

UCSF

UC San Francisco Previously Published Works

Title

Likelihood-based analysis of longitudinal data from outcome-related sampling designs

Permalink

<https://escholarship.org/uc/item/1418b2jp>

Journal

Biometrics, 70(1)

ISSN

0006-341X

Authors

Neuhaus, John M
Scott, Alastair J
Wild, Christopher J
[et al.](#)

Publication Date

2014-03-01

DOI

10.1111/biom.12108

Peer reviewed

Published in final edited form as:

Biometrics. 2014 March ; 70(1): 44–52. doi:10.1111/biom.12108.

Likelihood-based analysis of longitudinal data from outcome-related sampling designs

John M. Neuhaus^{1,*}, Alastair J. Scott², Christopher J. Wild², Yannan Jiang², Charles E. McCulloch¹, and Ross Boylan¹

¹Division of Biostatistics, University of California, San Francisco, California 94143-0560, U.S.A.

²Department of Statistics, The University of Auckland, Auckland, New Zealand

Summary

Investigators commonly gather longitudinal data to assess changes in responses over time and to relate these changes to within-subject changes in predictors. With rare or expensive outcomes such as uncommon diseases and costly radiologic measurements, outcome-dependent, and more generally outcome-related, sampling plans can improve estimation efficiency and reduce cost. Longitudinal follow up of subjects gathered in an initial outcome-related sample can then be used to study the trajectories of responses over time and to assess the association of changes in predictors within subjects with change in response. In this paper we develop two likelihood-based approaches for fitting generalized linear mixed models (GLMMs) to longitudinal data from a wide variety of outcome-related sampling designs. The first is an extension of the semi-parametric maximum likelihood approach developed in and applies quite generally. The second approach is an adaptation of standard conditional likelihood methods and is limited to random intercept models with a canonical link. Data from a study of Attention Deficit Hyperactivity Disorder in children motivates the work and illustrates the findings.

Keywords

Conditional likelihood; Retrospective sampling; Subject-specific models

1. Introduction

Investigators commonly gather longitudinal data to assess changes in responses over time and to relate these changes to within-subject changes in predictors. With rare diseases or expensive outcomes (e.g., costly imaging), outcome-dependent sampling plans can improve estimation efficiency and reduce cost. For example, Hartung et al. (2002) examined determinants of the time course of Attention Deficit Hyperactivity Disorder (ADHD) symptom expression in children. ADHD is a relatively rare disorder and the investigators used an outcome-related sampling plan whereby subjects were sampled on the basis of whether a teacher or parent suspected that the child exhibited ADHD symptoms instead of directly on ADHD status, as they would in a standard case-control study. The study also recruited a sample of children who were not so suspected. Thus, the study data form a longitudinal study of ADHD outcomes following sampling dependent on a variable

*john@biostat.ucsf.edu.

associated with the baseline ADHD outcome. We call this type of sampling *outcome-related*.

Another example comes from the Osteoarthritis Initiative (OAI), a multi-center, longitudinal, prospective observational study of knee osteoarthritis (OA). The objective of the study is to understand risk factors for knee OA, OA progression and the natural history of the disease course. The data set includes clinical evaluation data, radiological (x-ray and magnetic resonance) images and a biospecimen repository gathered from 4796 men and women aged 45-79 years. Magnetic resonance images (MRIs) yield both binary and continuous measurements of OA status and are more accurate but also more expensive than X-rays. Because of the expense, MRI-based variables have still not been exhaustively evaluated. To reduce cost and improve estimation efficiency, investigators select a subset of longitudinal MRIs to evaluate based on X-ray-based variables and clinical data (e.g. pain). Sets of longitudinal MRIs form an outcome-related cluster sample.

The ADHD and OAI study data, and data from outcome-related samples in general, exhibit several features that complicate the statistical analysis. Firstly, the chance a cluster is included in the sample varies from cluster to cluster. Since ADHD is relatively rare, investigators sampled children suspected of ADHD at a greater rate than children not suspected in order to yield a sample containing enough outcomes of interest. Secondly, the referral/sampling variables in the ADHD and OAI studies are not exactly the outcomes of interest and may be longitudinal themselves. Finally, as with any longitudinal study, the data exhibit correlation within subjects over time.

We assume that the data of interest consist of clustered or longitudinal responses Y_{ij} along with p -dimensional covariates x_{ij} where i indexes clusters (subjects) ($i = 1, \dots, m$) and j indexes units within clusters ($j = 1, \dots, n_i$), and that we want to assess the association of within-cluster changes in x with a known function of $E(Y)$. We further assume that we have auxiliary quantities $Z_i = (Z_{i1}, \dots, Z_{ik_i})$ that are associated with both Y_i and x_i and that the i^{th} subject (cluster) is chosen for the study with a probability based on Z_i (and possibly x_i). We assume that the objectives of the study are to assess the *individual-specific* (within-subject or within-cluster) association of X with Y using all of the data and to examine within-subject (cluster) aggregation. The most common way to handle individual-specific effects is to use generalized linear mixed models (GLMMs) since these models enable us to estimate individual-specific covariate effects (McCulloch et al., 2008). We also seek an approach that accommodates a wide variety of outcome-related sampling schemes from relatively simple designs where subject selection depends on a single outcome to more complicated designs where selection depends on several outcomes, perhaps through subject-specific trajectories.

In a prospective longitudinal study, the standard way to fit generalized linear mixed models is through maximum likelihood. In the special case of models with only random intercepts, an alternative method is to condition on the sum of the responses within a cluster and then use conditional maximum likelihood. This conditioning eliminates the random intercept from the conditional likelihood. Neither of these approaches is generally valid with data from an outcome-related sample, however (Neuhaus and Jewell, 1990). In this paper, we adapt both these likelihood-based methods so that they can handle longitudinal data from outcome-related sampling designs such as those used in the ADHD and OAI studies. In particular, we extend a profile likelihood approach developed in a series of papers by Scott, Wild and Neuhaus (Scott and Wild, 1997, 2001; Neuhaus et al., 2002, 2006) to accommodate longitudinal data for these sampling designs. We will also correct standard conditional likelihood methods to provide consistent estimation in canonical link, random intercept model settings. A key ingredient that enables us to “undo” the effect of the sampling design in both cases is a model, $pr(z / y, x)$, for the variable z that determines the

probability of inclusion in the study in terms of the response variables and the covariates. We illustrate our approaches using data from the ADHD study (Hartung et al., 2002) and simulation studies.

While this paper focuses on subject-specific models for longitudinal data and likelihood-based methods, we note that Schildcrout and Rathouz (2010) proposed population-averaged methods to analyze longitudinal data gathered using outcome sampling designs. In particular, Schildcrout and Rathouz (2010) assumed that interest lies in fitting a model for the *population-averaged*, or marginal, mean $E(Y_i | \mathbf{x}_i)$ and they developed an approach based on generalized estimating equations to accomplish this.

