Received: 30 March 2023 | Revised: 15 October 2023 | Accepted: 29 November 2023

DOI: 10.1002/pst.2357

WILEY

MAIN PAPER

Evaluation of a flexible piecewise linear mixed-effects model in the analysis of randomized cross-over trials

Moses Mwangi[1,2] | **Geert Verbeke**[1,3] | **Edmund Njeru Njagi**[4] | **Alvaro Jose Florez**[5,6] | **Samuel Mwalili**[7] | **Anna Ivanova**[1,3] | **Zipporah N. Bukania**[2] | **Geert Molenberghs**[1,3]

[1]I-BioStat, Universiteit Hasselt, Diepenbeek, Belgium

[2]Center for Public Health Research, Kenya Medical Research Institute, Nairobi, Kenya

[3]L-BioStat, Katholieke Universiteit (KU) Leuven, Leuven, Belgium

[4]Non-Communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK

[5]School of Statistics, Universidad del Valle, Cali, Colombia

[6]Data Science Institute, I-BioStat, Universiteit Hasselt, Diepenbeek, Belgium

[7]Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture & Technology, Nairobi, Kenya

Correspondence
Moses Mwangi, I-BioStat, Universiteit Hasselt, Agoralaan Gebouw D, 3590 Diepenbeek, Belgium.
Email: moses.mwangi@uhasselt.be; mmwangi@kemri.org

Funding information
Canadian International Food Security Research Fund (CIFSRF), Grant/Award Number: 106510

Abstract

Cross-over designs are commonly used in randomized clinical trials to estimate efficacy of a new treatment. They have received a lot of attention, particularly in connection with regulatory requirements for new drugs. The main advantage of using cross-over designs over conventional parallel designs is increased precision, thanks to within-subject comparisons. In the statistical literature, more recent developments are discussed in the analysis of cross-over trials, in particular regarding repeated measures. A piecewise linear model within the framework of mixed effects has been proposed in the analysis of cross-over trials. In this article, we report on a simulation study comparing performance of a piecewise linear mixed-effects (PLME) model against two commonly cited models—Grizzle's mixed-effects (GME) and Jones & Kenward's mixed-effects (JKME) models—used in the analysis of cross-over trials. Our simulation study tried to mirror real-life situation by deriving true underlying parameters from empirical data. The findings from real-life data confirmed the original hypothesis that high-dose iodine salt have significantly lowering effect on diastolic blood pressure (DBP). We further sought to evaluate the performance of PLME model against GME and JKME models, within univariate modeling framework through a simulation study mimicking a 2×2 cross-over design. The fixed-effects, random-effects and residual error parameters used in the simulation process were estimated from DBP data, using a PLME model. The initial results with full specification of random intercept and slope(s), showed that the univariate PLME model performed better than the GME and JKME models in estimation of variance–covariance matrix (G) governing the random effects, allowing satisfactory model convergence during estimation. When a hierarchical view-point is adopted, in the sense that outcomes are specified conditionally upon random effects, the variance–covariance matrix of the random effects must be positive-definite. The PLME model is preferred especially in modeling an increased number of random effects, compared to the GME and JKME models that work equally well with random intercepts only. In some cases, additional random effects could explain much variability in the data, thus improving precision in estimation of the estimands (effect size) parameters.

KEYWORDS
cross-over design, longitudinal, piecewise model, repeated measures

1 | INTRODUCTION

Cross-over designs are commonly used in randomized clinical trials to estimate efficacy of a new treatment with reference to a standard or placebo treatment. They have received a lot of attention, particularly, in connection with regulatory requirements for new drugs.[1,2] The main advantage of using cross-over designs over conventional parallel designs is for their increased precision, thanks to within-subject comparisons.[3] Every subject becomes its own control, thereby reducing confounding effects. For this reason, cross-over trials require smaller samples, hence are time and/or cost effective.[4] While there are some disadvantages, such as a requirement for patients to begin each new treatment period in a state comparable to those completed, cross-over trials are the design of choice in many settings, resulting in them accounting for 22% of all published trials in December 2000 (see[5]). They are well established for long-term chronic and stable illnesses such as hypertension.[6]

Jones and Kenward[7] discuss in detail more recent developments in the analysis of cross-over trials. In particular, they consider the situation in which each subject produces a set of profiles of repeated measurements, one from each treatment period, and the aim of the analysis is to examine the effect of treatments on these profiles. One simple approach is to reduce each profile of repeated measurements to a small number of summary statistics that represent aspects of the individual profiles that are relevant to the trial under analysis. Of course, this reduces the amount of information available for analysis. Three commonly used summary statistics include: (1) a particular end point; (2) the area under the profile; and (3) the slope of the profile. Two data settings may exist: one with a baseline and a single endpoint measurement, the other with a baseline and several repeated measurements within treatment periods. Baseline measurements can be used in three different ways: (1) in the change-from-baseline analyses[8–11]; (2) as covariates[12–14]; and (3) as outcomes.[15] In the case of highly unbalanced designs and/or missing values, use of baseline in the change-from-baseline analysis and as covariates, induces cross-level bias.[16] Use of baseline measurements as outcomes is preferred, for reduced bias and increased precision.[7]

Mwangi et al.[17] discuss the limitations of single endpoint (adjusting for baseline covariate) and change-from-baseline analysis. These methods are applied to designs with two time points (baseline and final). The designs potentially miss to capture variability of the outcome measurements between the baseline and the final time points in each period, hence, may lead to loss of information that otherwise would be captured during follow-up. This limitation affects precision in estimation of the actual treatment effect. Use of repeated measures designs overcomes this limitation.

In the statistical literature, models for the analysis of longitudinal repeated measures within the framework of mixed effects have been proposed. Methodologies have been developed for continuous, Gaussian data, as well as for non-Gaussian settings, such as binary, count, and ordinal data. Overviews can be found in Verbeke and Molenberghs[18] for the Gaussian case and in Molenberghs and Verbeke[19] for the non-Gaussian setting. In pharmacokinetics and pharmacodynamics studies, analysis of cross-over trials with nonlinear mixed-effect models have been widely documented.[20–23]

While estimating variance–covariance matrices in mixed-effects models, one may encounter convergence problems due to violations of positive-definiteness. This includes but is not limited to negative variance components. The concept of negative variance components in linear mixed-effects models has received considerable attention in the literature.[24,25] Broadly, negative variance components in linear mixed models are allowable if inferences are restricted to the implied marginal model.[26] Hierarchical models arise when outcomes are modeled conditional upon covariates and random effects.[19,27] When a hierarchical view-point is adopted, in the sense that outcomes are specified conditionally upon random effects, the variance–covariance matrix of the random effects must be positive-definite (positive-semi-definite is also possible, but raises issues of degenerate distributions). Molenberghs and Verbeke[28] further reviewed negative variance components in linear mixed models. Oliveira et al.[29] highlights that whenever inference for variance components is required, one has to choose between a hierarchical and a marginal view. Under a marginal interpretation, the variance components can be negative as long as the resulting marginal variance–covariance matrix of the observations is positive definite. On the other hand, when a hierarchical view is adopted, random effects retain their interpretation and, hence, their variances must be nonnegative. The usual action when a negative variance component estimate is

obtained for a random coefficient would be to constrain the variance component to zero or to refit the model with the random coefficient removed.[30]

