $V(t_0) = 0$, power increases as θ increases provided $s > 1$ unit (if $s < 1$ unit the correlation between two measures separated by one unit will be larger under DEX than under CS). The effect of departures from CS, i.e. $\theta > 0$, on power and sample size can be assessed by computing the asymptotic relative efficiency (ARE) of the test statistic under CS and under DEX. Since the numerator of the test statistic is the same in both cases, the ARE is equivalent to the variance ratio, which in turn is equivalent to the inverse of the ratio of required sample sizes to achieve the same power (Dawson and Lagakos, 1993). Figure 3 shows the percent reduction in the required sample
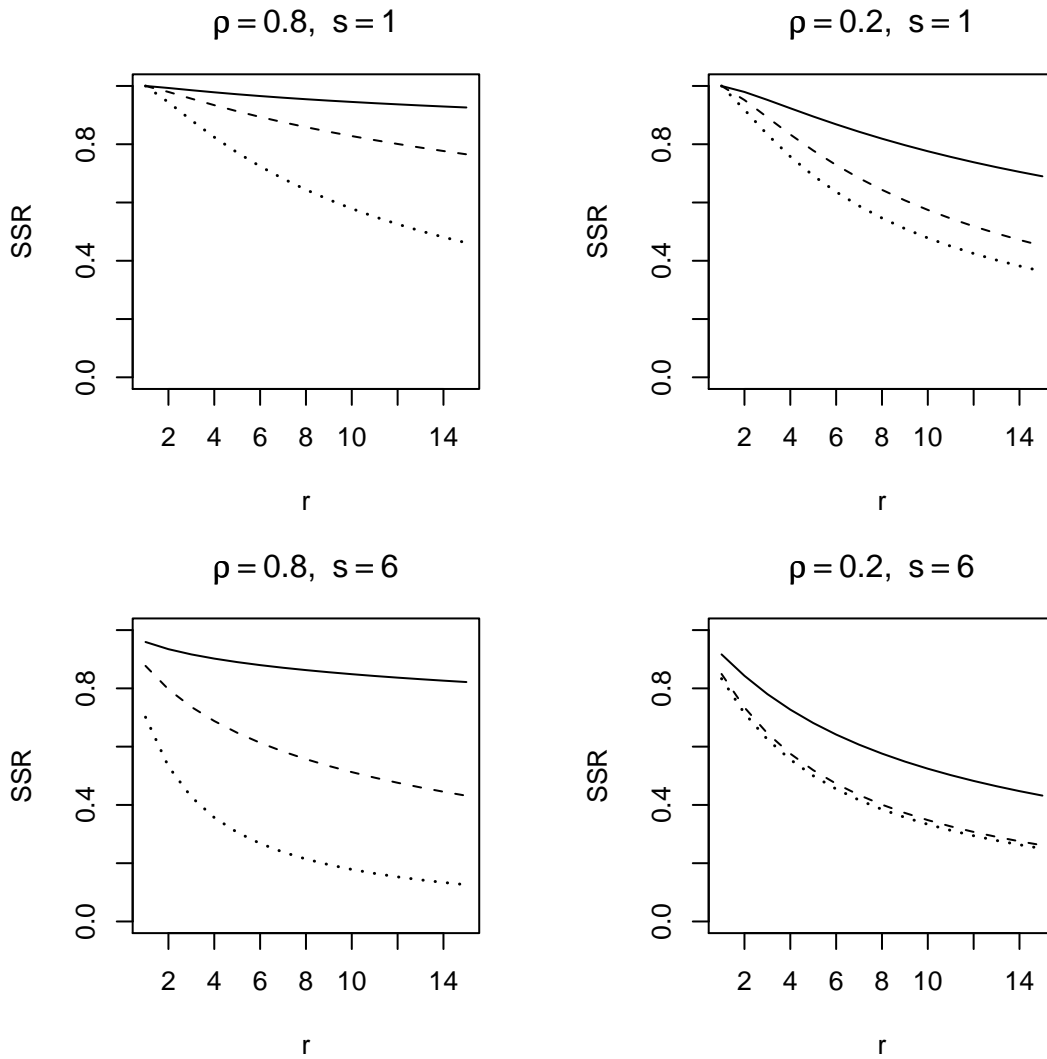
Figure 3: Ratio of required sample sizes (SSR) to achieve the same power comparing DEX and CS ($\theta = 0$) for fixed values of r, ρ, θ, and $V(t_0) = 0$ under CMD $\left(SSR = \frac{N_{DEX}}{N_{CS}} \right)$. The groups are: $\theta = 0.2$ (——), $\theta = 0.5$ (- - -), and $\theta = 1$ ($\cdots\cdots$).

size at a fixed power, when the covariance structure is DEX compared to CS under CMD for fixed s and $V(t_0) = 0$. Similar graphs are obtained for the fixed τ case. The reduction can be considerable and it is bigger when one takes many repeated measurements and the time between measurements s is large.

The required r to achieve a power π when N is fixed can be computed with our program. To assess the effect of θ on the required r, we computed it over a grid of values of the parameters, restricting to $N \in [400, 2000]$, $p_1 \in [0.1, 0.2]$, $\sigma/\mu_{00} \in [0.5, 2]$, $\tau \in [2, 60]$ and $V(t_0) = 0$. Similar to the effect of θ on power and number of participants, we observed in this region of the parameter space that the required r decreased as θ increased, provided $s > 1$ unit.

3.3.2 LDD

The formula for Σ_{B} does not have a general simple expression for all values of θ and r, and therefore neither do the formulas for power or for N with fixed r. For AR(1) and $V(t_0) = 0$, Table 1 shows the formula for the variance of $\hat{\gamma}_3$ needed by equation (3.1) to obtain the power of a study. Under LDD and DEX, we observed through a grid search over a wide range of the design space that the power to detect an effect increases as r increases. We computed the limit of power when r goes to infinity for the AR(1) case. For fixed s, this limit goes to one (Appendix A.6.4). Therefore, any pre-specified power can be reached by increasing the number of repeated measures. For fixed τ, this is not the case, and for example when $V(t_0) = 0$ the limit is (Appendix A.6.4)

$$\Phi \left[\frac{\sqrt{N p_e (1 - p_e) \left[12 + \tau (\log \rho)(\tau \log \rho - 6) \right]} \, |\gamma_3|}{\sqrt{-24\sigma^2 \log \rho}} - z_{1-\alpha/2} \right] \Bigg($$

The effect of ρ and θ on power depend on each other and it is not always monotone, even when $V(t_0) = 0$. We computed power for a grid of values for the parameters, restricting $r \leqslant 15$ and $s \leqslant 6$ units, and observed that when $V(t_0) = 0$, as θ departs from 0, the power starts to decline, but it can increase again when θ approaches 1 for large values of s and r (data not shown). However, in the range of values we investigated,
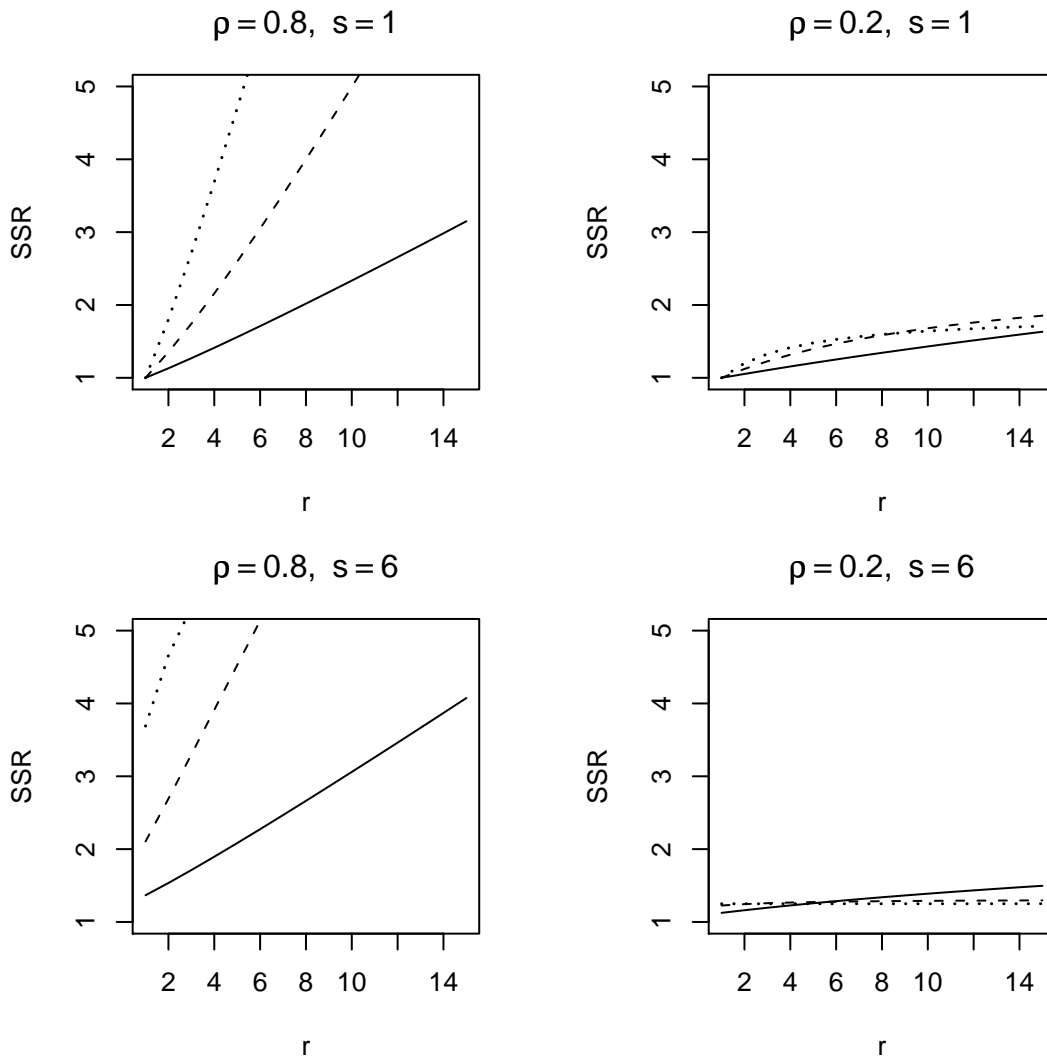
Figure 4: Ratio of required sample sizes (SSR) to achieve the same power comparing DEX and CS ($\theta = 0$) for fixed values of r, ρ and θ, and under LDD with $V(t_0) = 0$ $\left(SSR = \frac{N_{DEX}}{N_{CS}} \right)$. The groups are $\theta = 0.2$ (—), $\theta = 0.5$ (- - -), and $\theta = 1$ ($\cdots\cdots$).

power was maximized at $\theta = 0$ (i.e. CS). When $V(t_0) > 0$, power is not necessarily maximized at $\theta = 0$ when $V(t_0)/s^2$ is large, around 20 or larger, and θ is near one, coupled with small values of r. Figure 4 shows the increase in the required number of participants to achieve a certain power when there is covariance decay compared to CS, for LDD with $V(t_0) = 0$ and fixed s. For the case of fixed τ, a similar pattern was observed. The increase in N as θ increases can be quite large when the intraclass correlation, ρ, is high and r is large. For example, if the true covariance is AR(1) ($\theta = 1$) and ρ is large, one may have to enroll more than three times more participants than if the true covariance is CS, i.e. ($\theta = 0$).

The required r to achieve a power π when N is fixed can be computed with our program. Computing the required r for several values of the parameters we observed that the effect of θ on r is not necessarily monotone. In general, r increases as θ departs from 0, but it may decrease again as θ gets larger.

3.4   Random intercepts and slopes

In this section, we consider another generalization of CS. We consider the covariance structure obtained when an additional random effect (b_{1i}) associated with time is assumed (i.e. random intercepts and slopes, denoted RS), leading to model

$$Y_{ij} = \beta_0 + f(t_{ij}; \beta_1) + \beta_2 k_i + b_{0i} + t_{ij} b_{1i} + e_{ij}$$

under CMD and

$$Y_{ij} \quad = \quad \gamma_0 \quad + \quad f(t_{ij}, \gamma_1) \quad + \quad \gamma_2 k_i \quad + \quad \gamma_3 (t_{ij} \times k_i) \quad + \quad b_{0i} \quad + \quad t_{ij} b_{1i} \quad + \quad e_{ij}$$

under LDD, where $f(t_{ij}, \gamma_1)$ is a function of time that includes a linear term and is otherwise arbitrary. In mixed models notation, the residual covariance matrix is often written as $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}'_i + \sigma_w^2 \mathbf{I}$, where $\mathbf{Z}_i$ contains a subset of columns of the design matrix for participant i, and $\mathbf{D}$ is the covariance matrix of the random effects (e.g.

Fitzmaurice et al. (2004), p. 199). Here, the matrix $\mathbf{Z}_i$ contains a column of ones and the column of times for participant i, and

$$\mathbf{D} = \begin{pmatrix} \sigma_{b_0}^2 & \rho_{b_0 b_1} \sigma_{b_0} \sigma_{b_1} \\ \rho_{b_0 b_1} \sigma_{b_0} \sigma_{b_1} & \sigma_{b_1}^2 \end{pmatrix} \Bigg($$

where $\sigma_{b_0}^2$ and $\sigma_{b_1}^2$ are the variance of the random effect associated with the intercept and slope, respectively, and $\rho_{b_0 b_1}$ is the correlation between them. When there is only a random effect associated with the intercept, the matrix $\mathbf{Z}_i$ contains only a column of ones, and the resulting matrix, $\mathbf{\Sigma}_i$, follows a CS structure. Likewise, when $\sigma_{b_1}^2 = 0$, RS reduces to CS. If $V(t_0) = 0$ then $\mathbf{\Sigma}_i = \mathbf{\Sigma}$, i.e. the covariance matrix is the same for all participants, as it was the case when CS or DEX was assumed, even when $V(t_0) > 0$ for CS and DEX. The RS covariance structure is heteroscedastic, with the residual variance of the responses (the diagonal elements of $\mathbf{\Sigma}$) assumed to change as a quadratic function of time, with positive curvature $\sigma_{b_1}^2$. In addition, this covariance structure assumes that the correlation between repeated measures changes with time and with increasing duration between visits - in either scenario, it can either increase or decrease. When pilot data are available, the parameters of $\mathbf{D}$ can be estimated and used directly as inputs into our program to perform design calculations.