2. Basic Theory

2.1 Semiparametric Approach

Suppose that we have a cohort or finite population of N clusters with values (z_i, y_i, \mathbf{x}_i) generated independently from some joint distribution. Recall that \mathbf{y} contains the longitudinal responses (e.g. ADHD) and that our model of interest is $f(\mathbf{y} | \mathbf{x}; \boldsymbol{\theta})$, the conditional distribution of \mathbf{Y} given \mathbf{x} in the process that generated the cohort. We have information on all of the z_i s and based upon z_i we either observe (y_i, \mathbf{x}_i) (set $R_i = 1$) or do not (set $R_i = 0$). For example, in the ADHD study \mathbf{Z} is a simple binary variable coding whether or not parents or teachers suspected that the child was exhibiting ADHD symptoms prior to the start of the study. Following standard practice in outcome-related sampling (Scott and Wild, 1997, 2001), we work with the likelihood conditional on $R_i = 1$, since we assume that the marginal distribution of R_i contains no information about the parameters of interest.

The resulting likelihood is

$$\prod_{i=1}^N pr(z_i) pr(R_i | z_i) pr(\mathbf{y}_i, \mathbf{x}_i | R_i, z_i)^{R_i} = \prod_{i=1}^N pr(z_i, \mathbf{y}_i, \mathbf{x}_i)^{R_i} pr(z_i)^{(1-R_i)} pr(R_i | z_i), \quad (1)$$

since $pr(\mathbf{y}_i, \mathbf{x}_i | R_i, z_i)$ does not depend on R_i . We can insert the model of interest into the likelihood using the following decomposition of the joint distribution into conditional distributions:

$$pr(z_i, \mathbf{y}_i, \mathbf{x}_i) = pr(z_i | \mathbf{y}_i, \mathbf{x}_i) pr(\mathbf{y}_i | \mathbf{x}_i; \boldsymbol{\theta}) g(\mathbf{x}_i).$$

Here, $g(\mathbf{x}_i)$ denotes the marginal density of the covariates \mathbf{x}_i . In prospective regression the term in the likelihood involving $pr(\mathbf{y}_i | \mathbf{x}_i; \boldsymbol{\theta})$ is orthogonal to the term involving $g(\mathbf{x}_i)$ so the latter can be ignored; here they are inextricably inter-related. Since the marginal distribution of \mathbf{X} is of no direct interest and may be very complicated we treat $g(\mathbf{x})$ nonparametrically as a (potentially infinite-dimensional) nuisance parameter. We will, however, use a parametric model, $r(z | \mathbf{y}, \mathbf{x}; \boldsymbol{\theta})$, for $pr(z | \mathbf{y}, \mathbf{x})$. In our example we model \mathbf{Z} (determined prior to the study) in terms of Y_1 , the ADHD state at the first visit, in analyses reported in Section 4. In other analyses, not reported here, we allowed the distribution of \mathbf{Z} also to vary, for example, with gender. No distributional assumptions were needed because we could use saturated models. The likelihood (1) now becomes

$$L(\boldsymbol{\gamma}, \boldsymbol{\theta}, \mathbf{g}) = L(\boldsymbol{\delta}, \mathbf{g}) = \prod_{i=1}^N \{pr(z_i | \mathbf{y}_i, \mathbf{x}_i; \boldsymbol{\gamma}) pr(\mathbf{y}_i | \mathbf{x}_i; \boldsymbol{\theta}) g(\mathbf{x}_i)\}^{R_i} pr(z_i)^{(1-R_i)}, \quad (2)$$

where we use $\delta = (\boldsymbol{\gamma}, \boldsymbol{\theta})$ and have omitted $\prod_{i=1}^N pr(R_i|z_i)$ since we assume that it does not involve any of $(\boldsymbol{\gamma}, \boldsymbol{\theta}, g)$. We note that if we write $\tilde{\boldsymbol{y}} = (\boldsymbol{z}, \boldsymbol{y})$ and

$$pr(\tilde{\boldsymbol{y}}|\boldsymbol{x};\delta) = f(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}) r(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{x};\boldsymbol{\gamma})$$

then (2) falls into the class of likelihoods in Scott and Wild (2006).

From now on we restrict our attention to the subclass in which \boldsymbol{Z} can take only a finite set of values, say $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_L\}$. Let $q_\ell = pr(\boldsymbol{Z} = \boldsymbol{v}_\ell)$ and suppose that there are N_ℓ clusters in the cohort, and n_ℓ in the sample, for which $\boldsymbol{z}_i = \boldsymbol{v}_\ell$. Then (2) can be written in the form

$$L_1(\delta, \boldsymbol{g}) = \prod_{\ell=1}^L \left\{ \prod_{i \in \mathcal{D}_\ell} pr(\tilde{\boldsymbol{y}}_i|\boldsymbol{x}_i;\delta) g(\boldsymbol{x}_i) \right\} q_\ell^{N_\ell - n_\ell}, \quad (3)$$

where \mathcal{D}_ℓ denotes the set of sample clusters with $\boldsymbol{z}_i = \boldsymbol{v}_\ell$ (i.e. clusters with $R_i = 1$ and $\boldsymbol{z}_i = \boldsymbol{v}_\ell$ for $\ell = 1, \dots, L$). The semiparametric maximum likelihood methods of Scott and Wild (1997, 2001), expressed in most generality in Section 5 of Scott and Wild (2006), apply precisely to likelihoods of this form. If we let $\ell_p(\delta) = \sup_g \log \{L_1(\delta, \boldsymbol{g})\}$ denote the profile log-likelihood of $\delta = (\boldsymbol{\theta}, \boldsymbol{\gamma})$ after maximizing over all possible values \boldsymbol{g} , then Scott and Wild (2001) show that we can obtain the maximum profile likelihood estimate, $\hat{\delta}$, by solving the score equations derived from

$$\ell^*(\delta, \boldsymbol{q}) = \sum_{\ell=1}^L \left\{ \sum_{D_\ell} \log pr_\ell^*(\tilde{\boldsymbol{y}}_{\ell k}|\boldsymbol{x}_{\ell k};\delta, \boldsymbol{q}) \right\} + (N_\ell - n_\ell) \log q_\ell, \quad (4)$$

with

$$pr_\ell^*(\tilde{\boldsymbol{y}}|\boldsymbol{x};\delta, \boldsymbol{q}) = \frac{\rho_\ell pr(\tilde{\boldsymbol{y}}|\boldsymbol{x};\delta)}{\sum_{i=1}^L \sum_{\boldsymbol{y} \in \mathcal{S}_i} \rho_i pr(\tilde{\boldsymbol{y}}|\boldsymbol{x};\delta)}, \quad (5)$$

