Exploratory assessment of treatment-dependent random-effects distribution using gradient functions

Takumi Imai
Department of Medical Statistics,
Graduate School of Medicine,
Osaka City University
Osaka, Japan

Shiro Tanaka
Department of Clinical Biostatistics,
Graduate School of Medicine,
Kyoto University
Kyoto, Japan

Koji Kawakami
Department of
Pharmacoepidemiology,
Graduate School of Medicine and
Public Health,
Kyoto University
Kyoto, Japan

ABSTRACT

In analyzing repeated measurements from randomized controlled trials with mixed-effects models, it is important to carefully examine the conventional normality assumption regarding the random-effects distribution and its dependence on treatment allocation in order to avoid biased estimation and correctly interpret the estimated random-effects distribution. In this paper, we propose the use of a gradient function method in modeling with the different random-effects distributions depending on the treatment allocation. This method can be effective for considering in advance whether a proper fit requires a model that allows dependence of the random-effects distribution on covariates, or for finding the subpopulations in the random effects.

1 INTRODUCTION

In randomized controlled trials (RCTs), some subject characteristics are monitored longitudinally and measured repeatedly. Subject-specific data features are often present, and a statistical model that accommodates these features is preferable for the analysis. One popular method for analyzing repeated measurements is mixed-effects modeling, where each subject-specific feature is explicitly taken into account as a random effect in the model. The inference is based on the marginal likelihood where the random effects are integrated over an assumed parametric distribution. Since the random effects are unobservable quantities, it is difficult to make assumptions about their distributions. The normality assumption is commonly made for practical reasons such as theoretical simplicity or analysis feasibility using standard software packages. However, mis-specifying the random-effects distribution is sometimes problematic [1, 2, 4, 13, 14, 17, 21] and can result in biased estimates of fixed effects [9] or lead to improper predictions regarding subject-specific features. [20] For example, Verbeke and Lesaffre (1996) reported simulations of repeated measurements in which the random-effects predictions under the normality assumption reflected neither the true ones nor the true random-effects distribution due to the so-called shrinkage effect.

To check the normality assumption on random-effects distributions, Verbeke and Molenberghs (2013) proposed an exploratory diagnostic tool based on the so-called gradient function. Although this tool is used for exploratory assessment, the gradient function offers insight on how the parametric distribution assumption is violated and what types of models would improve the fit in terms of likelihood. The calculation of the gradient function is based on the model fitting results and does not involve any technical computations.[15, 22] Verbeke and Molenberghs (2013) illustrated the use of the gradient function method combined with a finite mixture of normals, proposed by Verbeke and Lesaffre (1996), instead of using the simple normality assumption on random-effects distribution. Although mixtures of normals are very flexible and may offer very simple interpretations,[16, 22] it is often difficult to obtain stable solutions. Therefore, it was suggested that a finite mixture of normals for random-effects distributions could be incorporated into a general sensitivity analysis.[14] Insights from the gradient function can be utilized in model construction in this framework,[16, 22] and can also be applied to research involving joint modeling of longitudinal measurements and survival time.[3]

In addition to the normality assumption, dependence of the random-effects variance on covariates is also an important issue. The dependence on covariates is usually ignored in the modeling unless valid evidence is available; however, when fitting generalized linear mixed-effects models, the estimates of fixed effects may be sensitive to the assumption that random effects do not depend on the covariates. [8–10] Several models flexible enough to accommodate this dependence are available. [5, 7, 11, 12] However, it is ideal to consider in advance whether such complex models are indeed required. Furthermore, dependence of the random-effects distribution on covariates can occur concurrently with violation of the normality assumption. For example, due to treatment adherence or to the heterogeneity of the treatment effect, some non-trivial subpopulations with subject-specific features can appear in either treatment group. In these cases, modeling is usually difficult because there will be many complex model candidates. The insight from the gradient function will be useful in this context as well. In this paper, we apply the gradient function method in modeling with different random-effects distributions depending on the treatment allocation.

The remainder of the paper is organized as follows. In section 2, the details of two real data examples and their corresponding mixed-effects modeling strategies will be introduced. In section 3, the gradient function method by Verbeke and Molenberghs (2013) will be extended to the case in which the random-effects distribution may differ depending on the treatment allocation. In section 4, the utilities of the proposed methods will be investigated in several study settings for normal and binary outcomes using simulated datasets. In section 5, the methods will be applied to real datasets. In section 6, some practical issues regarding use of the gradient

function method for random-effects distribution assessment will be discussed.

2 MIXED-EFFECTS MODELS FOR REPEATED MEASUREMENTS

The application of the gradient function will be explained following the notation of Verbeke and Molenberghs (2013). Although the general form of random effects $\mathbf{b}_i$ with q components is adopted in sections 2 and 3, in this paper we mainly consider the random intercept models ($q = 1$). The application for random slope models ($q = 2$) is discussed in the supplementary materials. Here, i is the label for a total of N subjects. The repeated outcomes are expressed as $\mathbf{y}_i = (y_{i1}, ..., y_{in_i})^{\mathrm{T}}$, where n_i is the number of observations in subject i.

2.1 The mixed-effects model framework

In the mixed-effects model framework, the subject-specific feature is explicitly presented in the distribution function of the outcomes as $\mathbf{y}_i \mid \mathbf{b}_i \sim F_i(\mathbf{y}_i \mid \mathbf{b}_i)$. The corresponding density function is denoted as $f_i(\mathbf{y}_i|\mathbf{b}_i)$. Furthermore, the distribution function for the population of subject-specific features is explicitly expressed as $G(\mathbf{b}_i)$, where $\mathbf{b}_i$s are assumed to be randomly sampled according to this random-effects distribution. The corresponding density function is denoted as $g(\mathbf{b}_i)$.

For the likelihood-based inference for the parameters in $F(\mathbf{y}_i|\mathbf{b}_i)$ and $G(\mathbf{b}_i)$, the $\mathbf{b}_i$ is marginalized over $G(\mathbf{b}_i)$, denoted as,

$$f_i(\mathbf{y}_i|G) \equiv \int f_i(\mathbf{y}_i|\mathbf{b}_i)dG(\mathbf{b}_i) \,,$$

and the logarithm of marginal likelihood is given by

$$l[G] = \sum_{i=1}^{N} \ln\{f_i(\mathbf{y}_i|G)\} \,.$$

Thus, the forms of f_i or l highly depend on the choice of G.

Regarding model fitting, the candidates for G can be compared based on the likelihoods or other fit statistics. One of the important features of possible choices of G in the RCT setting is the dependence of G on the treatment allocation T_i (for simplicity, two groups of control ($T_i = 0$) and treated ($T_i = 1$) subjects are considered in this paper). Because the subject-specific features exist before the treatment allocation, the original random-effects distribution should be independent of the allocation result. Consistent with this point, two types of random effects are introduced in this paper: $\mathbf{b}_{0i}$ and $\mathbf{b}_{1i}$, whose distributions G_0 and G_1, respectively, are independent of the allocation. Either $\mathbf{b}_{0i}$ or $\mathbf{b}_{1i}$ is assumed to be included in the model if the subject i is allocated to the control group or treated group. The resulting random effects $\mathbf{b}_i$ in the model and its distribution can be written as

$$\mathbf{b}_i = \mathbf{b}_{0i}(1 - T_i) + \mathbf{b}_{1i}T_i \,,\; G = G_0(1 - T_i) + G_1 T_i \,, \tag{1}$$

where $\mathbf{b}_i$ and G are dependent on T_i.