Often, however, longitudinal pilot data are not available, and a more intuitive parameterization is needed so that investigators can propose plausible values on which to base designs. To make the parameters more intuitive, we defined $\sigma_{t_0}^2 = \sigma_w^2 + \sigma_{b_0}^2$ as the residual variance at baseline (or at the mean initial time). Then, we define

$$\rho_{t_0} = \frac{\sigma_{b_0}^2}{\sigma_w^2 + \sigma_{b_0}^2}$$

as the reliability coefficient at baseline (or at the mean initial time), i.e. the percentage of residual variance at baseline that is due to between-subject variation. One additional parameter is needed, to fix the between-subjects variance in slopes. Following a parameterization proposed for characterizing the relative variability in slopes from several studies compared to their within-study variance in the context of meta-analysis (Takkouche et al., 1999), we defined the slope reliability as the percentage

of variation in the estimated coefficient $\hat{\gamma}_3$ that is due to between-subjects variation. When s is fixed, we define $\rho_{b_1,s,\tilde{r}}$ as the slope reliability with $\tilde{r}$ repeated measures, where $\tilde{r}$ is a hypothetical or trial value of r. When $V(t_0) = 0$, this quantity is

$$\rho_{b_1,s,\tilde{r}} = \frac{\sigma_{b_1}^2 s^2 \tilde{r}(\tilde{r}+1)(\tilde{r}+2)}{12(1-\rho_{t_0})\sigma_{t_0}^2 + \sigma_{b_1}^2 s^2 \tilde{r}(\tilde{r}+1)(\tilde{r}+2)}.$$

For the case of fixed τ, we define the equivalent quantity

$$\rho_{b_1,\tau,\tilde{r}} = \frac{\sigma_{b_1}^2 \tau^2(\tilde{r}+1)(\tilde{r}+2)}{12\tilde{r}(1-\rho_{t_0})\sigma_{t_0}^2 + \sigma_{b_1}^2 \tau^2(\tilde{r}+1)(\tilde{r}+2)}.$$

The variance matrix can now be expressed in terms of these new intuitive, parameters $\sigma_{t_0}^2$, $\rho_{t_0} \in [0, 1]$, $\rho_{b_0 b_1} \in [-1, 1]$, and $\rho_{b_1,s,\tilde{r}} \in [0, 1]$ or $\rho_{b_1,\tau,\tilde{r}} \in [0, 1]$. In the case of fixed τ, the covariance matrix can be expressed as

$$\Sigma_i = \sigma_{t_0}^2 (1 - \rho_{t_0})\mathbf{I} + \mathbf{Z}_i \left(\begin{pmatrix} \rho_{t_0} & \rho_{b_0 b_1}\sqrt{\frac{12\rho_{t_0}(1-\rho_{t_0})\tilde{r}}{\tau^2(\tilde{r}+1)(\tilde{r}+2)}\left(\frac{\rho_{b_1,\tau,\tilde{r}}}{1-\rho_{b_1,\tau,\tilde{r}}}\right)} \\ & \frac{12(1-\rho_{t_0})\tilde{r}}{\tau^2(\tilde{r}+1)(\tilde{r}+2)}\left(\frac{\rho_{b_1,\tau,\tilde{r}}}{1-\rho_{b_1,\tau,\tilde{r}}}\right) \end{pmatrix} \right) \mathbf{Z}'_i, \quad (3.9)$$

and for the fixed s case, one just needs to substitute $\rho_{b_1,\tau,\tilde{r}}$ by $\rho_{b_1,s,\tilde{r}}$ and τ by $s\tilde{r}$. If the value of r is known a priori, $\tilde{r}$ will take the value of r. Otherwise, for design problems where r is not fixed (i.e. when finding r for fixed N, or when finding (N_{opt}, r_{opt})), the investigator needs to provide the slope reliability together with a trial value of $\tilde{r}$ associated with it, and then find r or (N_{opt}, r_{opt}). In the calculations that follow, $\tilde{r}$ will act as a constant. If the value of r that solves the design problem is different from the one used to define the initial $\rho_{b_1,s,\tilde{r}}$ or $\rho_{b_1,\tau,\tilde{r}}$, the investigator should recalculate $\rho_{b_1,s,\tilde{r}}$ or $\rho_{b_1,\tau,\tilde{r}}$ with the new value of r to ascertain that the resulting values of $\rho_{b_1,s,r}$ or $\rho_{b_1,\tau,r}$ are realistic. Our software automatically recalculates $\rho_{b_1,s,r}$ or $\rho_{b_1,\tau,r}$ with the value of r that is the solution to the design problem. In the figures shown in this paper and, otherwise, when grid searches were performed, we chose $\tilde{r} = 5$ and $\rho_{b_1,s,r=5}$ or $\rho_{b_1,\tau,r=5}$ are used.

3.4.1 CMD

Formula (3.3) under RS and CMD results in a complex formula for $\mathbf{c}'\Sigma_\mathrm{B}\mathbf{c}$ which we do not provide here. However, in practice, a RS correlation structure will be usually not

fitted under CMD. For particular cases, calculations can be performed with our program by entering the intuitive parameters or by using formula (3.3) directly. Unlike the analogous CS and DEX scenarios, when $V(t_0) > 0$, formula (3.3) cannot be used because it is based on all participants having the same covariance matrix Σ. When $V(t_0) > 0$ under RS, Σ_i is different for each participant (equation (3.9)). We will still compute Σ_B as $\mathbb{E}_X^{-1}\left(\mathbf{X}'_i\Sigma_i^{-1}\mathbf{X}_i\right)$ as we did in section 2, but this calculation will now require correctly specifying the full distribution of (k_i, t_{0i}) and not just the first two moments. In this paper and in the software (section 6), we assume that t_0 is normally distributed within each exposure group, with the same variance, $V(t_0)$, but a different mean that will depend on ρ_{e,t_0}, and that k_i follows a Bernoulli with probability p_e. We then compute $\mathbb{E}_X^{-1}\left(\mathbf{X}'_i\Sigma_i^{-1}\mathbf{X}_i\right)$ by numerical integration (see Appendix A.8 for more details). We assessed the sensitivity of results to the normality assumption for t_0 by comparing to results obtained with t_0 assumed to be uniform, a four-parameter Beta with several values for the shape parameters and lognormal with several values of the shape parameter, with the mean and variance of each distribution matched to the mean and variance of the normal case, over a grid of values of the covariance parameters. We found that the resulting variance depended on the distribution assumed, and depending on the values of the other parameters, the distributions we considered can provide variances that smaller or larger than the normal case, with no clear pattern. So, the results given in this paper for CMD, RS and $V(t_0) > 0$ will rely on the times being normally distributed. Unlike the case where $\Sigma_i = \Sigma$, we found empirically that even when t_0 is normally distributed, the value of p_e that maximizes power is not necessarily 0.5 when $\rho_{\mathrm{e},t_0} \neq 0$.

The limit of the power when r goes to infinity and $V(t_0) = 0$ is (Appendix A.6.5)

$$\Phi\left[\frac{\sqrt{N\,p_e\,(1-p_e)}\,|\beta_1|}{\sqrt{\sigma_{t_0}^2\rho_{t_0}(1-\rho_{b_0b_1}^2)}} - z_{1-\alpha/2}\right]\Bigg($$

The effect of the covariance parameters ρ_{t_0}, $\rho_{b_0b_1}$, and $\rho_{b_1,s,r=5}$ or $\rho_{b_1,\tau,r=1}$ on power, number of participants, and number of repeated measures is not monotone and de-

pends upon the values of more than one parameter. Through a grid search, we found that for the same value of ρ_{t_0}, power can be either larger or smaller than in the CS case. Similarly, through a grid search, we found that the effect of $V(t_0)$ on power did not follow a monotone pattern throughout our grid search.

3.4.2 LDD

When $V(t_0) = 0$, power does not depend upon the parameter $\rho_{b_0 b_1}$. Table 1 shows the terms needed to compute power and number of participants (N) for fixed r, using equations (3.1) and (3.5). These formulas are equivalent to those reported by Schlesselman (1973), Raudenbush and Xiao-Feng (2001), Yi and Panzarella (2002), Galbraith and Marschner (2002) and Fitzmaurice et al. (2004). When $V(t_0) > 0$, formula (3.4) cannot be used because it requires that the response of all participants have the same covariance matrix, $\mathbf{\Sigma}$. As discussed above in the CMD case (section 3.4.1), under RS, when $V(t_0) > 0$, $\mathbf{\Sigma}_i$ is different for each participant. In this paper and in the software (section 6), to compute $\mathbf{\Sigma}_{\mathrm{B}}$ we assumed that t_0 is normally distributed within each exposure group, with the same variance $V(t_0)$ and a different mean depending on ρ_{e,t_0}, and that k_i follows a Bernoulli with probability p_e. We then computed $\mathbf{\Sigma}_{\mathrm{B}}$ as $\mathbb{E}_X^{-1}\left(\mathbf{X}'_i\mathbf{\Sigma}_i^{-1}\mathbf{X}_i\right)$, by numerical integration (see Appendix A.8 for details). As is section 3.4.1 for the CMD case, we compared the results assuming normality for t_0 to results assuming other distributions. The resulting variances were not materially different from the normal case for symmetric or moderately skewed distributions, but were greater than the normal case in situations where the distribution of t_0 was very skewed. Since our program would then provide underestimates of the true variance, if it is believed that t_0 is severely skewed, it might make sense to use formulas with $V(t_0) = 0$, which does not require assumptions on the distribution of t_0 and appears to provide conservative estimates of the variance compared to the $V(t_0) > 0$ case, as observed through the grid searches over all the distributions of t_0 we studied. As in the CMD case, when $\mathbf{\Sigma}_i \neq \mathbf{\Sigma}$, we found empirically that the value of p_e that maximizes

31

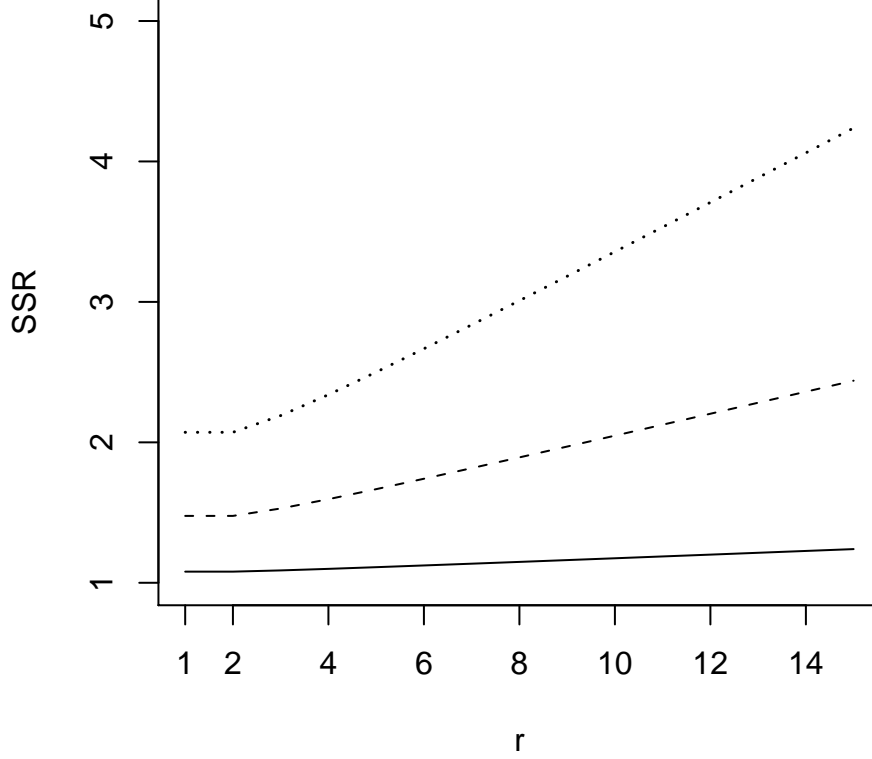Figure 5: Ratio of required sample sizes (SSR) to achieve the same power comparing RS and CS ($\rho_{b_1,\tau,r=5} = 0$) for fixed values of r, ρ_{t_0} and $\rho_{b_1,\tau,r}$ and $V(t_0) = 0$ under LDD $\left(SSR = \frac{N_{RS}}{N_{CS}}\right)$. The groups are $\rho_{b_1,\tau,r=5} = 0.1$ (—), $\rho_{b_1,\tau,r=5} = 0.4$ (- - -), and $\rho_{b_1,\tau,r=5} = 0.6$ (······).

power is not necessarily 0.5 when $\rho_{e,t_0} \neq 0$.

Power is an increasing function of r. However, the limit of the power when r goes to infinity is not one, so there can be instances where a pre-specified power cannot be achieved by just increasing r. When $V(t_0) = 0$ the limit is (Appendix A.6.6)

$$\Phi\left[\left(\frac{\sqrt{N\, p_e\,(1-p_e)}\,|\gamma_3|}{\sqrt{\frac{12\sigma_{t_0}^2\,(1-\rho_{t_0})\tilde{r}}{\tau^2(\tilde{r}+1)(\tilde{r}+2)}\left(\frac{\rho_{b_1,\tau,\tilde{r}}}{1-\rho_{b_1,\tau,\tilde{r}}}\right)}} - z_{1-\alpha/2}\right)\right]$$

for the fixed τ case, and the equivalent expression substituting $\rho_{b_1,\tau,\tilde{r}}$ by $\rho_{b_1,s,\tilde{r}}$ and τ

by $s\tilde{r}$ for the fixed s case. As in the CS case, the effect of ρ_{t_0} on power depends on other parameters in a complicated fashion, but when $V(t_0)= 0$ increasing ρ_{t_0} always increases power. To examine the effect of $\rho_{b_1,s,\tilde{r}}$ or $\rho_{b_1,\tau,\tilde{r}}$, that is, the effect of departures from compound symmetry towards a random slopes covariance structure, we calculated the ratio of sample sizes required to achieve the same power (ARE) comparing RS and CS. It is easily proven that more participants are required when either $\rho_{b_1,s,\tilde{r}}$ or $\rho_{b_1,\tau,\tilde{r}}$ are greater than zero. Figure 5 shows the ARE as a function of $\rho_{b_1,\tau,r=5}$ when $V(t_0) = 0$. It can be seen that the increase in number of participants can be quite large for large values of r and $\rho_{b_1,\tau,r=5}$. This is, for fixed σ_w^2, $\sigma_{b_0}^2$ and $\sigma_{b_1}^2$, as r increases, the within-subjects variance component of $Var(\hat{\gamma}_3)$ is reduced, and it becomes very small for large values of r. Thus, the percentage of variance due to the between-subjects component is much greater, and the only way to reduce the between-subjects variance component is to recruit more subjects. When $V(t_0) > 0$, we computed power over a grid of values of the other parameters, with restrictions $r \leqslant 15$, $s \leqslant 6$ units, and $\sqrt{V(t_0)}/s \leqslant 10$ for the fixed s case and $\tau \leqslant 60$ units and $\sqrt{V(t_0)}/\tau \leqslant 10$ for the fixed τ case. Over this wide region of the parameter space, the power decreased as either $\rho_{b_1,s,r=1}$ or $\rho_{b_1,\tau,r=1}$ increased, even when $V(t_0) > 0$.