$\rho_i = (N_i - n_i)/q_i - 1$, and $\mathcal{S}_1, \dots, \mathcal{S}_L$ are strata defined by the values of \boldsymbol{v} . Here we treat \boldsymbol{q} as a set of unknown parameters (for subtleties in handling \boldsymbol{q} , see Scott and Wild (2006)). Although \boldsymbol{q} is known, previous research has shown (e.g. Kalbfleisch et al., 1999) that estimating it improves efficiency, so we follow that approach here. In other words, we can obtain the maximum likelihood estimator of δ by solving $\partial \ell^*(\phi) / \partial \phi = \mathbf{0}$ where $\boldsymbol{\phi} = (\boldsymbol{\delta}, \boldsymbol{q})$. The semiparametric efficiency of the resulting estimator has been established by Breslow et al. (2003) and Lee and Hirose (2010). We can also obtain an estimate of $Var\{\hat{\delta}\}$ by treating $\ell^*(\phi)$ as if it were a likelihood. More specifically, if we let $\mathcal{J}(\phi) = -\partial^2 \ell^*(\phi) / \partial \phi^2$ be the observed information matrix, then the appropriate block of $\mathcal{J}(\hat{\phi})^{-1}$ gives a consistent estimate of $Cov\{\hat{\delta}\}$ under fairly general conditions (see Lee and Hirose (2010) for details).

We have previously developed a profile-likelihood approach that addresses both the clustering and sampling at differential rates based directly on the outcome variable \boldsymbol{Y} (Scott and Wild, 1997; Neuhaus et al., 2002, 2006) so that one role of this paper is to extend the work of Neuhaus et al. (2006) to longitudinal designs that select clusters based on an auxiliary variable \boldsymbol{Z} which may be related to \boldsymbol{Y} .

2.2 Generalized linear mixed models

While the previous theory applies generally, for the rest of this paper we shall assume that our regression model, $f(\mathbf{y} / \mathbf{x}; \boldsymbol{\theta})$, is a generalized linear mixed model (McCulloch et al., 2008). Such models specify that, given a random vector \mathbf{b}_i of parameters specific to the i^{th} cluster the conditional density of Y_{ij} , the response for the j^{th} unit in the i^{th} cluster, is

$$f_Y(y_{ij} | \mathbf{b}_i, \mathbf{x}_{ij}) = \exp \left[\{y_{ij} \Delta_{ij} - c(\Delta_{ij})\} \phi + d(y_{ij}, \phi) \right] \quad (6)$$

where c and d are functions of known form, ϕ is a scale parameter and Δ_{ij} is a function of $\mu_{ij} = E(Y_{ij} / \mathbf{b}_i, \mathbf{w}_{ij}, \mathbf{x}_{ij})$ and hence depends on covariates \mathbf{x}_{ij} through the assumption that

$$\mu_{ij} = g^{-1} \left(\mathbf{w}_{ij}^T \mathbf{b}_i + \mathbf{x}_{ij}^T \boldsymbol{\beta} \right) \equiv g^{-1}(\eta_{ij}). \quad (7)$$

Here \mathbf{x}_{ij}^T and \mathbf{w}_{ij}^T are specified covariate row vectors relating the fixed and random effects, respectively, to the observations and g is a link function. Given \mathbf{b}_i , we assume that the responses Y_{i1}, \dots, Y_{in_i} are independent. Although the theory applies broadly, our example and simulations focus mainly on models with random intercepts and slopes, as well as models with random intercepts only. In such cases, $\mathbf{b}_i = (b_{0i}, b_{1i})^T$ is a two dimensional vector, as is \mathbf{w}_{ij} and we can write the covariate function in (7) as

$$\eta_{ij} = \beta_0 + \sigma_{b0} b_{0i} + \mathbf{x}_{1ij}^T (\beta_1 + \sigma_{b1} b_{1i}) + \mathbf{x}_{2ij}^T \beta_2, \quad (8)$$

where $E(b_{0i}) = E(b_{1i}) = 0$, $\text{var}(b_{0i}) = \text{var}(b_{1i}) = 1$, $\text{corr}(b_{0i}, b_{1i}) = \rho$ and where we have separated $\boldsymbol{\beta} = (\gamma_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2)^T$, into an intercept, a regression parameter associated with a random slope and a regression parameter with no corresponding random slope term. In this formulation the random effects have variances equal to 1 and are scaled by σ_{b0} and σ_{b1} in the model equation to obtain the desired magnitude of variability.

2.3 Conditional likelihood methods

With a canonical link, the parameter Δ_{ij} in (6) is equal to η_{ij} . In the special case of a random intercept only GLMM (equation 8 with $\sigma_{b1} = 0$), this leads to

$$\Delta_{ij} = \beta_0 + \sigma_{b0} b_{0i} + \mathbf{x}_{1ij}^T \beta_1 + \mathbf{x}_{2ij}^T \beta_2.$$

The conditional likelihood approach treats the intercepts $\sigma_{b0} b_{0i}$ as fixed constants and eliminates them from the likelihood by conditioning on their sufficient statistics

$\bar{y}_i = (1/n_i) \sum_{j=1}^{n_i} y_{ij}$. Thus, if we have a random sample of clusters, the conditional likelihood has the form

$$CL(\boldsymbol{\beta}_W) = \prod_{i=1}^m f(\mathbf{y}_i | \mathbf{x}_i, \bar{y}_i), \quad (9)$$

which depends only on \mathbf{x}_{Wi} and $\boldsymbol{\beta}_W$, where \mathbf{x}_{Wi} is the portion of the covariates in (7) that varies within clusters and $\boldsymbol{\beta}_W$ is the corresponding regression coefficient. Since $CL(\boldsymbol{\beta}_W)$ is a valid likelihood, we can use standard likelihood theory to make inferences about $\boldsymbol{\beta}_W$.

Conditional likelihood has a number of advantages over full maximum likelihood. It is simpler to compute since no numerical integration is required and there is no need to specify a distribution for the random intercepts, b_{0i} . No model is needed for components of \mathbf{x} that are constant within clusters and so it is robust against their misspecification. It gives

consistent estimates when the random effects are correlated with one or more of the predictors (Neuhaus and McCulloch, 2006). Moreover, there is little loss of efficiency in estimating coefficients of quantities that vary mostly within clusters (Neuhaus and Lesperance, 1996).