2.2 Real data example

In this section, two longitudinal datasets from RCTs will be presented as motivating examples. The framework of the analysis using mixed-effects models is described here.

2.2.1 Cholesterol data. The first example is the repeated measurements data from the "Ezetimibe Lipid-Lowering Trial on Prevention of Atherosclerotic Cardiovascular Disease in 75 or Older" study (EWTOPIA 75),[19] where subjects were asked to undergo treatment consisting of dietary counseling plus ezetimibe or dietary counseling alone. As secondary outcomes, several types of cholesterol, including low-density lipoprotein cholesterol (LDLC), were measured annually from baseline to a maximum of 7 years of follow-up. In this paper, we analyze the LDLC data up to the third year of follow-up. In total, 1520 and 1507 subjects in the ezetimib and control groups, respectively, had cholesterol measurements at baseline and at a minimum of one follow-up visit. Among these individuals, 1477 and 1454 subjects in the ezetimib and control groups, respectively, had first-year measurements; 1236 and 1220 subjects, respectively, had second-year measurements; and 1000 and 1018 subjects, respectively, had third-year measurements. The mean LDLC value at each time point and a histogram showing the changes in follow-up values from baseline are presented in Figure 2 (a) and (b). The LDLC reduction was about 20 mg/dL greater in the treatment group than in the control group. Although the LDLC reductions in both groups had single and almost symmetric modes, the tail of the side with greater reduction was slightly heavier in the control group.

For follow-up LDLC values, the following linear mixed-effects model will be fitted:

$$\begin{aligned} y_{ij} &\sim \mathrm{Normal}(\mu_{ij}, \tau^2) \,, \\ \mu_{ij} &= \beta_0 + \beta_1 \mathrm{LDLC}_{\mathrm{base}\,i} + \beta_2 T_i + \beta_3 \delta_{2j} + \beta_4 \delta_{3j} \\ &\quad + \beta_5 T_i \delta_{2j} + \beta_6 T_i \delta_{3j} + b_i \,, \\ b_i &\sim g(b_i) \,, \end{aligned} \tag{2}$$

in which $j = 2, 3$, $\delta_{kj} = 1$ if $k = j$ and $\delta_{kj} = 0$ if $k \neq j$. Although a normal distribution with a 0 mean is usually fitted for a random-effects distribution g, other models will be examined based on evaluation by gradient functions. For a model that allows the dependence of random-effects distribution on the treatment allocation, the following model can also be fitted:

$$\begin{aligned} y_{ij} &\sim \mathrm{Normal}(\mu_{ij}, \sigma^2) \,, \\ \mu_{ij} &= \beta_0 + \beta_1 \mathrm{LDLC}_{\mathrm{base}\,i} + (\beta_2 + b_{1i} - b_{0i})T_i + \beta_3 \delta_{2j} + \beta_4 \delta_{3j} \\ &\quad + \beta_5 T_i \delta_{2j} + \beta_6 T_i \delta_{3j} + b_{0i} \\ &= \beta_0 + \beta_1 \mathrm{LDLC}_{\mathrm{base}\,i} + \beta_2 T_i + \beta_3 \delta_{2j} + \beta_4 \delta_{3j} \\ &\quad + \beta_5 T_i \delta_{2j} + \beta_6 T_i \delta_{3j} + b_{0i}(1 - T_i) + b_{1i} T_i \,, \\ b_{0i} &\sim g_0(b_{0i}) \,,\; b_{1i} \sim g_1(b_{1i}) \,. \end{aligned} \tag{3}$$

Again, the proper model for g_0 and g_1 will be examined based on evaluation by gradient functions.

2.2.2 Toenail data. The second example involves repeated measurements data from a RCT by De Backer et al. (1996) that compared two oral treatments for toenail dermatophyte onychomycosis.[6] These data were also analyzed in Verbeke and Molenberghs (2013) as an illustration of gradient function use. For comparison with their results, we use the same subset of data, consisting of at most 7 measurements of the infection severity, categorized as severe or not severe, from 146 and 148 subjects for each of the two treatments, respectively. The observed percentages of severe infections during the follow-up periods are shown in Figure 5 (a).

For the binary outcome of the severity of toenail infection, the following logistic mixed-effects model will be fitted:

$$\begin{aligned} y_{ij} &\sim \text{Bernoulli}(\pi_{ij}) , \\ \text{logit}(\pi_{ij}) &= \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij} + b_i , \\ b_i &\sim g(b_i) , \end{aligned} \tag{4}$$

where t_{ij} is the time point (in months) at which the j th measurement is obtained for the subject i. The appropriate model for g was described in Verbeke and Molenberghs (2013), and based on their discussion, we will start the model with a mixture of three normals for g. For the model that allows the dependence of random-effects distribution on the treatment allocation, the following model can also be fitted:

$$\begin{aligned} y_{ij} &\sim \text{Bernoulli}(\pi_{ij}) , \\ \text{logit}(\pi_{ij}) &= \beta_0 + (\beta_1 + b_{1i} - b_{0i}) T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij} + b_{0i} \\ &= \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij} \\ &\quad + b_{0i}(1 - T) + b_{1i} T , \\ b_{0i} &\sim g_0(b_{0i}) , \; b_{1i} \sim g_1(b_{1i}) . \end{aligned} \tag{5}$$

3 THE GRADIENT FUNCTION METHOD

In this section, we build on the introduction of the gradient function by Verbeke and Molenberghs (2013) and create an extension for the case in which G depends on the treatment allocation T_i. First, we consider the directional derivative Φ of the log-likelihood at G into the direction H, where Φ expresses the change in log-likelihood between the cases where the assumed distributions for random effects are G or $(1-\alpha)G + \alpha H$ with an infinitesimal weight α.

$$\begin{aligned} \Phi[G, H] &= \lim_{\alpha \to +0} \frac{l[(1-\alpha)G + \alpha H] - l[G]}{\alpha} \\ &= \left. \frac{\partial l[(1-\alpha)G + \alpha H]}{\partial \alpha} \right|_{\alpha = 0} \end{aligned} \tag{6}$$

If we have an estimate $\hat{G}$ and $\Phi[\hat{G}, H] \leq 0$ for all H, no better random-effects distribution than our parametric fit $\hat{G}$ can be found. Furthermore, noting that $d\{(1-\alpha)G + \alpha H\} = (1-\alpha)dG + \alpha dH$, we have

$$\begin{aligned} \frac{1}{N}\Phi(G, H) &= \left. \frac{\partial \sum_i \ln\{(1-\alpha) f_i(\mathbf{y}_i|G) + \alpha f_i(\mathbf{y}_i|H)\}}{\partial \alpha} \right|_{\alpha = 0} \\ &= \frac{1}{N} \sum_i \frac{f_i(\mathbf{y}_i|H) - f_i(\mathbf{y}_i|G)}{f_i(\mathbf{y}_i|G)} \\ &= \frac{1}{N} \sum_i \frac{f_i(\mathbf{y}_i|H)}{f_i(\mathbf{y}_i|G)} - 1 \\ &= \frac{1}{N} \sum_i \frac{\int f_i(\mathbf{y}_i|\mathbf{b}) dH(b)}{f_i(\mathbf{y}_i|G)} - 1 \\ &= \int \frac{1}{N} \sum_i \frac{f_i(\mathbf{y}_i|\mathbf{b})}{f_i(\mathbf{y}_i|G)} dH(\mathbf{b}) - 1 . \end{aligned} \tag{7}$$

Here, the integrand of (7) is defined as the gradient function Δ.