With N and s fixed, there is no closed-form solution for the minimum value of r to satisfy a specified power, but calculations to solve this non-linear equation can be performed with our program. As noted previously, there may be situations where the pre-specified power cannot be reached by simply increasing r. The effect of the covariance parameters on r in this setting is not monotone. However, when $V(t_0) = 0$, we show that as the correlation ρ_{t_0} increases, the required r decreases (appendices A.5.4-5). Conversely, as $\rho_{b_1,s,\tilde{r}}$ or $\rho_{b_1,\tau,\tilde{r}}$ increase, the required r increases, both when s is fixed and when τ is fixed (provided $r > 1$ for the latter case) (appendices A.5.6.1-2).

Table 2: Summary of the results in section 4, CMD hypothesis.

	r_{opt} **when** $V(t_0) = 0$	**Characteristics** *	r_{opt} **when** $V(t_0) \neq 0$*
CMD (Fixed s)			
CS	$r_{opt} = \sqrt{\dfrac{(\kappa-1)(1-\rho)}{\rho}} - 1$	r_{opt} decreases as ρ increases	r_{opt} not a monotone function of $V(t_0)$
DEX	r_{opt} greater than for CS	r_{opt} decreases as ρ increases r_{opt} increases as θ increases	r_{opt} not a monotone function of $V(t_0)$
RS	r_{opt} can be either greater or smaller than for CS	r_{opt} not a monotone function of ρ_{t_0} r_{opt} not a monotone function of $\rho_{b_1, s, r=5}$	r_{opt} not a monotone function of $V(t_0)$
CMD (Fixed τ)			
CS	$r_{opt} = \sqrt{\dfrac{(\kappa-1)(1-\rho)}{\rho}} - 1$	r_{opt} decreases as ρ increases	r_{opt} not a monotone function of $V(t_0)$, although r_{opt} is only affected by $V(t_0)$ for large values of ρ_{e, t_0}
DEX	• if τ small then r_{opt} smaller than for CS * • if τ large then r_{opt} greater than for CS *	r_{opt} decreases as ρ increases r_{opt} not a monotone function of θ	same as in the CS case
RS	r_{opt} can be either greater or smaller than for CS	r_{opt} not a monotone function of ρ_{t_0} r_{opt} not a monotone function of $\rho_{b_1, \tau, r}$	r_{opt} decreases as $V(t_0)$ increases.

Table 3: Summary of the results in section 4, LDD hypothesis.

	r_{opt} when $V(t_0) = 0$	Characteristics *	r_{opt} when $V(t_0) \neq 0$*
LDD (Fixed s)			
CS	r as large as possible	ρ does not affect r_{opt}	still 15 in our restricted space, except for large $V(t_0)/s^2$ and small ρ_{e,t_0}
DEX	r as large as possible	ρ does not affect r_{opt} θ does not affect r_{opt}	θ does not affect r_{opt}
RS	r_{opt} solves $$\kappa = \frac{r_{opt}^2 \left(-(3+2r_{opt})\tilde{r}(\tilde{r}+1)(\tilde{r}+2)+(r_{opt}+1)^2(r_{opt}+2)^2 \frac{\rho_{b_1,s,\tilde{r}}}{1-\rho_{b_1,s,\tilde{r}}} \right)}{\left(2+6r_{opt}+3r_{opt}^2\right)\tilde{r}(\tilde{r}+1)(\tilde{r}+2)}$$	ρ_{t_0} does not affect r_{opt} r_{opt} decreases as $\rho_{b_1,s,r=5}$ increases	• r_{opt} still smaller than CS • r_{opt} decreases as either $\rho_{b_1,s,r=5}$ or $V(t_0)$ increase
LDD (Fixed τ)			
CS	• if $\kappa < 5$ then $r_{opt} = 1$ • if $\kappa > 5$ then choose $r > \frac{2(\kappa+1)}{\kappa-5}$	ρ does not affect r_{opt}	• can take repeated measures even with $\kappa < 5$ • r_{opt} not a monotone function of $V(t_0)$
DEX	r_{opt} is smaller or equal than for CS	r_{opt} decreases as θ increases r_{opt} not a monotone function of θ	r_{opt} not a monotone function of $V(t_0)$
RS	• if $\kappa > 5, r > \frac{2(\kappa+1)}{\kappa-5}$ and $$\rho_{b_1,\tau,\tilde{r}} < \frac{[-2(\kappa+1)+(\kappa-5)r](\tilde{r}+1)(\tilde{r}+2)}{6\tilde{r}(r+2)+[-2(\kappa+1)+(\kappa-5)r](\tilde{r}+1)(\tilde{r}+2)}$$ then r_{opt} solves $$\kappa = \frac{r_{opt}(4+3r_{opt})(\tilde{r}+1)(\tilde{r}+2)+\tilde{r}(r_{opt}+1)^2(r_{opt}+2)^2 \frac{\rho_{b_1,\tau,\tilde{r}}}{1-\rho_{b_1,\tau,\tilde{r}}}}{(r_{opt}^2-2)(\tilde{r}+1)(\tilde{r}+2)}$$ • Otherwise $r_{opt} = 1$	ρ_{t_0} does not affect r_{opt} r_{opt} decreases as $\rho_{b_1,\tau,\tilde{r}}$ increases	• it can be advisable to take repeated measures even with $\kappa < 5$ • r_{opt} decreases as $\rho_{b_1,\tau,r=5}$ increases. The effect of $V(t_0)$ is not monotone.

* Results derived from a grid search. The restrictions in the parameters where $r \leqslant 15$, $\kappa \leqslant 40$, $s \leqslant 6$ units, and $\frac{\sqrt{V(t_0)}}{s} \leqslant 10$ for the fixed s case and $\tau \leqslant 60$ units, and $\frac{\sqrt{V(t_0)}}{\tau} \leqslant 10$ for the fixed τ case.

4 Optimal Allocation

In planning a study, one often needs to consider cost. If the cost of recruiting a participant is c_1 monetary units and the first measurement for each participant is $\kappa \geqslant 1$ times more expensive than the rest, then the total cost of the study is

$$COST = N c_1 \left(1 + r/\kappa\right). \tag{4.1}$$

Then, when the budget is fixed at cost C, we need to choose the combination (N, r) that maximizes the power to detect the hypothesized effect, subject to this cost constraint. The imposition of the cost constraint determines a unique solution in (N, r) that solves the optimization problem, unlike in section 3, where a discrete two-dimensional 'curve' in (N, r) provides the desired power. Using a Lagrange multiplier, we maximize the power equation (3.1) with respect to r, subject to constraint (4.1). Once the optimal r is obtained, it can be plugged into equation (4.1) to obtain the corresponding optimal N. The constrained problem reduces to the following unconstrained problem (Appendix A.9)

$$\underset{r}{Min} \; (\kappa + r) \left(\mathbf{c}' \mathbf{\Sigma}_{\mathrm{B}} \mathbf{c}\right) \tag{4.2}$$

It turns out that the value of r_{opt} that maximizes power subject to a fixed cost is the same one that minimizes the cost of a study subject to a fixed power (Appendix A.9). Of course, N_{opt} will be different, depending on the nature of the constraint. Since r is not fixed by design, we must consider the two different scenarios as above: when there is a fixed frequency of measurement (fixed s) and when there is a fixed follow-up time (fixed τ). A summary of the results of this section is given in Tables 2 and 3. The exposure prevalence, p_e, does not have any effect on r_{opt} for any of the two constraints, and neither has an effect on N_{opt} for the cost constraint problem. For the power constraint problem, N_{opt} depends on p_e as in the case of deriving N for fixed r described in section 3.1.

Throughout section 4, when the analytical solutions could not be derived we com-

puted the optimal r for a grid of values of the other design parameters, with the restrictions $r \leqslant 15$, $\kappa \leqslant 40$ and $s \leqslant 6$ units and $\sqrt{V(t_0)}/s \leqslant 10$ for the fixed s case, and $\tau \leqslant 60$ units and $\sqrt{V(t_0)}/\tau \leqslant 10$ for the fixed τ case.

4.1 Fixed frequency of measurement, s

In this section, the frequency of measurements is fixed, for example, to yearly visits. Then increasing the number of repeated measures, r, implies increasing the length of follow-up. If a correlation decay is assumed, e.g. under DEX, increasing r will keep the correlation between adjacent measurements the same, but the first and last observations will be less correlated. Under RS, increasing the length of follow-up can either increase or decrease the variances of successive responses and their correlations with each other.

4.1.1 CMD

For CMD under CS and $V(t_0) = 0$, the optimal r is

$$r_{opt} = \sqrt{\left(\frac{(\kappa - 1)(1 - \rho)}{\rho}\right) - 1},\qquad(4.3)$$

and using the cost constraint (4.1),

$$N_{opt} = \frac{\kappa}{c_1 \left(\kappa + r_{opt}\right)} COST.$$

This result has been given in the context of cluster randomized trials (Raudenbush, 1997; Cochran, 1977). As can be seen in figure 6, the greater the correlation between measurements of the same person, the smaller the optimal number of repeated measures; and the bigger the cost of the first measurement compared to the rest, the greater the optimal number of repeated measures. If all measurements have the same cost ($\kappa = 1$), the optimal design takes no repeated measurements and recruits as many participants as the cost constraint allows. Large values of κ are needed to justify taking more than a small number of repeated measures. For the case where $V(t_0) > 0$
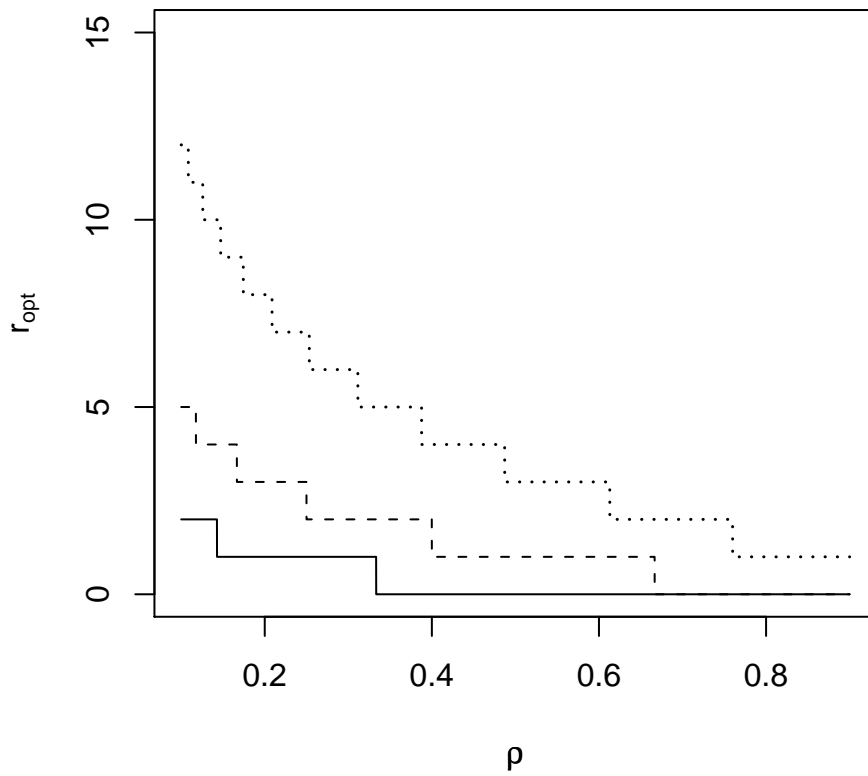
37

Figure 6: Optimal number of repeated measures, r_{opt}, as a function of ρ, under CMD, CS and $V(t_0) = 0$ for different cost ratios ($\kappa = 2$ (——), $\kappa = 5$ (- - -), and $\kappa = 20$ ($\cdots\cdots$)).
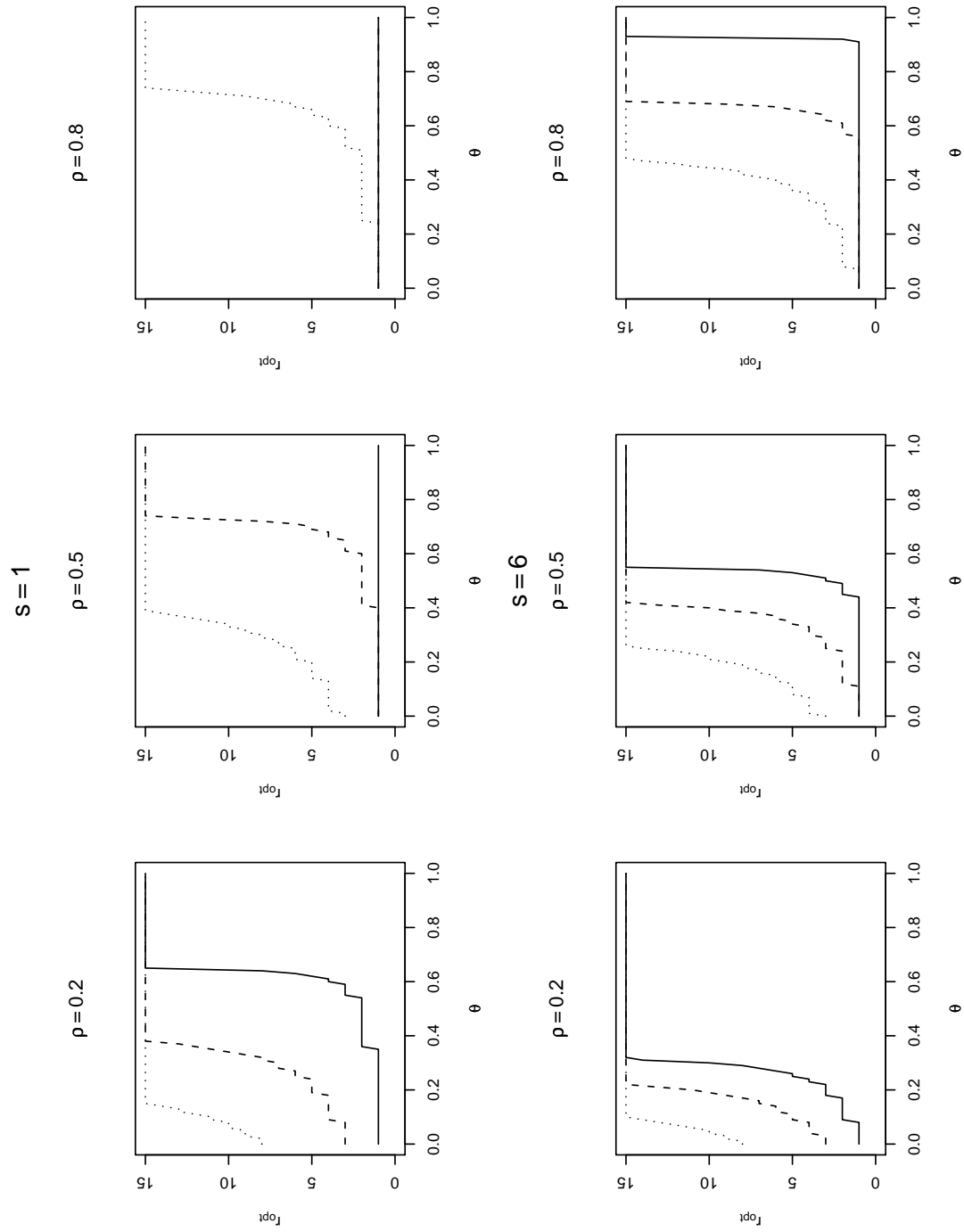
Figure 7: Optimal r as a function of θ under DEX, CMD, fixed frequency of measurement s and $V(t_0) = 0$, for $r \leq 15$ and different cost ratios ($\kappa = 2$ (———), $\kappa = 5$ (- - -), and $\kappa = 20$ (········)).

and $\rho_{\mathrm{e},t_0} \neq 0$, we computed the optimal r for a range of values of the other design parameters and observed that the effect of $\sqrt{V(t_0)}/s$ on r_{opt} was not monotone. In general, r_{opt} increased as $\sqrt{V(t_0)}/s$ increases, but it decreased again for large values of $\sqrt{V(t_0)}/s$, especially for cases with small κ.