However, the conditional likelihood approach must be corrected to accommodate outcome-related sampling. Thus, in addition to conditioning on \bar{y}_i as above, we also need to condition on the fact that the cluster has been selected for study based upon z_i . Our corrected conditional likelihood takes the form

$$CL_C(\beta_W) = \prod_{i=1}^m f(z_i, \mathbf{y}_i | \mathbf{x}_i, \bar{y}_i, R_i=1), \quad (10)$$

where, as in §2.1, R_i is a binary variable taking the value 1 if the i th cluster is selected for the study and 0 otherwise. Using Bayes theorem we have,

$$f(z_i, \mathbf{y}_i | \mathbf{x}_i, \bar{y}_i, R_i=1) = \frac{\text{pr}(R_i=1 | z_i, \mathbf{y}_i, \mathbf{x}_i, \bar{y}_i) \text{pr}(z_i, \mathbf{y}_i | \mathbf{x}_i, \bar{y}_i)}{\text{pr}(R_i=1 | \mathbf{x}_i, \bar{y}_i)} \quad (11)$$

$$= \frac{\text{pr}(R_i=1 | z_i) r(z_i | \mathbf{y}_i, \mathbf{x}_i; \gamma)}{\text{pr}(R_i=1 | \mathbf{x}_i, \bar{y}_i)} \times f(\mathbf{y}_i | \mathbf{x}_i, \bar{y}_i; \beta_W), \quad (12)$$

where we assume that the probability of selecting cluster i depends only on the value of z_i (which could partly contain \mathbf{x}_i). Note that $f(\mathbf{y} | \mathbf{x}, \bar{y}_i)$ is the standard conditional likelihood term in (9) so the first term is the correction for outcome-related ascertainment. Calculation of the correction term requires $r(z | \mathbf{y}, \mathbf{x}; \gamma)$ (as in §2.1) and the sampling probabilities $\text{pr}(R_i = 1 | z_i)$.

For the ADHD data, z_i is the binary referral/selection variable at baseline and one just needs to specify a simple binary regression model for $z_{i1} | y_{i1}$. For the OAI data, z_i can be more complicated, e.g., a longitudinal course of X-ray outcomes, and one may need to specify more complex models. For the ADHD data, $\text{pr}(R_i = 1 | z_i)$ is unknown but Schildcrout and Rathouz (2010) provide a range of estimates. For the OAI data and other studies that subsample existing cohorts, $\text{pr}(R_i = 1 | z_i)$ will be known since investigators will specify sampling rates. However, since the profile likelihood (4) also depends on $\text{pr}(R_i = 1 | z_i)$ implicitly through the specification of the $N_{\ell S}$, it is worthwhile to investigate the potential effects of misspecifying the sampling rates. Having specified both $\text{pr}(R_i = 1 | z_i)$ and a binary regression model for the sampling variable Z , we can then make inferences about β_W by applying standard likelihood methods to $CL_C(\beta_W)$.

Note that in one special case, we can ignore the correction term. If z_i depends only on \bar{y}_i as in Neuhaus and Jewell (1990), then $f(\mathbf{y}_i | \mathbf{x}_i, \bar{y}_i, R_i = 1)$ reduces to $f(z_i | \bar{y}_i) f(\mathbf{y}_i | \mathbf{x}_i, \bar{y}_i)$. Assuming the first factor does not involve β_W , standard conditional likelihood methods then give valid inferences without the need for any correction.

3. Simulations

We conducted simulation studies to illustrate the magnitudes of bias resulting from ignoring an outcome-related sampling plan in a cluster-specific analysis and to assess the efficiency

of estimators based on the corrected conditional likelihood approach (12) with respect to full, profile likelihood estimators (4). The first set of simulations generated data from simple models containing only random intercepts and assessed the performance of conditional likelihood methods. Such models and methods might be appropriate in longitudinal studies with few repeated measures for each subject since they may be approximately correct. The second set of simulation studies generated data using more appropriate models for longitudinal settings, i.e. containing both random intercepts and slopes, and fit such models to the generated data. Since both profile likelihood (4) and conditional likelihood approaches (12) depend on specification of $pr(R_i = 1 | z_i)$, we conducted additional simulations to examine the effect of misspecifying the sampling rate $pr(R_i = 1 | z_i)$ of the profile likelihood (4) and conditional likelihood approaches (12). As in Schildcrout and Rathouz (2010), the simulations generated longitudinal data to resemble the ADHD data. Specifically, the simulations generated populations of 5000 subjects with auxiliary variable $Z = 1$ or $Z = 0$, along with repeated binary ADHD responses from 4 or 8 visits. We gathered samples of approximately 150 subjects from each population using an outcome-related sampling design where sampling depended on the value of an auxiliary variable measured at time 1.

In our initial simulations we generated longitudinal binary responses from simple mixed-effects logistic models with random intercepts :

$$\begin{aligned} y_{ij} | b_i, \mathbf{x}_{ij} &\stackrel{\text{ind}}{\sim} f_{y|b,x}, \\ \text{logit} \{ \text{pr} (Y_{ij}=1 | b_i, \mathbf{x}_{ij}) \} &= \beta_0 + \sigma_b b_i + \mathbf{x}_{ij}^T \beta, \quad (13) \\ b_i &\stackrel{\text{ind}}{\sim} N(0, 1). \end{aligned}$$

The first set of simulations included two within-cluster covariates: 1) x_t , a variable taking on equally spaced values in $(1, n)$ for $n=4$ or 8 ; and 2) x_{bin} , a binary variable taking on values 0 and 1 each with probability 0.5. Additional simulations also included x_{norm} , a standard normal variable. The parameter values for the simulations were $\beta_0 = -5.5$, $\beta_t = 0.2$, $\beta_{bin} = -1.0$, $\beta_{norm} = 0.5$ and $\log \sigma_b = 1.0$, values similar to those obtained from fits of mixed-effects logistic models using (4) to the ADHD data (Table 3).

We sampled the i^{th} cluster from the sub-populations defined by $z_{i1} = 0$ and $z_{i1} = 1$, where the sampling variable z_{i1} was associated with the outcome Y through the model:

$$\text{logit} \{ \text{pr} (Z_{i1}=1 | y_{i1}) \} = \gamma_0 + \gamma_1 y_{i1}. \quad (14)$$

We set $\gamma_0 = -3.2$ and $\gamma_1 = 3.0$ to create strong dependence between Z_{i1} and Y_{i1} and a realistic prevalence of $Z_{i1} = 1$. The sampling design kept all clusters with $Z_{i1} = 1$, giving from 250 to 300 clusters per simulation repetition, and an equal number of clusters with $Z_{i1} = 0$.

We fit three approaches to these data:

1. a profile likelihood approach (4) using a mixed-effects logistic model (13) along with a logistic model (14) to relate the outcome Y to the sampling variable Z and the observed rates of sampling the subpopulations with $Z_{i1} = 1$ and $Z_{i1} = 0$;
2. a corrected conditional likelihood approach (12) using the logistic model (14) and sampling rates above;
3. a standard conditional likelihood approach that ignored the outcome-related sampling.