$$\Delta(G, \mathbf{b}) \equiv \frac{1}{N} \sum_i \frac{f_i(\mathbf{y}_i|\mathbf{b})}{f_i(\mathbf{y}_i|G)} . \tag{8}$$

In general, the random-effects distribution H is allowed to depend on treatment allocation. For example, H may be expressed as $H = H_0(1 - T_i) + GT_i$ and the directional derivative at G into this direction is written as follows,

$$\begin{aligned} \frac{1}{N}\Phi(G, H) &= \int \frac{1}{N} \sum_i \frac{f_i(\mathbf{y}_i|\mathbf{b})}{f_i(\mathbf{y}_i|G)} dH(\mathbf{b}) - 1 \\ &= \int \frac{1}{N} \sum_i (1 - T_i) \frac{f_i(\mathbf{y}_i|\mathbf{b})}{f_i(\mathbf{y}_i|G)} dH_0(\mathbf{b}) - \frac{1}{N} \sum_i (1 - T_i) \\ &= \frac{N_0}{N} \left\{ \int \frac{1}{N_0} \sum_{T_i = 0} \frac{f_i(\mathbf{y}_i|\mathbf{b})}{f_i(\mathbf{y}_i|G)} dH_0(\mathbf{b}) - 1 \right\} \\ &\equiv \frac{N_0}{N} \left\{ \int \Delta_0(G, \mathbf{b}) dH_0(\mathbf{b}) - 1 \right\} , \end{aligned} \tag{9}$$

where $N_0 \equiv \sum_i (1 - T_i)$. Here, $\Phi(G, H)$ represents the change in log-likelihood obtained by the infinitesimal change of G toward H_0, but only for $T_i = 0$ subjects; this will be evaluated by comparing Δ_0 and 1 following the original gradient function method described in Verbeke and Molenberghs (2013). Similarly, by considering $H = G(1 - T_i) + H_1 T_i$, Δ_1 is also obtained as

$$\begin{aligned} \frac{1}{N}\Phi(G, H) &= \int \frac{1}{N} \sum_i \frac{f_i(\mathbf{y}_i|\mathbf{b})}{f_i(\mathbf{y}_i|G)} dH(\mathbf{b}) - 1 \\ &= \int \frac{1}{N} \sum_i T_i \frac{f_i(\mathbf{y}_i|\mathbf{b})}{f_i(\mathbf{y}_i|G)} dH_1(\mathbf{b}) - \frac{1}{N} \sum_i T_i \\ &= \frac{N_1}{N} \left\{ \int \frac{1}{N_1} \sum_{T_i = 1} \frac{f_i(\mathbf{y}_i|\mathbf{b})}{f_i(\mathbf{y}_i|G)} dH_1(\mathbf{b}) - 1 \right\} \\ &\equiv \frac{N_1}{N} \left\{ \int \Delta_1(G, \mathbf{b}) dH_1(\mathbf{b}) - 1 \right\} , \end{aligned} \tag{10}$$

where $N_1 \equiv \sum_i T_i$. Here, $\Phi(G, H)$ represents the change in log-likelihood obtained by the infinitesimal change of G toward H_1, but only for $T_i = 1$ subjects; this will be evaluated as described above for equation (9).

If the areas for $\mathbf{b}$ where $\Delta_0 > 1$ or $\Delta_1 > 1$ are different from each other, it is suggested that assuming a different random-effects distribution depending on the treatment allocation will provide a better fit. This evaluation should be performed in the common support intervals $[\mathbf{b}_{\min}, \mathbf{b}_{\max}]$ of Δ_0 and Δ_1. Using the model (1), $\Delta_0(G, \mathbf{b}) = \Delta_0(G_0, \mathbf{b}_0)$ and $\Delta_1(G, \mathbf{b}) = \Delta_1(G_1, \mathbf{b}_1)$, which enables us to search separately for better $\hat{G}_0$ and $\hat{G}_1$ following the procedure based on the original gradient function. This will achieve a better $\hat{G}$ overall. The method can easily be extended to the situation where more than two treatment groups exist.

Fitting by Model A				Scenario 1		Scenario 2	
				Bias	Coverage %	Bias	Coverage %
Normal outcome	$\beta_1 = 0$	$n_i = 5$	$N = 300$	0.00	0.95	0.00	0.94
			$N = 1000$	0.00	0.95	0.00	0.95
		$n_i = 20$	$N = 300$	0.00	0.95	0.00	0.95
			$N = 1000$	0.00	0.96	0.00	0.96
	$\beta_1 = 1$	$n_i = 5$	$N = 300$	0.00	0.95	0.00	0.95
			$N = 1000$	0.00	0.95	0.00	0.95
		$n_i = 20$	$N = 300$	0.00	0.94	0.00	0.94
			$N = 1000$	0.00	0.97	0.00	0.95
Binary outcome	$\beta_1 = 0$	$n_i = 5$	$N = 300$	0.00	0.96	-0.06	0.93
			$N = 1000$	0.00	0.95	-0.06	0.89
		$n_i = 20$	$N = 300$	0.00	0.95	-0.04	0.92
			$N = 1000$	0.00	0.95	-0.04	0.90
	$\beta_1 = 1$	$n_i = 5$	$N = 300$	-0.07	0.91	-0.17	0.79
			$N = 1000$	-0.06	0.87	-0.16	0.48
		$n_i = 20$	$N = 300$	-0.04	0.93	-0.12	0.81
			$N = 1000$	-0.04	0.88	-0.12	0.52
Fitting by Model B				Scenario 1		Scenario 2	
				Bias	Coverage %	Bias	Coverage %
Binary outcome	$\beta_1 = 0$	$n_i = 5$	$N = 300$	0.00	0.94	-0.06	0.92
			$N = 1000$	0.00	0.93	-0.06	0.88
		$n_i = 20$	$N = 300$	0.01	0.95	-0.03	0.94
			$N = 1000$	0.00	0.95	-0.03	0.94
	$\beta_1 = 1$	$n_i = 5$	$N = 300$	0.01	0.95	-0.08	0.90
			$N = 1000$	0.00	0.95	-0.08	0.83
		$n_i = 20$	$N = 300$	0.00	0.94	-0.04	0.92
			$N = 1000$	0.00	0.94	-0.04	0.90

Table 1: Summary of the performances of treatment effect estimators under Scenarios 1 and 2 by mixed-effects models with Models A and B.

4 SIMULATION STUDY

4.1 Simulation settings

To understand the behavior and utility of the gradient functions Δ_0 and Δ_1 where random-effects distributions differed depending on treatment allocation, datasets were generated and analyzed under several basic study settings with repeated measurements. The normal and binary outcomes were generated from the models below.

$$\text{Normal outocomes} : y_{ij} \sim \mathrm{N}(\mu_{ij}, 0.5^2),\ \mu_{ij} = \beta_1 T_i + b_i$$
$$\text{Binary outocomes} : y_{ij} \sim \text{Bernoulli}(\pi_{ij}),\ \text{logit}(\pi_{ij}) = \beta_1 T_i + b_i$$

We considered a number of measurements in one subject n_i of 5 or 20, a total sample size N of 300 or 1000, and a treatment effect β_1 of 0 or 1. For both types of outcomes, the treatment allocation T_i for subject i was generated from Bernoulli(0.5). For random-effects in control group b_{0i} and treated group b_{1i}, the following scenarios were considered.

$$\text{All scenarios} : b_{0i} \sim \mathrm{N}(0, 0.5^2)$$
$$\text{Scenario 0} : b_{1i} \sim \mathrm{N}(0, 0.5^2)$$
$$\text{Scenario 1} : b_{1i} \sim \mathrm{N}(0, 1)$$
$$\text{Scenario 2} : b_{1i} \sim \frac{1}{2}\mathrm{N}(-1, 0.5^2) + \frac{1}{2}\mathrm{N}(1, 1)$$

Under Scenario 2, the random-effects distribution for the treated group is skewed. Shapes of random-effects distributions in all scenarios are shown in the figures of the gradient functions.