Mwangi et al.[17] proposed the use of a piecewise linear model within the framework of mixed-effects models in the analysis of cross-over trials. They analyzed systolic and diastolic blood pressure (DBP) repeated measures and demonstrated that a PLMEs model is a useful tool for analyzing cross-over trials with repeated measurements within period. Its application allowed smooth transition from one period to the other even when there is no washout between the treatment periods. This was important because of the associated health benefits of the administered treatment. Contrary to standard cross-over design studies, absence of a washout period was informed by ethical issues. It was considered unethical denying participants iodized household salts because of potential health risks associated with non-consumption of iodine fortified salt.[31,32]

Piecewise linear models are easy to interpret, hence widely cited in the literature, following extensive application owing to their flexibility in modeling non-linear trends. In this article, we report on the results of a simulation study comparing performance of a PLMEs model against two commonly cited models—Grizzle[33,34]; Jones and Kenward[7]—used in the analysis of cross-over trials. We simulate two treatments by two periods (2×2) cross-over design datasets using parameter estimates generated from the PLMEs model (see[17]). The model was applied to blood pressure data described in the next section. These are continuous variables assumed to arise from a Gaussian multivariate distribution. We analyze the simulated data by estimating the average slope of each profile, adopting the recommendation of Jones and Kenward,[7] where baseline measurements are used as outcomes. We later compare the performance measures of PLMEs model against the Grizzle and Jones & Kenward models, in estimating treatment effect (difference in marginal means) in cross-over trials.

The remainder of the article is organized as follows. Section 2 describes the motivating case study, Section 3 gives an overview of methods applied, Section 4 presents the analysis results, Section 5 outlines the simulation study, and finally Section 6 offers concluding remarks.

2 | MOTIVATING CASE STUDY

This study concerns a two-arm, double blind, randomized 6 weeks cross-over trial to compare the effect of high-dose (84 mg/kg) and low-dose (50 mg/kg) Iodine in household salt, on DBP in women of reproductive age (15–49 years) conducted in Kenya between October 22, 2013 and November 29, 2013, with a first paper already published by Bukania et al.[35] The aim of the study was to investigate the role of Iodine intake in modulating blood pressure. The two Iodine dosages (high and low) were derived from the extreme range values (50–84 mg/kg) acceptable in the Kenyan mandatory salt iodization process.[36] A total of 174 women of reproductive age, randomized into two independent treatment sequences were followed for 6 weeks, constituting two treatment periods of 3 weeks each with no washout between the treatment periods. In the first sequence, 85 participants received high-dose Iodine salt during the first period and later swapped to low-dose Iodine salt during the second period. In the second sequence, 89 participants received low-dose Iodine salt during the first period and later swapped to high-dose Iodine salt during the second period. Treatment 1, denoted as Tr_1, represents high-dose Iodine salt and Treatment 2, denoted as Tr_0, represents low-dose Iodine salt. Each participant was expected to use the assigned treatment every day. Table 1 summarizes the sample design.

The primary outcome measurement is DBP. To determine the duration effect (blood pressure lowering), the outcomes for each individual were measured at seven equally spaced time points: baseline (week 0), week 1, and then through to week 6. The expected total number of measurements across the seven time points was 1218. However, due to drop-out and failure to attend some scheduled appointments, 90 measurements were missing. The sample data from four individuals are shown in Appendix S1.

TABLE 1 Cross-over design.

		Period	
Sequence	**Subjects**	**1**	**2**
1	$n = 85$	Tr_1	Tr_0
2	$n = 89$	Tr_0	Tr_1

3 | MODELING CONCEPTS

3.1 | General model framework

Longitudinal data are (often non-uniformly) ordered in time, and missing data are very common. Furthermore, serial measurements of one subject are potentially correlated, and the between-subject variance is not constant over time due to possibly diverging trajectories.[37] The linear mixed-effects model[38] is a flexible model to handle such data.

The proposed model formulation for *continuous DBP repeated measurements* is anchored within the framework of the general linear mixed-effects models (LMMs), and is expressed algebraically as follows:

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i, \tag{1}$$

where Y_i is an $n_i \times 1$ dimensional vector of observed responses for the ith individual, $i = 1,...,N$. X_i is a $n_i \times p$ matrix of known covariates associated with a p-dimensional vector β of unknown fixed-effects parameters. Z_i is a $n_i \times q$ matrix of known covariates, $b_i \sim N(0,G)$ is the q-dimensional vector of unknown random-effects parameters, $\epsilon_i \sim N(0,R_i)$ is the vector of unknown error/residual components. R_i is an $n_i \times n_i$ covariance matrix that depends on i only through its dimension n_i (in case of conditional independence, we have $R_i = \sigma^2 I_{n_i}$). See Bernal-Rusiel et al.,[37] Verbeke and Molenberghs[18,26] for a number of applications of the LME model. Restricted maximum likelihood (ReML) is used to estimate all parameters in the marginal model. Computational details are found in Lindstrom and Bates,[39] Laird and Ware,[38] and Verbeke and Molenberghs.[26] Details on hypothesis testing can be found in Bernal-Rusiel et al.[37]

3.2 | Modeling cross-over data

This article considers the simplest design of cross-over studies, with a two-sequence and two-period design for comparing two treatments (or 2 × 2). In cross-over clinical trials with one active (A) and one placebo/standard (B) treatment, participants are randomly assigned to a sequence of AB or BA.[33] That is, all participants are supposed to take one treatment during the first period and the other treatment during the second period in a random order. Let S_i denote a randomization assignment sequence indicator for the ith participant ($i = 1,2,...,n$), with $S_i = 0$ denoting sequence AB and $S_i = 1$ denoting sequence *BA*. Furthermore, let $T_{ij} = 0$ if the ith participant is assigned to treatment A at time t_j ($j = 1,2,...,n_i$), and $T_{ij} = 1$ if assigned to treatment B. $P_j = 0$ for all measurements during the first period and $P_j = 1$ during the second period. Y_{ij} is the response variable for the i^{th} participant assigned to treatment T_{ij}, during period P_j at time t_j.

Next, we implement the three candidate models (Grizzle's, Jones & Kenward's and PLMEs) in the identification and estimation of parameters under the following assumptions.[4]

I. Random assignment: Pre-treatment variables including potential outcome and baseline variables are independent of randomization.
II. Absence of carryover effects.
III. Fixed compliance: A subject would have the same compliance behavior across the two periods in each sequence.
IV. Fixed treatment effect: Each treatment has a fixed effect across the two periods, that is, there is no interaction between treatment and period.

Assumptions I, II, and III, apply to all the candidate models. Assumption IV applies to Grizzle's and Jones & Kenward's models, but can be relaxed for the PLMEs model.