We fit approaches (1) and (2) using R routines written by two of the authors (CW and YJ) and fit approach (3) using a standard conditional likelihood approach in the R package.

Table 1 presents relative biases in parameter estimates from the three approaches as means of the simulation estimates minus true values divided by true values, $(\bar{\beta} - \beta) / \beta \times 100\%$. As expected, Table 1 shows that all estimators from the profile likelihood (4) and corrected conditional likelihood (12) approaches exhibit essentially no bias; nearly all estimated biases are less than 1.0%. Also as expected, ignoring the outcome-related sampling design by fitting a standard conditional likelihood approach produced large biases in estimates of β_t , the effect of x_t , a variable strongly associated with the response at time 1. However, the standard conditional likelihood estimators of both β_{bin} and β_{norm} , effects of variables independent of x_t exhibited essentially no bias.

We also calculated observed estimation efficiencies of the corrected conditional likelihood (12) estimators relative to the profile likelihood (4) estimators of β_t , β_{bin} and β_{norm} and report full results in Web Appendix A. The results indicate that the corrected conditional likelihood (12) estimators of β_t were highly efficient and are consistent with those of Neuhaus and Lesperance (1996) who showed that conditional likelihood estimators are fully efficient with respect to full maximum likelihood for covariates such as x_t that are maximally different within clusters and that estimation efficiency increases with cluster size.

The next set of simulations generated longitudinal binary responses from a design that would often be more appropriate for longitudinal studies, namely a mixed-effects logistic model with random intercepts and slopes:

$$\begin{aligned} y_{ij}|b_i, x_{ij} &\stackrel{\text{ind}}{\sim} f_{y|b,x}, \\ \text{logit} \{ \text{pr} (Y_{ij}=1|b_i, x_{ij}) \} &= \beta_0 + \sigma_{b0} b_{0i} + x_{1ij}^T (\beta_1 + \sigma_{b1} b_{1i}) + x_{2ij}^T \beta_2, \\ \begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} &\stackrel{\text{ind}}{\sim} \text{Normal} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{b0}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{b1}^2 \end{bmatrix} \right). \end{aligned} \quad (15)$$

The simulations included three covariates to model a longitudinal study of two groups followed over time: 1) x_t , with associated parameter β_t , a “time” covariate taking on 8 equally spaced values in (0,1) ; 2) x_G , a “group” covariate, with associated parameter β_G , a binary variable taking on values 0 or 1 for half the population; and 3) $x_I = x_G \times x_t$, an “interaction” covariate, with associated parameter β_I . The parameter values for the simulations were $\beta_0 = -4.5$, $\beta_t = \beta_G = \beta_I = 1.0$, $\log \sigma_{b0} = 1.0$, $\log \sigma_{b1} = 0$ and $\text{corr}(b_{0i}, b_{1i}) = 0.5$.

We sampled the i^{th} cluster from the sub-populations defined by $z_{i1} = 0$ and $z_{i1} = 1$, where the sampling variable z_{i1} was associated with the outcome Y through the model:

$$\text{logit} \{ \text{pr} (Z_{i1}=1|y_{i1}) \} = \gamma_0 + \gamma_1 y_{i1}. \quad (16)$$

We set $\gamma_0 = -4.5$ and $\gamma_1 = 4.0$ to create strong dependence between Z_{i1} and Y_{i1} and a reasonable prevalence of $Z_{i1} = 1$. The sampling design kept all clusters with $Z_{i1} = 1$, giving from 250 to 300 clusters per simulation repetition, and an equal number of clusters with $Z_{i1} = 0$.

We fit four approaches to these data:

1. a profile likelihood approach (4) using a mixed-effects logistic model (15) with random slopes and intercepts along with a logistic model (16) to relate the outcome Y to the sampling variable Z and the observed rates of sampling the subpopulations with $Z_{i1} = 1$ and $Z_{i1} = 0$;

2. a standard mixed-effects logistic model (15) with random slopes and intercepts that ignored the outcome-related sampling;
3. a corrected conditional likelihood approach (12) using the logistic model (14) and sampling rates above;
4. a standard conditional likelihood approach that ignored the outcome-related sampling. We fit approaches (1) and (3) using R routines written by authors CW, YJ and RB. We fit approach (2) using the NLMIXED procedure in SAS and fit approach (4) using a standard conditional likelihood approach in the R package.

The first set of simulations correctly specified the logistic model (16) and used the actual subpopulation sampling rates. In the second set of simulations, which we report in Web Appendix A, we misspecified the sampling rates $pr(R_i = 1 | Z_{i1} = 0)$ to be one-half or twice the observed rates. That is, we correctly specified that we sampled all subjects with $z_{i1} = 1$, but we misspecified the sampling rates for subjects with $z_{i1} = 0$.

Table 2 presents relative biases in parameter estimates as means of the simulation estimates minus true values divided by true values. As expected, Table 2 shows that all estimators from the profile likelihood (4) approach exhibit essentially no bias; nearly all estimated biases are less than 1.0%. Also as expected, ignoring the outcome-related sampling design by fitting a standard mixed effects logistic model produced large biases in estimates of all parameters. In particular, the biases in estimates of β_0 , β_t and β_G all exceeded 30%. The conditional likelihood estimators also exhibited large bias, as expected since they remove the effects of random intercepts, but not slopes, from the likelihood.

Web Appendix A presents estimated biases from a modification of the Table 2 setting where we misspecify the $Z = 0$ subpopulation sampling rates in the profile likelihood approach. These simulation studies also assessed the performance of 95% confidence intervals. To summarize the findings, misspecifying the sampling rates produced mild bias in estimators of β_t , but essentially no bias in estimates of both β_G and β_t , effects of variables not as strongly connected to time 1 as is x_t . Coverage rates for β_t were poor in settings with sampling rates misspecified to be one-half the actual rate, but were close to nominal for the the other two regression parameters, β_G and β_t . Coverage rates for all three regression parameters were close to nominal in settings with sampling rates misspecified to be twice the actual rate.

We ran analogous simulations to assess the effects of misspecifying sampling rates with data generated from models with only random intercepts and report full results in Web Appendix A. In general, the results from models with only random intercepts closely corresponded to those from models with both random intercepts and slopes.

