Research Article

Statistics in Medicine

Received 26 January 2009, Accepted 22 September 2009 Published online 6 November 2009 in Wiley Interscience

(www.interscience.wiley.com) DOI: 10.1002/sim.3772

Power and sample size calculations for longitudinal studies comparing rates of change with a time-varying exposure‡

X. Basagaña[a,b,c] and D. Spiegelman[d]*†

Existing study design formulas for longitudinal studies have assumed that the exposure is time-invariant. We derived sample size formulas for studies comparing rates of change by exposure when the exposure varies with time within a subject, focusing on observational studies where this variation is not controlled by the investigator. Two scenarios are considered, one assuming that the effect of exposure on the response is acute and the other assuming that it is cumulative. We show that accurate calculations can often be obtained by providing the intraclass correlation of exposure and the exposure prevalence at each time point. When comparing rates of change, studies with a time-varying exposure are, in general, less efficient than studies with a time-invariant one. We provide a public access program to perform the calculations described in the paper (http://www.hsph.harvard.edu/faculty/spiegelman/optitxs.html).

Keywords: longitudinal study; observational study; rate of change; repeated measures; study design; time-dependent

1. Introduction

Studies involving repeated measurements of a binary exposure typically aim to compare the rate of change in response in the two exposure groups, or equivalently, the interest is in an exposure-by-time interaction. Study design formulas for such studies have mostly been derived in the context of clinical trials, and therefore assume that the exposure is time-invariant [1–11]. When the exposure varies within a subject, some methods have been developed when the interest is in the main effect of exposure, for example in the crossover designs [12, 13] and in the multicenter clinical trials with randomization at the patient level [14]. However, these methods do not apply to observational studies, where exposure or treatment is not assigned by the investigator. When the interest is in the exposure-by-time interaction, no study design formulas have previously been developed for longitudinal studies with a time-varying exposure, to the best of our knowledge.

In this paper, we derive study design formulas that are valid when the aim is to compare rates of change in response in relation to a time-varying exposure. Our formulas are derived for a binary exposure and for equidistant time points. The formulas we derive here are motivated by applications in the observational studies, where the exposure is not assigned by design and many exposure patterns may be observed, with variation in the number of exposed periods per participant and changes in the cross-sectional prevalence of the exposure over time. However, all results readily apply to the controlled setting as a simple special case. We then compare the formulas we derive for the time-varying case with the existing ones for the time-invariant case, so that the investigators can anticipate when they will need to recruit more or less participants because their exposure is time-varying instead of time-invariant. This paper is structured as follows. In Section 2, we introduce notation and several models

[a]*Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain*

[b]*Municipal Institute of Medical Research (IMIM-Hospital del Mar), Barcelona, Spain*

[c]*CIBER Epidemiologia y Salud Publica (CIBERESP), Barcelona, Spain*

[d]*Departments of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, U.S.A.*

**Correspondence to: D. Spiegelman, Departments of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, U.S.A.*

†*E-mail: stdls@channing.harvard.edu*

‡*Supporting information may be found in the online version of this article.*

Contract/grant sponsor: NIH; contract/grant number: CA06516

that extend the rates of change comparison to the time-varying exposure setting. In Section 3, we derive expressions for the variance of the coefficient of interest for each model considered, in order to obtain the formula for the test statistic upon which power and sample size calculations are based. In Section 4, we assess the effect of changing some of the assumptions, including time-varying vs time-invariant exposure, on sample size. In Section 5, the methods are applied to the design of a real study. In Section 6, the results and their implications are discussed. In addition, we provide public access software to perform all the calculations discussed in this paper (`http://www.hsph.harvard.edu/faculty/spiegelman/optitxs.html`).

2. Notation and models

2.1. Notation

Let Y_{ij} be the outcome of interest for the measurement taken at the jth ($j=0,\ldots,r$) time for the ith ($i=1,\ldots,N$) participant, and E_{ij} represent the exposure for the period between the measurements of $Y_{i,j-1}$ and Y_{ij}. Thus, r is the number of post-baseline measurements of the response per participant, or, equivalently, the total number of measurements per participant is $r+1$. We consider studies that obtain repeated measures every s time units, as is the usual design in epidemiologic studies. Let t_{i0} be the initial time for participant i and $V(t_0)$ be its variance. When $V(t_0)=0$, all participants have the same time vector, as when using time since enrollment in the study as the time variable of interest. However, when age is the time metameter of interest, as is often the case in epidemiology, and when, in addition, participants enter the study at different ages, $V(t_0)>0$. We base our results on linear models of the form $\mathbf{Y}_i=\mathbf{X}_i\boldsymbol{\Gamma}$ ($i=1,\ldots,N$), for some covariate matrix $\mathbf{X}_i$ ($(r+1)\times q$), where q is the number of variables in the model, and $\boldsymbol{\Gamma}$ is a vector of unknown regression parameters. The $(r+1)\times(r+1)$ residual covariance matrix is $\mathrm{Var}(\mathbf{Y}_i|\mathbf{X}_i)=\boldsymbol{\Sigma}_i$ ($i=1,\ldots,N$). Note that $\boldsymbol{\Sigma}_i$ can be any valid covariance matrix, and can include terms associated with between-subjects variability as well as within-subjects variation. We base our development on the generalized least-squares (GLS) estimator of $\boldsymbol{\Gamma}$, which has the form

$$\hat{\boldsymbol{\Gamma}}=\left(\frac{1}{N}\sum_i \mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right)^{-1}\left(\frac{1}{N}\sum_i \mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{Y}_i\right)$$

Since the design matrix is not known *a priori*, following Whittemore [15] and Shieh [16], study design calculations use $(1/N)\boldsymbol{\Sigma}_\Gamma$ as the variance of $\hat{\boldsymbol{\Gamma}}$, where

$$\boldsymbol{\Sigma}_\Gamma=(\mathbb{E}_X[\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i])^{-1} \tag{1}$$

As long as $\boldsymbol{\Sigma}_i$ does not depend on the covariates, (1) can be fully specified by the first- and second-order moments of the covariate distribution [17].

2.2. Extending the model for the time-varying exposure

When the exposure is time-invariant, the usual model is

$$\mathbb{E}(Y_{ij}|X_{ij})=\gamma_0+\gamma_t t_{ij}+\gamma_e E_i+\gamma_{te}(t_{ij}\times E_i) \tag{2}$$

where the interest is in the exposure-by-time interaction. Figures 1(a)–(c) illustrate some possible trajectories that could occur when the rate of change depends on the exposure. The left panels show the trajectories for a time-invariant exposure, which occurs with a linear exposure-by-time interaction as described in (2).

The model can be extended to the time-varying exposure case in at least two ways: one that assumes that the effect of exposure on the response is cumulative and the other that assumes that it is acute and transient. The new model equations for these two generalizations are given below. For both options, the response trajectory of participants whose exposure status does not vary with time is equal to the trajectories shown in the left panels of Figures 1(a)–(c), and this trajectory will be the same whether we assume that the exposure has a cumulative or an acute effect. The trajectories for participants with changes in exposure over time are different in the two models. Since all possible trajectories implied by the models are encompassed by the two extreme trajectories of those with the time-invariant exposure, Singer and Willett [18] referred to these as the *envelope* trajectories.