With DEX covariance structure, there is no closed-form solution for the optimal (N, r). Computing the optimal r for a grid of values of the other design parameters, we observed that the optimal r was larger for DEX than for the compound symmetry case ($\theta = 0$), the remaining parameters being equal. Figure 7 shows r_{opt} as a function of θ for several cases with $V(t_0) = 0$. The effect of $V(t_0) > 0$ was very similar to its effect in the CS case, with r_{opt} increasing with $\sqrt{V(t_0)}/s$ but decreasing again for large values of $\sqrt{V(t_0)}/s$.

With RS covariance structure, there is no closed form solution for the optimal (N, r). We computed the optimal design for a grid of combinations of the other parameters. Few patterns appeared. The optimal r increased as κ increased, but the effects of ρ_{t_0}, $\rho_{b_1,s,r=5}$, $\rho_{b_0 b_1}$ and $V(t_0)$ strongly depended on values of the other parameters. Calculations for specific situations can be performed with our program.

4.1.2 LDD

Under LDD, and with fixed frequency of measurements, $V(t_0) = 0$ and CS, it is always advisable, in terms of maximizing the power subject to cost constraint, to choose r as large as possible, regardless of κ (Appendix A.10.1). In contrast, with $V(t_0) > 0$ and CS, under LDD, there are situations where choosing r as large as possible is not the optimal design. We computed the optimal r for a grid of combinations of the other parameters. The optimal was $r = 15$ for almost all cases investigated, except for a few combinations characterized mainly by large values of $\sqrt{V(t_0)}/s$ (close to 10) and small ρ_{e,t_0}. Using our program, (N_{opt}, r_{opt}) can be obtained for specific cases.

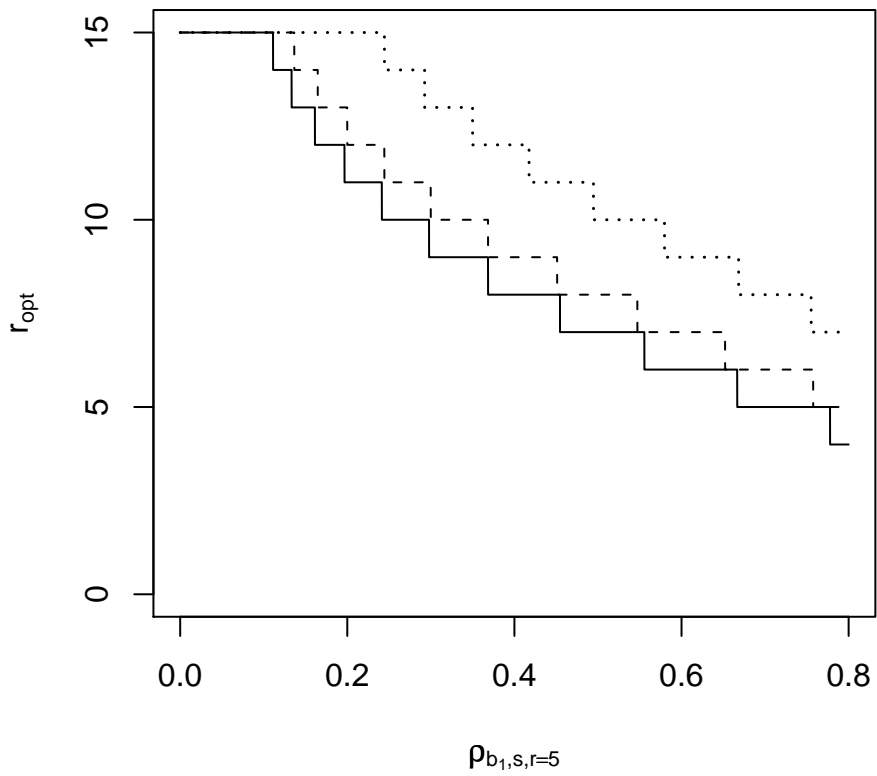For DEX, we computed the optimal r over a grid of values of the parameters. We

Figure 8: Optimal number of repeated measures, r_{opt}, under LDD and RS for fixed frequency of measurement assuming $V(t_0) = 0$ as a function of $\rho_{b_1,s,r=5}$ for different cost ratios ($\kappa = 2$ (—), $\kappa = 5$ (- - -), and $\kappa = 20$ ($\cdots\cdots$)).

observed that, under DEX and $V(t_0) = 0$, we end up choosing the maximum r (15 in our case), which agrees with the CS case. However, when $V(t_0) > 0$, there were cases where choosing r as large as possible did not give the highest power design at a fixed cost, in particular for large values of $\sqrt{V(t_0)}/s$ (close to 10), small values of κ, large values of ρ and small ρ_{e,t_0}.

RS behaves differently than CS and DEX since the length of follow-up has an effect on the variance of the observations. When $V(t_0) = 0$, for a given κ, s and $\rho_{b_1,s,\tilde{r}}$ we show in Appendix A.10.2 that the optimal r solves the equation

$$
\kappa = r_{opt}^2 \left[\left(-(3 + 2r_{opt})\tilde{r}(\tilde{r} + 1)(\tilde{r} + 2) \right. \right.
$$
$$
\left. + (r_{opt} + 1)^2 (r_{opt} + 2)^2 \frac{\rho_{b_1,s,\tilde{r}}}{1 - \rho_{b_1,s,\tilde{r}}} \right] \left(
$$
$$
\left\{ \left(2 + 6r_{opt} + 3r_{opt}^2 \right) \tilde{r}(\tilde{r} + 1)(\tilde{r} + 2) \right\}^{-1} .
$$

Figure 8 shows the optimal r as a function of κ, and $\rho_{b_1,s,r=5}$ when $V(t_0) = 0$. The optimal r is smaller than for CS, where $\rho_{b_1,s,r=5} = 0$, and it decreases as $\rho_{b_1,s,r=5}$ increases.

When $V(t_0) > 0$, with RS, the optimal r, r_{opt}, depends additionally on $\sqrt{V(t_0)}/s$, $\rho_{b_0 b_1}$ and ρ_{e,t_0}. We computed r_{opt} for a grid of values of the other parameters. Within the range of the parameter space investigated, the optimal r decreased as either $\rho_{b_1,s,r=5}$ or $\sqrt{V(t_0)}/s$ increased, i.e. less repeated measures and more participants were needed as the variation of slopes between participants increased and as the variance of the baseline time variable increased. For specific cases, r_{opt} can be computed with our program.

4.2   Fixed follow-up period, τ

Under this scenario, increasing r involves increasing the frequency of measurements during a fixed time period, τ. The interval between measurements, s, varies in this setting.

4.2.1 CMD

For CMD under CS and $V(t_0) = 0$, the problem is equivalent to the fixed frequency of measurement setting, since the correlations are not affected by the duration of follow-up. Therefore, the optimal r is given by equation (4.3). When $V(t_0) > 0$ and $\rho_{e,t_0} \neq 0$, we computed the optimal r for a grid of values of the other parameters and observed that $\sqrt{V(t_0)}/\tau$ only changed r_{opt} for large values of ρ_{e,t_0}, and in that case the pattern is similar to the fixed s case, where r_{opt} increased as $\sqrt{V(t_0)}/\tau$ separates from zero, but it decreased again for large values of $\sqrt{V(t_0)}/\tau$.

When there is a correlation decay ($\theta > 0$), we computed r_{opt} for a grid of values of the other parameters and found instances where the optimal r was smaller than in the CS case when τ was small and bigger than the CS case when τ was large (data not shown). Increasing $\frac{\sqrt{V(t_0)}}{\tau}$ produced no changes in r_{opt} in most situations. For RS, we investigated the dependency of r_{opt} as a function of the parameters of the RS covariance structure using a grid of values of the other parameters. Few patterns appeared. We observed that the optimal r increased with κ, the optimal being $r = 1$ or $r = 2$ when $\kappa = 2$, and ranging from 1 to 15 depending on the values of the other parameters when $\kappa = 40$. Figure 9 shows the relationship between r_{opt} and $\rho_{b_1,\tau,r=5}$ for several values of ρ_{t_0} and κ and $V(t_0) = 0$. The effects of ρ_{t_0}, $\rho_{b_1,\tau,r=5}$ and $\rho_{b_0b_1}$ strongly depended on the values of all other parameters. We observed that increasing $\frac{\sqrt{V(t_0)}}{\tau}$ reduced r_{opt}, whith r_{opt} being one for most cases with $\sqrt{V(t_0)}/\tau > 3$.

4.2.2 LDD

Under CS with $V(t_0) = 0$, then $r_{opt} = 1$ if $\kappa < 5$. Recall that when LDD is the alternative hypothesis of interest, we need at least one repeated measure to identify the parameters. For $\kappa > 5$, r_{opt} is not one. Any combination of (N, r) where $r > \frac{2(\kappa+1)}{\kappa-5}$ would improve the power achieved by the combination where $r = 1$. The optimal would involve taking r as large as possible, even though fewer participants would be
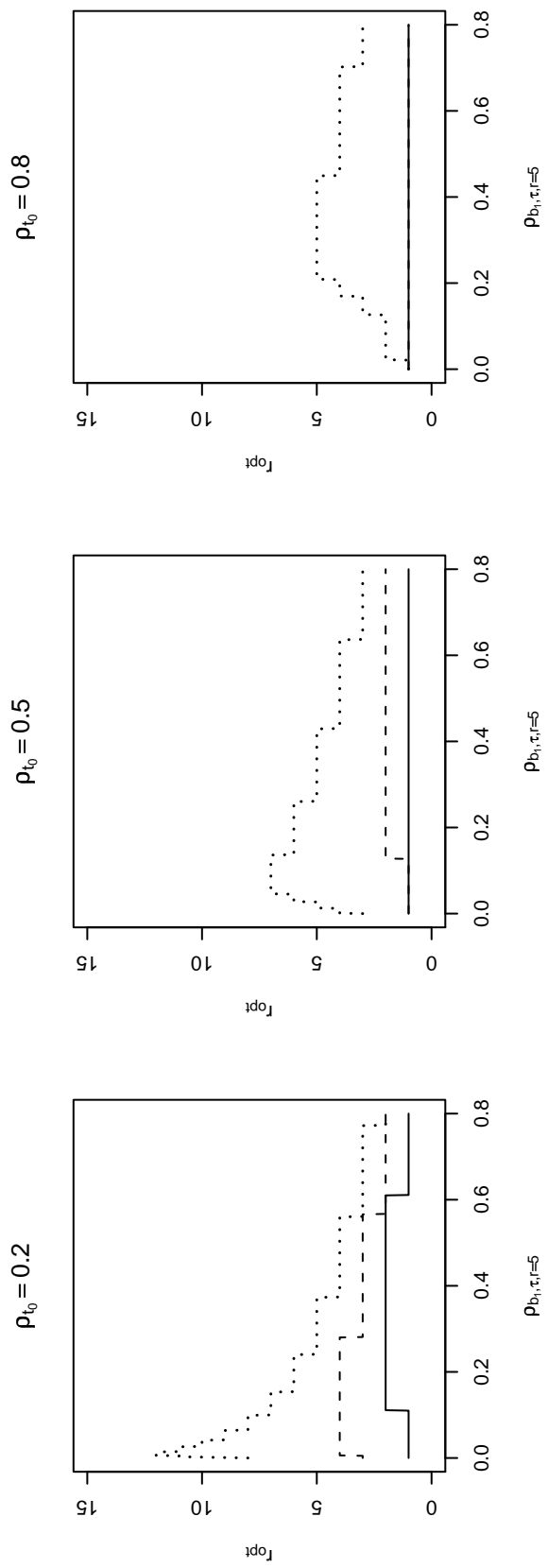
Figure 9: Optimal r as a function of $\rho_{b_1,\tau,r=5}$ under RS, CMD and fixed follow-up period $\tau = 5$, for $r \leq 15$, $\rho_{b_0 b_1} = -0.5$, $V(t_0) = 0$ and different cost ratios ($\kappa = 2$ (—), $\kappa = 5$ (- - -), and $\kappa = 20$ (······)).
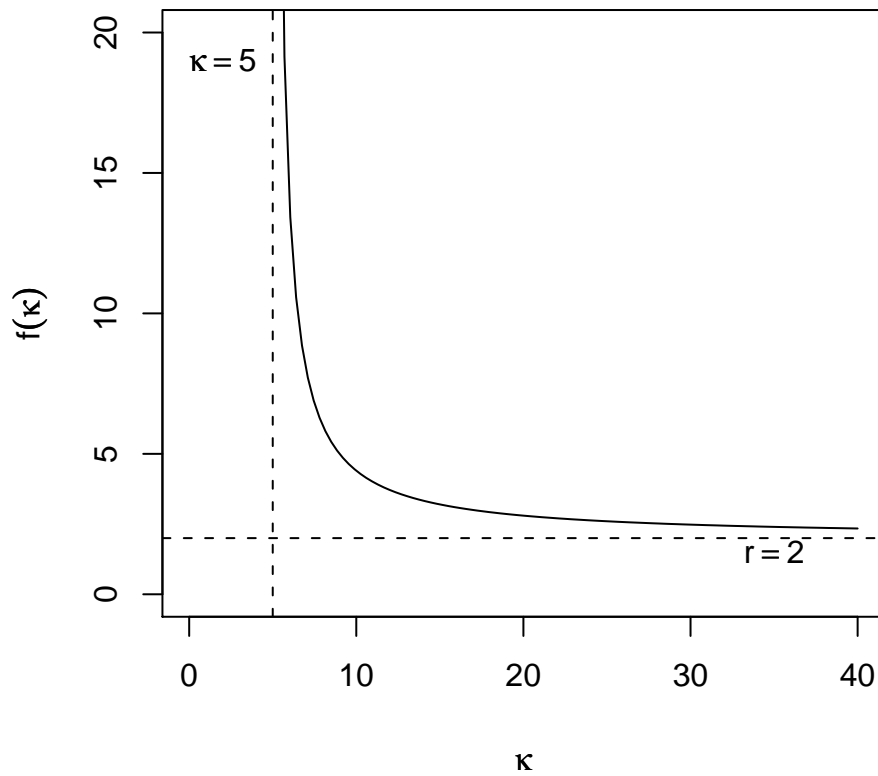
Figure 10: Function $f(\kappa) = \frac{2(\kappa+1)}{\kappa-5}$. For a particular κ, r needs to be greater than this function in order to have more power (for a given cost) than the power of the design with $r = 1$ (under LDD and CS).

recruited (Appendix A.10.3). If, for example, $\kappa = 10$ then with $r \geqslant 5$ measures one would have more power for the same cost than with $r = 1$ (figure 10). Note that large values of κ are needed to justify taking $r = 3, 4$ or 5, values that are common in many studies.