The following mixed-effects models were fitted to the simulated datasets.

$$\text{Model for normal outocomes} : y_{ij} \sim \mathrm{N}(\mu_{ij}, v^2),\ \mu_{ij} = \beta_0 + \beta_1 T_i + b_i$$
$$\text{Model for binary outocomes} : y_{ij} \sim \text{Bernoulli}(\pi_{ij}),\ \text{logit}(\pi_{ij}) = \beta_0 + \beta_1 T_i + b_i$$
$$\text{Model A} : b_i \sim \mathrm{N}(0, \sigma^2)$$
$$\text{Model B} : b_i \sim \mathrm{N}(0, (1 - T_i)\sigma_0^2 + T_i \sigma_1^2)$$

More complicated settings, with fit by random slope models, are discussed in the supplementary materials.

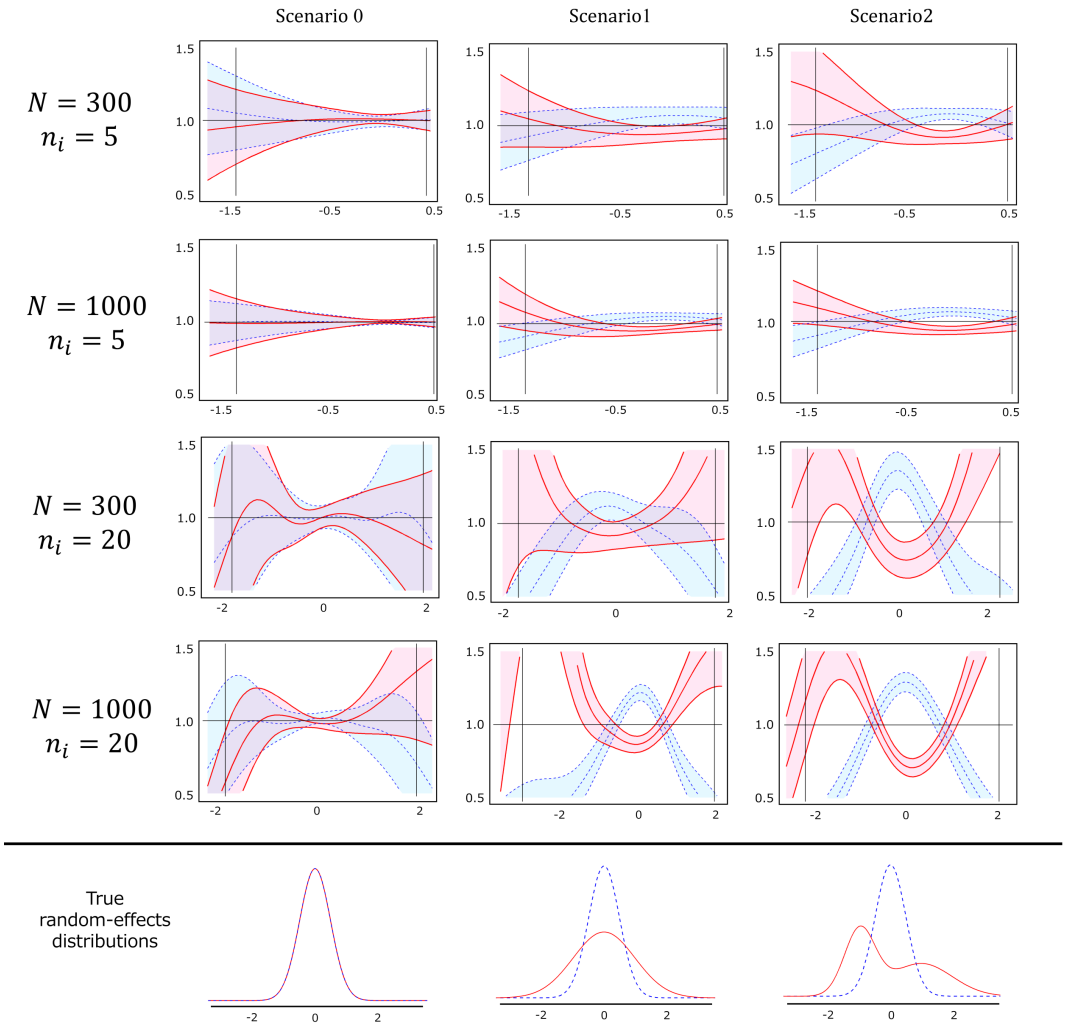

Figure 1: Gradient functions Δ_0 (blue dashed lines) and Δ_1 (red solid lines) with 95% confidence bands from mixed-effects models with Model A for binary outcomes and the true random-effects distributions for the control (blue dashed lines) and treated (red solid lines) groups.

4.2 Results of simulation studies

4.2.1 Performances of estimators in mixed-effects models. First, we investigated the performances of the treatment effect estimators under Scenarios 1 and 2 by mixed-effects models with Models A and B. Biases of the estimator and the proportion of 95% confidence intervals that covered the true value were calculated by averaging the results from 1000 simulated datasets (Table 1).

In normal outcomes with Model A, ignoring the dependence of random-effects distributions on treatment allocation did not produce biases in treatment effect estimates and did not harm the coverage proportions under both Scenarios 1 and 2. In binary outcomes with Model A, biased estimates were observed in $\beta_1 = 1$ cases under Scenario 1. Under Scenario 2, biased estimates were observed in both $\beta_1 = 0$ and $\beta_1 = 1$; the latter had a larger bias. When fitting by Model B, biases disappeared in Scenario 1, but were still present in Scenario 2 (the amount of bias was reduced when $\beta_1 = 1$).

As described in Verbeke and Molenberghs (2013), the gradient method is calculated under the assumption that the conditional distribution $f_i(\mathbf{y}_i \mid \mathbf{b}_i)$ is correctly specified. Therefore, gradient functions from mixed-effects models for binary outcomes should be carefully interpreted due to biases in the estimators.

4.2.2 Gradient functions Δ_0 and Δ_1. Based on the performance of treatment effect estimators under Scenarios 1 and 2, we focused on the setting of $\beta_1 = 1$ to assess the behavior of gradient functions. Gradient functions Δ_0 and Δ_1 from mixed-effects models with Model A for normal outcomes are plotted in Figure S1 in the supplementary materials. Clear differences between Δ_0 and Δ_1 were observed, and the shapes roughly indicated the types of violations of the assumptions. Gradient functions Δ_0 and Δ_1 from mixed-effects