When the effect of exposure is cumulative, we define a new variable, the cumulative exposure variable, E_{ij}^*, where $E_{ij}^*=\sum_{k\leqslant j}E_{ik}$, and we assume that Y_{ij} depends on the exposure only through E_{ij}^* (see Checkoway *et al.* [19] for the motivation for and examples of cumulative exposure). Note that if a participant is exposed for the entire study period, the cumulative exposure, since entering the study, is proportional to time in the study. We denote the cumulative exposure before entering the study for subject i as $E_{i,-1}^*$. Then, $E_{ij}^*=E_{i,-1}^*+\sum_{k=0}^{j}E_{ik}$. Often, $E_{i,-1}^*$ is unknown. The right panel of Figure 1(a) shows a possible trajectory for one participant with $E_{i,-1}^*=0$ and exposure $\mathbf{E}_i=(0,1,1,0,1)$, which gives $\mathbf{E}_i^*=(0,1,2,2,3)$. This pattern is consistent with a cumulative

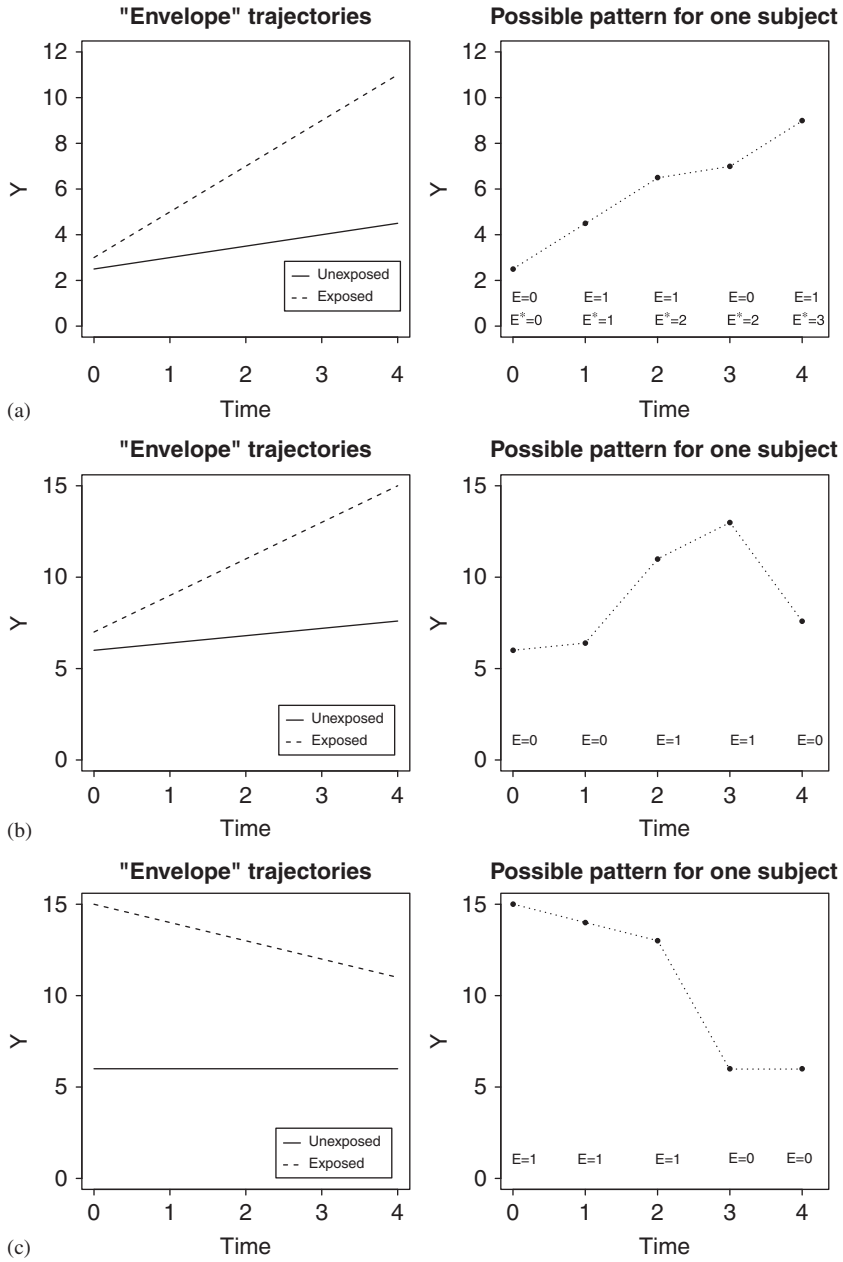

Figure 1. Envelope trajectories (trajectory for participants with time-invariant exposure) and possible individual patterns according to: (a) models (3) and (4); (b) and (c) models (5) and (6).

effect of exposure as well as an independent effect of time for both the exposed and the unexposed (e.g. due to ageing), and it can be modeled as

$$\mathbb{E}(Y_{ij}|\mathbf{X}_i) = \gamma_0 + \gamma_t t_{ij} + \gamma_{e*} E^*_{ij} \tag{3}$$

For example, in a study of the relationship between changes in lung function and smoking status in a cohort followed from 13 to 27 years of age, it was found that the rate of increase in lung function was smaller in periods with active smoking [20, p. 118]. Since cumulative exposure is, by definition, the sum of the product of point exposure by exposure duration, model (3) does not need to include an interaction term to be a generalization of model (2) for the time-varying case. A response trajectory is defined for each possible exposure history, $\mathbf{E}_i$, or equivalently, for each cumulative exposure history, $\mathbf{E}^*_i$. Model (3) assumes that the within- and between-subject effects of cumulative exposure (and time) are equal, that is, there is no confounding by between-subject effects [21]. If this assumption is unreasonable, one may want to fit the following change model:

$$\mathbb{E}(Y_{ij} - Y_{i,j-1}|\mathbf{X}_i) = \gamma^W_t + \gamma^W_{e*}(E^*_{ij} - E^*_{i,j-1}) = \gamma^W_t + \gamma^W_{e*} E_{ij} \tag{4}$$

This model results from applying the first difference operator

$$\mathbf{\Delta}=\begin{pmatrix}-1 & 1 & 0 & \dots & 0\\ 0 & -1 & 1 & \ddots & \vdots\\ \vdots & \ddots & \ddots & \ddots & 0\\ 0 & \dots & 0 & -1 & 1\end{pmatrix}$$

to model (3), so that $\mathbf{\Delta Y}_i$ is the vector with elements $Y_{i,j+1}-Y_{ij}$, $j=1,\dots,r$, and $\text{Var}(\mathbf{\Delta Y}_i)=\mathbf{\Delta\Sigma}_i\mathbf{\Delta}'$. For a multivariate normal response with known $\mathbf{\Sigma}_i$, fitting model (4) by GLS is equivalent to fitting model (3) by the conditional linear regression (Web Appendix A). If there is no confounding by between-subjects determinants of response, $\hat{\gamma}_{e*}$ will estimate the same parameter as $\hat{\gamma}^W_{e*}$, otherwise not [21]. In observational studies, model (4) is often preferred, since each participant serves as his or her own control, fully controlling for confounding by all between-subject (time-invariant) effects, including cumulative exposure at entry and age at entry. However, the trade-off is that model (4) is less efficient than (3) [21].

When the effect of exposure is acute, we assume that the response depends on exposure only through the exposure in the previous period. The right panel of Figure 1(b) shows a possible trajectory for one participant with exposure $\mathbf{E}_i=(0,0,1,1,0)$. This situation can be modeled as

$$\mathbb{E}(Y_{ij}|\mathbf{X}_i)=\gamma_0+\gamma_t t_{ij}+\gamma_e E_{ij}+\gamma_{te}(E_{ij}\times t_{ij}) \tag{5}$$

and we are interested in the parameter γ_{te}. Note that under this model, a participant shifts trajectories when exposure changes, the 'jumps' being larger or smaller as time increases. Although this model is considered in the literature on longitudinal data analysis, the situation it implies may be harder to find in real life, and very often models with only a main effect of exposure may be more appropriate when the effect of exposure is acute. Singer and Willett [18] fit model (5) in a study on the effect of time since unemployment in relation to the occurrence of depression symptoms. The trajectory implied by their analysis is illustrated in Figure 1(c). Immediately after layoff (time 0), participants had high depression index values. Over time, they acclimated to their new status and the values for the depression index decreased over time, without reaching the levels of the employed. Once a formerly unemployed individual found a job and kept it, the depression score dropped to the level of the employed and remained constant over time. Like model (3), model (5) assumes that there is no between-subject confounding [21]. As above, the within-subject effects of exposure and time can be estimated by using the model for change that is obtained after applying the first difference operator to model (5),

$$\mathbb{E}(Y_{i,j+1}-Y_{ij}|\mathbf{X}_i)=\gamma^W_t+\gamma^W_e(E_{i,j+1}-E_{ij})+\gamma^W_{te}[(E_{i,j+1}-E_{ij})t_{ij}+E_{i,j+1}] \tag{6}$$

Again, under multivariate normality, fitting model (6) by GLS is algebraically equivalent to fitting model (3) by conditional likelihood (Web Appendix A).