When $V(t_0) > 0$ under CS covariance structure, we computed the optimal r over a grid of values of the other parameters. We found that when $V(t_0) > 0$, r_{opt} can be greater than one even when $\kappa < 5$. In addition, the optimal r rarely reached 15 except for small values of $\sqrt{V(t_0)}/\tau$ coupled with large κ. So, for most of the combinations considered, the optimal r was usually an intermediate values between one and fifteen. The effect of $\sqrt{V(t_0)}/\tau$ on r_{opt} was not monotone – in some cases it was found to increase r_{opt}, and in others to decrease it. Figure 11 when $\theta = 0$ exemplified this for particular values of the parameters. For particular cases, the optimal (N, r) can be obtained with our program.

Under DEX covariance structure, with $V(t_0) = 0$, the optimal value of r that maximizes the power for a given cost also has a complicated expression. We computed the optimal over a grid of possible values of the other parameters, and observed that compared to CS, the optimal r was smaller when $\theta > 0$ (figure 12). So, larger values of κ are needed to justify taking the same number of repeated measures as would optimize the design under CS. We found in our grid search that the optimal r increased as τ or κ increased, and when ρ decreased. However, the effect of θ was found to not be monotone for large values of τ (Figure 12). The optimal r and N for different values of the parameters can be computed with our program.

Under DEX and $V(t_0) > 0$, we computed the optimal design for a grid of values of the other parameters. We found that the effect of $V(t_0)$ on r_{opt} was not monotone – it both increased r_{opt} and decreased it, depending on values of the other parameters. Figure 11 exemplifies some cases for particular values of the parameters. When $V(t_0) > 0$, the optimal r was not always smaller than in the otherwise analogous situation but under
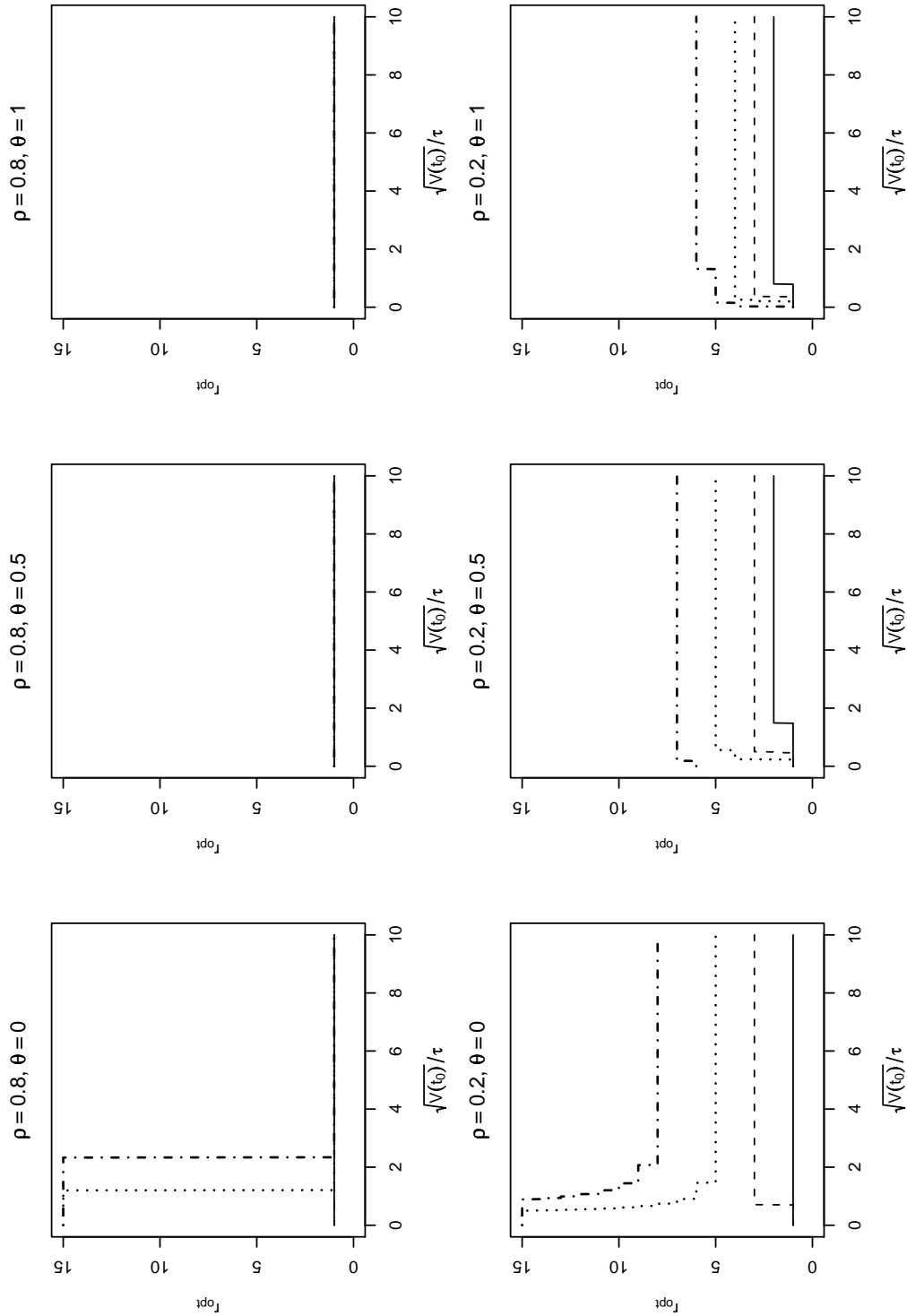
46

Figure 11: Optimal r as a function of $\sqrt{\hat{V}(t_0)}/\tau$ under LDD, DEX, fixed $\tau = 5$, $\rho_{e,t_0} = 0$. Lines indicate $\kappa = 2$ (—), $\kappa = 5$ (- - -), $\kappa = 10$ (·······), $\kappa = 20$ (·-·-·-·).
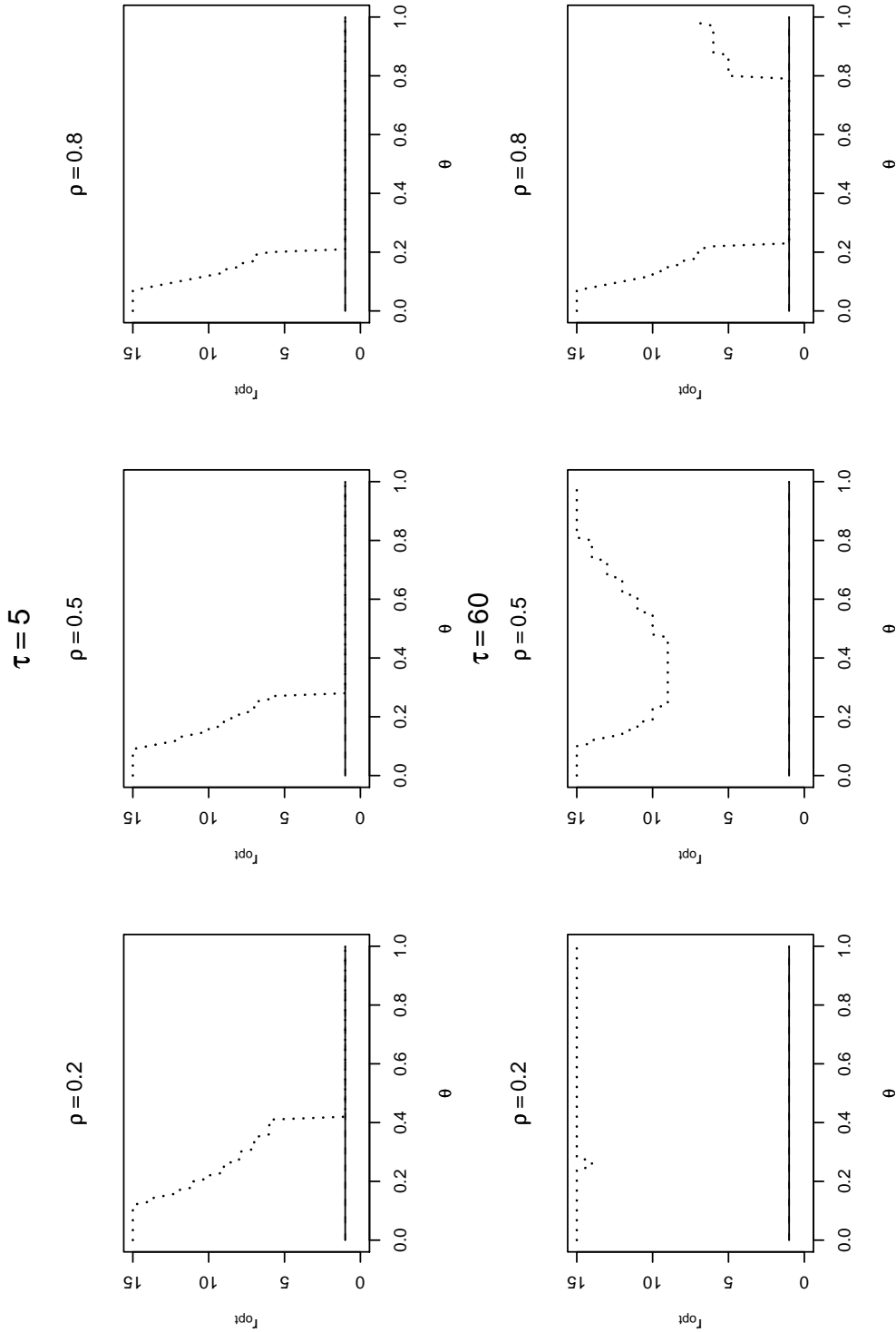
Figure 12: Optimal r as a function of θ under DEX, LDD and fixed follow-up period τ, for $r \leq 15$, $V(t_0) = 0$ and different cost ratios ($\kappa = 2$ (—), $\kappa = 5$ (- - -), and $\kappa = 20$ ($\cdots\cdots$)).

CS (i.e. when $\theta = 0$). Our program can compute the optimal value for given values of the parameters.

Under RS covariance with $V(t_0) = 0$ involves the following condition must be met for the optimal r to be greater than one,

$$\rho_{b_1,\tau,\tilde{r}} < [-2(\kappa + 1) + (\kappa - 5)r](\tilde{r} + 1)(\tilde{r} + 2)$$
$$\left\{6\tilde{r}(r + 1)(r + 2) + [-2(\kappa + 1) + (\kappa - 5)r](\tilde{r} + 1)(\tilde{r} + 2)\right\}^{-1}$$

(Appendix A.10.4). Noting that this can only be true when the right hand side of the inequality is positive, we can deduce that $\kappa > 5$ and $r > \frac{2(\kappa+1)}{\kappa-5}$, as in the corresponding CS case (Appendix A.10.4). When the condition is met, then the optimal r is the solution to the following equation

$$\kappa = \left\{r_{opt}(4 + 3r_{opt})(\tilde{r} + 1)(\tilde{r} + 2) + \tilde{r}(r_{opt} + 1)^2(r_{opt} + 2)^2\frac{\rho_{b_1,\tau,\tilde{r}}}{1 - \rho_{b_1,\tau,\tilde{r}}}\right\}\bigg($$
$$\left\{(r_{opt}^2 - 2)(\tilde{r} + 1)(\tilde{r} + 2)\right\}^{-1}.$$

Otherwise, the optimal is $r = 1$. Regions of the design space can be calculated for which at given values of $\rho_{b_1,\tau,r=5}$ and κ, (N_{opt}, r_{opt}) provides a design with more power than the design which takes $r = 1$. The shaded regions in Figure 13 show these regions in some examples. Note that if, for example, $\kappa = 10$ and $\rho_{b_1,\tau,r=5} = 0.3$, there are no values of r that improve the power over that attained at $r = 1$. The optimal r is also plotted in Figure 13. For example, when $\kappa = 20$ and $\rho_{b_1,\tau,r=5} = 0.3$, the optimal r is greater than 15. Our program can calculate the optimal value for a given κ and $\rho_{b_1,\tau,\tilde{r}}$.