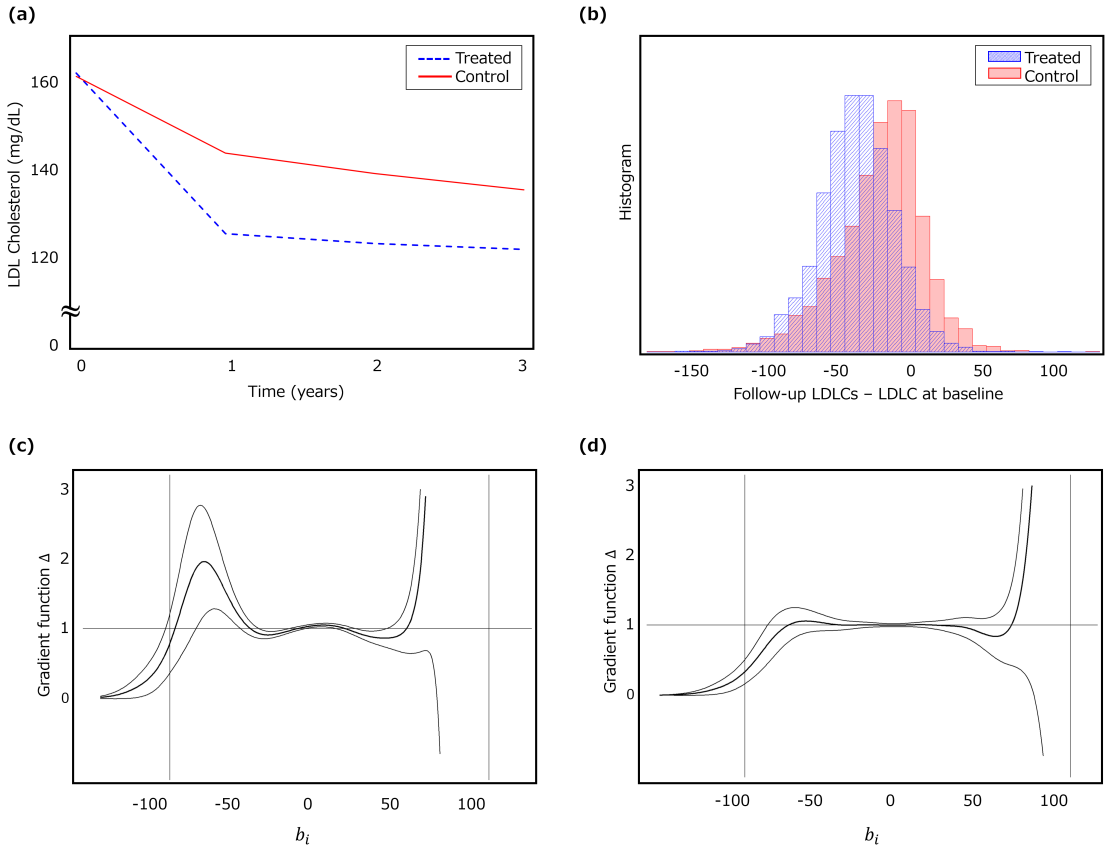

Figure 2: Cholesterol data: (a) The mean value of LDL cholesterol at each time point. (b) Histogram of changes in follow-up LDLC from baseline LDLC. (c) Gradient function Δ and 95% confidence bands for fitting with (2) and Model 1. (d) Gradient function Δ and 95% confidence bands for fitting with (2) and Model 2.

models with Model A for binary outcomes are plotted in Figure 1. Again, with biases in the estimators, clear differences between Δ_0 and Δ_1 were observed, signaling violations of the assumptions. Under Scenario 0, where the mixed-effects models with Model A corresponded to the true models, differences between Δ_0 and Δ_1 were not observed for either type of outcome. For normal outcomes, shapes of gradient functions became wavy as n_i increased, resulting in clearer differences when $n_i = 5$. On the other hand, the degrees of differences of gradient functions Δ_0 and Δ_1 became larger as n_i increased for binary outcomes.

Gradient functions Δ_0 and Δ_1 from mixed-effects models with Model B for both types of outcomes are plotted in Figure S2 in the supplementary materials. Under Scenario 1, where the mixed-effects models with Model B corresponded to the true models, no violation was suggested in either type of outcome. For cases with normal outcomes, the shapes roughly indicated the types of violations of model assumptions under Scenario 2, which could suggest that more complicated models should be fit, e.g., a model with a finite mixture of normals [20]. For cases with binary outcomes, clear differences between Δ_0 and Δ_1 were observed in $n_i = 20$ cases under Scenario 2, but differences were not obvious in $n_i = 5$ cases.

To roughly check the variations of gradient functions between simulated datasets, gradient functions were plotted for three independent datasets in some settings, as shown in Figure S3 in the supplementary materials. Under correct model specifications, gradient functions were plotted near 1 in all three cases. Under incorrect model specifications, some differences were present between Δ_0 and Δ_1.

5 REAL DATA EXAMPLES

5.1 Cholesterol data

The cholesterol data were first modeled by equation (2) with

$$\text{Model 1} : b_i \sim \mathrm{N}(0, \sigma^2) \ .$$

Based on the estimates of equation (2) and Model 1, the gradient function shown in Figure 2 (c) was obtained. This model suggests that a more flexible random-effects distribution that allows moving the probability mass from the region $b = [-30; -20]$ toward the region $b = [-80; -50]$ can improve the model in terms of likelihood. Furthermore, in the the region $b = [80; 100]$, the value of the gradient function is quite large and the corresponding pointwise interval is also very wide, which makes it difficult to interpret the

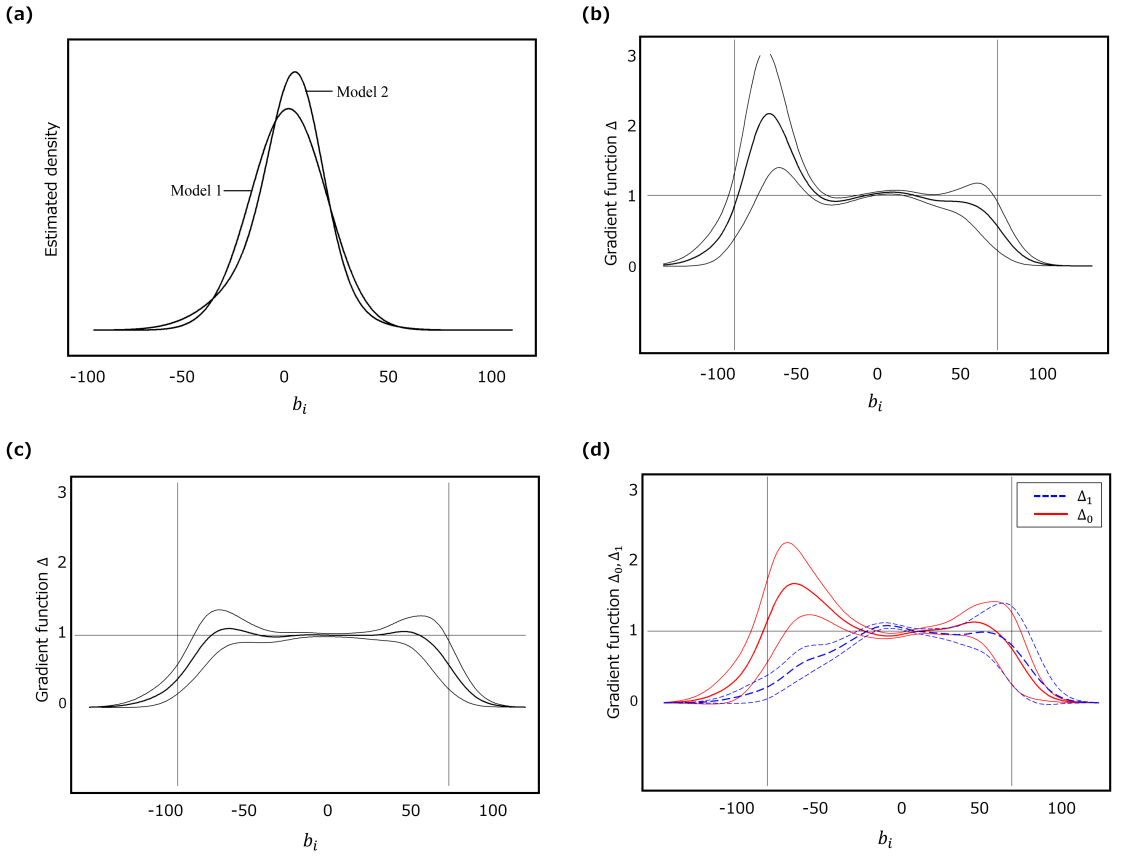

Figure 3: Cholesterol data: Results of fitting to dataset without the six subjects with the largest marginal residuals averaged within subjects. (a) Fitted random-effects distributions based on (2) **with Model 1 and Model 2. (b) Gradient function** Δ **and 95% confidence bands for fitting with** (2) **and Model 1. (c) Gradient function** Δ **and 95% confidence bands for fitting with** (2) **and Model 2. (d) Gradient function** Δ_0 **(solid),** Δ_1 **(dashed), and 95% confidence bands for fitting with** (2) **and Model 2.**

