Acta Scientiarum
http://periodicos.uem.br/ojs
ISSN on-line: 1807-8664
Doi: 10.4025/actascitechnol.v46i1.61512

Bayesian approach for a 2 x 2 crossover design with repeated measures: a simulation study

Yaciled Paola Miranda Lopez[1*], Denismar Alves Nogueira[2] and Luiz Alberto Beijo[2]

[1]Programa de Pós-Graduação em Estatística Aplicada e Biometria, Universidade Federal de Alfenas, Rua Gabriel Monteiro da Silva, 700, 37130-000, Alfenas, Minas Gerais, Brazil. [2]Departamento de Estatísticas, Universidade Federal de Alfenas, Alfenas, Minas Gerais, Brazil. *Author for correspondence. E-mail: yac12mir@hotmail.com

ABSTRACT. In crossover designs, the subjects receive all treatments, according to the groups of sequences formed. Therefore, if carryover effects are present in the model, inferences about the treatments effects become difficult. Furthermore, repeated measures of the response variable can be taken over time in the same experimental unit; however, these measures may be correlated. In this way, we aimed to analyze a 2 x 2 crossover design with repeated measures within the treatment period, using a Bayesian approach. A simulation study was performed to evaluate the performance. The posterior estimates of the model parameters were obtained under non-informative prior distributions and the normal likelihood function. The model performed well with a sample size of 20 subjects, showing even better results with samples of 100 subjects. With larger samples, exact tests for the differences in carryover effects and time effects were obtained. However, the test of time effect proved to be powerful even with small samples. In turn, considering carryover effects different from zero did not influence the estimates of treatment differences, although biased estimates of the period effect were obtained.

Keywords: carryover effects; longitudinal data; MCMC; prior distribution.

Received on November 11, 2022.
Accepted on March 17 2023.

Introduction

In clinical, psychological and agricultural investigations, studies with crossover designs are commonly performed, with the purpose of comparing the effects of treatments based on the information provided by each subject. In this sense, the main feature of crossover designs is that each subject receives different treatments in different periods of time (Jones & Kenward, 2014; Krogh et al., 2019).

In health area, crossover designs are suitable to study treatments involving chronic diseases such as asthma, migraine, epilepsy, hypertension, rheumatism, among others (Gomes & Dantas, 2017; Niazi, Hashempur, Taghizadeh, Heydari, & Shariat, 2017; Hermans et al., 2018; Samsonsen et al., 2018; Mcloughlin et al., 2019). Drugs with a quick, short and irreversible effect are used to treat these diseases, since the main objective is not the total cure, but to minimize the effects through treatments (Senn, 2002).

One of the advantages of crossover studies is that each subject acts as its own control (Jones & Kenward, 2014), providing a direct comparison of the treatments received, and the difference between these measurements minimizes the effects of the subject and, consequently, the experimental variability. Furthermore, a smaller number of subjects is needed compared to other types of designs to obtain efficient results with a good precision of the treatments effects. These crossover designs remove any biological and methodological variation, which provides ethical and economic advantages (Senn, 2003; Li & Sivaganesan 2016; Krogh et al., 2019).

However, the possible carryover effects may be a disadvantage in crossover designs. These effects are the treatment effects that persist in the next period and may hinder the obtaining of unbiased estimates of treatments and other parameters of the crossover model. To deal with the possibility of the presence of carryover effects, studies propose performing a period of cleaning up the effects of the current treatment, which is called washout (Jones & Kenward, 2014; Krogh et al., 2019). However, in some cases, this period is not enough.

A procedure that can be applied in experimental designs is called repeated measures, in which several measurements are taken from the same experimental unit over time or space. As in the crossover designs, repeated measures allow the evaluation of the differences in responses in a more efficient way, since they allow to analyze the response variation, over time, requiring a smaller number of subjects (West, Welch, &

Galecki, 2022). However, the measurements taken in the same experimental unit may be correlated, due to the common contribution of each subject.