3.2.1 | Grizzle mixed-effects model

The algebraic notation for the Grizzle's model for DBP measurements, repeated within each period is given by:

$$Y_{ij} = \beta_0 + \beta_1 t_j + \beta_2 P_j + \beta_3 T_{ij} + b_{0i} + b_{1i} t_j + \epsilon_{ij}. \quad (2)$$

Here, b_{0i} is the random intercept and b_{1i} is the random slope (averaged over first and second period) for the ith participant, ϵ_{ij} is the random error in the measurement for the ith participant assigned to treatment T_{ij}, during period P_j, at time t_j. The assumptions for the random elements are:

$$\begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01}\rho_{01} \\ \sigma_{01}\rho_{01} & \sigma_1^2 \end{bmatrix} \right), \quad (3)$$

and $\epsilon_{ij} \sim N(0, \sigma^2)$. The components in $[b_{0i}, b_{1i}]'$ and ϵ_{ij} are independent. The vector $[\beta_0, \beta_1, \beta_2, \beta_3]'$ of fixed effects describes the average evolution of the response variable (DBP), and the vector $[b_{0i}, b_{1i}]'$ of random effects describes how the profile of the ith participant deviates from the average profile.

3.2.2 | Jones and Kenward mixed-effects model

The Jones & Kenward model is an extention of Grizzle's with inclusion of two interaction terms. The algebraic notation for the model is given by:

$$Y_{ij} = \beta_0 + \beta_1 t_j + \beta_2 P_j + \beta_3 T_{ij} + \beta_4 P_j t_j + \beta_5 T_{ij} t_j + b_{0i} + b_{1i} t_j + \epsilon_{ij}. \quad (4)$$

The parameter β_4 represents the fixed effect corresponding to the period-by-time interaction and is the net period effect. It implies that the effect of the treatment received at two different time periods may differ. β_5 denotes the fixed effect due to treatment-by-time interaction, estimating the net treatment effect commonly known as the difference-in-difference (DiD) estimate. The Jones & Kenward model notation for the random effects and random error is similar to Grizzle's.

3.2.3 | A flexible piecewise linear mixed-effects model

The piecewise linear mixed-effects (PLMEs) model has an extended parametrization beyond the Jones & Kenward model. The fact that the model does not assume fixed treatment effects across periods makes it flexible in modeling non-linear trends (in different periodic phases), without putting constraints on the curve evolution. The algebraic notation for the PLMEs model is given by:

$$\begin{aligned} Y_{ij} = {} & \beta_0 + \beta_1 t_j + \beta_2 P_j + \beta_3 T_{ij} + \beta_4 P_j t_j + \beta_5 T_{ij} t_j + \beta_6 P_j T_{ij} + \beta_7 P_j T_{ij} t_j \\ & + b_{0i} + b_{1i} t_j (1 - P_j) + b_{i2} t_j P_j + \epsilon_{ij}, \end{aligned} \quad (5)$$

β_6 and β_7 parameters represent the fixed effect due to treatment-by-period and treatment-by-period-by-time interactions, respectively, aliased with the carryover effect in two-period or three-period designs.[40] Armitage and Hills[3] prefer to speak in general of a treatment-by-period interaction rather than a carry-over effect or difference in carry-over effects because such an interaction may exist in the absence of any direct carry-over effect of a pharmacological or psychological nature. b_{0i} is the random intercept for the ith participant, b_{1i} is the random pre-change slope for the ith participant during the first period ($P_j = 0$), b_{i2} is the random post-change slope for the ith participant during the second period ($P_j = 1$), ϵ_{ij} is the random error in the measurement for the ith participant assigned to treatment T_{ij}, during period P_j, at time t_j. Note that the fixed intercepts and slopes are period-specific, but for the random effects only the slopes are period specific. The assumptions for the random elements are:

$$\begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01}\rho_{01} & \sigma_{02}\rho_{02} \\ \sigma_{01}\rho_{01} & \sigma_1^2 & \sigma_{12}\rho_{12} \\ \sigma_{02}\rho_{02} & \sigma_{12}\rho_{12} & \sigma_2^2 \end{bmatrix} \right), \tag{6}$$

and $\epsilon_{ij} \sim N(0, \sigma^2)$. The random effects and measurement errors are independent.

We further examine the time, treatment and period effects on the mean outcome profile for model (5). At $t_j = t_4$ it can be shown that $\beta_2 = -t_4\beta_3$ and $\beta_6 = -t_4\beta_7$. For implementation of a two piecewise linear spline model, the linear mean profile plot was fitted over time t_j with a knot at time t_4, where the curve was expected to change, hence a time spline t_j^* for all $t_j > t_4$ was created:

$$t_j^* = \begin{cases} 0, & \text{if } t_j \le t_4, \\ t_j\text{-}t_4 & \text{if } t_j > t_4. \end{cases} \tag{7}$$

Using two-time variables, t_j and t_j^*, two separate straight lines were fitted before and after t_4. The final model is constructed in four parts. The model for the reference group ($T_{ij} = 0$) is:

$$Y_{ij} = t_j^* = \begin{cases} \beta_0 + \beta_1 t_j + b_{i0} + b_{i1} t_j + \epsilon_{ij} & \text{if } t_j \le t_4, \\ (\beta_0 - t_4\beta_3) + (\beta_1 + \beta_3) t_j^* + b_{i2} t_j^* + \epsilon_{ij} & \text{if } t_j > t_4. \end{cases} \tag{8}$$

The model for the comparison group is:

$$Y_{ij} = \begin{cases} (\beta_0 + \beta_4) + (\beta_1 + \beta_5) t_j + b_{i0} + b_{i1} t_j + \epsilon_{ij} & \text{if } t_j \le t_4, \\ (\beta_0 + \beta_4) - t_4(\beta_3 + \beta_7) + (\beta_1 + \beta_3 + \beta_5 + \beta_7) t_j^* + b_{i2} t_j^* + \epsilon_{ij} & \text{if } t_j > t_4. \end{cases} \tag{9}$$

4 | ANALYSIS OF BLOOD PRESSURE DATA FROM A RANDOMIZED CROSS-OVER TRIAL

Next, we analyze the continuous DBP data for a cross-over trial described in Section 2. The time variable t_j used in the analysis is of a continuous type, common to the three models. At the point of swapping treatment, the GME and JKME models are implemented on a time scale t_j recoded as follows; $t = t_j$ for all $t_j \le t_4$, and $t = t_j - t_4$ for all $t_j > t_4$. The PLME model is implemented on two-time variables, t_j and t_j^*. The second time scale is recoded as shown in (7). The choice of treating time as a factor or a continuous variable depends on the research goal. If one prefers to see at what time point the change was significant, across particular time points, treating time as a categorical variable would be recommended. However, if one is interested in studying the functional relationship between the outcome and time, it is appropriate to treat time as a continuous variable. Also, if time points are not fixed by design implying different subjects to be measured at different time points, it may no longer be feasible to treat time as a categorical variable.