gradient function around this area. Thus, as an attempt to improve the situation, the mixture of two normals was fitted.

$$\begin{aligned} \text{Model 2} : b_i \quad &\sim (1-p_2)\,\mathrm{N}(\mu_1, \sigma_1^2) + p_2 \mathrm{N}(\mu_2, \sigma_2^2) \\ &\text{with } (1-p_2)\mu_1 + p_2\mu_2 = 0 \end{aligned}$$

The finite mixture of normals for the random-effects distribution, proposed by Verbeke and Lesaffre (1996), offers flexible models and was also used as an alternative to the normal distribution in Verbeke and Molenberghs (2013), where the model was fitted by the method shown in Liu and Yu (2007) using the NLMIXED procedure in SAS. Based on the estimates of equation (2) and Model 2, the gradient function shown in Figure 2 (d) was obtained. Although the value of the gradient function around $b < -20$ moved toward 1, the value was still quite large in the region $b = [80; 100]$.

By carefully investigating each subject's contribution to the gradient function, it was found that a very small portion of subjects had huge influences that dominated the shape around the region $b = [80; 100]$. According to the definition of the gradient function (8), these subjects exhibited a large gap between $f_i(\mathbf{y}_i|\mathbf{b})$ and $f_i(\mathbf{y}_i|G)$ in the region $b = [80; 100]$. Two choices were considered in order to overcome this issue. The first was to use a more flexible model with three normals to accommodate these influential subjects. This could alter $f_i(\mathbf{y}_i|G)$ and might decrease the gaps. The second choice was to exclude these subjects and attempt the modeling for the rest of the population. However, exclusion based on the influence on gradient function must be carefully considered, because the interpretation of each subject's contribution $f_i(\mathbf{y}_i|\mathbf{b})/f_i(\mathbf{y}_i|G)$ is not entirely clear. In the modeling of cholesterol data with (2) and Model 1 or 2, 6 subjects (0.2% of the whole population, three in the treated group and three in the control group) were identified as influential subjects in the right side of gradient function. Furthermore, these subjects accounted for the top six largest marginal residuals averaged within subjects in both Models 1 and 2. Based on these facts, these six subjects were excluded in the following analyses to simplify the illustration of the methodology.

Equation (2) with Models 1 and 2 were again fitted to the dataset, this time excluding the aforementioned six subjects. The estimated random-effects distributions are plotted in Figure 3 (a). These results

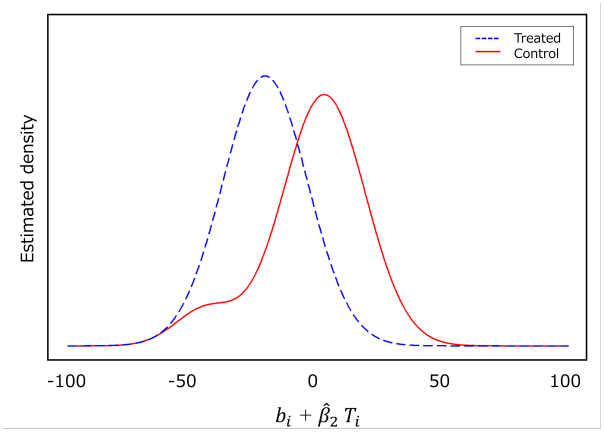

Figure 4: Cholesterol data: Estimated random-effects distributions for b_{oi} (solid) and b_{1i} (dashed) plotted against $b_i + \hat{\beta}_2 T_i$ based on (3) with Model 4 fitted to the dataset without the six subjects with the largest marginal residuals averaged within subjects.

were used to obtain gradient functions Δ for Models 1 and 2, shown in Figure 3 (b) and (c), respectively. Figure 3 (c) suggests no evidence of model improvement from Model 2 based on Δ. Indeed, Model 2 achieved a better fit in terms of $-2l$ and the Akaike information criteria (AIC) (Table 1).

Next, the gradient functions Δ_0 and Δ_1 were investigated given the estimates of equation (2) and Model 2, and plotted in Figure 3 (d). According to these plots, the trends of deviation from 1 in Δ_0 and Δ_1 were totally different in the region $b = [-100; 10]$, suggesting that a better fit will be achieved using the model with dependence on the random-effects distribution on treatment allocation. Therefore, as the next candidate, equation (3) with the following random-effects distribution was adopted to fit,

$$\text{Model 3} : b_{0i} \sim \mathrm{N}(0, \sigma_0^2),\ b_{1i} \sim \mathrm{N}(0, \sigma_1^2)\ .$$

The estimated random-effects distributions are plotted in Figure S5 (a) in the supplementary materials. Given the estimates of equation (3) and Model 3, the gradient functions Δ_0 and Δ_1 were obtained and are shown in Figure S5 (b) and (c), respectively, in the supplementary materials. According to Figure S5 (b), a more flexible random-effects distribution that allows moving the probability mass from the region $b = [-30; -20]$ toward the region $b = [-80; -50]$ can improve the model in the control group. On the other hand, Figure S5 (c) suggests no evidence of model improvement in the treated group. The fit of Model 3 was not as good as that of Model 2 (Table 1). Therefore, as the next candidate, the following model was adopted and fitted:

$$\begin{aligned}\text{Model 4} : b_{0i} \quad &\sim (1 - p_{02})\ \mathrm{N}(\mu_{01}, \sigma_{01}^2) + p_{02}\ \mathrm{N}(\mu_{02}, \sigma_{02}^2), \\ b_{1i} &\sim \mathrm{N}(0, \sigma_1^2) \\ \text{with } (1 - p_{02})\mu_1 &+ p_{02}\mu_{02} = 0\ .\end{aligned}$$

The estimated random-effects distributions are plotted in Figure S5 (d) in the supplementary materials. Given the estimates of equation (3) and Model 4, the gradient functions Δ_0 and Δ_1 were obtained and are shown in Figure S5 (e) and (f), respectively, in the supplementary materials. This time, both Δ_0 and Δ_1 suggest no evidence of model improvement in terms of likelihood. Model 4 achieved the best model fit among the models attempted (Table 2). The estimated model parameters were similar among all models and are summarized in Table S1 in the supplementary materials.

	Model 1	Model 2	Model 3	Model 4
$-2l$	17804	17762	17768	17722
AIC	17822	17786	17788	17748

Table 2: Cholesterol data: Summary of model fit.