2.3. Power and sample size formulas

Let γ be the parameter of interest, which is γ_{e*} for model (3), γ^W_{e*} for model (4), γ_{te} for model (5) or γ^W_{te} for model (6). Let $\tilde{\sigma}^2$ be the diagonal element of the matrix $\mathbf{\Sigma}_\Gamma$, defined in (1), that is associated with the parameter γ. The Wald test statistic to test if $\hat{\gamma}$ is different from zero is $T=\sqrt{N}\hat{\gamma}/\tilde{\sigma}$ [9], and the formula for the power of a study to detect an effect γ is $\Phi[\sqrt{N}|\gamma|/\tilde{\sigma}-z_{1-\alpha/2}]$, where α is the significance level, and z_u and $\Phi(\cdot)$ are the uth quantile and the cumulative density of a standard normal. The formula for the required sample size to detect an effect γ with power π is $N=\tilde{\sigma}^2(z_\pi+z_{1-\alpha/2})^2/\gamma^2$. For both power and sample size calculations, we need to derive $\tilde{\sigma}^2$ following (1) and the model of choice from (3)–(6). Note that $\tilde{\sigma}^2$ will depend on r and on several other parameters associated with the distributions of the response, the exposure and time.

3. Derivation of $\tilde{\sigma}^2$

3.1. Compound symmetry of both the response and the exposure process and other simplifications

We start by deriving formulas for $\tilde{\sigma}^2$ when it is assumed that both the response and the exposure process have a CS covariance and the exposure prevalence, p_e, is constant over time, in which case a closed-form formula for $\tilde{\sigma}^2$ can be derived. Under CS of the response, $\mathbf{\Sigma}_i=\mathbf{\Sigma}$ has diagonal terms equal to σ^2, where $\sigma^2=\text{Var}(Y_{ij}|X_{ij})$ is the residual variance of the response given the covariates, and the off-diagonal terms equal to $\sigma^2\rho$, where ρ is the correlation between the two measurements from the same participant, known as the reliability coefficient or the intraclass correlation coefficient [11]. Then, assuming CS of exposure, the covariance matrix of exposure has diagonal elements equal to $p_e(1-p_e)$ and off-diagonal elements equal to $\rho_x p_e(1-p_e)$, where ρ_x is the common correlation between exposure at different time points.

When the prevalence of exposure is constant over time, the correlation ρ_x is equal to the intraclass correlation of exposure, ρ_e, defined as the percentage of variation in exposure that is due to between-subject variation [22]. Otherwise, one can still derive a relationship between the intraclass correlation, ρ_e, and the common correlation under compound symmetry, ρ_x (Web

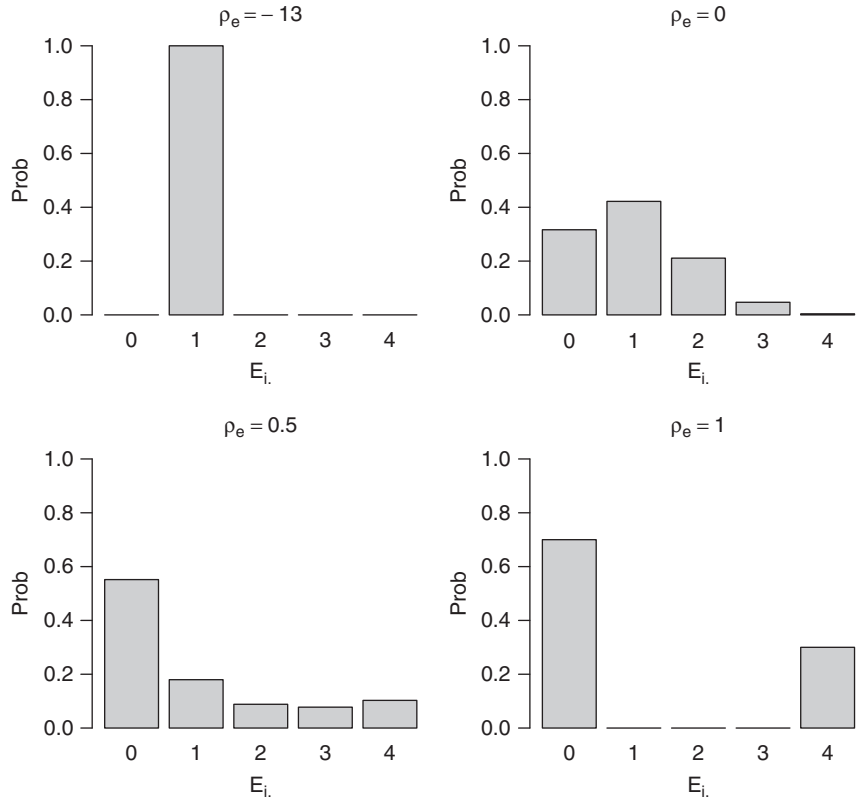

Figure 2. Distribution of $E_{i\cdot}$ for $r=3$, $\bar{p}_e=0.25$ and different values of ρ_e.

Appendix B). Because of the equality of ρ_e and ρ_x in the constant exposure prevalence case, and the fact that ρ_e has an intuitive interpretation and well-defined properties as detailed below, we will parameterize $\tilde{\sigma}^2$ in terms of ρ_e henceforth. The intraclass correlation, ρ_e, is bounded below by $-1/r$, and for binary variables, as here, there is a correction factor that needs to be added to this bound [22]. The upper bound is one when the exposure prevalence is constant over time, and below one otherwise. An expression for the upper bound in the latter case was derived in Web Appendix C. These bounds are calculated by our program and displayed to the user after r and the prevalence at each time point have been provided.

Apart from being the percentage of between-subject variation in exposure and being equal to the common correlation if the exposure prevalence is fixed, the intraclass correlation of exposure can also be regarded as a measure of imbalance in the number of exposed periods per subject, $E_{i\cdot}$. When $E_{i\cdot}$ is balanced across subjects, then everyone is exposed for the same number of periods as, for example, in some crossover studies. Then, $\rho_e=-1/r$. Conversely, when the exposure is time-invariant, the imbalance is maximal since $E_{i\cdot}$ is either zero with probability $(1-p_e)$ or $r+1$ with probability p_e, and $\rho_e=1$. In observational studies, intermediate values between the bounds $\rho_e=-1/r$ (same number of exposed periods for all participants) and $\rho_e=1$ (time-invariant exposure) will often be observed, and when pilot data are not available, the investigator can assess the sensitivity of the study design over a range of plausible values for ρ_e. To help the investigator assess what value of ρ_e is appropriate for his or her exposure, our program provides the distribution of $E_{i\cdot}$ once r and p_e are fixed and a CS covariance of exposure is assumed. Examples of distributions of $E_{i\cdot}$ by varying ρ_e are shown in Figure 2.

Table I shows the expressions for $\tilde{\sigma}^2$ for models (3)–(6). For models (3), (5) and (6), all the participants are assumed to enter the study at the same time ($V(t_0)=0$). For model (3), all the participants are assumed to be unexposed at baseline ($E^*_{i,-1}=0\ \forall i$). Expressions are derived in Web Appendix D. In Section 4, we will assess the effects of departures from the scenarios assumed in this section that were used to derive the expressions in Table I.

3.2. General covariance of the response and the exposure process

As shown above, the main difficulty in computing power and sample size for longitudinal studies with the time-varying exposures is characterizing $\tilde{\sigma}^2$, the diagonal element of the matrix $\mathbf{\Sigma}_\Gamma$, defined in (1), that is associated with the parameter γ. Its calculation involves the specification of the matrix $\mathbb{E}_X[\mathbf{X}_i'\mathbf{\Sigma}_i^{-1}\mathbf{X}_i]$ and the computation of its inverse. A simple general expression for $\tilde{\sigma}^2$ is difficult to obtain (Web Appendix D), but it can be easily computed with our program once $\mathbb{E}_X[\mathbf{X}_i'\mathbf{\Sigma}_i^{-1}\mathbf{X}_i]$ has been defined. So, in this section, we only state the input parameters needed for each one of the models (3)–(6) to perform such calculations.