Figures 1 and 2 show the observed and predicted DBP individual and mean profiles, respectively. The predicted individual profiles for Grizzle's and Jones & Kenward's models in both periods are overlayed on the same time scale. The predicted mean profile is an average over individual profiles across both periods. Under the PLMEs model, the predicted individual profiles in both periods are estimated correctly, even though the mean plot depicts a break-point. This phenomenon is attributed to swapping of treatment arms coupled with the period data coding mechanism. The treatment effect is defined within the fixed-effects part of the model, at either side of the knot. At this point (of swapping treatment), individual profiles transition from last measurement in period 1 (at week 3) to the first measurement in period 2 (at week 4) at varying gradients. Individual plots are defined at the random-effects part of the model, which does not include the treatment effect, thus connected continuously across all time points. Considering the main effect of treatment, participants receiving High-dose and Low-dose Iodine treatment started with comparable means in DBP at baseline (week 0). There was a general decrease in DBP in both treatment groups. However, the rate of decrease was higher for those receiving High-dose compared to Low-dose Iodine household salt.

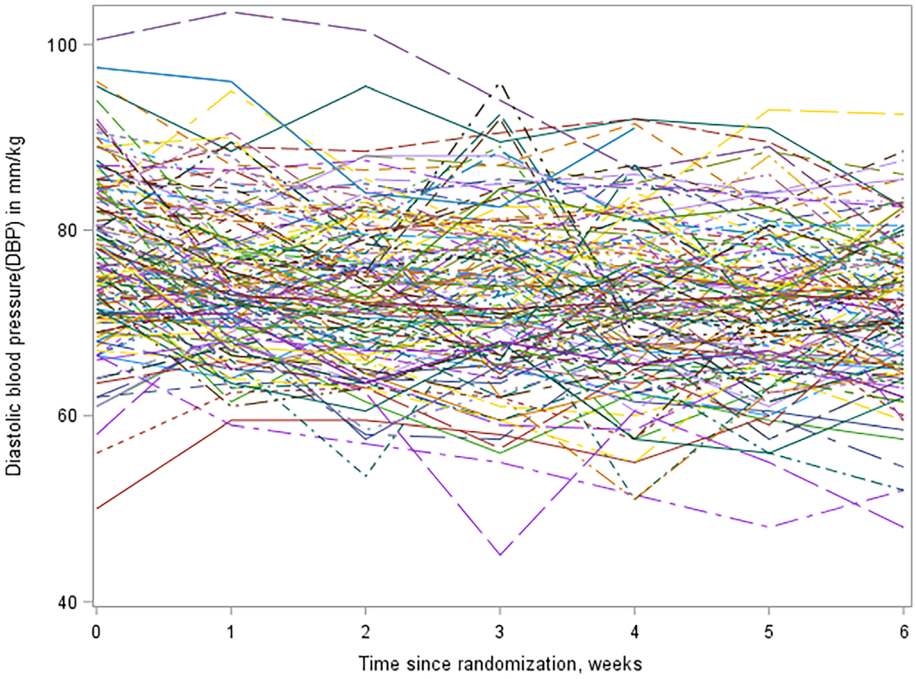

FIGURE 1 Observed DBP individual profiles.

With additional coding, the estimated parameters for the fixed effects were used to compute directly interpretable results (see SAS code in Appendix S1). There was a certain amount of intermittent missingness in the DBP outcome (7.4%). Hence, we used ReML for unbiased estimation of parameters, thanks to its validity under missing at random.[41–43] Table 2 shows the ReML estimates for the fixed effects and the variance components. The rate of decrease in DBP was statistically significant for all the candidate models (Grizzle's, Jones & Kenward's and PLMEs). However, the magnitude of change was highest for the PLMEs model (-0.81, s.e. $= 0.32$; $p = 0.0111$) followed by Jones & Kenward's model (-0.68, s.e. $= 0.30$; $p = 0.0225$) and lowest for Grizzle's model (-0.60, s.e. $= 0.30$; $p = 0.0480$). The residual error variance for DBP was lowest for the PLMEs model (21.02, s.e. $= 1.19$), followed by Jones & Kenward's (23.79, s.e. $= 1.20$) and highest for Grizzle's model (23.97 s.e. $= 1.21$).

Results for the likelihood-based measures for assessing how best each model fitted the data show that PLME model performed better, yielding the lowest estimates (-2LL $= 7248.1$, AIC $= 7252.1$, and BIC $= 7258.4$) compared to GME model which yielded the highest (-2LL $= 7263.8$, AIC $= 7267.8$, and BIC $= 7274.1$).

5 | SIMULATION STUDY

Our simulation study aims at mirroring real-life situation by deriving true underlying parameters from empirical data. We adopted continuous repeated measures evaluation setting. Assuming a vector Y_i of a single outcome repeatedly measured seven times for individual i,

$$Y_i = [y_{1i}, y_{2i}, \ldots, y_{7i}]^T,$$

follows a multivariate normal distribution, within the general linear mixed effects framework. The simulation of the continuous repeatedly measured DBP was implemented using the following set of ingredients. Let $X_i \in (0,1)$ be an indicator denoting assignment to two treatment sequences with an equal allocation ratio, where assignment is generated using $X_i \sim \text{Bern}(0.5)$; vector of parameters for the fixed effects

$$\beta = [75.9594 \quad -0.6328 \quad -0.0443 \quad 1.2296 \quad -1.0757 \quad 1.6276]^T;$$