For the analysis of crossover studies with repeated measures, a mixed model approach is proposed in the present study, which is suitable for the adjustment of the correlation produced by repeated measures (Littell, Milliken, Stroup, Wolfinger, and Schabenberger, 2006). The denomination of mixed models comes from the fact that systematic and random effects are considered in the model. In the crossover model, the effects of treatments, periods, sequences, time and the simple interaction between time and treatment are commonly considered systematic, whereas the effect of the experimental unit can be considered both systematic or random.

The Bayesian approach can be applied in the analysis of mixed models in crossover studies with repeated measures. In this approach, previous information or initial beliefs about the model are implemented through prior distributions, and the likelihood function is also specified, to be further represented by the posterior distribution. Specifically in crossover designs, Bayesian analysis allows inferences about the model parameters (such as treatment and period effects) considering the uncertainties about the presence of carryover effects (Li & Sivaganesan, 2016).

In this sense, we aimed to analyze a simple crossover study with repeated measures within treatment period through data simulation with the presence of time effect, with no time-treatment interaction and with equal carryover effects for the treatments, using a Bayesian approach. A second objective was to verify if the posterior estimates of the model parameters correspond to the simulated values, and therefore, evaluate the performance of the estimates through the mean squared error, accuracy and precision.

Material and methods

2 × 2 Crossover designs

A crossover design is an experimental design in which each subject randomly receives all treatments according to determined sequences; in this way, each subject acts as its own control. There are different sets of sequences depending on the number of treatments. In the simplest *AB/BA* design, there are two treatments (*A* and *B*) and it is possible to make two sequences (*AB* and *BA*) in two distinct periods. According to Senn (2002) and Jones and Kenward (2014), this simpler design is called 2×2 and the subjects are randomly distributed in each sequence; sequence *AB* receives treatment A in the first period and treatment B in the second period, while sequence *BA* receives treatment B in the first period and treatment A in the second period.

To minimize any relevant interference in subject responses or carryover effects, a washout period is applied between the application of each treatment (Jones & Kenward, 2014; Krogh et al., 2019). This washout period can be passive, in which no treatment is applied to the subject, a rest period, or active (non-stop).

The statistical model that considers the carryover effect is based on the studies of Chow and Lui (2009) and Jones and Kenward (2014). If y_{ijk} is the observed value of the response variable Y, then, in a traditional way, the linear model of the crossover design is described as:

$$y_{ijk} = \mu + \pi_j + \tau_{t[i,j]} + \lambda_{t[i,j-1]} + S_k + e_{ijk}, \quad (1)$$

in which μ is the intercept; π_j is the effect associated with the period j with $j = 1, 2$; $\tau_{t[i,j]}$ is the direct effect of treatment applied in the period j of the sequence i, $i = 1, 2$; $\lambda_{t[i,j-1]}$ is the carryover effect for the treatment t, in which $\lambda_{t[i,0]} = 0$; S_k is the effect associated to the $k\text{-}th$ subject, $k = 1, \cdots, n_i$, in which n_i is the amount of subjects in the sequence i and e_{ijk} is the random error, that follows a normal distribution, with mean 0 and variance σ_e^2.

The carryover effect in the model can be considered as the sequence effect or as the interaction between treatment and period, and it is only present in the second period.

2 X 2 Crossover designs with repeated measures

Repeated measures is a procedure that consists in collecting data of a response variable in several occasions over time or in different study conditions of the same experimental unit (West, et al., 2022). According to Littell et al. (2006), the interest of studies with repeated measures is to verify if there are main effects of treatment or time and if there is an interaction of time and treatment.

A common practice in crossover experiments is to collect a sequence of observations from each subject, under the same treatment, in each period (Jones & Kenward, 2014). Thus, the traditional 2×2 crossover model with repeated measures of each treatment in each period of time can be written as:

$$y_{ijkl} = \mu + \pi_j + \tau_{t[i,j]} + \lambda_{t[i,j-1]} + \gamma_l + \gamma_l\tau_{t[i,j]} + S_k + e_{ijkl}, \quad (2)$$

note that the difference from the simple crossover model (1) is that it incorporates the systematic effect of the l-th time as a repeated measure (γ_l) and the simple interaction between time and treatment ($\gamma_l\tau_{t[i,j]}$). Furthermore, in this model it cannot be assumed that the errors e_{ijkl} for the same subject are independent; although the treatments have been randomly assigned to each subject, the levels of the repeated measures factor, in this case, the time, are not randomly assigned to the subjects (Littell et al., 2006). Therefore, it can be assumed that $e_{ijkl} \sim N(0,R)$, in which R is a variance-covariance matrix. The other model effects are defined as in model (1).