The parameters defining the covariance of the response need to be provided for all four models. Examples of two common covariance structures different from CS are discussed in Section 4.1. In addition, one needs to provide the exposure prevalence

Table I. Power and sample size equations for models (3)–(6) assuming CS covariance of both response and exposure, and constant exposure prevalence.

Power formula: $\pi=\Phi[\sqrt{N}|\gamma|/\tilde{\sigma}-z_{1-\alpha/2}]$
Sample size formula: $N=\tilde{\sigma}^2(z_\pi+z_{1-\alpha/2})^2/\gamma^2$

Model	Parameter of interest	$\tilde{\sigma}^2$
(3)	γ_{e*}	$\frac{12\sigma^2(1-\rho)(1+r\rho)}{p_e(1-p_e)(r+1)(r+2)[6+2(r-3)\rho+r(4+(r-5)\rho)\rho_e]}$
(4)	γ^W_{e*}	$\frac{12\sigma^2(1-\rho)}{p_e(1-p_e)s^2r(r+2)[2+(r-1)\rho_e]}$
(5)	γ_{te}	$\frac{12\sigma^2(1-\rho)(1+r\rho)}{p_e(1-p_e)s^2r(r+1)(r+2)[1+r\rho-\rho(1-\rho_e)]}$
(6)	γ^W_{te}	$\frac{12\sigma^2(1-\rho)}{p_e(1-p_e)s^2r(r+\rho_e)(r+2)}$

For models (3), (5) and (6), all participants are assumed to enter the study at the same time ($V(t_0)=0$). For model (3), all participants are assumed to be unexposed at baseline ($E^*_{i,-1}=0\ \forall i$).

at each time point, p_{ej}, $j=0,\ldots,r$, and the correlation between exposure at the jth and j'th measurements, $\rho_{e_j,e_{j'}}$ $\forall j\neq j'$, for all four models.

If a cumulative exposure effect is assumed and model (4) is assumed, no additional parameters apart from those discussed in the previous paragraph are needed (Web Appendix D.2). However, if model (3) is assumed, age is the time metameter for the study and the participants enter at different ages, i.e. $V(t_0)>0$, and/or if the participants enter the study with different values of cumulative exposure, i.e. $V(E^*_{-1})>0$, then additional parameters are needed for an exact calculation of $\tilde{\sigma}^2$ (Web Appendix D.1). Some of these parameters involve correlations between pairs of variables, or equivalently, the expected values of products of variables. These parameters are: the baseline mean and variance of time metameter, $\mathbb{E}(t_0)$ and $V(t_0)$; the baseline mean and variance of cumulative exposure, $\mathbb{E}(E^*_{-1})$ and $V(E^*_{-1})$; the correlation between the baseline cumulative exposure and the baseline time metameter, or $\mathbb{E}[E^*_{-1}t_0]$; correlation between the time metameter at baseline and the exposure at all times henceforth, or $\mathbb{E}[E_jt_0]$ $\forall j$; and the correlation between the baseline cumulative exposure and the exposure at all subsequent measurement times, or $\mathbb{E}[E^*_{-1}E_j]$ $\forall j$. These quantities are difficult to provide *a priori* unless longitudinal pilot data are available; hence, one option is to base the study design on model (4), which estimates only within-subject effects. In that case, none of these additional parameters are needed and a conservative study design will result, i.e. more participants than needed will be recruited. When it is reasonable to assume that all participants are unexposed at baseline, i.e. $E^*_{i,-1}=0\ \forall i$, and time in the study is the time metameter of interest, i.e. $V(t_0)=0$, then only the exposure prevalence at each time point, p_{ej} $\forall j$, and the intraclass correlation of exposure, $\rho_{e_j,e_{j'}}$ $\forall j,j'$, are needed, even for a study designed to fit model (3) (Web Appendix D.1.1). However, in an observational study, one still may want to base the study design calculations on model (4), which estimates the cumulative exposure during the study period but controls for all measured and unmeasured time-invariant confounders.

If a cumulative effect of exposure is assumed, when age is the time metameter of interest and participants enter the study at different ages, i.e. $V(t_0)>0$, to exactly compute $\tilde{\sigma}^2$, the investigator would need to specify the following additional parameters for both models (5) and (6) (Web Appendix D.3): the variance of the time metameter at baseline, $V(t_0)$; the correlation between the time metameter at baseline and exposure at all time henceforth, or $\mathbb{E}[E_jt_0]$ $\forall j$; the expected value of the crossproduct of exposure at all pairs of time points and baseline time metameter, $\mathbb{E}[E_jE_{j'}t_0]$ $\forall j,j'$; and the expected value of the crossproduct of exposure at all pairs of time points and the baseline time metameter squared $\mathbb{E}[E_jE_{j'}t_0^2]$ $\forall j,j'$. Clearly, these quantities will be impossible to provide *a priori* in most study settings unless longitudinal pilot data are available; hence, some simplifications need to be implemented. One option is to perform the calculations for the case $V(t_0)=0$, which seemed numerically to always provide conservative estimates, although we were unable to prove this analytically. When $V(t_0)=0$ is assumed, only p_{ej} $\forall j$ and $\rho_{e_j,e_{j'}}$ $\forall j,j'$ need to be provided (Web Appendix D.3.1).

4. Efficiency comparisons

In Section 3.1, we obtained formulas for $\tilde{\sigma}^2$ under several scenarios likely to be encountered in practice. In this section, we compare the required sample size under those assumptions with the required sample size when those assumptions are not met. We will investigate the direction of the differences to identify under what circumstances incorrect assumptions lead to conservative (overpowered) designs, and under what circumstances incorrect assumptions will lead to underpowered studies. In practice, the investigator can also do these comparisons by performing the calculations under various scenarios using our software.

4.1. Effect of departures from the assumption of CS response covariance

In this section, we consider two response covariance structures that are commonly used and are more general than CS, but that include CS as a particular case: damped exponential (DEX) and random intercepts and slopes (RS). Under the DEX covariance structure [23], the $[j,j']$ element of $\mathbf{\Sigma}_i=\mathbf{\Sigma}$ has the form $\sigma^2\rho^{|j-j'|^\theta}$, where the correlation between two measurements decays exponentially as the separation between the measurements increases, but the parameter θ attenuates this decay. Thus, when $\theta=0$, the CS covariance structure is obtained, and when $\theta=1$, the AR(1) covariance structure is given. The RS covariance structure is the one that arises in mixed models, when there is a random effect associated with the intercept and the one associated with time. The covariance matrix is typically given as $\mathbf{\Sigma}_i=\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i'+\sigma_w^2\mathbf{I}$, where $\mathbf{Z}_i$ contains a column of ones and the column of times for participant i, and $\mathbf{D}$ is the 2×2 covariance matrix of the random effects, with elements $(\sigma^2_{b_0},\rho_{b_0b_1}\sigma_{b_0}\sigma_{b_1},\rho_{b_0b_1}\sigma_{b_0}\sigma_{b_1},\sigma^2_{b_1})$ [11]. The RS covariance structure is heteroscedastic and non-stationary [11]. When all the participants are observed at the same times, $\mathbf{Z}_i=\mathbf{Z}\ \forall i$ and, therefore, $\mathbf{\Sigma}_i=\mathbf{\Sigma}\ \forall i$. When that is not the case, $\mathbf{\Sigma}_i$ depends on the covariates, and expression (1) can only be calculated if a distribution for the time variable is assumed. We define the total variance at baseline as $V(Y_{i0}|\mathbf{X}_i)=\sigma^2_{t_0}=\sigma^2_{b_0}+\sigma^2_w$, the intraclass correlation at baseline as $\rho_{t_0}=\sigma^2_{b_0}/\sigma^2_{t_0}$, and the slope reliability as

$$\rho_{b_1}=\sigma^2_{b_1}\Big/\left(\frac{12(1-\rho_{t_0})\sigma^2_{t_0}}{s^2r(r+1)(r+2)}+\sigma^2_{b_1}\right)$$

which is the percentage of variation in the slopes that is between subjects. It does not depend on the response units and it has a closed range (between zero and one) [18]. Then, the RS covariance can be expressed in terms of $(\sigma^2_{t_0},\rho_{t_0},\rho_{b_1},\rho_{b_0b_1})$. When $\rho_{b_1}=0$ the CS covariance is obtained. The slope reliability needs to be defined for a particular r. Throughout this paper, the values of ρ_{b_1} will be calculated at the value $r=5$. Although this choice was arbitrary, it would be consistent with a typical longitudinal study funded by the U.S. National Institutes of Health. These studies can be funded for no more than 5 years, and if a measurement was to be taken at the end of each year of funding, $r=5$.