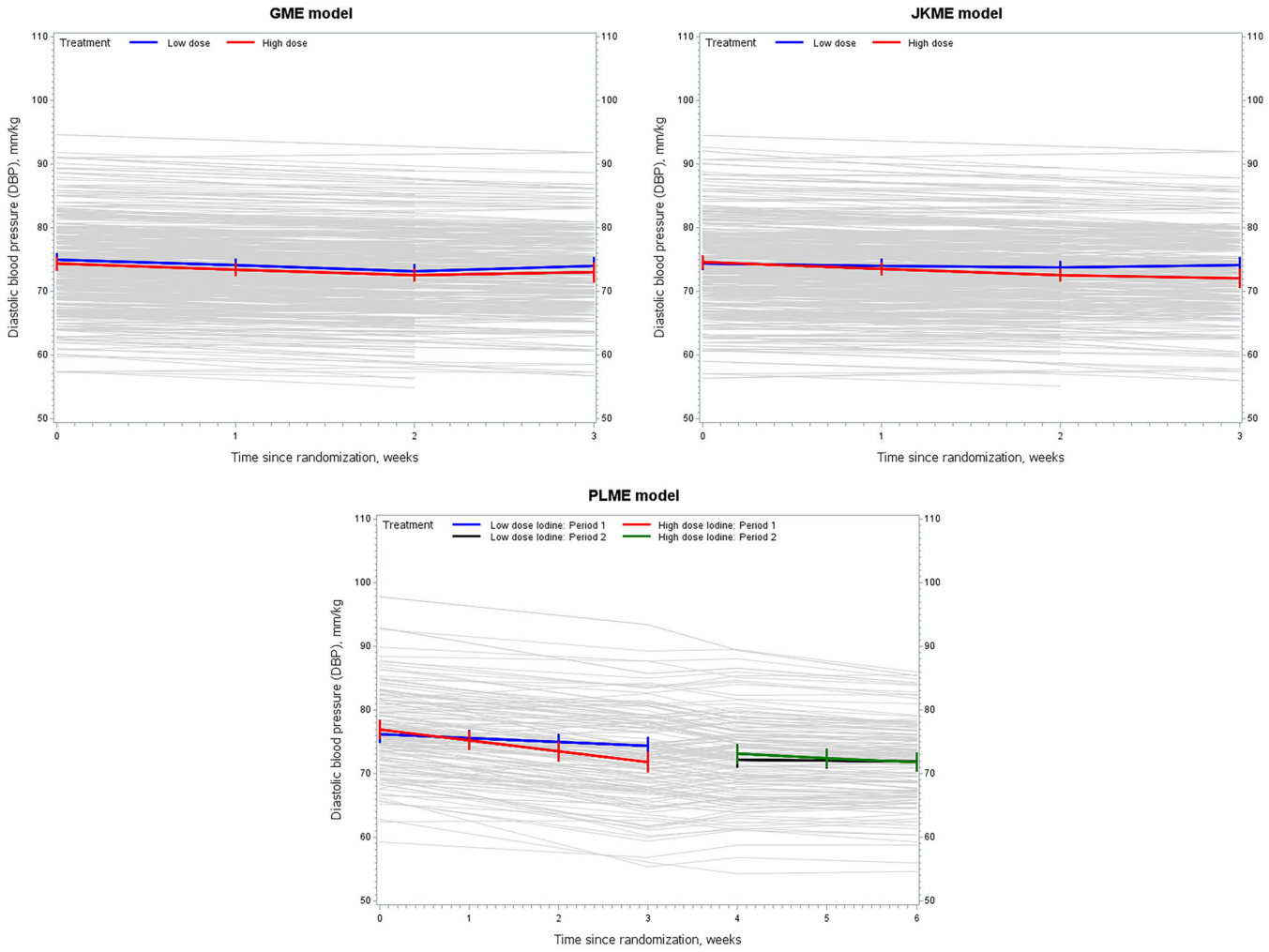

FIGURE 2 Predicted DBP individual and mean profiles by treatment.

variance–covariance matrix of the random effects

$$G = \begin{bmatrix} 48.7846 & -2.1354 & 0.0418 \\ -2.1354 & 1.8259 & -2.2249 \\ 0.0418 & -2.2249 & 3.7287 \end{bmatrix};$$

and the residual error variance $\sigma^2 = 21.0231$.

The data-generating mechanism adopted the following scenarios: small ($n = 20$), medium ($n = 100$) and large ($n = 300$) sample sizes with unstructured covariance for the random effects. Five hundred datasets were simulated ($c_{\text{sim}} = 500$) under each sampling scenario. The fixed-effects, random-effects and residual error parameters used in the simulation process were estimates obtained from fitting PLME model (5) to the data introduced in Section 2, with 2-treatments × 2-periods constituting two treatment periods of 3 weeks each with no washout between the treatment periods. Starting from baseline (week 0), seven measurements were repeatedly simulated, four measurements within period 1, and three measurements within period 2.

5.1 | Modeling simulation data

Modeling of the simulated data was done in various ways. In general, the process was adjusted in order to address two commonly encountered modeling challenges, namely, (1) problem of the data and/or (2) complexity in model specification. To start, we fitted a mixed-effects model with unstructured covariance of the random effects, across the three

TABLE 2 Model for continuous DBP with fixed- and random-effects.

		GME model		JKME model		PLME model	
Effect	**Parameter**	**Est (s.e)**	***p*-Value**	**Est (s.e)**	***p*-Value**	**Est (s.e)**	***p*-Value**
Fixed effects							
Intercept: *L*	β_0	74.96 (0.58)	<0.0001	74.38 (0.61)	<0.0001	75.96 (0.77)	<0.0001
Intercept: *H*	$\beta_0 + \beta_3$	74.37 (0.58)	<0.0001	74.62 (0.60)	<0.0001	77.19 (0.79)	<0.0001
Intercept: *H* – *L*	β_3	−0.60 (0.30)	0.0480	0.24 (0.47)	0.6159	1.23 (0.98)	0.2113
Slope (time) before week 3: *L*	β_1			−0.34 (0.23)	0.1484	−0.63 (0.26)	0.0139
Slope (time) before week 3: *H*	$\beta_1 + \beta_5$			−1.02 (0.23)	<0.0001	−1.71 (0.26)	<0.0001
Slope (time) before week 3: *H* – *L*	β_5			−0.68 (0.30)	0.0225	−1.08 (0.37)	0.0035
Slope (time) after week 3: *H*	$\beta_1 + \beta_4$					−0.68 (0.30)	0.0248
Slope (time) after week 3: *L*	$\beta_1 + \beta_4 + \beta_5 + \beta_7$					−0.13 (0.29)	0.6643
Slope (time) after week 3: *L* – *H*	$\beta_5 + \beta_7$					0.55 (0.48)	0.2462
Slope (time) change: *LtoH*	β_4					−0.04 (0.45)	0.9313
Slope (time) change: *HtoL*	$\beta_4 + \beta_7$					1.58 (0.44)	0.0004
Difference-in-difference: ((*HtoL*)-(*LtoH*))	β_7					1.63 (0.64)	0.0111
Average effect of H on DBP: (−(*HtoL*)-(*LtoH*))/2						−0.81 (0.32)	0.0111
Random effects							
Intercept	σ_0^2	44.55 (5.80)		44.56 (5.79)		48.78 (6.88)	
Intercept × time	$\sigma_{01}\rho_{01}$	−1.06 (1.25)		−1.00 (1.23)		−2.14 (1.62)	
Time	σ_1^2	0.88 (0.50)		0.80 (0.49)		1.83 (0.66)	
Intercept × timespl1	$\sigma_{02}\rho_{02}$					0.04 (2.59)	
Time × timespl1	$\sigma_{12}\rho_{12}$					−2.22 (1.05)	
timespl1	σ_2^2					3.73 (1.87)	
Residual	σ^2	23.97 (1.21)		23.79 (1.20)		21.02 (1.19)	
Model fit criteria	$-2logL\left(\widehat{\theta}\right)$	7263.8		7251.9		7248.1	
	AIC	7267.8		7255.9		7252.1	
	BIC	7274.1		7262.2		7258.4	

Abbreviations: $-2\log L\left(\widehat{\theta}\right)$, Negative 2 log-likelihood; AIC, Akaike information criterion; BIC, Bayesian information criterion; Est, estimate; GME, Grizzle's mixed-effects; H, high-dose Iodine; JKME, Jones & Kenward's mixed-effects; L, low-dose Iodine; PLME, piecewise linear mixed-effects; s.e, standard error.

candidate models (Grizzle, Jones & Kenward and PLMEs), for which estimates were computed. The fixed effects included time, treatment and period, with random intercepts and slopes. Convergence was satisfied across all candidate models in real-life data settings. The same models were fitted to the simulated datasets where convergence problems and other warnings were checked. Convergence criteria were met for all sampled datasets. However, in some of them the estimated *G* matrix turned non-positive definite. To overcome the problem associated with the *G* matrix, we reduced the model complexity by specifying one random effect (intercept only), and no problems were encountered across all scenarios in fitting the three candidate models (see SAS code in the Appendix S1).