The assumption of errors independence of the classic models of analysis of variance is not guaranteed. In addition, repeated measures variances can often change over time. These situations lead to losses in precision and predictive capacity of models fitted according to classical assumptions (Gómez, Torres, García, & Navarro, 2012). However, mixed models allow capturing the correlation between measurements. Furthermore, these models can handle missing and imbalanced data well.

2 x 2 Crossover designs with repeated measures as a mixed model

The term mixed models comes from the fact that these models incorporate systematic and random effects to describe the relationship between the response variable (dependent) and the explanatory variables (independent, predictors or covariates) (Seoane, 2014). The systematic effects are, in most cases, the factors of interest in the study. They are restricted to the observed levels of the variable and there is no interest in generalizing the results to other levels. On the other hand, random effects are explanatory variables which levels were sampled from a population about which there is interest in inferring (Littell et al., 2006).

According to West et al. (2022), the mixed model is given in matrix by:

$$\begin{aligned} &\mathrm{Y} = \mathrm{X}\beta + \mathrm{Zu} + \mathrm{e}, \\ &\mathrm{u} \sim \mathrm{N}(0, \mathrm{G}) \\ &\mathrm{e} \sim \mathrm{N}(0, \mathrm{R}), \end{aligned} \quad (3)$$

in which $Y \sim N(X\beta, ZGZ' + R)$ is the response vector of dimension $n \times 1$; X is the incidence matrix associated with the systematic effects of dimension $n \times$ b; β is the vector of systematic effects of dimension b × 1; Z is the incidence matrix associated with the random effects of dimension $n \times$ q; u is the vector of random effects of dimension q × 1; and e is the vector of errors of dimension q × 1; in which n is the number of observations; b is the number of systematic effects; and q is the number of random effects. The G e R matrices are positive definite and, therefore, not singular.

The mixed models are suitable for the analysis of crossover designs with repeated measures. Considering S_k as the random effect of the k-th subject in the 2 × 2 crossover model with repeated measures (2), the marginal mixed model can be considered as:

$$\begin{aligned} &\mathrm{y}_{\mathrm{ijkl}} = \mu + \pi_{\mathrm{j}} + \tau_{\mathrm{t[i,j]}} + \lambda_{\mathrm{t[i,j-1]}} + \gamma_{\mathrm{l}} + \gamma_{\mathrm{l}}\tau_{\mathrm{t[i,j]}} + \epsilon_{\mathrm{ijkl}} \\ &\epsilon_{\mathrm{ijkl}} = \mathrm{S}_{\mathrm{k}} + \mathrm{e}_{\mathrm{ijkl}} \end{aligned} \quad (4)$$

it is assumed that $S_k \sim N(0,\sigma_s^2)$, in which σ_s^2 is the variance of the subjects, and e_{ijkl} is the vector of errors with $e_{ijkl} \sim N(0, R)$. Note that model (4) can be written in a matrix form as: $Y = X\beta + e^*$, where e^* is the random vector of errors, in which the random effects of subjects S_k are absorbed by the residual part of the model, $e^* \sim N(0, V)$, in which $V = \mathrm{var}(Y) = ZGZ' + R$.

Simulation of a 2 x 2 crossover data with 3 repeated measures

To evaluate the simulation behavior of the data obtained from the 2 × 2 crossover experiment, 28 scenarios were considered, which were simulated N = 1,000 times each, stablishing sample sizes of n = 20 and 100, and considering equal number of subjects in each sequence. The variances $\sigma^2 = (\mu * cv)^2$ =25 and 400 were defined as a function of the general mean μ = 100, and the coefficient of variation of the subject cv = 5% and 20%. Furthermore, null covariance between the subjects was assumed.