Under the assumption that the effect of exposure is cumulative, we assessed departures by assuming CS response covariance based on model (4) (similar results were obtained using model (3)). The covariance of the exposure is still assumed to be CS. Through a grid search, we found that, for two studies with the same ρ_e, departures from CS in the response, i.e. of $\theta>0$ if DEX is assumed or $\rho_{b_1}>0$ if RS is assumed, increased $\tilde{\sigma}^2$ and therefore increased the required sample size. Figure 3 illustrates this increase when DEX response covariance, CS exposure covariance and constant exposure prevalence are assumed, for several values of ρ, θ and ρ_e (including independence, i.e. $\rho_e=0$, and time-invariant exposure, i.e. $\rho_e=1$). The increase in the required sample size with θ is larger for larger values of ρ_e and ρ. We performed the same assessment when the effect of exposure is acute based on models (5) and (6), and the effect was similar to that described in Figure 3, but with lower SSRs as the within-subject exposure correlation decreased.

4.2. Effect of departures from the assumption of CS exposure covariance

To assess departures from CS in the exposure covariance, we adopted a slightly different approach than when assessing departures from CS of the response covariance. We performed a numerical analysis to evaluate the accuracy of the CS exposure covariance assumption as an approximation to $\tilde{\sigma}^2$ when the exposure process had some other correlation structure, i.e. when the exposure covariance was misspecified. To compute the true and misspecified $\tilde{\sigma}^2$, the exposure prevalence vector and the correlation matrix of exposure are needed. For values of r equal to 2, 5 and 10, we generated 10 000 arbitrary prevalence vectors and correlation matrices using a process described in Web Appendix E. Then, the SSR based upon the misspecified and true $\tilde{\sigma}^2$ was calculated. The response covariance was assumed to be known here, and the calculations were repeated for CS, DEX and RS response covariance to determine whether the results were dependent upon the assumed response covariance structure. The values of the response covariance parameters were $\rho=(0.8,0.5,0.2)$ for CS; the same values of ρ and $\theta=(0.2,0.5,0.8,1)$ for DEX; and the same values of ρ_{t_0} that we used for ρ, and $\rho_{b_1}=(0.05,0.1,0.2,0.5,0.8)$ for RS.

Figure 4 summarizes the results for some values of the response covariance for $r=5$. Results for $r=2$ and $r=10$ were similar. Results from model (3) were very similar to those of model (4), and results from model (6) were similar to those of model (5). For the cumulative exposure effect models (3) and (4), most of the scenarios considered (around 90 per cent) fell into the interval (0.9, 1.1) with the CS response. As θ increased, the results were better than in the CS case. When RS of the response was assumed, the results were similar for all combinations of ρ_{t_0} and ρ_{b_1} considered, and less than 5 per cent of the scenarios had SSR$<$0.9. We are particularly concerned with those scenarios with low SSR ($<$0.9), since when SSR is greater than one, conservative designs are obtained. We observed that the scenarios with the lowest SSRs were characterized by small, negative correlations for pairs of exposures close in time, and large, positive correlations for pairs of exposures distant from each other. To confirm this numerically, we regressed all correlation coefficients from each of the 10 000 exposure correlation matrices against time separation; hence 10 000 slopes were obtained. A positive slope indicates that the correlations increase with time separation. Then, when the SSRs were regressed against these slopes, a strong linear negative relationship was obtained for all response covariances used in Figure 4, with the higher slopes presenting the smallest SSRs. The R^2 of these regressions ranged from 0.63 to 0.79.

For the acute exposure effect model with CS response, the SSR was between 0.9 and 1.1 for more than 95 per cent of the scenarios. For DEX response, the same pattern as with the CS response was observed when θ or ρ were small, but when both θ and ρ were large, a higher percentage of scenarios with SSR less than 0.9 were observed. For example, when the response was AR(1) with $\rho=0.8$, 15 per cent of the scenarios were below that value. Similarly, for RS covariance, only for high ρ_{t_0} and

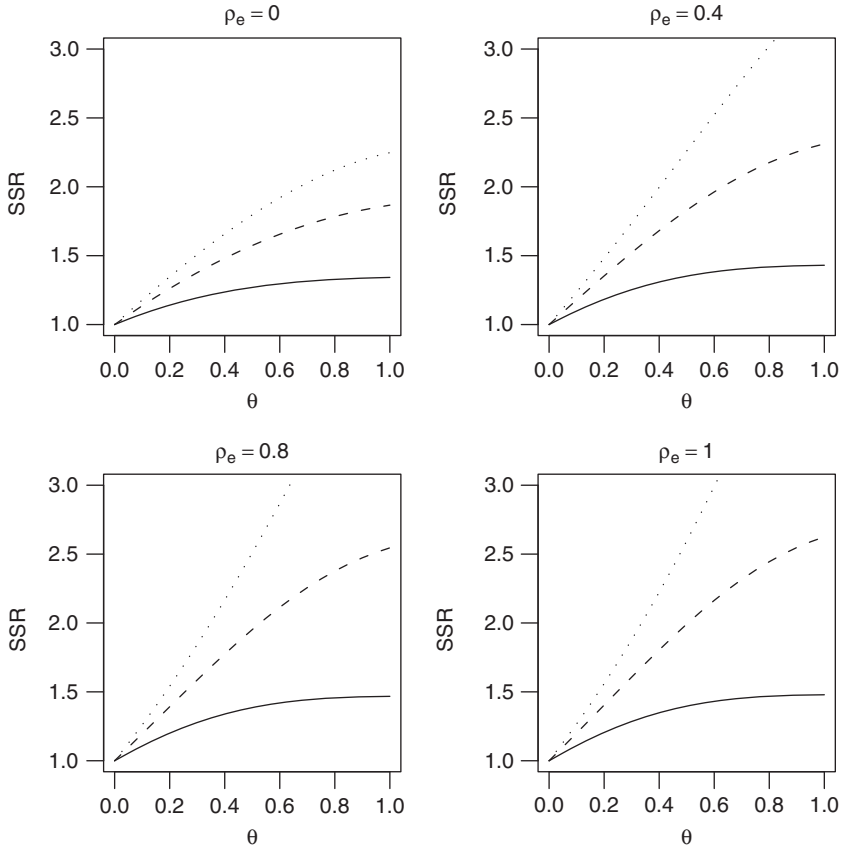

Figure 3. Sample size ratio (SSR$=N_\theta/N_{\theta=0}$) comparing the required sample size of a study with the CS response covariance (i.e. $\theta=0$) to a study with the DEX response covariance (i.e. $\theta>0$) when the cumulative exposure effect model (4), CS exposure covariance, constant exposure prevalence and $r=5$ are assumed, for several values of ρ and ρ_e. Lines indicate: (—) $\rho=0.2$, (- - -) $\rho=0.5$, (···) $\rho=0.8$.