4. Example

We illustrate our results by fitting profile likelihood and conditional likelihood approaches to data from the ADHD study (Hartung et al., 2002). The data set consists of 138 children suspected to have ADHD, 117 not suspected to have ADHD, followed up to 8 annual visits. Covariates of interest included time, gender, ethnicity, and several interactions. The augmented response included the measured ADHD symptom expression outcome, Y_i , as well as the referral/sampling group variable Z_i . Preliminary model fits indicated that $\text{logit}\{\text{pr}(Y_{ij} = 1 | \text{visit}, b)\}$ followed a non-linear trajectory in the visit variable which we could appropriately describe using binary indicators for visits 1 and 2, along with a linear effect of visit. We fit four cluster-specific approaches to these data: 1) the profile likelihood approach (4) with a binary mixed-effects logistic model with random intercepts and slopes (15) and the augmented response $(Y_i, Z_{i1})^T$; 2) the profile likelihood approach in 1) that only

included random intercepts; 3) the corrected conditional likelihood (12) approach; and 4) a standard conditional likelihood approach that ignored the outcome-related sampling design. Additional covariates included indicators for female gender, African American and Other ethnicity, as well as interactions of visit by female gender and visit by African American ethnicity. We also fit a marginal (population-averaged) model with the same set of covariates using the approach of Schildcrout and Rathouz (2010) and software they generously provided. The fit followed the recommendation of Schildcrout and Rathouz (2010) to fit a very flexible model for the sampling variable Z . The model included all the predictors above, the ADHD responses, Y_{ij} and interactions of Y_{ij} with each of the predictors.

Our general approach is to fit flexible models to describe the relationship between the referral/selection variable Z and the vector of measured ADHD symptoms, Y . The variable Z was strongly related to the first outcome, Y_1 , the closest outcome in time to Z , but only weakly related to the later outcomes, Y_2, \dots, Y_8 , conditional on Y_1 . For example, a Wald test of $H_0 : \gamma_2 = \dots = \gamma_8 = 0$ based on a logistic model yielded a χ^2 statistic of 10.6 on 7 degrees of freedom, $p=0.16$, along with estimates $\hat{\gamma}_j$ that were much smaller than $\hat{\gamma}_1$. This makes sense since Z is an attempt to measure the ADHD outcome near the beginning of the study. Outcomes Y_2, \dots, Y_8 are progressively farther away from Z and more weakly related.

We modeled the relationship between the referral/selection variable, Z and the measured ADHD symptom outcome, Y using $\text{logit } pr(Z_{ij} = 1 | y_{i1}) = \gamma_0 + \gamma_1 y_{i1}$. To implement both profile likelihood and conditional likelihood approaches we must also specify the ratio

$$\lambda = \frac{pr(R_i=1|z_{i1}=1)}{pr(R_i=1|z_{i1}=0)},$$

which we can compute using specifications of $pr(Z_i)$, the observed $pr(Z | R = 1)$ and Bayes theorem. Following Schildcrout and Rathouz (2010), we assume that approximately 5% of girls in the population would qualify for referral and consider three different prevalence rates for the boys: 5%; 10%; and 15%. We calculate the required ratio to be 22.6 for the girls, and 22.4, 10.6 and 6.7, for the boys and prevalences of 5%, 10%, and 15%, respectively. Table 3 presents the results for sampling ratios $\lambda(girls) = 22.6$, $\lambda(boys) = 22.4$. Web Appendix A presents analogous results for the other two sampling ratio sets.

The estimate of $\log \hat{\sigma}_1$ for the PL(slp) fit is much larger than its standard error (Table 3), indicating the need to include random slopes in the model for the ADHD outcomes. As in Table 1, Table 3 shows that ignoring the outcome-related sampling plan using a standard conditional likelihood approach produced a very different estimate of the Visit 1 effect than the profile and corrected conditional likelihood approaches that accommodate the sampling design. To get a clearer sense of the differences in the model fits we plotted the estimated longitudinal change for the reference group (Caucasian boys) versus visit for the various model fits. For the mixed-effects models, the fit is for the average value (0) of the random effects and, for plotting purposes, we set the intercept for the conditional maximum likelihood fits (which do not supply an intercept estimate) to that from the PL(int) model fit.

Figure 1 gives those plots and shows qualitative differences in the estimated change over time. As noted above, the uncorrected conditional likelihood estimator, which fails to accommodate the outcome-related selection at visit 1 gives a very different estimate there. Otherwise, the profile likelihood fit assuming only random intercepts and the two conditional likelihood fits give very similar results with the odds of ADHD increasing up to visit 3 and beginning to decline. However, those fits differ substantially from the profile

likelihood fit allowing random slopes as well as random intercepts. In that model, the longitudinal trend increases and then essentially levels off over visits 3 through 8. The population averaged approach gives a trend similar to that of the random intercepts and slopes fit, but with values attenuated compared to the mixed effects model. Given the different targets of inference from the subject-specific and marginal approaches, such attenuation is to be expected.

The example illustrates several important points. First, methods such as standard conditional likelihood that ignore the outcome-related sampling can give drastically incorrect fits. Second, ignoring random slopes by fitting only random intercept models (and thus specifying an incorrect dependence structure) can lead to somewhat different qualitative fits with nonlinear models such as the logistic. Third, we can uncover such omissions in the dependence structures by fitting the more elaborate model including random slopes as well as random intercepts. Web Appendix A shows that approaches using the sampling ratios 10.6 and 6.7 for the boys produced noticeable changes in the coefficients of visit 1, female and the visit by female interaction, variables all related to the sampling process.

5. Discussion

The most common approaches to the analysis of longitudinal data are generalized estimating equations or GEEs (which typically fit marginal models) and mixed-effects or conditional regression models (which typically fit conditional models). The advantages and disadvantages of each approach have been widely debated (Neuhaus et al., 1991; McCulloch et al., 2008). Key aspects of that debate are: 1) robustness to dependence model misspecification for GEE methods, 2) cluster-specific interpretations for mixed-effects and conditional regression approaches (which are often more natural in longitudinal data settings), 3) more explicit modeling (and correspondingly more detailed information) for mixed-effects models, and 4) avoidance of distributional assumptions for cluster-specific intercepts for conditional analysis methods. The work of Schildcrout and Rathouz (2010) extended GEE approaches to handle outcome-related sampling whereas our work provides that extension for analysts who prefer to use mixed-effects and conditional approaches.

Our profile and conditional likelihood approaches can accommodate data from simple schemes, such as the design of the ADHD study, as well as more complicated schemes that involve several outcomes, such as a longitudinal sample of MRI outcomes sampled based on trajectories of X-ray measurements in the OAI study. Indeed, OAI investigators are very interested in comparing subjects who exhibit large within-subject changes in MRI outcomes with subjects who have little change, but the investigators need to efficiently select subjects based on, e.g., X-ray measurements.

Although the simulation studies and example data analyses in this paper focus on a particular binary-outcome, longitudinal model with random intercepts and slopes, our profile likelihood approach is much more general and applies to a wide variety of longitudinal responses and auxiliary variables. Given sampling rates, our approach applies in any setting where we specify fully parametric models to relate the outcome of interest to the covariates and to relate the sampling variable to the outcomes. Our corrected conditional likelihood approach also applies more broadly to any canonical link generalized linear model with a random intercept, for example, it applies to repeated Poisson outcomes.