With RS and $V(t_0) > 0$, the optimal design depends on $\frac{\sqrt{V(t_0)}}{\tau}$, $\rho_{b_0b_1}$ and ρ_{e,t_0} as well. We computed the optimal design for a grid of values of the other parameters. As $\rho_{b_1,\tau,r=5}$ increased, r_{opt} decreased, i.e. less repeated measurements and more partici- pants were needed. The effect of $\sqrt{V(t_0)}/\tau$ on r_{opt} was observed to not be monotone – it was found to both increase r_{opt} and decrease it, depending on other values of the pa- rameters. Figure 14 shows how r_{opt} varies as a function of $\sqrt{V(t_0)}/\tau$ in some particular
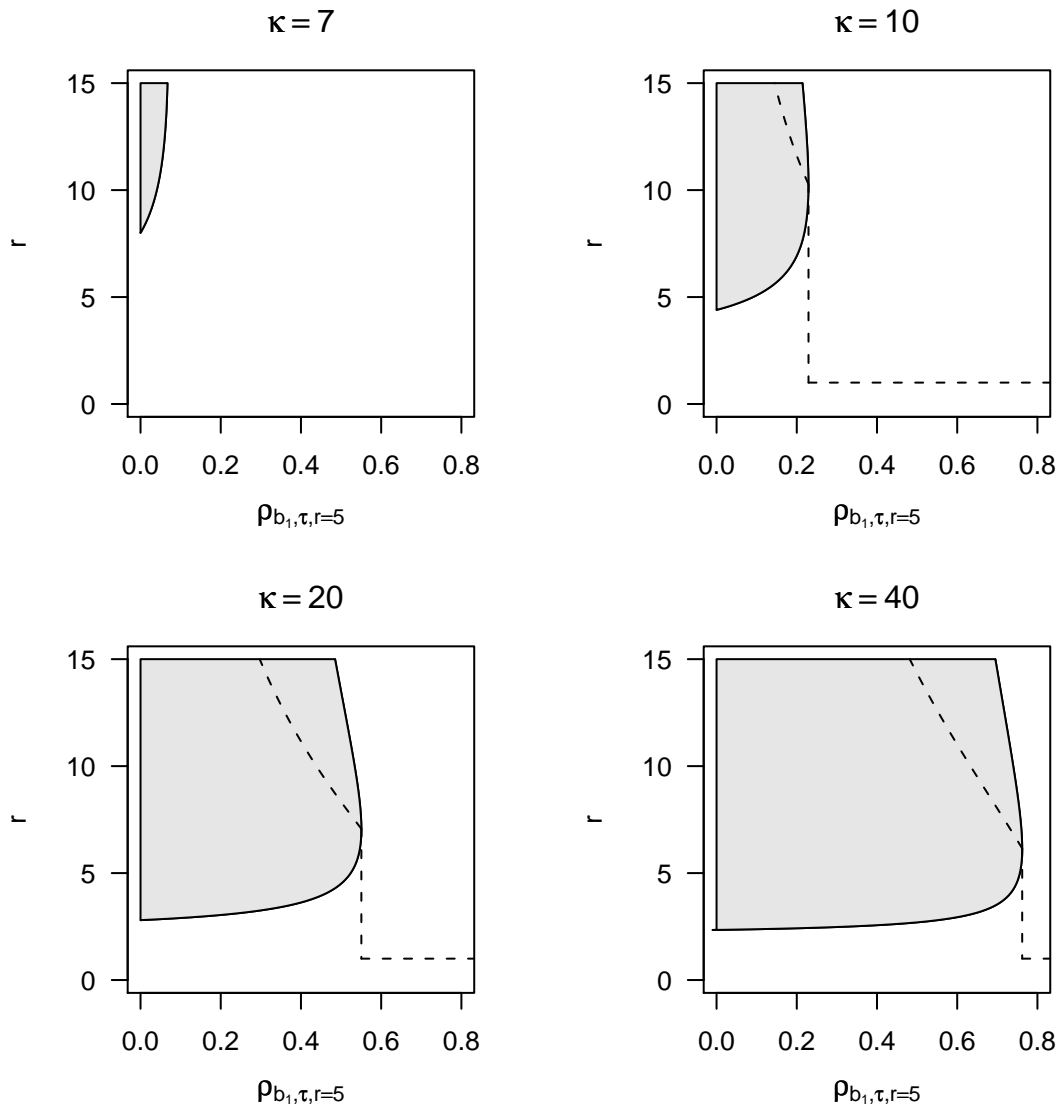
Figure 13: Optimal r (dashed line) and values of r that improve the power attained with $r = 1$ for a fixed cost and same κ and $\rho_{b_1,\tau,r=5}$ (shaded area) under LDD, RS and fixed follow-up time, τ
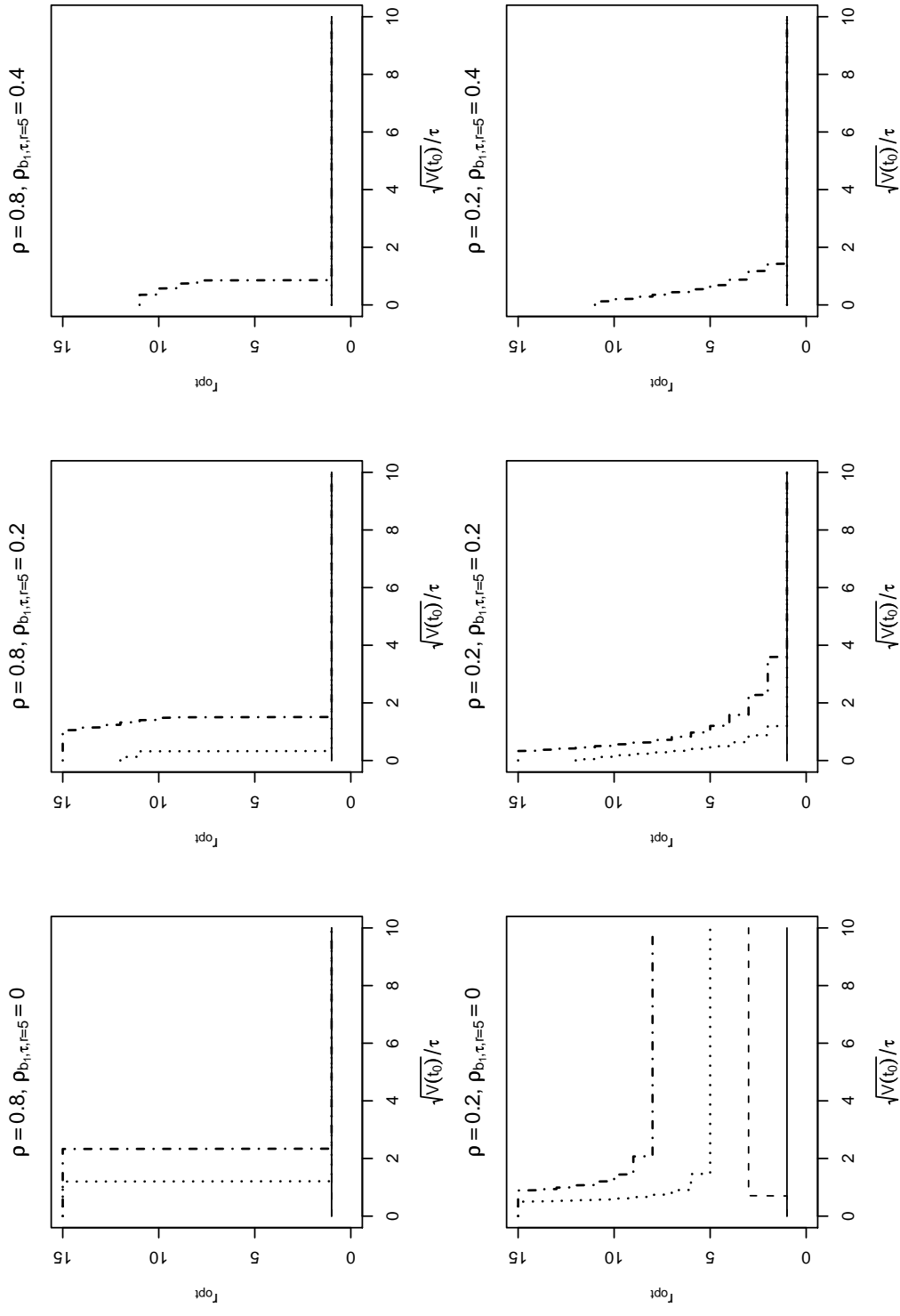
Figure 14: Optimal r as a function of $\sqrt{V(t_0)}/\tau$ under RS and fixed $\tau = 5$, with $\rho_{b_0,b_1} = -0.5$ and $\rho_{e,t_0} = 0$. Lines indicate $\kappa = 2$ (——), $\kappa = 5$ (— —), $\kappa = 10$ (·······), $\kappa = 20$ (·—·—·).

cases cases. When $V(t_0) > 0$, the optimal can be greater than one even when $\kappa < 5$. Our program can compute the optimal value for given values of the parameters.

5 Illustrative Example

To illustrate the methods used in this paper, we consider the subset of the Vlagtwedde-Vlaardingen study (Rijcken et al., 1987; Van der Lende et al., 1981) that was made available on the website of a recent textbook on longitudinal analysis (Fitzmaurice et al., 2004) (`http://biosun1.harvard.edu/~fitzmaur/ala/`) as a pilot study, and use it as the basis of an investigation of options for the design of an expanded longitudinal study of the effect of smoking on lung function. Briefly, these pilot data consist of 133 men and women from rural Vlagtwedde, The Netherlands, aged 36 or older at baseline, who were followed every three years for up to 19 years for evaluation of their lung function, through spirometric measurement of forced expiratory volume (FEV_1). The exposure of interest, current smoking at baseline, was defined as smoking at least one cigarette per day at baseline.

First, as discussed previously, the design of longitudinal study depends on up to nine parameters. To get a realistic idea about the likely range of design input parameters that apply to this study population and its anticipated extension, we fitted a linear model for FEV_1 as a function of smoking, time on study (in years) and the interaction of smoking with time to the pilot data, using CS, DEX and RS covariance structure assumptions. The estimated coefficients of this model assuming a DEX covariance structure model were $\hat{\gamma}_0 = 3.5086$, the average FEV_1 (liters) at baseline among non-smokers, $\hat{\gamma}_1 = -0.2760$, the average yearly rate of decline of FEV_1 (liters) among non-smokers, $\hat{\gamma}_2 = -0.0337$, the average difference in FEV_1 (liters) at baseline between smokers and non-smokers, and $\hat{\gamma}_3 = -0.0045$, the average difference in the yearly rate of decline of FEV_1 (liters) of smokers compared to non-smokers, corresponding to $\mu_{00} = 3.5086$, the average FEV_1 (liters) at baseline among non-smokers, $p_1 = -7.86\%$, the percent differ-

ence in FEV_1 (liters) between smokers and non-smokers at baseline, $p_2 = -18.2\%$, the percent change FEV_1 (liters) from baseline to end of follow-up among non-smokers, and $p_3 = 13.35\%$, the percent difference between the change FEV_1 (liters) from baseline to end of follow-up in smokers and non-smokers.

Getting a good estimate of the residual variance from the model of interest at the design stage of a study is not easy. In order of increasingly likely accuracy, we suggest directly estimating σ^2 from longitudinal pilot data when available. Here, the data are available, and the estimated value of σ^2 from the regression of FEV_1 on baseline smoking status, time in years from start of the study, and their cross-product, was 0.3214 and 0.3179 under assumptions of CS and DEX, respectively. Under RS, $\sigma_{t_0}^2$ was 0.3400. When longitudinal pilot data are not available, as will typically be the case, we suggest using cross-sectional pilot data if available. Here, the estimated value of σ^2 among the exposed only (since the majority are exposed) was 0.3403. If the time metameter for analysis is time since start of study, this value of the residual variance is likely to generate accurate design calculations. If the time metameter for analysis varies at the start of study, e.g. if time is age, it is best to estimate the residual variance over a restricted age range. The pilot data available do not permit estimation of the variance over a restricted age range - hence, in this example, the option is not available, as it often would not be in practice. Often at the design stage, variance values such as those discussed just above, over presumably comparable subjects but over a range of times, however time may be defined, and perhaps pooled across exposed and unexposed subjects, may be all that is available, from the literature or from pilot data. Here, the variance of FEV_1 using all of the measurements for all of the available subjects ($N = 133$) was 0.3837, and the analogous value given in the publication for 1607 subjects, pooled across gender, was 0.3740. Then, conservatively assuming that the ultimate model will explain no more than 10% of this total variation, the investigator may use a value for σ^2 of $0.3837 \times 0.90 = 0.3453$, nearly identical to the analogous value obtained from the fit of the model to the pilot data. For the published value of

the marginal variance, the projected value for the residual variance is 0.3366, also very similar to that obtained directly from the pilot data and to the one obtained from the model fit to the pilot data. When N must be found subject to fixed power and a fixed number of measurements per person (r), it can be seen from equation (3.5) that the percent over- or under-estimation of N is directly proportional to the percent over- or under-estimation of σ^2. When power needs to be calculated as a function of (N, r) or r must be found subject to fixed power and a fixed number of participants (N), we can see from equations (3.1) and (3.6) that the effect of over- or under-estimation of σ^2 on design cannot be easily described. Interestingly, (N_{opt}, r_{opt})does not depend upon σ^2 when the design is constrained by a minimum acceptable power, only when it is constrained by a maximum cost.

Values for other parameters characterizing the covariance structure are needed as well before design calculations can be conducted. If no pilot data are available to estimate them, it is suggested that sensitivity analysis be conducted over what is believed to be a realistic range. Here, we were able to estimate these values from the available longitudinal pilot data. Under the assumption of CS, ρ was 0.857 and 0.896 under DEX, where θ was 0.18. Assuming RS, ρ_{t_0}, was 0.877; $\rho_{b_1, s=3, r=6}$, the slope reliability for $r = 6$ measurements per participant was 0.36, indicating a moderate amount of between-subjects variation in slopes; alternatively, $\rho_{b_1, \tau=18, r=1}$, the slope reliability at the end of follow-up with $r = 1$, was 0.27; and ρ_{b_0, b_1} was -0.32. These RS covariance parameters correspond to $\sigma_w^2 = 0.0418$, $\sigma_{b_0}^2 = 0.2982$, $\sigma_{b_1}^2 = 0.000095$ and $\sigma_{b_0, b_1} = -0.0017$. Finally, most of these pilot study participants were smokers, i.e. p_e, was 0.79, and the published value of the standard deviation of age at entry into the study was 10 years ($\sqrt{V(t_0)}$) (Rijcken et al., 1987).

In Table 4, we show the minimum detectable effects obtained for these trial parameter values under CMD and LDD for $r = 6$. In all the tables given in this section, we assumed, for the RS covariance structure, that t_{0i} is normally distributed. As noted in Section 3.4.2, we found that unless the distribution of t_{0i} was extremely skewed, re-

Table 4: Minimum detectable effects in the pilot study ($N = 133$, $r = 6$, $\tau = 18$, $p_e = 0.79$, $\mu_{00} = 3.5086$, $p_2 = -18.2\%$, $V(t_0) = 100$, $\rho_{e,t_0} = 0$).

Correlation	CMD (p_1)		LDD (p_3)	
Power	80%	90%	80%	90%
CS[1]	$\pm$ 9%	$\pm$ 10%	$\pm$ 22%	$\pm$ 25%
DEX[2]	$\pm$ 9%	$\pm$ 10%	$\pm$ 26%	$\pm$ 30%
RS[3*]	$\pm$ 9%	$\pm$ 10%	$\pm$ 26%	$\pm$ 30%

[1] $\sigma^2 = 0.3214$, $\rho = 0.857$
[2] $\sigma^2 = 0.3179$, $\rho_1 = 0.896$, $\theta = 0.18$
[3] $\sigma_{t_0}^2 = 0.3400$, $\rho_{t_0} = 0.877$, $\rho_{b_1,s,r=6} = 0.36$, $\rho_{b_0,b_1} = -0.32$
[*] t_0 assumed normally distributed with variance $V(t_0)$

Table 5: Minimum number of participants (N) to detect a 10% effect ($p_1 = 0.1$ or $p_3 = 0.1$) with 90% power in the pilot study ($r = 6$, $\tau = 18$, $p_e = 0.79$, $\mu_{00} = 3.5086$, $p_2 = -18.2\%$).