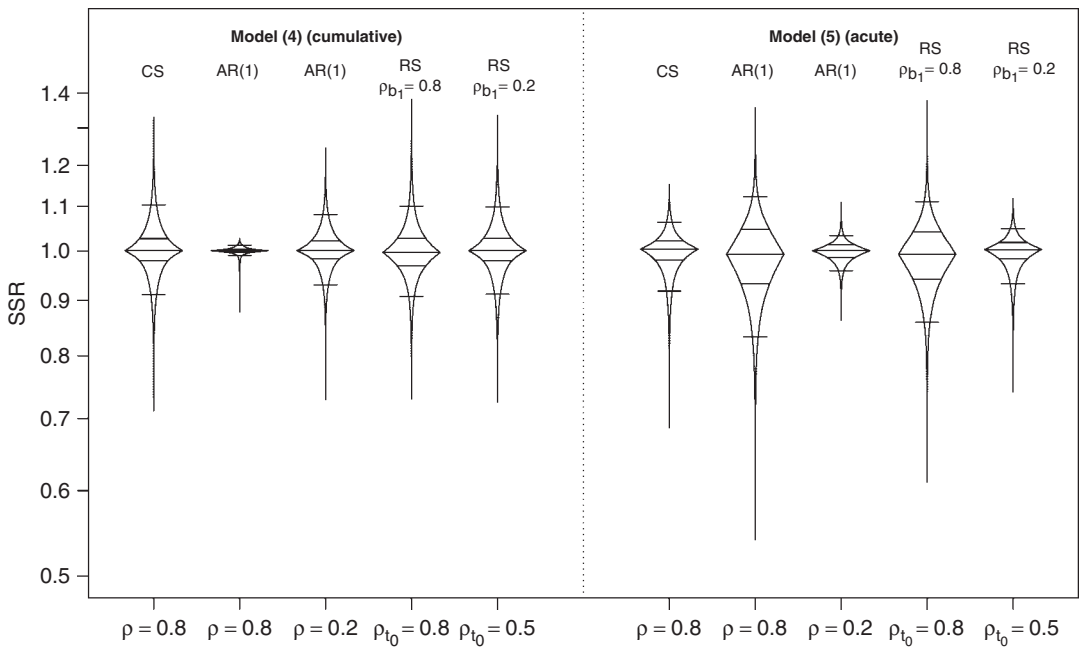

Figure 4. Box-percentile plots of the ratio of required sample sizes obtained when incorrectly assuming CS covariance of exposure divided by the required sample size obtained using the true exposure covariance in 10 000 scenarios generated to have an arbitrary exposure covariance, for $r=5$ and for several correlation structures of the response. For RS models, $\rho_{b_0,b_1}=-0.5$ is assumed. At any height, the width of the irregular 'box' is proportional to the percentile of that height. Horizontal lines indicate the 5th, 25th, 50th, 75th and 95th percentiles. *Y*-axis on logarithmic scale.

ρ_{b_1} did we find a high percentage of scenarios with SSRs less than 0.9 (12 per cent for $\rho_{t_0}=0.8$, $\rho_{b_1}=0.5$). We observed that the scenarios with the lowest SSRs were characterized by having high correlations for pairs of exposures that were both either at the beginning or at the end of the study, while the remaining correlations were negative. This implied a convex quadratic relationship between correlations and the sum of times of each pair. To numerically confirm this, we regressed the correlation coefficients from each of these 10 000 exposure correlation matrices against the sum of times and the sum of times squared, and obtained 10 000 coefficients associated with the quadratic term. Then, when the SSRs were regressed against these coefficients, a strong linear negative relationship was obtained for all response covariances used in Figure 4, with the higher coefficients presenting the lowest SSRs. The R^2 of these regressions ranged from 0.54 to 0.77.

In conclusion, both for the cumulative and the acute exposure effect models, the cases with SSR<0.9 had exposure correlation matrices that are unlikely to be found in practice, and assuming CS of exposure resulted in good approximations of the required sample size in most cases considered.

4.3. Effect of ρ_e on sample size

In this section, we assessed the effect of ρ_e on sample size. This will also allow comparing the required sample size obtained in a study with the time-invariant exposure ($\rho_e=1$) to a study with the time-varying exposure ($\rho_e<1$). This comparison has never been done, precluding investigators to anticipate whether they will need to recruit more or less participants if they have a time-varying exposure instead of a time-invariant one.

When assuming that the effect of exposure is cumulative, the results were based on model (4) (results were verified to be similar using model (3)). For model (4), we show in Web Appendix D.2.2 that if $w^{jj'}\geqslant 0\ \forall j\neq j'$, where $w_{jj'}$ is the $[j,j']$ element of $(\mathbf{\Delta\Sigma}_i\mathbf{\Delta}')^{-1}$, then $\tilde{\sigma}^2$ is minimal when ρ_e takes its upper bound (i.e. $\rho_e=1$, the time-invariant exposure case, if the prevalence is constant over time) and maximal when ρ_e takes its lower bound, regardless of the form of the covariance of the exposure. The condition $w^{jj'}\geqslant 0\ \forall j\neq j'$ holds when the response has CS or DEX covariance, but it does not necessarily hold for RS (Web Appendix D.2.2). Thus, under CS or DEX covariance of the response, and when $w^{jj'}\geqslant 0\ \forall j\neq j'$ for RS, efficiency is always lost when the exposure varies over time. To illustrate this, Figure 5 shows the ratio of the required sample sizes, $\text{SSR}=N_{\rho_e}/N_{\rho_e=1}$, comparing a study with a time-varying exposure with CS covariance to a study with a time-invariant exposure, for the case where the response variable has RS covariance. It can be seen that, for large r and ρ_{b_1}, a study with a time-varying exposure can end up being more efficient than the one with a time-invariant exposure. In contrast, for DEX response, the SSR keeps increasing as r increases, with a slower rate than for CS response, which is shown in the first panel of Figure 5 ($\rho_{b_1}=0$ case). When both the response and the exposure process can be assumed to follow CS and the prevalence of exposure is constant over time, we can derive an explicit formula for the ratio of variances, or equivalently, the ratio of the required sample sizes, comparing the time-varying exposure case to the time invariant one, $\text{SSR}=N_{\rho_e}/N_{\rho_e=1}=(r+1)/(2+(r-1)\rho_e)$. The SSR is always greater than or equal to one. So, as discussed above, in this particular case, efficiency is lost when the exposure varies over time.

When the exposure is acute, we based our results on model (5) (results were verified to be similar using model (6)). For model (5), Figure 6 shows the ratio of the required sample sizes, $\text{SSR}=N_{\rho_e}/N_{\rho_e=1}$, comparing a study with a time-varying exposure with CS covariance to a study with a time-invariant exposure, where both studies have an RS covariance of the response. When the response follows CS ($\rho_{b_1}=0$ case), a study with a time-varying exposure is less efficient than a study with a time-varying one, although for large values of r the differences become negligible. For this particular case ($\rho_{b_1}=0$), the SSR has the expression $N_{\rho_e}/N_{\rho_e=1}=(1+r\rho)/(1+r\rho-\rho(1-\rho_e))\geqslant 1$. Thus, as with the cumulative effect model, in this simplest case, efficiency is lost when the exposure varies over time, and the loss increases as the response correlation increases, as the within-subject exposure correlation decreases and as r decreases. However, when the response has RS covariance ($\rho_{b_1}>0$), a study with a time-varying exposure can become more efficient than a study with a time-invariant exposure as r increases, and this gain in efficiency can be substantial. If the response follows DEX, results similar to the RS case (Figure 6) are obtained. Thus, it is important to correctly specify the covariance of the response when assessing the effect of a time-varying exposure, since assuming CS vs RS or DEX can lead to very different study designs and studies that can be substantially under- or over-powered.

5. Example: the MSCM study [24]

In this section, we applied our methods to the setting of a study of the effects of maternal depression on child health (the MSCM study, data set available online at `http://faculty.washington.edu/heagerty/Books/AnalysisLongitudinal/datasets.html` [7]), to assess with real data the performance of our formulas. Participants ($N=142$ with complete data) were followed for $r+1=28$ consecutive days. The prevalence of the depression fluctuated from a maximum of 0.26 at baseline to a minimum of 0.06 at the second to the last day, with a mean of $\bar{p}_e=0.13$, and ρ_e was 0.10. Although the response variable in this study (child health) was binary, we considered the planning of a new study where child health is measured continuously and wanted to perform sample size calculations for it. We performed several calculations under the assumption of different scenarios for the response covariance. As in Section 4.2, we will illustrate, this time with exposure data from a real study, that by computing sample size under an assumed CS exposure covariance gives a required sample size that closely approximates the result obtained using the observed exposure covariance. In addition, we will show that using the formulas for a time-invariant exposure, poor approximations are obtained.