5.2 | Estimands

The difference in mean DBP between two treatments was considered as the estimand (θ), representing treatment effect in a randomized cross-over trial. The parameter was based on effect size, defined by differences in DBP marginal means for participants consuming low-dose ($T_{ij} = 0$) and high-dose ($T_{ij} = 1$) Iodized salt, during the first and second periods in a randomized sequence. Appendix S1 shows the estimand for the three candidate models, estimated from real-life data.

5.3 | Performance measures

The term "performance measure" describes a numerical quantity used to assess the performance of a method. The performance measures required in a simulation study depend on the aims and targets for the study. When the target is an estimand, the most obvious performance measure to consider is bias. Precision and coverage of confidence intervals will also be of interest. The empirical standard error is similar to root mean square error. A simulation study targeting an estimand may of course also assess power and type I error.[44] Five performance measures, namely, bias, root mean square error (RMSE), coverage probability (C_p), empirical power (E_p) and likelihood-based measures of model fit were used to evaluate the performance of the PLMEs model against Grizzle and Jones & Kenward models.

RMSE is a quantitative measure of unexplained variability for a given response model. It is a metric that indicates how far apart (on average) the predicted values are from observed values. The closer the value to zero the better, indicating how well the model fits the data. In a longitudinal data setting, the RMSE of predicted values $\widehat{y}_t$ for times t of a regression's outcome variable y_t, with variables observed over T times, is computed for T different predictions as the square root of the mean of the squares of the deviations, expressed as,

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \widehat{y}_t)^2}. \tag{10}$$

C_p is the proportion of samples drawn from a sampling distribution for which the (known) population parameter is contained in the specified confidence interval (CI). The CI often has the form $\xi \pm \tau(\alpha, n)$, where ξ is an unbiased estimate of the parameter and $\tau(\alpha, n)$ is a width that depends on the significance level α, the sample size n, and the standard error of the estimate. The degree of certainty pre-specified by the analyst, referred to as the confidence level or confidence coefficient of the constructed interval, is effectively the nominal coverage probability of the procedure for constructing confidence intervals. The nominal coverage probability is often set at 0.95. A C_p estimate closer to the nominal coverage probability is desired for a given model. If all assumptions used to derive the confidence interval are met, nominal coverage equals C_p. If any of the assumptions are not met, the C_p may be smaller or larger than the nominal probability of application. When the C_p is greater than the nominal coverage probability, the interval is called a conservative (confidence) interval; when it is less than the nominal coverage probability, the interval is called a "non-conservative" or "acceptable" interval.

E_p of the design is the fraction of datasets with p-values smaller than or equal to 0.05 across the simulation runs (Arnold et al.[45]). This measure is key when the simulation study targets hypothesis testing. The higher the fraction the better for a given design. From the analysis of empirical (real-life) data, the direction of p-value corresponding to specific estimand (effect size) are shown in Appendix S1. This direction forms the basis of E_p calculation in our study.

Three likelihood-based measures of model fit used in the evaluation include: Negative 2 log-likelihood ($-2\log L\left(\widehat{\theta}\right)$), Akaike information criterion (AIC) and Bayesian information criterion (BIC). Let $\log L\left(\widehat{\theta}\right)$ denote the value of the maximized log-likelihood objective function for a model with k parameters fit to T data points. The AIC compares models from the perspective of information entropy, as measured by Kullback–Leibler divergence.[46] The AIC for a given model is expressed as,

$$\text{AIC} = \left[-2\log L\left(\widehat{\theta}\right) + 2k\right].$$

The BIC compares models from the perspective of decision theory, as measured by expected loss. The BIC for a given model is expressed as,

$$\text{BIC} = \left[-2\log L\left(\widehat{\theta}\right) + k\log T\right].$$

5.4 | Simulation results

The analysis was commenced by fitting models (2), (4), and (5) for Grizzle, Jones & Kenward and PLMEs models, respectively, with random intercept and slope(s). Even though convergence criteria were met in all 500 simulated datasets across three sampling scenarios, the specified structure was complex to fit, in a relatively high proportion of

simulated datasets. While estimating variance covariance matrix (G) for this datasets, some variance components turned negative hence forced to zero.

Table 3 presents the proportion of simulated datasets meeting convergence criteria with estimated G matrix turning positive definite. From the results, the PLMEs model sufficiently fitted all the datasets generated using medium and large sample scenarios, with a small proportion (9.2%) showing non-convergence to a hierarchical model in the small sample scenario. Both Grizzle and Jones & Kenward models turned complex in fitting relatively high proportion of the datasets ranging from 35% to 52.6% across all sampling scenarios. Contrary to expectation, the PLME model achieved stable convergence in a high proportion of simulated small samples ($n = 20$) compared to GME and JKME models. Even though our proposed PLME model is complex compared to GME and JKME, intuitively, its flexibility in modeling data in each phase, allowed satisfactory convergence in the majority of samples. Due to observed model complexities in a number of datasets, the performance measures were not computed for comparison across the three candidate models.

We fitted a reduced version of models (2), (4), and (5), with one random effect, namely random intercept. The models converged to a hierarchical view in all simulated datasets across the three sampling scenarios. Performance measures were therefore computed for comparison across the three candidate models. Table 4 presents the results for the performance measures for the three candidate models.

TABLE 3 % of datasets with G matrix turning positive-definite.

n	GME model	JKME model	PLME model
20	47.4	48.6	90.8
100	51.0	55.2	100.0
300	53.6	65.0	100.0

TABLE 4 Performance measures for the GME, JKME, and PLME models.

Performance measures	n	GME model	JKME model	PLME model
Bias	20	−0.0102	0.0034	−0.0050
	100	0.0040	0.0185	0.0070
	300	−0.0171	−0.0028	−0.0148
RMSE	20	2.639	2.600	2.575
	100	2.693	2.668	2.642
	300	2.705	2.681	2.656
C_p (%)	20	85.8	85.8	84.0
	100	87.8	87.8	87.6
	300	85.2	84.8	85.2
E_p (%)	20	31.6	32.2	29.0
	100	75.0	75.6	68.4
	300	99.2	99.2	97.4
$-2\log L\left(\widehat{\theta}\right)$	20	756.67	752.87	750.53
	100	3842.24	3832.26	3820.49
	300	11541.10	11512.05	11477.46
AIC	20	760.67	756.87	754.53
	100	3846.24	3836.26	3824.49
	300	11545.10	11516.05	11481.46
BIC	20	762.66	758.86	756.52
	100	3851.45	3841.47	3829.70
	300	11552.51	11523.46	11488.87

Abbreviations: $-2\log L\left(\widehat{\theta}\right)$, Negative 2 log-likelihood; AIC, Akaike information criterion; BIC, Bayesian information criterion; GME, Grizzle's mixed-effects; JKME, Jones & Kenward's mixed-effects; PLME, piecewise linear mixed-effects.

A PLMEs model yielded low RMSE estimates compared to Grizzle and Jones & Kenward models in the analysis of DBP continuous outcome. The pattern was consistent across all sampling scenarios.