Two population types were thought to exist in the control group based on the estimated results of Model 4. In detail, around 10% of the population seemed to show a remarkable reduction in LDL cholesterol during the follow-up period. This trend persisted even after excluding subjects who received antidyslipidemic drugs during the follow-up period. To properly interpret this observation, estimated random-effects distributions are plotted against $b_i + \hat{\beta}_2 T_i$ in Figure 4. One possible explanation is that a small portion of the population may have been influenced by the dietary counseling provided to both groups. This might have contributed to the presence of the left tails of the distributions, which could have been masked by treatment effects in the treated group.

5.2 Toenail data

These data were analyzed in Verbeke and Molenberghs (2013), and the appropriate random effects distribution was fully discussed based on the gradient function Δ. The population was heterogeneous and was divided into three groups, namely no responders, partial responders, and full responders. As a result, equation (4) with

$$\text{Model 5} : b_i \sim (1 - p_2 - p_3)\ \mathrm{N}(\mu_1, \sigma_1^2) + p_2 \mathrm{N}(\mu_2, \sigma_2^2) + p_3 \mathrm{N}(\mu_3, \sigma_3^2)$$

was adopted. Here, we also used this model to derive the gradient functions. Given the estimates of equation (4) and Model 5, the estimated random-effects distribution, the gradient function Δ, and the gradient functions Δ_0 and Δ_1 were plotted in Figure 5 (b), (c) and (d), respectively. Neither Figure 5 (c) nor (d) suggest any evidence

of model improvement. There seems to be no difference in trends in Δ_0 vs Δ_1, suggesting that the model in which the random-effects distribution is dependent on treatment allocation might not offer a better model fit.

6 DISCUSSION

We applied the gradient function method for modeling different random-effects distributions depending on the treatment allocation. To do this properly, we considered the directional derivative with the direction depending on treatment allocation, then defined the gradient functions Δ_0 and Δ_1, and compared these in their common support. The method with Δ_0 and Δ_1 was illustrated for both normal and binary outcomes using simulated and real data examples, and its utility in detecting the violation of assumptions for random-effects distributions was investigated in random intercept and slope models.

Scenario 1 for binary outcomes in the simulation study was also considered in Heagerty and Kurland (2001), and we observed biases in the treatment effect due to ignoring the dependence of different random-effects distributions on treatment allocation, as they calculated. We further observed additional types of biases in modeling binary outcomes when only one of two treatment groups had a skewed random-effects distribution, as in Scenario 2 in the simulation study. In these situations, the gradient function should be used with caution because the conditional model $f_i(\mathbf{y}_i|b_i)$ can be misspecified, and the signals from the gradient function may not always be interpreted for random-effects distributions. In their original paper, Verbeke and Molenberghs (2013) suggest that the use of gradient functions in these situations will be similar in spirit to residual and influence function plots often used in linear regression analysis, in which the unknown model parameters are replaced by their estimates. Based on this spirit, we showed that the existence of these biases due to ignoring the dependence of different random-effects distributions on treatment allocation could be detected by comparing the plots of Δ_0 and Δ_1, as shown in Figure 1. This led to a more flexible Model B in the simulation study, and the amount of bias was reduced in both Scenario 1 and 2. After fitting by Model B, comparison of the plots of Δ_0 and Δ_1 still detected the violation of model assumptions in Scenario 2 for $n_i = 20$ cases, but not for $n_i = 5$ cases. This might be because signals became more distinct as n_i increased in modeling binary outcomes, as shown in Figure 1.

When the amount of biases in model parameters may seem small, as in the cases of modeling normal outcomes in the simulation study, the signals from gradient functions will be interpreted more straightforwardly for random-effects distributions. In this situation, the purpose of plotting gradient functions is to explore the true distributions. We proposed extending the use of the gradient function in this context to detect the dependence of random-effects distributions on treatment allocation, which led to a deeper understanding of the nature of random-effects in the cholesterol data example. One practical note regarding gradient functions in modeling normal outcomes is that very small samples that are potentially not normally distributed can have a huge influence on the gradient function, as was the case in the cholesterol data. In such cases, researchers must decide whether to fit a more flexible model to accommodate the sample near outliers or to exclude them with reason. This decision should be made regardless of the treatment allocation before calculating Δ_0 and Δ_1.

We have several practical notes on the use of the gradient function method described in this paper. First, because Δ_0 and Δ_1 depend on the reference distribution $\hat{G}$, it is recommended that the initial modeling step should be to evaluate the random-effects distribution for the entire sample using Δ, and then to check the shapes of Δ_0 and Δ_1. This might make it possible to avoid creating an unnecessarily complex model. Second, as shown in Figure 5 of the toenail data example, the 95% bands of gradient functions Δ_0 and Δ_1 were wider than that of Δ because the calculations of the gradient functions Δ_0 and Δ_1 involve dividing the sample population into two parts. This will be more problematic in cases with much smaller sample sizes. Under the settings considered in the simulation study, we observed that a sample size of roughly $N < 10$ could give a gradient function plot with 95% bands wider than 2 in their support area, making it practically difficult to evaluate departures from 1 in their support area (gradient functions are bounded above 0). Therefore, as a rule of thumb for plotting Δ_0 and Δ_1, we recommend roughly 20 or more sample size with 1:1 allocation. Third, the evaluation of Δ_0 and Δ_1 can be performed for observational data, replacing treatment allocation with exposure status. However, the assumption that the conditional distribution $f_i(\mathbf{y}_i \mid \mathbf{b}_i)$ is correctly specified is often unrealistic in observational study settings. For example, if the probability of exposure is associated with the random effects, which in principle is not directly testable, then the basic assumption in random effects is violated and can produce a biased estimate for the exposure effect. It is unclear what influence this will have on Δ, and of course on Δ_0 and Δ_1. On the other hand, the association between the probability of exposure and random effects can be expressed as a random-effects distribution depending on exposure status. [18] Therefore, interpretations of Δ_0 and Δ_1 under the conditional distribution with biased estimates might be important so as to permit insights regarding directly untestable assumptions. Further research is needed on interpreting the gradient functions Δ, Δ_0 and Δ_1 in observational study settings.

ACKNOWLEDGEMENTS

The authors thank the EWTOPIA 75 study group for providing the cholesterol data. This work was supported by the Project Promoting Clinical Trials for Development of New Drugs/Support Program for Biostatisticians of the Japan Agency for Medical Research and Development (19lk0201061t0004).

REFERENCES

[1] A Agresti, B Caffo, and P Ohman-Strickland. 2004. Examples in which misspecification of a random effects distribution reduces effciency, and possible remedies. *Comput Stat Data Anal.* 47 (2004), 639–653.
[2] P Albert. 2012. A linear mixed model for predicting a binary event from longitudinal data under random effects misspecification. *Stat Med.* 31 (2012), 143–154.
[3] T Baghfalakia, M Ganjalib, and G Verbeke. 2017. A shared parameter model of longitudinal measurements and survival time with heterogeneous random-effects distribution. *J Appl Stat.* 44, 15 (2017), 2813–2836.
[4] K Cheon, P Albert, and Z Zhang. 2012. The impact of random-effect misspecification on percentile estimation for longitudinal growth data. *Stat Med.* 31 (2012), 3708–3718.
[5] M Daniels and Y Zhao. 2003. Modelling the random effects covariance matrix in longitudinal data. *Stat Med.* 22, 10 (2003), 1631–1647.
[6] M De Backer, P De Keyser, C De Vroey, and E Lesaffre. 1996. A 12-week treatment for dermatophyte toe onychomycosis terbinafine 250mg/day vs. itraconazole 200mg/day-a double-blind comparative trial. *Br J Dermatol.* 134, 46 (1996), 16–17.