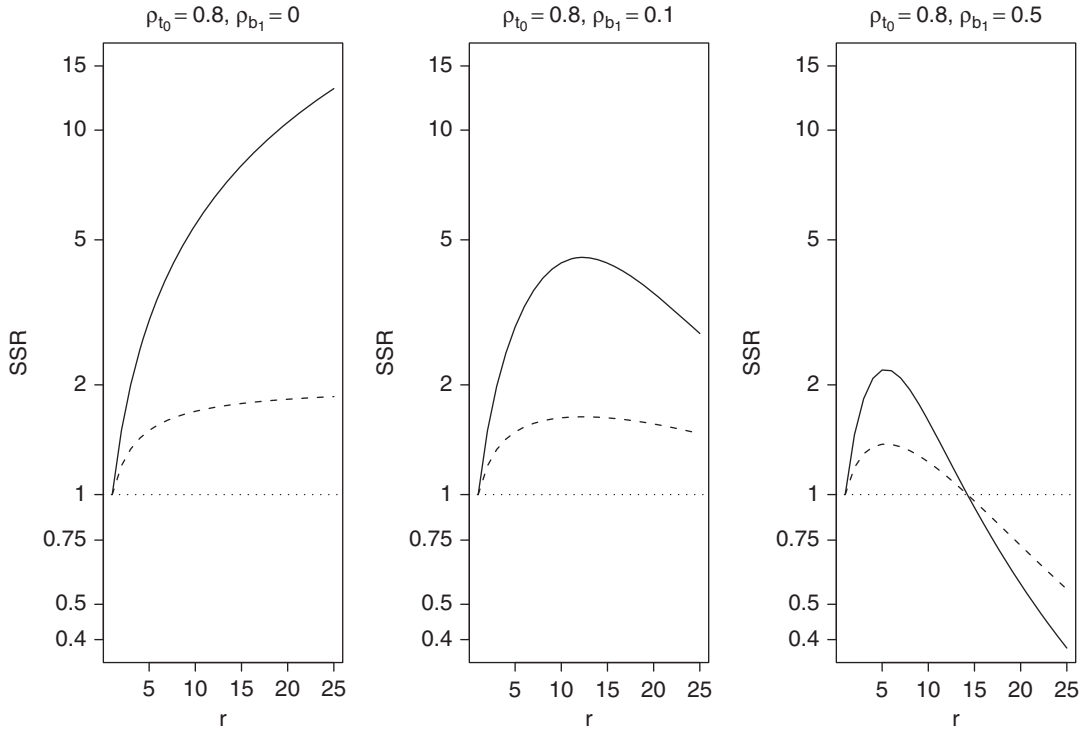

Figure 5. Sample size ratio (SSR$=N_{\rho_e}/N_{\rho_e=1}$) comparing the required sample size of a study with the time-varying exposure to a study with the time-invariant exposure when the cumulative exposure effect model (4), RS covariance of the response, CS covariance of exposure, constant exposure prevalence and $\rho_{b_0,b_1}=-0.5$ are assumed, for several values of r, ρ_{t_0}, ρ_{b_1} and ρ_e. Y-axis on logarithmic scale. Lines indicate: (—) $\rho_e=0$, (- - -) $\rho_e=0.5$, (···) $\rho_e=1$.

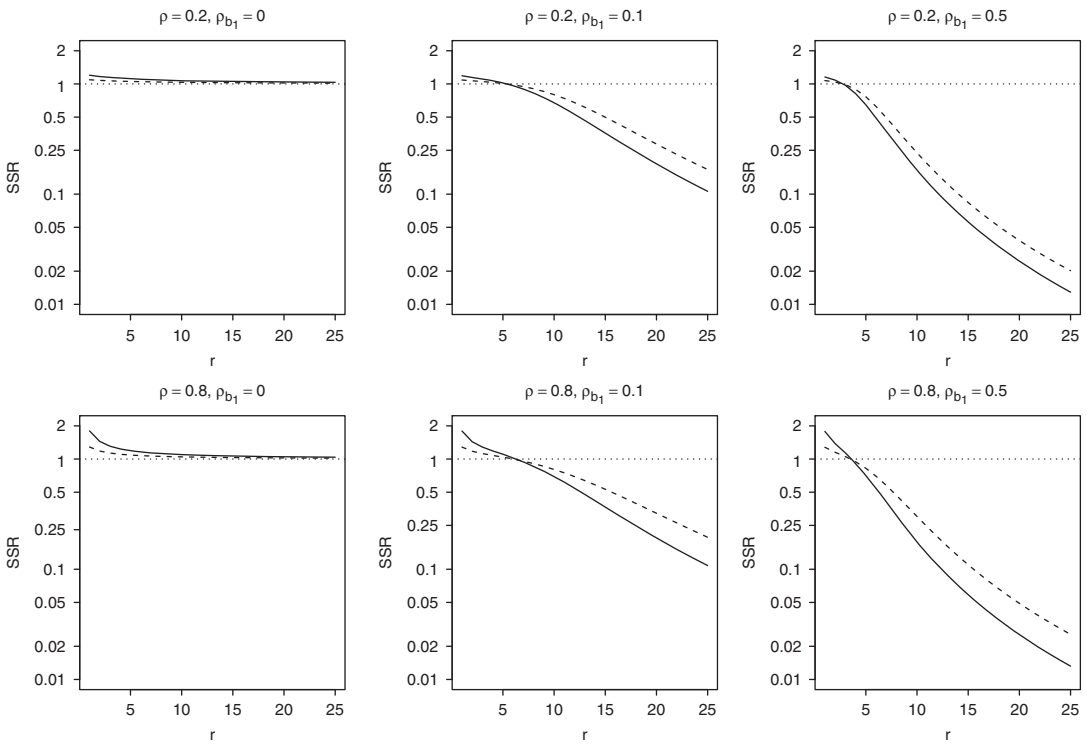

Figure 6. Sample size ratio (SSR$=N_{\rho_e}/N_{\rho_e=1}$) comparing the required sample size of a study with time-varying exposure to a study with the time-invariant exposure under the acute exposure effect model (5), RS covariance of the response, CS covariance of exposure, constant exposure prevalence, $V(t_0)=0$ and $\rho_{b_0,b_1}=-0.5$ are assumed, for several values of r, ρ, θ and ρ_e. Y-axis on logarithmic scale. Lines indicate: (—) $\rho_e=0$, (- - -) $\rho_e=0.5$, (···) $\rho_e=1$.

Table II shows the ratio between the required sample size obtained when CS or time-invariant exposure are used and the required sample size obtained when the observed exposure process is used, i.e. when equation (1) is replaced by $\left((1/N)\sum_i \mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right)^{-1}$ using the observed $\mathbf{X}_i$ values in the MSCM study. Calculations are done for both the cumulative exposure effect model (4) and the acute exposure effect model (5). The results for models (3) and (6) were almost identical. In this study, under the cumulative exposure effect model, incorrectly using the sample size formula for time-invariant exposure led to a large underestimate of the required sample size, except when RS was assumed and ρ_{b_1} was large. Assuming CS of exposure led to reasonably accurate estimates of the required sample size except when the response was RS, in which case slight overestimations were obtained. In order to consider the situation where fewer repeated measurements are planned, we repeated these calculations using data only from the first three periods of the study ($r=2$). In that case, the approximation of assuming a CS exposure process provided accurate calculations for all the response covariances considered.

For the acute exposure effect model (5), as the response covariance departed from CS, the formulas that assumed a time-invariant exposure led to a substantial overestimation of the required sample size, in part due to the large r of the study. Assuming CS covariance of exposure led to reasonably accurate calculations, with underestimation of sample size no less than 10 per cent in all the scenarios considered. As before, we repeated the same calculations but only using the first three periods of the study. Then, using the formulas for a time-invariant exposure led to underestimations of 10 to 20 per cent, as opposed to the overestimations obtained when the full study was considered, while assuming that CS of exposure still provided good approximations. In conclusion, assuming CS covariance of exposure gave very accurate calculations in this example, whereas the existing formulas for a time-invariant exposure led to seriously flawed calculations.