C_p estimates across all the candidate models were $>80\%$, ranging from 84.0% to 87.8%. Compared to the nominal coverage probability of 95% used in our simulation study, the actual coverage probability across the three candidate models were relatively below the nominal level. Nonetheless, the estimated C_p from the PLME model were similar to those estimated from Grizzle and Jones & Kenward models, consistently across all simulation scenarios.

Results from the analysis of E_p revealed inconsistent inference across different scenarios. The estimates were relatively low for small samples ($n=20$) and highest for large samples ($n=300$), across the three candidate models. E_p estimates for PLME model were low compared to Grizzle and Jones & Kenward models, consistently across all simulation scenarios.

The results of the likelihood-based measures for assessing model fit were consistent with those from the analysis of real-life data presented in (Table 2), the likelihood-based measures for assessing how best each model fitted the data show that PLME model performed better by yielding the lowest estimates compared to GME model which yielded the highest, across the three sampling criteria.

In the case of a simple linear (marginal) model, we conducted the analysis of the simulated data to see the impact of the simplified model on the performance of the three candidate models. From the results of the performance measures (Table 5), it is clear that with a simple model (without random effects), the three models performed equally well, with PL marginal model showing marginal improvements compared to G marginal and JK marginal models. Comparing results in Table 4 and Table 5, it is imperative to note that additional random effects could explain much variability in the data, thus improving precision in estimation of the estimands (effect size) parameters.

TABLE 5 Performance measures for G marginal, JK marginal, and PL marginal models.

Performance measures	*n*	G marginal	JK marginal	PL marginal
Bias	20	0.0412	−0.0461	−0.0403
	100	−0.0173	−0.0222	−0.0160
	300	−0.0353	−0.0402	−0.0348
RMSE	20	6.857	6.843	6.726
	100	7.167	7.158	7.125
	300	7.178	7.170	7.149
C_p (%)	20	99.0	99.2	98.8
	100	99.8	99.8	99.8
	300	99.4	99.4	99.4
E_p (%)	20	98.0	98.0	97.6
	100	90.8	90.8	90.4
	300	52.2	52.0	52.0
$-2\log L\left(\widehat{\theta}\right)$	20	925.67	922.68	919.57
	100	4741.06	4738.42	4733.68
	300	14240.69	14235.84	14225.72
AIC	20	927.67	924.68	921.57
	100	4743.06	4740.42	4735.68
	300	14242.69	14237.84	14227.72
BIC	20	930.57	927.58	924.47
	100	4747.60	4744.96	4740.22
	300	14248.34	14243.49	14233.36

Abbreviations: $-2\log L\left(\widehat{\theta}\right)$, Negative 2 log-likelihood; AIC, Akaike information criterion; BIC, Bayesian information criterion; G, Grizzle's; JK, Jones & Kenward's; PL, piecewise linear.

6 | CONCLUDING REMARKS

We fitted a univariate PLMEs model to our simulation data problem. This was achieved by incorporating the piecewise linear spline model proposed by Fitzmaurice et al.[47] The performance of a PLMEs model was evaluated against two commonly cited models—Grizzle[33,34] and Jones and Kenward[7]—used in the analysis of cross-over trials. Its extended parametrization beyond Grizzle and Jones & Kenward models—without assuming a fixed treatment effect across periods—makes it more flexible in modeling non-linear trends (in different periodic phases), without putting constraints on the curve evolution. This allows a better fit to the data, which may result in improved accuracy in modeling treatment effects. From the results of the simulation study, the PLMEs model sufficiently fitted a significantly high proportion of the simulated datasets (across all sampling scenarios) to a complex hierarchical model with random intercept and slope(s) compared to Grizzle and Jones & Kenward models. When a hierarchical view-point is adopted, in the sense that outcomes are specified conditionally upon random effects, the variance–covariance matrix of the random effects must be positive-definite.[29] The PLME model is preferred especially in modeling increased numbers of random effects, compared to GME and JKME models, which work equally well with random intercepts only. Under a marginal viewpoint, all models achieved comparable performance, with some observed improvement in piecewise linear model. In some cases, additional random effects could explain much variability in the data, thus improving precision in estimation of the estimands (effect size) parameters.

Upon refitting the model with random intercept only, the three candidate models converged successfully for all simulated datasets (across all sampling scenarios) with varying performance measures. PLMEs model yielded low RMSE estimates compared to Grizzle and Jones & Kenward models. The C_p values were comparable across the three models, with mixed findings in the inference for E_p values.

A key conclusion from this investigation is that the univariate JKME model is a powerful tool, especially when fitting mixed models with random intercepts only. The addition of random slopes may lead to model complexities in most cases, resulting in unsatisfactory model convergence during estimation. To circumvent convergence pitfalls, extension of JKME model to PLME model allows a more flexible fit to the data (generated from cross-over design settings).

It was desirable to compare the three candidate models using the complex hierarchical model, with random intercept and slope(s), as fitted to the motivating case study data. Convergence in the proportion of datasets with non-positive definite Hessian matrix, occurred because of negative variance component(s), resulting in the reduction of the dimensions in the random effects matrix to random intercept only, whose convergence was successful across all sampling scenarios. Reduction in the dimensions of the random effects matrix did not limit the simulation study in a substantial way, our focus was in the comparison of three candidate models with application of a mixed-effects model, which was achieved successfully.

The data generating mechanism used in this paper captures two time periods for the simple cross-over design (2-treatments $\times$ 2-periods). Its application is extendible to more complex cross-over designs with multiple periods. In addition, it is important to note that, even for single response models, adding more random effects increases the complexity of the model and thus may be difficult or impossible to fit in some cases.

The time variable t_j used in our analysis is of a continuous type, in which case only monotonic increase (or decrease) over time was considered. In a simple linear mixed model, it is possible to include higher-order terms or time points as categorical variables. The illustration of a linear relationship is without loss of generality as one can easily extend the models with, for example, polynomial terms or add dummy variables to allow for non-linear trends. However, with a complex model such as the PLME model, this may be difficult from the viewpoint of model stability. Hence the assumption of the model may be too strong where non-monotonic increase (or decrease) over time t_j are considered. This could be an interesting topic for further research.

For implementation, the proposed method allows efficient use of available resources, such as the SAS procedure MIXED. Additional coding effort is required in order to fit PLMEs models and extract the summary tables, leading to generating results that are directly interpretable. See Appendix S1 for the SAS code used to analyze the DBP data.

ACKNOWLEDGMENTS

Special thanks to the team of investigators who participated in the original research work.[35] The authors acknowledge financial support provided by the Government of Canada through Foreign Affairs, Trade and Development Canada (DFATD). The fund was channeled through Canadian International Food Security Research Fund (CIFSRF), a program of Canada's International Development Research Centre (IDRC), project number 106510.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Since the data were collected from humans, the study proposal was approved by the Kenya Medical Research Institute (KEMRI) Ethical Review Board. The study participants/subjects gave an informed consent by signing a consent form. This manuscript has no financial/commercial conflicts of interests. Permission for data sharing can be sought through KEMRI Ethical Review Board.