	$V(t_0) = 0$				$V(t_0) = 100$			
	CMD		LDD		CMD		LDD	
ρ_{e,t_0}	0	0.8	0	0.8	0	0.8	0	0.8
CS[1]	151	151	918	918	151	155	863	897
DEX[2]	144	144	1330	1330	144	152	1215	1286
RS[3*]	144	144	1305	1305	147	160	1260	1289

[1] $\sigma^2 = 0.3214$, $\rho = 0.857$
[2] $\sigma^2 = 0.3179$, $\rho = 0.896$, $\theta = 0.18$
[3] $\sigma_{t_0}^2 = 0.3400$, $\rho_{t_0} = 0.877$, $\rho_{b_1,s,r=6} = 0.36$, $\rho_{b_0,b_1} = -0.32$
[*] t_0 assumed normally distributed with variance $V(t_0)$.

sults would be quite insensitive to departures from this assumption. Under CMD, the minimum detectable effect was the same for the three alternate covariance structures, but under LDD, the patterns that allow for the covariance structure to vary with time on study had lower power and larger minimum detectable effects than under CS, as discussed in sections 3.3.2 and 3.4.2. We repeated the table assuming $\rho_{e,t_0} = 0.8$ and obtained almost identical results with the minimum detectable effects slightly higher. Suppose one wants to design a study with seven repeated measures ($r = 6$) taken every three years ($s = 3$). Using the parameters estimated from the pilot data, we computed the number of participants (N) needed to detect a 10% difference in the parameter of interest (p_1), the percent difference in exposure group means which is constant over time, and p_3, the percent difference in the exposure group slopes characterizing their change over time, with 90% power, for values of $V(t_0)$ of zero and 100 and values of ρ_{e,t_0} of zero and 0.8. Results are presented in Table 5. As expected, under CMD we need fewer participants when we have departures from CS, while under LDD departures from CS lead to having to recruit more participants. The departures from CS in these data did not appear to be large (e.g., $\theta = 0.18$), but even this degree of decay had a considerable influence on the required sample sizes. This example suggests that it will often be important for investigators to consider even small departures from CS in their design calculations, and report the maximum departures from CS they are prepared to accommodate in their proposed study. We can also see in Table 5 a small increase in sample size under CMD when both $V(t_0)$ and ρ_{e,t_0} are greater than zero. On the other hand, larger $V(t_0)$ lead to reduction in sample under LDD, which is going to be maximum when $\rho_{e,t_0} = 0$.

Now suppose that neither N nor r are fixed by design and we have a budget of 15,000 monetary units (denoted without loss of generality as \$15,000) for CMD and \$100,000 monetary units for LDD, the cost of recruiting each participant and recording their first measurements is \$80 and the subsequent measures are κ times cheaper. We must distinguish between two possible situations to proceed here: one in which the frequency

of measurements is fixed, and the other, in which the follow-up time is fixed. We restricted consideration for reasons of feasibility to a maximum of $r = 10$ for the fixed frequency case, which would be equivalent to 30 years of follow-up, and to $r = 18$ for the fixed follow-up case, corresponding to one measurement per year. The optimal (N, r) under CMD when $\kappa = 5$ to detect $p_1 = 10\%$ was $(N, r) = (187, 0)$ for all three covariance structures, i.e. the most powerful design for the least amount of money is a cross-sectional study that recruits as many participants as possible. With $\kappa = 20$, the optimal design took one post-baseline measurement from 178 participants, and was again invariant to assumptions about the covariance structure. We obtained a small r_{opt} even for large values of κ because the correlation between measurements on the same participant is large, as seen previously in Figure 6 with CS. For LDD and $s = 3$, the optimal design to detect a 10% difference between exposure group slopes was at the maximum feasible r, here, $r = 10$, with $N = 146$ for all three covariance structures considered, fixed at the same values of the covariance parameters as in Table 5. At this fixed cost of \$100,000, the power for $(N_{opt}, r_{opt}) = (10, 416)$ differed depending on the assumed covariance structure, with a power of 99% for CS, 88% DEX with a small dampening coefficient ($\theta = 0.18$) and 71% for RS with a 69% slope reliability with $r = 10$ ($\rho_{b_1, s=3, r=10}$). With r_{opt}=10 and $s = 3$, the study is planned to be of 30 years duration and in many cases would not be a realistic choice. The optimal combination of r and N was the same when we assumed $V(t_0) = 0$, and the resulting power was only slightly smaller.

Table 6 shows the optimal (N, r) under LDD when the duration of the study is fixed at $\tau = 18$, for $\kappa = 5$ and $\kappa = 20$ and for $V(t_0) = 0$ and $V(t_0) = 100$ and $\rho_{e,t_0} = 0$. We repeated the results with $\rho_{e,t_0} = 0.8$ and they were almost identical. The optimal design for $\kappa = 5$ was at $r_{opt} = 1$, as we observed in section 4.2.2 for $\kappa \leqslant 5$, and this was independent of $V(t_0)$, which only changed slightly the resulting power. For $\kappa = 20$, the optimal design in (N, r) varied considerably depending on the covariance structure, and within the same covariance structure it varied depending on the value of $V(t_0)$.

Table 6: Optimal design (N_{opt}, r_{opt}) to maximize power for a study to detect a 10% difference in slopes ($p_3 = 10\%$) for fixed $\tau = 18$ under LDD at a cost of no more than \$100,000 with $c_1 = \$80$ and $\rho_{e,t_0} = 0$.

		$\kappa = 5$		$\kappa = 20$	
		(N, r)	Power	(N, r)	Power
CS	$V(t_0) = 0$	$(1041, 1)$	79%	$(657, 18)^4$	98%
	$V(t_0) = 100$	$(1041, 1)$	83%	$(657, 18)$	99%
DEX	$V(t_0) = 0$	$(1041, 1)$	73%	$(925, 7)$	79%
	$V(t_0) = 100$	$(1041, 1)$	77%	$(1190, 1)$	82%
RS	$V(t_0) = 0$	$(1041, 1)$	70%	$(757, 13)$	82%
	$V(t_0) = 100^*$	$(1041, 1)$	72%	$(781, 12)$	83%

[1] $\sigma^2 = 0.3214$, $\rho = 0.857$
[2] $\sigma^2 = 0.3179$, $\rho = 0.896$, $\theta = 0.18$
[3] $\sigma_{t_0}^2 = 0.3400$, $\rho_{t_0} = 0.877$, $\rho_{b_1,s,r=6} = 0.36$, $\rho_{b_0,b_1} = -0.32$
[4] Note that with $r = 18$, one measurement will be taken every year, three times more often than in the pilot study
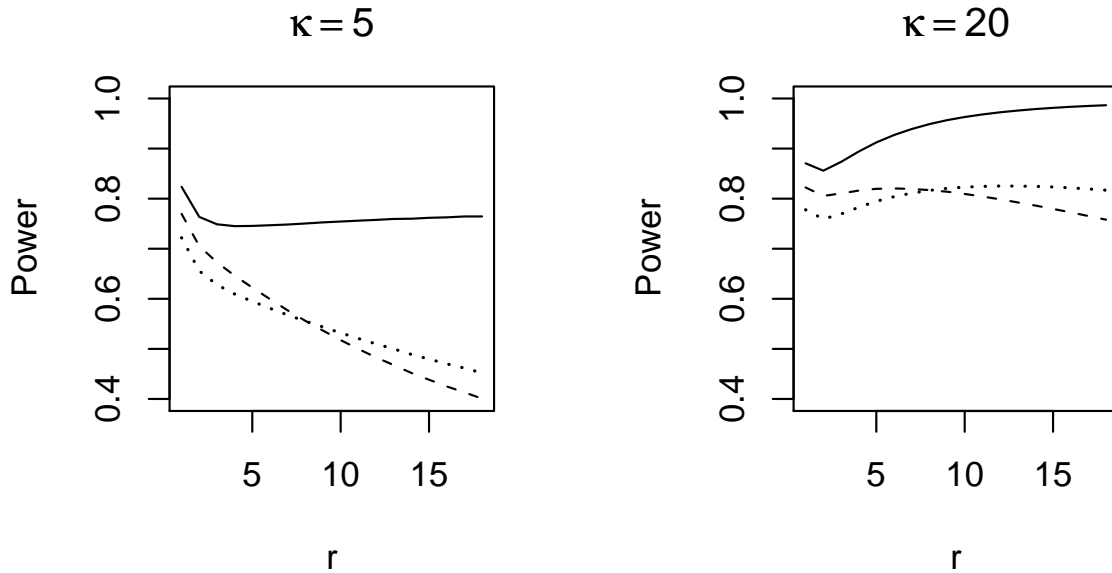* t_0 assumed normally distributed with variance $V(t_0)$.

Figure 15: Power as a function of r under LDD for $p_3 = 10\%$, fixed $\tau = 18$, $V(t_0) = 100$, $\rho_{e,t_0} = 0$ and a cost restriction of \$100,000. The values of the parameters are, for CS, $\sigma^2 = 0.3214$, $\rho = 0.857$; DEX, $\sigma^2 = 0.3179$, $\rho = 0.896$, $\theta = 0.18$; RS, $\sigma_{t_0}^2 = 0.3400$, $\rho_{t_0} = 0.877$, $\rho_{b_1,s,r=6} = 0.36$, $\rho_{b_0,b_1} = -0.32$. The lines indicate CS (——), DEX (- - -), and RS (······).

Figure 15 shows how power varied as a function of r_{opt} (and N_{opt}) for fixed study cost and different values of κ. When κ is large, the optimal combination (N, r) strongly depended on the covariance structure, and combinations that achieved a high power for one covariance structure were underpowered for others.

6 Software

Although others have provided public access software for longitudinal study design that is applicable to observational studies (Hedeker et al., 1999; Raudenbush et al., 2005; Snijders, 20033; Ziegler, 2004), nothing currently available comprehensively addresses all the cases considered in this paper, as they might arise in epidemiologic applications. Table 7 summarizes the features of the existing programs, as well as our own. Because program OD (Raudenbush, 1997; Raudenbush et al., 2005; Raudenbush and Xiao-Feng, 2001) only supports 'balanced designs' where $p_e = 0.5$, we do not consider it further here as such balance will rarely occur in an observational setting. Our program performs all the calculations described in this paper, and is publicly available, along with a user's guide, at the second author's website (`http://www.hsph.harvard.edu/faculty/spiegelman/optitxs.html`). The program runs in the R statistical package (R Development Core Team, 2006) which can be downloaded for free (`http://www.r-project.org`). Our program has an interactive user interface that queries the user for the optimal design scenario to consider, and for the relevant inputs for that scenario; no knowledge of R is required to run our program. The program has modules to calculate power, number of participants when number of repeated measures is fixed, number of repeated measures when number of participants is fixed, minimum detectable difference and optimal number of repeated measures and participants under budget constraints. A demonstration of its use to compute the optimal (N, r) under LDD and RS is shown in Appendix A.11.

Table 7: Summary of the features of existing programs.

Software	Reference	CMD	LDD	CS	RS	DEX	$V(t_0) > 0$	Exposure and time correlated	Optimal (N,r) for fixed cost and/or fixed power	Power	N for fixed r	r for fixed N	Minimum detectable effect
PINT	Snijders (1993, 2003)	√	√	√	√	×	×*	×	√ (fixed cost)	It computes the standard errors, $Var\left(\hat{\beta}_2\right)$ and $Var\left(\hat{\gamma}_3\right)$			
RMASS2	Hedeker (1999a, 1999b)	√	√	√	√	×	×	×	×	×	√	×	×
GEESIZE	Rochon (1998), Ziegler (2004)	√	√	√	×	√	×	×	×	×	√	×	×
OPTITXS	Basagaña and Spiegelman (2007)	√	√	√	√	√	√	√	√ (both constraints)	√	√	√	√

* Only considers B&W model, which reduces to the $V(t_0) = 0$ case (Appendix A.1.3)

7 Conclusions

The power of a longitudinal study varies with several factors. Independent of assumptions about the covariance structure, power increases as the number of participants, the number of repeated measurements and the length of follow-up increase, and when the exposure prevalence approaches 0.5 from either direction. However, in many situations, we found that one cannot always achieve a pre-specified power solely by increasing the number of repeated measurements, since power sometimes reaches an asymptote. In practice, the length of follow-up or the time interval between successive measurements may be fixed or be restricted to a limited range of values. Sometimes, N may be fixed (e.g. when planning longitudinal follow-up of an existing cross-sectional study) and r needs to be determined. In other cases, r may be fixed (e.g. when the study is to be based on data collected during monthly clinic visits in an ongoing randomized trial of multivitamin and vitamin A supplementation among HIV-infected women in Dar es Salaam, Tanzania, for which funding can be obtained for no more than five years (Villamor et al., 2002) and N will need to be determined. In still other situations, (N, r) is fixed (e.g. in the Nurses' Health Study where the number of nurses, the interval between successive measurements and the duration of follow-up is all fixed (Koh-Banerjee et al., 2003)), and one may need to determine power for a range of effect sizes given the data available. These problems can all be solved using the program developed in this paper to implement the calculations discussed.

The power of a longitudinal study, the required number of participants and the required number of repeated measures all vary with assumptions about the covariance structure of the outcome variable as it evolves over time. For a difference between exposure groups that is constant over time (CMD), the lower the correlations between observations within the same participant, the more powerful the study. That is, a study with observations that are less correlated would have more power for the same

number of participants and repeated measures; the study would need to recruit fewer participants to achieve the same power with a fixed number of repeated measures; and in most situations, a smaller number of repeated measures would be needed to achieve a pre-specified power for a fixed number of participants. If, in addition, there is a correlation decay compatible with DEX, the study would have more power for fixed N and r than a study with CS covariance, and it would require fewer participants and fewer repeated measures to achieve the same power. The effect of departures from CS compatible with RS cannot be summarized in a straightforward manner.

When there is a linear interaction between time and exposure (LDD), higher correlations between repeated measurements within participants lead to increased power for the same N and r, and require fewer participants and repeated measures to achieve the same power. We showed that departures from CS towards DEX or RS decreased power and increased the required number of participants and repeated measures, although the effect of θ on power and sample size was not monotone. These relationships under LDD may change when participants enter the study at different times (e.g. when age is the time variable of interest and participants begin follow-up at different ages). After providing tentative design parameters from pilot data, the literature, and intuition, investigators can perform a sensitivity analysis for departures from compound symmetry in their sample size or power calculations using our program.