6. Discussion

In this paper, we developed study design formulas that accommodate a time-varying exposure for longitudinal studies designed to compare rates of change in the response according to exposure. The paper is mainly focused on the design of observational studies, where the within-subject variation of exposure is not determined by design. We defined models for acute and cumulative effects of exposure and presented some simplifications so that the only additional parameters needed to calculate the required sample size when the exposure varies with time within a subject are the exposure prevalence and the intraclass correlation of exposure. When studied numerically and applied to a real-data example, our methods provided reasonable approximations for the required sample size, which in many cases greatly improved the results that would have been obtained by using the available methods for the time-invariant exposure. The formulas presented in this paper are implemented in a public access program that can be downloaded at the link provided in Section 1 (Web Appendix F).

We examined whether studies with a time-varying exposure require recruiting more or less participants than a study with a time-invariant exposure. Under the cumulative exposure effect models, a study with a time-varying exposure requires more participants than the one with a time-invariant exposure when the response covariance is CS or DEX, and in many cases with RS covariance. Under the acute exposure effect models, the results depend on the covariance of the response. If the response has a CS covariance, we find the same result as in the cumulative exposure case, and the study with a time-varying exposure requires more participants. However, in studies with several repeated measures and a DEX or RS covariance of the response, a significantly lower number of participants may be needed in the time-varying case compared with the time-invariant case.

Table II. Ratio of the required sample sizes obtained by dividing the required sample size assuming CS or time-invariant exposure by the required sample size obtained using the observed exposure distribution for the MSCM study ($r=27, \bar{p}_e=0.13, \rho_e=0.1$).

	Response covariance									
	$\rho=0.8$					$\rho=0.5$				
Exposure covariance assumption	CS	DEX $\theta=0.5$	AR(1)	RS* $\rho_{b_1}=0.1$	RS* $\rho_{b_1}=0.5$	CS	DEX $\theta=0.5$	AR(1)	RS* $\rho_{b_1}=0.1$	RS* $\rho_{b_1}=0.5$
Cumulative exposure effect (*model* (4))										
Time-invariant	0.16	0.26	0.27	0.53	3.48	0.16	0.20	0.18	0.53	3.48
CS	1.01	1.04	1.01	1.17	1.25	1.01	1.03	1.02	1.17	1.25
Acute exposure effect (*model* (5))										
Time-invariant	0.97	9.09	14.4	10.2	84.5	0.97	4.02	3.31	10.2	85.6
CS	0.98	0.91	0.90	0.97	0.97	0.98	0.93	0.92	0.97	0.97

The cumulative exposure effect model (4) or the acute exposure effect model (5) is assumed. Calculations are reported for different scenarios of the response covariance. The true prevalence of exposure at each time point is used except for the time-invariant exposure case, where $\bar{p}_e$ is used. For model (5), $V(t_0)=0$ is assumed.
*$\rho_{b_0,b_1}=-0.5$.

The influence of dropout in the required sample size has been studied previously in studies with the time-invariant exposure [4, 5, 8–10]. Galbraith and Marschner [8] suggested the method of computing N for 90 per cent power when 80 per cent power is intended and Fitzmaurice *et al.* [11] suggested inflating N by $1/(1-f)$, where f is the anticipated fraction of loss to follow-up. The performance of these approaches in longitudinal studies with time-varying exposures remains to be investigated. Another interesting topic for future research is to generalize the summary measure approach [4, 11] to the analysis of the longitudinal studies with a time-varying exposure, and compare the efficiency of that analysis with the GLS approach used here.

We presented methods that provide accurate estimates of the required sample size given the information on parameters that can often be obtained or guessed by the investigator *a priori*. It is advisable to perform sensitivity analysis by assessing the effects of different values of the required parameters before determining the final sample size. We hope that the software we provide, which can be downloaded at the link provided in Section 1 and implements all calculations presented in this paper, will be a helpful tool for planning study in these unavoidably complex settings. A demonstration of the program use can be found in Web Appendix F.

Acknowledgements

Research supported, in part, by NIH grant CA06516.

References

1. Schlesselman JJ. Planning a longitudinal study. II. Frequency of measurement and study duration. *Journal of Chronic Diseases* 1973; **26**(9):561–570.
2. Kirby AJ, Galai N, Munoz A. Sample size estimation using repeated measurements on biomarkers as outcomes. *Controlled Clinical Trials* 1994; **15**(3):165–172.
3. Frison LJ, Pocock SJ. Linearly divergent treatment effects in clinical trials with repeated measures: efficient analysis using summary statistics. *Statistics in Medicine* 1997; **16**(24):2855–2872.
4. Dawson JD. Sample size calculations based on slopes and other summary statistics. *Biometrics* 1998; **54**(1):323–330.
5. Hedeker D, Gibbons RD, Waternaux C. Sample size estimation for longitudinal designs with attrition: comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics* 1999; **24**(1):70–93.
6. Raudenbush SW, Xiao-Feng L. Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods* 2001; **6**(4):387–401.
7. Diggle P, Heagerty P, Liang KY, Zeger S. *Analysis of Longitudinal Data* (2nd edn). Oxford Statistical Science Series, vol. 25. Oxford University Press: Oxford, New York, 2002.
8. Galbraith S, Marschner IC. Guidelines for the design of clinical trials with longitudinal outcomes. *Controlled Clinical Trials* 2002; **23**(3):257–273.
9. Yi Q, Panzarella T. Estimating sample size for tests on trends across repeated measurements with missing data based on the interaction term in a mixed model. *Controlled Clinical Trials* 2002; **23**(5):481–496.
10. Jung SH, Ahn C. Sample size estimation for Gee method for comparing slopes in repeated measurements data. *Statistics in Medicine* 2003; **22**(8):1305–1315.
11. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience: Hoboken, NJ, 2004.
12. Jones B, Kenward MG. *Design and Analysis of Cross-over Trials* (1st edn). Monographs on Statistics and Applied Probability, vol. 34. Chapman & Hall: London, New York, 1989.
13. Julious SA. Sample sizes for clinical trials with normal data. *Statistics in Medicine* 2004; **23**(12):1921–1986.
14. Moerbeek M, Van Breukelen JP, Berger MPF. Optimal experimental designs for multilevel models with covariates. *Communications in Statistics—Theory and Methods* 2001; **30**(12):2683–2697.
15. Whittemore AS. Sample size for logistic regression with small response probability. *Journal of the American Statistical Association* 1981; **76**(373):27–32.
16. Shieh G. On power and sample size calculations for likelihood ratio tests in generalized linear models. *Biometrics* 2000; **56**(4):1192–1196.
17. Tu XM, Kowalski J, Zhang J, Lynch KG, Crits-Christoph P. Power analyses for longitudinal trials and other clustered designs. *Statistics in Medicine* 2004; **23**(18):2799–2815.
18. Singer JD, Willett JB. *Applied Longitudinal Data Analysis*: *Modeling Change and Event Occurrence*. Oxford University Press: Oxford, New York, 2003.
19. Checkoway H, Pearce N, Kriebel D. *Research Methods in Occupational Epidemiology* (2nd edn). Monographs in Epidemiology and Biostatistics, vol. 34. Oxford University Press: Oxford, New York, 2004.
20. Twisk JWR. *Applied Longitudinal Data Analysis for Epidemiology*. Cambridge University Press: Cambridge, U.K., 2003.
21. Neuhaus JM, Kalbfleisch JD. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 1998; **54**(2):638–645.
22. Ridout MS, Demetrio CG, Firth D. Estimating intraclass correlation for binary data. *Biometrics* 1999; **55**(1):137–148.
23. Munoz A, Carey V, Schouten JP, Segal M, Rosner B. A parametric family of correlation structures for the analysis of longitudinal data. *Biometrics* 1992; **48**(3):733–742.
24. Alexander CS, Markowitz R. Maternal employment and use of pediatric clinic services. *Medical Care* 1986; **24**(2):134–147.