Conditional likelihood methods are attractive in the setting of cluster-specific intercepts because they avoid any assumption about the distribution of the intercepts across clusters and may give robustness to other model violations (Neuhaus and McCulloch, 2006). As we demonstrated here, standard conditional likelihood methods do not work in the context of

outcome-related sampling. However, our corrected conditional likelihood method exhibits little to no bias in our simulation studies with cluster-specific intercepts.

While conditional methods have attractive properties, they only give information about time-varying predictors (variation within clusters) and discard information that might be usefully recovered between clusters. As an example, in the ADHD analysis we considered, we would not be able to characterize average effects for females or African-Americans using a conditional analysis. And our simulations showed loss of efficiency of up to 25% (see Web Appendix A) in situations where a predictor had both within- and between-subject variation, which might be concerning. Furthermore, when the models are more complicated than cluster-specific intercepts (e.g., cluster-specific intercepts *and* slopes), then both standard and corrected conditional analysis methods fail (see Table 2). Taking these drawbacks together and with demonstrated robustness to parametric assumptions in mixed effects models (Neuhaus and McCulloch, 2011), we generally prefer the flexibility afforded by mixed-effects regression methods over conditional approaches, except in extreme situations.

In summary, we have developed and evaluated two approaches to outcome-related designs for analysts who wish to fit cluster-specific models. For an outcome process with cluster-specific intercepts (only) and flexible accommodation of outcome-related sampling designs our corrected conditional likelihood method performed well. For more complicated longitudinal outcome settings (such as random intercepts and slopes) we recommend a profile likelihood approach.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Grants from the U.S. National Institutes of Health and the Marsden Fund of New Zealand. supported this research. We thank Dr. Jonathan Schildcrout of Vanderbilt University and Dr. Paul Rathouz of the University of Wisconsin for sharing the Attention Deficit Hyperactivity Disorder study data and the software they developed to fit their methods.

References

- Breslow NE, McNeney B, Wellner JA. Large sample theory for semiparametric regression models with two-phase, outcome dependent sampling. *Annals of Statistics*. 2003; 31:1110–1139.
- Hartung CM, Willcutt EG, Lahey BB, Pelham WE, Loney J, Stein MA, Keenan K. Sex differences in young children who meet criteria for attention deficit hyperactivity disorder. *J Clin Child Adolesc Psychol*. 2002; 31:453–464. [PubMed: 12402565]
- Kalbfleisch J, Lawless J, Wild C. Estimation for response-selective and missing data problems in regression. *Journal of the Royal Statistical Society, Series B*. 1999; 6:413–438.
- Lee AJ, Hirose Y. Semi-parametric efficiency bounds for regression models under response-selective sampling: the profile likelihood approach. *Annals of the Institute of Statistical Mathematics*. 2010; 62:1023–1052.
- McCulloch, CE.; Searle, SR.; Neuhaus, JM. *Generalized, Linear and Mixed Models*. Second Edition. Wiley; New York: 2008.
- Neuhaus JM, Jewell NP. The effect of retrospective sampling on binary regression models for clustered data. *Biometrics*. 1990; 46:977–990. [PubMed: 2085642]
- Neuhaus JM, Kalbfleisch JD, Hauck WW. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review*. 1991; 59:25–35.

- Neuhaus JM, Lesperance ML. Estimation efficiency in a binary mixed-effects model setting. *Biometrika*. 1996; 83:441–446.
- Neuhaus JM, McCulloch CE. Separating between- and within-cluster covariate effects using conditional and partitioning methods. *Journal of the Royal Statistical Society, Series B*. 2006; 68:859–872.
- Neuhaus JM, McCulloch CE. The effect of misspecification of random effects distributions in clustered data settings with outcome dependent sampling. *Canadian Journal of Statistics*. 2011; 39:488–497. [PubMed: 23204632]
- Neuhaus JM, Scott AJ, Wild CJ. The analysis of retrospective family studies. *Biometrika*. 2002; 89:23–37.
- Neuhaus JM, Scott AJ, Wild CJ. Family-specific approaches to the analysis of case-control family data. *Biometrics*. 2006; 62:488–494. [PubMed: 16918913]
- Schildcrout JS, Rathouz PJ. Longitudinal studies of binary response data following case-control and stratified case-control sampling: design and analysis. *Biometrics*. 2010; 66:365–373. [PubMed: 19673861]
- Scott AJ, Wild CJ. Fitting regression models to case-control data by maximum likelihood. *Biometrika*. 1997; 84:57–72.
- Scott AJ, Wild CJ. Maximum likelihood for generalised case-control studies. *J. Statist. Plan. Infer.* 2001; 96:3–27.
- Scott, AJ.; Wild, CJ. Calculating efficient semiparametric estimators for a broad class of missing-data problems.. In: Liski, EP.; Isotalo, J.; Niemela, J.; Puntanen, S.; Styan, GPH., editors. *Festschrift for Tarmo Pukkila on his 60th Birthday*. Univ. of Tampere; Tampere: 2006. p. 301-314.