In practice, it is often the case that neither N nor r are fixed a priori. Then, there is an infinite set of combinations of (N, r) that achieve the same power. However, by taking cost into account, one can find the optimal combination of (N, r) to maximize the power subject to a particular budget or to minimize the budget subject to a fixed power. We solved this problem when s is fixed, where taking more repeated measures increases the length of follow-up, and when τ is fixed, where increasing r reduces the frequency of measurement. For CMD, if all observations have the same cost, one would not take repeated measures, i.e. one will design a cross-sectional study recruiting as many participants as possible. If the subsequent measures of the same

participant are cheaper than the first, it may be advisable to take repeated measures and recruit fewer participants. However, no repeated measures or just a small number of them are necessary when the correlation between measures of the same participant is large, as is often the case with longitudinal data. When deviations from CS exist, it becomes more advisable to take at least one repeated measure. Under LDD, if the follow-up period is not fixed, we showed that the optimal design has the maximum length of follow-up possible, except when a RS covariance structure is assumed, in which case the optimal combination of (N, r) needs to be computed. If the follow-up period is fixed, we showed that the optimal design takes more than one repeated measure only when the subsequent measures are more than five times cheaper. When there are departures from CS, we showed that values of κ around 10 or 20 are needed to justify taking 3 or 4 repeated measures. It must be pointed out that the optimal (N, r) and the resulting power strongly depended on the covariance structure in the extensive cases we examined, and combinations that are optimal for one assumed covariance structure were often quite poor for another. For particular problems, investigators should perform sensitivity analysis for departures from the assumed values of covariance parameters.

Of course, these recommendations for optimizing (N, r) are based purely on considerations of power and cost. There are several other reasons to collect more repeated measurement than the optimal. For example, it might be useful to schedule regular visits during the whole follow-up period in order to minimize dropout and to have intermediate data for those who will eventually drop out. Another reason to collect many repeated measurements is when non-linearity of the response profile over time is anticipated. If time is mismodeled, and if time is not independent of exposure, i.e. if study participants do not adhere to visit schedules in a way that is associated with exposure, $\hat{\beta}_2$ and $\hat{\gamma}_3$ will be biased. Another situation where repeated measures are essential is when the exposure is time-varying, a situation that we will discuss in a subsequent paper and is found, for example, in crossover studies (Jones and Kenward,

1989; Senn, 2002).

We considered scenarios where the participants do not have the same initial time value, which can be the case, for example, when participants from different ages are recruited and age is the time variable of interest. This requires two new parameters to be provided, $V(t_0)$ and ρ_{e,t_0}. Under CMD, having $V(t_0) > 0$ reduces the power of the study, unless $\rho_{e,t_0} = 0$. Under LDD, however, as $V(t_0)$ increases so does the power of the study, and the effect is stronger the closer ρ_{e,t_0} is to zero. This increase in power is due to a bigger range of observation of data used to estimate the time by exposure interaction term. The situation where $V(t_0) > 0$ resembles the accelerated longitudinal design (Bell, 1953), where multiple age cohorts are sampled and followed. Then, the response profiles of the exposed and the unexposed are modeled. In some applications, when cohort or period effects are suspected to occur that cannot be adjusted for through standard individual-level covariate adjustments, or when between-subjects confounders (time-invariant confounders) are mis- or un-measured, B&W models that separate within- and between-subjects effects can be fit (Neuhaus and Kalbfleisch, 1998; Ware et al., 1990). If those models are to be used, the results for the LDD case with $V(t_0) = 0$ derived in this paper apply and the power gained from $V(t_0) > 0$ is lost.

We based our formulas on the asymptotic variance of the estimator, in which case only the first two moments of the covariates and not its full distribution need to be provided. Other approaches have been used. When power is the object of interest, calculations can be based upon the expected value of the non-centrality parameter over the distribution of the covariates $\mathbf{X}$ (Lachin, 2000, chapter 3), where the test uses the conditional variance of $\hat{B}$ given the covariates, i.e. computing

$$
\Phi\left[\sqrt{\mathbb{E}_X\left\{\frac{((\mathbf{c}'\mathbf{B})_{H_A})^2}{\mathbf{c}'\left(\sum_{i=1}^{N}\mathbf{X}'_i\mathbf{\Sigma}_i^{-1}\mathbf{X}_i\right)^{-1}\mathbf{c}}\right\}} - z_{1-\alpha/2}\right].
$$

64

When only exposure is random, this approach results in a $(1 - 1/N)$ correction over the approach used in this paper, which is negligible for large N, and equation (3.1) becomes

$$\Phi\left[\frac{\sqrt{N-1}\,\left|(\mathbf{c}'\mathrm{B})_{H_A}\right|}{\sqrt{\mathbf{c}'\Sigma_{\mathrm{B}}\mathbf{c}}} - z_{1-\alpha/2}\right]$$

(Appendix A.1.1). To find N for fixed r and power, or to find r for fixed N and power using Lachin's approach, one would need to solve the power equation provided above for the desired quantity. To find the optimal (N, r) one would need to maximize the power equation given above subject to the cost constraint. Another approach would be to compute the expected value of the power formula over the distribution of $\mathbf{X}$, where again the conditional variance of $\hat{\mathrm{B}}$ given the covariates is used, i.e.

$$\mathbb{E}_X\left(\Phi\left[\frac{\sqrt{N}\,\left|(\mathbf{c}'\mathrm{B})_{H_A}\right|}{\sqrt{\mathbf{c}'\left(\sum_{i=1}^{N}\mathbf{X}'_i\Sigma_i^{-1}\mathbf{X}_i\right)^{-1}\mathbf{c}}} - z_{1-\alpha/2}\right]\right)$$

(Glueck and Muller, 2003). Since the actual power for the particular sample that is finally recruited can be different from the average power, more conservative quantities such as the 0.025 quantile of power (Glueck and Muller, 2003) over the distribuition of $\mathbf{X}$ might be more appropriate, but the results require the development of a numerical algorithm for accurately computing these quantiles. Similarly, one can derive the required N by taking the expected value of the sample size formula using the variance of $\hat{\mathrm{B}}$ conditional on the covariates,

$$N = \frac{\mathbf{c}'\left(\sum_{i=1}^{N}\mathbf{X}'_i\Sigma^{-1}\mathbf{X}_i\right)^{-1}\mathbf{c}\,\left(z_\pi + z_{1-\alpha/2}\right)^2}{(\mathbf{c}'\mathrm{B}_{H_A})^2},$$

over the distribution of $\mathbf{X}$. This leads to computing

$$\mathbb{E}_{\mathbf{X}}\left[\mathbf{c}'\,\left(\sum_{i=1}^{N}\mathbf{X}'_i\Sigma^{-1}\mathbf{X}_i\right)^{-1}\mathbf{c}\right],$$

which, for the case where only exposure is random, is infinity, so this approach would not useful for this case. This approach would be hard to apply to find r, since no

explicit functions for r exist for most situations. To find the optimal (N, r), one would need to maximize the power equation given above subject to the cost constraint. All these approaches should produce very similar results to the approach followed in this paper, with the advantage that with our approach the distribution of the covariates does not need to be specified and that simpler formulas are obtained.

In this paper, we assumed that all study participants are observed at the same scheduled time points, even though their initial age can vary (i.e. only initial age is random), because this most closely resembles the experience typically encountered in epidemiological studies. Jung and Ahn (2003) studied the effect of dispersed times around the scheduled time points, and found that even in the random measurement times case, the sample size calculated under fixed measurement time assumption had an empirical power very close to the nominal power.

Another factor which could be considered when designing observational longitudinal studies is the dropout rate. Several studies looked at this issue (Dawson, 1998; Galbraith and Marschner, 2002; Hedeker et al., 1999; Jung and Ahn, 2003; Yi and Panzarella, 2002) It was reassuring that Galbraith and Marschner (2002) found in a simulation study that in a study of the power to detect group by exposure interactions (i.e. LDD) under RS, as long as the total percentage of lost to follow-up was no more than 30%, then a study designed to achieve 90% power ignoring dropout will generally achieve an actual power of at least 80%. Another simple method consists of inflating the number of participants by a factor of $1/(1 - f)$, where f is the anticipated fraction of lost to follow-up (Fitzmaurice et al., 2004, page 409). Further research is needed to assess how dropout would affect the optimal combination (N, r), since the presence of dropout alters the inverse relationship between number of participants and number of time points at the same power (Hedeker et al., 1999).

In summary, we provided formulas for power, sample size, number of repeated measures, and the optimal combination of participants and repeated measures are pro-

vided for longitudinal studies with a continuous response and a binary time-invariant exposure. Our results extend to the case where, for example, age is the time measure of interest, instead of time in the study. General results were derived for the effects of the parameters involved in the calculations whenever possible. The parameters are formulated in an intuitive way to facilitate the choice of appropriate values and to ease sensitivity analyses. Using our publicly available program, users can perform all the design calculations described in the paper and we encourage investigators to make use of this, rather than relying on generalizations which may or may not apply in a particular situation, given the complexity of what needs to be considered in exploring the optimal design.

References

R. Q. Bell. Convergence: an accelerated longitudinal approach. *Child Dev*, 24(2):145–52, 1953.

D. A. Bloch. Sample size requirements and the cost of a randomized clinical trial with repeated measurements. *Stat Med*, 5(6):663–7, 1986.

William Gemmell Cochran. *Sampling techniques*. Wiley, New York, 3d edition, 1977.

J. D. Dawson. Sample size calculations based on slopes and other summary statistics. *Biometrics*, 54(1):323–30, 1998.

J. D. Dawson and S. W. Lagakos. Size and power of two-sample tests of repeated measures data. *Biometrics*, 49(4):1022–32, 1993.

Peter Diggle, P. Heagerty, K. Y. Liang, and S. Zeger. *Analysis of longitudinal data*. Oxford statistical science series 25. Oxford University Press, Oxford; New York, 2nd edition, 2002.

Garrett M. Fitzmaurice, Nan M. Laird, and James H. Ware. *Applied longitudinal analysis*. Wiley series in probability and statistics. Wiley-Interscience, Hoboken, N.J., 2004.

L. Frison and S. J. Pocock. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Stat Med*, 11(13):1685–704, 1992.

L. J. Frison and S. J. Pocock. Linearly divergent treatment effects in clinical trials with repeated measures: efficient analysis using summary statistics. *Stat Med*, 16(24): 2855–72, 1997.

S. Galbraith and I. C. Marschner. Guidelines for the design of clinical trials with longitudinal outcomes. *Control Clin Trials*, 23(3):257–73, 2002.

D. H. Glueck and K. E. Muller. Adjusting power for a baseline covariate in linear models. *Stat Med*, 22(16):2535–51, 2003.

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(1):70–93, 1999.

Byron Jones and Michael G. Kenward. *Design and analysis of cross-over trials*. Monographs on statistics and applied probability; 34. Chapman and Hall, London; New York, 1st edition, 1989.

S. H. Jung and C. Ahn. Sample size estimation for gee method for comparing slopes in repeated measurements data. *Stat Med*, 22(8):1305–15, 2003.

A. J. Kirby, N. Galai, and A. Munoz. Sample size estimation using repeated measurements on biomarkers as outcomes. *Control Clin Trials*, 15(3):165–72, 1994.

P. Koh-Banerjee, N. F. Chu, D. Spiegelman, B. Rosner, G. Colditz, W. Willett, and E. Rimm. Prospective study of the association of changes in dietary intake, physical activity, alcohol consumption, and smoking with 9-y gain in waist circumference among 16 587 us men. *American Journal of Clinical Nutrion*, 78:719–27, 2003.

John M. Lachin. *Biostatistical methods: the assessment of relative risks*. Wiley series in probability and statistics. Wiley, New York, 2000.

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

A. Munoz, V. Carey, J. P. Schouten, M. Segal, and B. Rosner. A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics*, 48(3):733–42, 1992.

J. M. Neuhaus and J. D. Kalbfleisch. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics*, 54(2):638–45, 1998.

J. E. Overall and S. R. Doyle. Estimating sample sizes for repeated measurement designs. *Control Clin Trials*, 15(2):100–23, 1994.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006.

S. W. Raudenbush. Statistical analysis and optimal design for cluster randomized trials. *Psychol Methods*, 2(2):173–85, 1997.

S. W. Raudenbush and L. Xiao-Feng. Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychol Methods*, 6(4):387–401, 2001.

S. W. Raudenbush, J. Spybrook, L. Xiao-Feng, and R. Congdon. Optimal design for longitudinal and multilevel research. *http://www.ssicentral.com/otherproducts/othersoftware.html*, 2005.

B. Rijcken, J. P. Schouten, S. T. Weiss, F. E. Speizer, and R. van der Lende. The relationship of nonspecific bronchial responsiveness to respiratory symptoms in a random population sample. *Am Rev Respir Dis*, 136(1):62–8, 1987.

J. Rochon. Application of gee procedures for sample size calculations in repeated measures experiments. *Stat Med*, 17(14):1643–58, 1998.

J. J. Schlesselman. Planning a longitudinal study. ii. frequency of measurement and study duration. *J Chronic Dis*, 26(9):561–70, 1973.

H. J. Schouten. Planning group sizes in clinical trials with a continuous outcome and repeated measures. *Stat Med*, 18(3):255–64, 1999.

S. R. Searle. *Linear models*. Wiley, New York,, 1971.

Stephen Senn. *Cross-over trials in clinical research*. J. Wiley, Chichester, Eng.; New York, 2nd edition, 2002.

T.A.B. Snijders. Pint. *http://stat.gamma.rug.nl/snijders*, 20033.

T.A.B. Snijders and R.J. Bosker. Standard errors and sample sizes for two-level research. *Journal of Educational Statistics*, 18(3):237–259, 1993.

T.A.B Snijders and R.J Bosker. Modeled variance in two-level models. *Sociological methods and research*, 22(3):342–63, 1994.

B. Takkouche, C. Cadarso-Suarez, and D. Spiegelman. Evaluation of old and new tests of heterogeneity in epidemiologic meta-analysis. *Am J Epidemiol*, 150(2):206–15, 1999.

X. M. Tu, J. Kowalski, J. Zhang, K. G. Lynch, and P. Crits-Christoph. Power analyses for longitudinal trials and other clustered designs. *Stat Med*, 23(18):2799–815, 2004.

R. Van der Lende, T.J. Kok, R. Peset, P. H. Quanjer, J. P. Schouten, and N.G.M. Orie. Decreases in vc and fev1 with time: Indicators for effects of smoking and air pollution. *Bulletin of European Physiopathology and Respiration*, 17:775–92, 1981.

E. Villamor, G. Msamanga, D. Spiegelman, G. Antelman, K. E. Peterson, D. J. Hunter, and W. W. Fawzi. Effect of multivitamin and vitamin a supplements on weight gain during pregnancy among hiv-1-infected women. *Am J Clin Nutr*, 76:1082–90, 2002.

J. H. Ware, D. W. Dockery, T. A. Louis, X. P. Xu, Jr. Ferris, B. G., and F. E. Speizer. Longitudinal and cross-sectional estimates of pulmonary function decline in never-smoking adults. *Am J Epidemiol*, 132(4):685–700, 1990.

Q. 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. *Control Clin Trials*, 23(5):481–96, 2002.

A. Ziegler. Geesize. *http://www.imbs.uni-luebeck.de/pub/Geesize*, 2004.