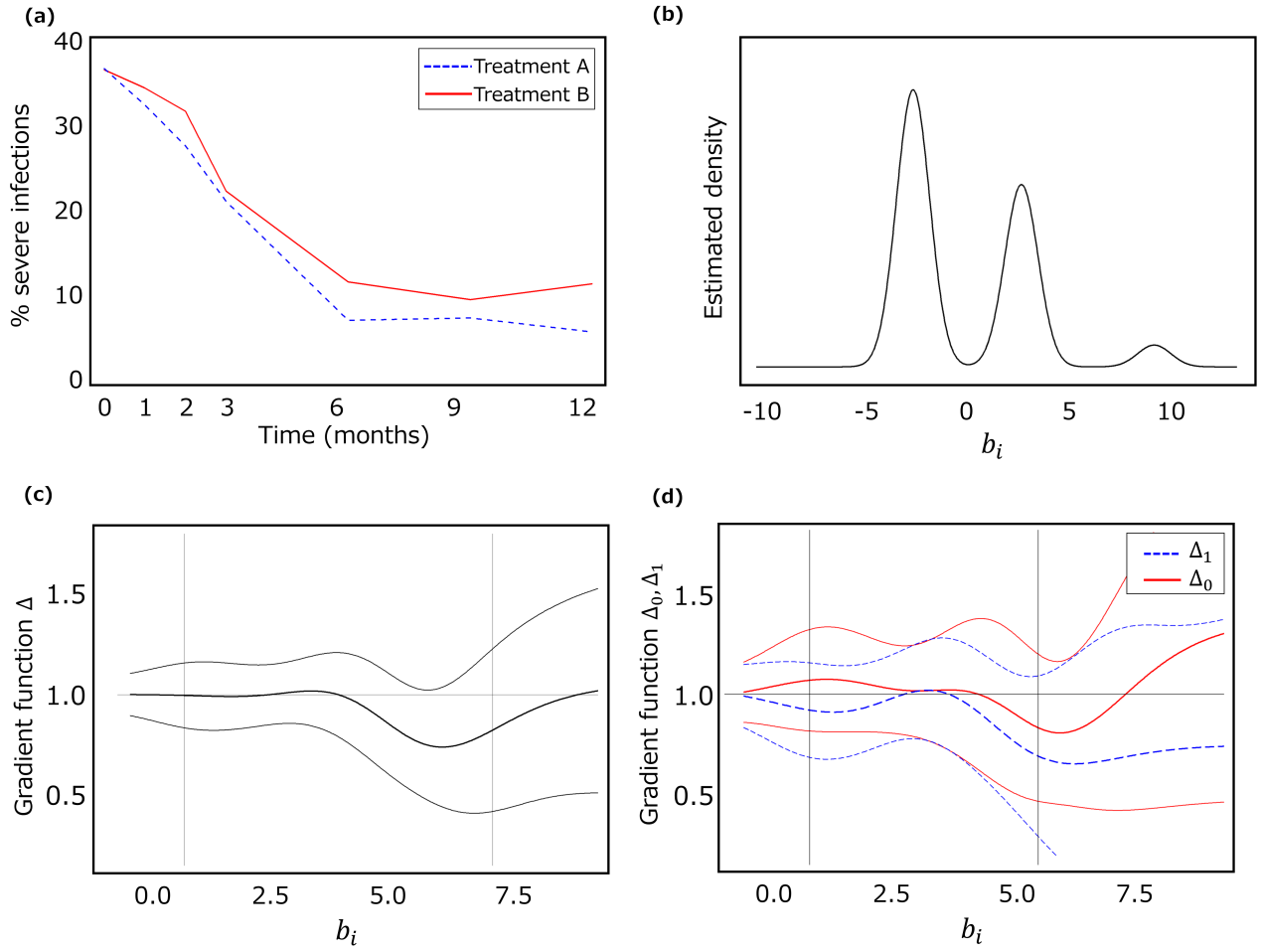

Figure 5: Toenail data: (a) Evolution of the percentage of severe toenail infections in the two treatment groups separately. (b) Fitted random-effects distributions based on (4) with Model 5. (c) Gradient function Δ and 95% confidence bands for fitting with (4) and Model 5. (d) Gradient function Δ_0 (solid), Δ_1 (dashed), and 95% confidence bands for fitting with (4) and Model 5.

[7] M Gosho and K Maruo. 2018. Effect of heteroscedasticity between treatment groups onmixed-effects models for repeated measures. *Pharm Stat.* 17 (2018), 578–592.

[8] P Heagerty. 1999. Marginally specified logistic-normal models for longitudinal binary data. *Biometrics.* 55 (1999), 688–698.

[9] P Heagerty and B Kurland. 2001. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika.* 88, 4 (2001), 973–985.

[10] P Heagerty and S Zeger. 2000. Marginalized multilevel models and likelihood inference. *Stat Sci.* 15, 1 (2000), 1–26.

[11] M Hoque and M Torabi. 2018. Modeling the random effects covariance matrix for longitudinal data with covariates measurement error. *Stat Med.* 37 (2018), 4167–4184.

[12] K Lee, J Lee, J Hagan, and J Yoo. 2012. Modeling the random effects covariance matrix for generalized linear mixed models. *Comput Stat Data Anal.* 55, 6 (2012), 1545–1551.

[13] S Litiere, A Alonso, and G Molenberghs. 2007. Type I and type II error under random-effects misspecification in generalized linear mixed models. *Biometrics.* 63 (2007), 1038–1044.

[14] S Litiere, A Alonso, and G Molenberghs. 2008. The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Stat Med.* 27 (2008), 3125–3144.

[15] L Liu and Z Yu. 2008. A likelihood reformulation method in non-normal random effects models. *Stat Med.* 27 (2008), 3105–3124.

[16] L Marquart and M Haynes. 2018. Misspecification of multimodal random-effect distributions in logistic mixed models for panel survey data. *J R Stat Soc Ser A Stat Soc.* 182, 1 (2018), 305–321.

[17] J Neuhaus, W Hauck, and J Kalbfleisch. 1992. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika.* 79, 4 (1992), 755–762.

[18] JM Neuhaus and CE McCulloch. 2006. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *J R Stat Soc Series B Stat Methodol.* 68, 5 (2006), 859–872.

[19] Y Ouchi, J Sasaki, H Arai, K Yokote, K Harada, Y Katayama, T Urabe, Y Uchida, M Hayashi, N Yokota, H Nishida, T Otonari, T Arai, I Sakuma, K Sakabe, M Yamamoto, T Kobayashi, S Oikawa, S Yamashita, H Rakugi, T Imai, S Tanaka, Y Ohashi, and M Hideki Ito Kuwabara. 2019. Ezetimibe Lipid-Lowering Trial on Prevention of Atherosclerotic Cardiovascular Disease in 75 or Older (EWTOPIA 75). *Circulation.* 140 (2019), 992–1003.

[20] G Verbeke and E Lesaffre. 1996. A linear mixed-effects model with heterogeneity in the random-effects population. *J Am Stat Assoc.* 91, 433 (1996), 217–221.

[21] G Verbeke and E Lesaffre. 1997. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Comput Stat Data Anal.* 23 (1997), 541–556.

[22] G Verbeke and G Molenberghs. 2013. The gradient function as an exploratory goodness-of-fit assessment of the random-effects distribution in mixed models. *Biostatistics.* 14, 3 (2013), 477–490.