ORCID

Moses Mwangi https://orcid.org/0000-0001-6172-2507

REFERENCES

1. Brown BW Jr. The crossover experiment for clinical trials. *Biometrics*. 1980;36:69-79.
2. Hills M, Armitage P. The two-period cross-over clinical trial. *Br J Clin Pharmacol*. 1979;8(1):7-20.
3. Armitage P, Hills M. The two-period crossover trial. *J R Stat Soc*. 1982;31(2):119-131.
4. Soltanian AR, Faghihzadeh S. A generalization of the grizzle model to the estimation of treatment effects in crossover trials with non-compliance. *J Appl Stat*. 2012;39(5):1037-1048.
5. Mills EJ, Chan A-W, Wu P, Vail A, Guyatt GH, Altman DG. Design, analysis, and presentation of crossover trials. *Trials*. 2009;10(1):1-6.
6. Jones B, Kenward MG. *Design and Analysis of Cross-over Trials*. Chapman and Hall; 1989.
7. Jones B, Kenward MG. *Design and Analysis of Cross-over Trials*. 3rd ed. Chapman and Hall; 2015.
8. Allison PD. Change scores as dependent variables in regression analysis. *Sociol Methodol*. 1990;20:93-114.
9. Glymour MM, Weuve J, Berkman LF, Kawachi I, Robins JM. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. *Am J Epidemiol*. 2005;162(3):267-278.
10. Liu GF, Lu K, Mogg R, Mallick M, Mehrotra DV. Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials? *Stat Med*. 2009;28(20):2509-2530.
11. Vickers AJ. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC Med Res Methodol*. 2001;1(1):1-4.
12. Clifton L, Clifton DA. The correlation between baseline score and post-intervention score, and its implications for statistical analysis. *Trials*. 2019;20(1):1-6.
13. Lee U, Garcia TP, Carroll RJ, Gilbreath KR, Wu G. Analysis of repeated measures data in nutrition research. *Front Biosci (Landmark Edition)*. 2019;24:1377-1389.
14. Nunes EV, Pavlicova M, Hu M-C, et al. Baseline matters: the importance of covariation for baseline severity in the analysis of clinical trials. *Am J Drug Alcohol Abuse*. 2011;37(5):446-452.
15. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. John Wiley & Sons; 2004:141-147.
16. Sheppard L. Insights on bias and information in group-level studies. *Biostatistics*. 2003;4(2):265-278.
17. Mwangi M, Verbeke G, Njagi EN, et al. Improved longitudinal data analysis for cross-over design settings, with a piecewise linear mixed-effects model. *Commun Stat*. 2021;7(3):413-431.
18. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer Science & Business Media; 2009.
19. Molenberghs G, Verbeke G. *Models for discrete longitudinal data*. Springer; 2005.
20. Nguyen TT, Bazzoli C, Mentré F. Design evaluation and optimization in crossover pharmacokinetic studies analysed by nonlinear mixed effects models. *Stat Med*. 2012;31(11–12):1043-1058.
21. Panhard X, Mentré F. Evaluation by simulation of tests based on non-linear mixed-effects models in pharmacokinetic interaction and bioequivalence cross-over trials. *Stat Med*. 2005;24(10):1509-1524.
22. Wang J, Jones B. Optimal cross-over designs for nonlinear mixed models using a first-order expansion. *J Stat Plan Inference*. 2009;139(2):203-212.
23. Wang X, Chinchilli VM. Analysis of crossover designs for longitudinal binary data with ignorable and nonignorable dropout. *Stat Methods Med Res*. 2022;31(1):119-138.
24. Chernoff H. On the distribution of the likelihood ratio. *Ann Math Stat*. 1954;25:573-578.
25. Nelder J. The interpretation of negative components of variance. *Biometrika*. 1954;41(3/4):544-548.
26. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer-Verlag; 2000.
27. Diggle PJ, Heagerty P, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. Oxford University Press; 2002.
28. Molenberghs G, Verbeke G. A note on a hierarchical interpretation for negative variance components. *Stat Model*. 2011;11(5):389-408.
29. Oliveira I, Molenberghs G, Verbeke G, Demétrio C, Dias C. Negative variance components for non-negative hierarchical data with correlation, over-, and/or underdispersion. *J Appl Stat*. 2017;44(6):1047-1063.
30. Brown H, Prescott R. *Applied Mixed Models in Medicine*. John Wiley & Sons; 2015.
31. Hetzel B. *The Nature and Magnitude of the Iodine Deficiency Disorders*. Oxford University Press; 2004.
32. Jooste PL, Weight MJ, Lombard CJ. Iodine concentration in household salt in South Africa. *Bull World Health Organ*. 2001;79:534-540.

33. Grizzle JE. The two-period change-over design and its use in clinical trials. *Biometrics*. 1965;21:467-480.
34. Grizzle JE. Correction to grizzle (1965). *Biometrics*. 1974;30(4):727.
35. Bukania Z, Mwangi M, Kaduka L, Kimiywe J, van der Haar F. Difference in blood pressure between arms of normotensive female farmers in the semi-arid midlands of Kenya. *Epidemiology (Sunnyvale)*. 2015;5(2):1165-2161.
36. Kenji GM, Nyirenda KK, Kabwe GC. Iodine levels in edible salt sold in Malawi, Kenya and Zambia. *Afr J Food Agric Nutr Dev*. 2003;3(2).
37. Bernal-Rusiel JL, Greve DN, Reuter M, Fischl B, Sabuncu MR. Statistical analysis of longitudinal neuroimage data with linear mixed effects models. *NeuroImage*. 2013;66:249-260.
38. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982;38(4):963-974.
39. Lindstrom MJ, Bates DM. Newton—raphson and em algorithms for linear mixed-effects models for repeated-measures data. *J Am Stat Assoc*. 1988;83(404):1014-1022.
40. Li B, Zhou Z, Zhang L, Yang Y. Testing treatment-by-period interaction in four-period crossover trials. *Pharm Stat*. 2020;19(2):145-163.
41. Little RJ, Rubin DB. *Statistical Analysis with Missing Data*. Vol 793. John Wiley & Sons; 2019.
42. Molenberghs G, Kenward M. *Missing data in clinical studies*. John Wiley & Sons; 2007.
43. Raudenbush SW, Bryk AS. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Vol 1. Sage Publications, Inc; 2002: Sage.
44. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019;38(11):2074-2102.
45. Arnold BF, Hogan DR, Colford JM, Hubbard AE. Simulation methods to estimate design power: an overview for applied research. *BMC Med Res Methodol*. 2011;11(1):1-10.
46. Pérez-Cruz F. Kullback-leibler divergence estimation of continuous distributions. In *2008 IEEE International Symposium on Information Theory*; 2008:1666-1670).
47. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. Vol 998. John Wiley & Sons; 2012.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Mwangi M, Verbeke G, Njagi EN, et al. Evaluation of a flexible piecewise linear mixed-effects model in the analysis of randomized cross-over trials. *Pharmaceutical Statistics*. 2023;1-15. doi:10.1002/pst.2357