Log odds of ADHD for five estimation methods

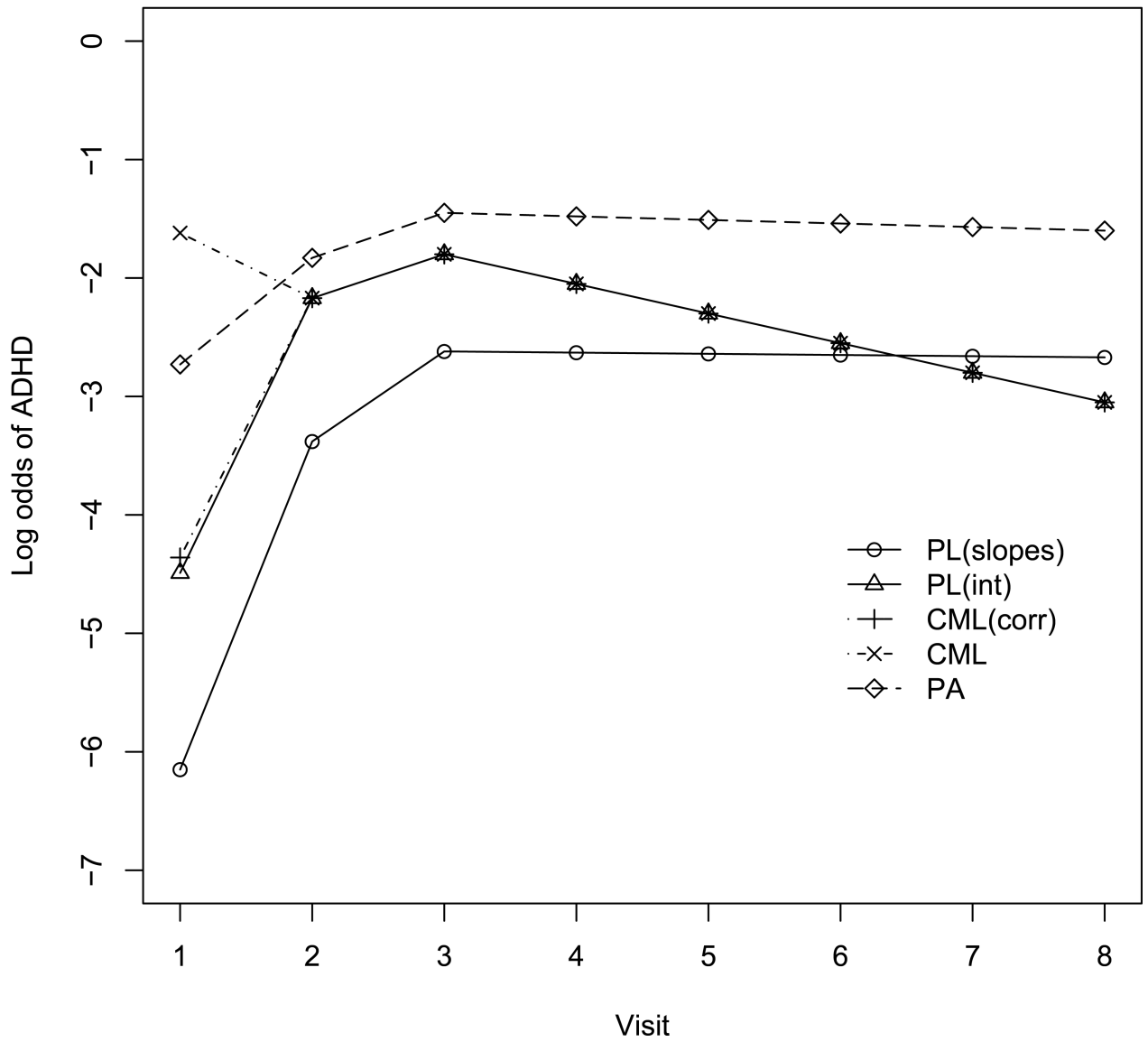


Figure 1. Log odds of ADHD for five estimation methods.

Table 1

Percentage bias in parameter estimates from three methods fit to simulated longitudinal data from outcome-related sampling designs: 1) Profile Likelihood (4) with random intercepts (PL(int)); 2) corrected conditional maximum likelihood (CML_{corr}) (12); 3) standard conditional maximum likelihood (CML).

Parameter	PL(int) (4)	CML _{corr} (12)	CML
a. $n_i = 8, x_t, x_{bin}$			
β_0	0.2		
β_t	0.8	0.6	-72.2
β_{bin}	0.3	0.7	0.5
γ_0	<0.1	<0.1	
γ_1	0.7	0.7	
$\log \sigma_b$	-0.3		
b. $n_i = 4, x_t, x_{bin}$			
β_0	0.6		
β_t	-0.5	-0.7	-258.4
β_{bin}	-0.4	<0.1	-0.7
γ_0	0.2	0.2	
γ_1	1.5	1.6	
$\log \sigma_b$	-0.2		
c. $n_i = 8, x_t, x_{bin}, x_{norm}$			
β_0	0.3		
β_t	1.4	1.2	-72.1
β_{bin}	-0.1	0.3	-0.7
β_{norm}	-0.2	0.1	-0.7
γ_0	<0.1	<0.1	
γ_1	2.0	2.0	
$\log \sigma_b$	-0.5		

Table 2

Percentage bias in parameter estimates from four methods fit to simulated longitudinal data from outcome-related sampling designs: 1) Profile Likelihood (4) with random intercepts & slopes (PL(slopes)); 2) standard mixed effects logistic random intercepts & slopes (STD mixed); 3) corrected conditional maximum likelihood (CML_{corr}) (12); 4) standard conditional maximum likelihood (CML).

Parameter	PL(slopes) (4)	STD mixed	CML _{corr} (12)	CML
β_0	-0.4	-39.6		
β_t	1.4	-42.0	50.5	-13.7
β_G	-0.6	31.8		
β_I	-0.6	-9.8	-36.2	-43.0
$\log \sigma_{b0}$	-0.8	20.8		
σ_{b1}^*	1.1	-14.7		
γ_0	0.1			
γ_1	-0.1			

* Percentage bias computed for σ_{b1} to avoid division by zero.

Table 3

Parameter estimates, $\hat{\beta}$, with standard errors as subscripts, from four subject-specific methods and one population-averaged (PA) methods fit to ADHD data. Subject specific methods included Profile Likelihood, as given in (4), with either random intercepts (PL(int)) or random intercepts and slopes (PL(slp)), standard conditional maximum likelihood (CML) and corrected conditional maximum likelihood (CML_{corr}) as given in (12)). The population averaged fit (PA(Y_i)) used the method of Schildcrout and Rathouz (2010). Sampling ratios: $\lambda_{\text{girls}} = 22.6$, $\lambda_{\text{boys}} = 22.4$.

parameter	PL(int)	PL(slopes)	CML _{corr}	CML	PA(Y _i)
Intercept	-1.05 _{0.37}	-2.59 _{0.63}			-1.36 _{0.30}
visit 1	-3.43 _{0.32}	-3.55 _{0.36}	-3.30 _{0.34}	-0.56 _{0.29}	-1.36 _{0.29}
visit 2	-0.62 _{0.26}	-0.77 _{0.30}	-0.62 _{0.26}	-0.62 _{0.26}	-0.41 _{0.24}
visit	-0.25 _{0.04}	-0.01 _{0.09}	-0.25 _{0.05}	-0.25 _{0.05}	-0.03 _{0.04}
Female	0.10 _{0.35}	0.03 _{0.48}			-0.33 _{0.36}
African Amer	-0.05 _{0.37}	0.18 _{0.53}			0.21 _{0.25}
Other	0.72 _{0.80}	1.13 _{0.73}			0.05 _{0.52}
visit*Female	-0.18 _{0.07}	-0.16 _{0.09}	-0.19 _{0.07}	-0.19 _{0.07}	-0.12 _{0.05}
visit*AfrAmer	0.27 _{0.06}	0.25 _{0.08}	0.27 _{0.06}	0.27 _{0.06}	0.15 _{0.04}
$\hat{\gamma}_0$	-5.46 _{0.30}	-5.46 _{0.30}	-5.46 _{0.32}		
$\hat{\gamma}_1$	6.55 _{0.80}	6.53 _{0.80}	6.50 _{0.78}		
$\log \hat{\sigma}_0$	0.87 _{0.07}	1.00 _{0.07}			
$\log \hat{\sigma}_1$		-1.03 _{0.12}			
corr(b_0, b_1)		-0.32 _{0.18}			