The difference between the treatments means $\tau = \tau_1 - \tau_2 = w \cdot SE$, with w = 1 and 2, and SE the standard error for the specific comparison (frequentist reference). It was assumed that there was no difference between the periods when $\pi = 0.1$, a value close to zero, and there was no significant effect of the sequence when establishing that the carryover effects of the treatments were equal, $\lambda_1 = \lambda_2 = 0$. In some cases, the carryover

effects $\lambda_1 = \lambda_2 = 4SE$, SE between the treatments means, were considered to evaluate the impact on the estimates when the carryover effects of each treatment were equal, but different from zero.

After establishing the values of the parameters, the data simulation from two bivariate normal distributions was performed, corresponding to the two groups of sequences. The data of AB sequence followed a bivariate normal distribution with mean $\mu_{AB} = (\mu + \pi_1 + \tau_1, \mu + \pi_2 + \tau_2 + \lambda_1)$ and variance and covariance matrix, Variance Components, $\Sigma_{AB} = I\sigma^2$. The data of BA sequence followed a bivariate normal distribution with mean $\mu_{BA} = (\mu + \pi_1 + \tau_2, \mu + \pi_2 + \tau_1 + \lambda_2)$ and the same structure for the variance and covariance matrix, $\Sigma_{BA} = \Sigma_{AB}$.

Consequently, the linear regression effect was inserted into the subjects within each period, through the following model:

$$y_{ijkl} = b_{ijk} + \gamma x_l + \epsilon_{ijkl},$$

in which b_{ijk} is the intercept, considered as the response of the k-th subject to the i-th sequence in the j-th period of the simulated 2 × 2 crossover data; γ is the regression coefficient associated with time; x_l is the effect of the l-th time, with $l = 1, 2, 3$, in which $x_1 = 0, x_2 = 1$ and $x_3 = 2$; and $\epsilon_{ijkl} \sim N(0, \boldsymbol{\Sigma}_k)$ is the vector of errors associated with the observation for the k-th subject, with $\boldsymbol{\Sigma}_k = \boldsymbol{I}\sigma^2$, in which $\boldsymbol{I}$ is the identity matrix and $\sigma^2 = 20$ is the variance between errors. The data obtained corresponded to the errors matrix of the k-th subject in the l-th time.

Since there was an interest in verifying whether the proposed simulation adequately captured the established time regression coefficient, for each simulated scenario, the time regression coefficients (γ) were varied according to their standard errors. In this way, γ were established as $\gamma = r \cdot SE_\gamma$, with $r = 0$, 2 and 6. For the scenarios with samples of 20 subjects, the $SE_\gamma = 0.5$, thus γ resulted in the 0, 1 and 3 values. For the scenarios with 100 subjects, the $SE_\gamma = 0.22$, therefore, γ were established in the simulation as 0, 0.44 and 1.32. Furthermore, it was assumed that there was no significant interaction between time and treatment, $\gamma_l \tau_{t[i,j]} = 0$.

The description of the simulated scenarios for the 2 × 2 crossover design with 3 repeated measures is presented in Tables 1 and 2. The data were simulated using an adaptation of the function *getSimulationData()* from the *Reproducer* package (Madeyski, 2017) of R software (R Core Team, 2018), in which the way of introducing the means of the sequences (μ_{AB} and μ_{BA}) was modified, as well as the possibility of entering with distinct n_i. Therefore, repeated measures could be generated through the effect of linear regression, as described in the link https://github.com/yacimiranda/Crossover_codex.git.

Table 1. Scenarios (C1 to C24) simulated for the 2 x 2 crossover design with 3 repeated measures, considering the difference between the treatment means (τ) equal to 1*SE* or 2*SE*, carryover effects $\lambda_1 = \lambda_2 = 0$, difference between the periods effects π=0.*1* and interaction between time and treatment, $\gamma_l \tau_{t[i,j]} = 0$.

Sce	n	cv(%)	τ	γ	Sce.	n	cv(%)	τ	γ
C1	20	5	1.581 (1SE)	0	C13	100	5	0.707 (1SE)	0
C2	20	5	1.581 (1SE)	1	C14	100	5	0.707 (1SE)	1
C3	20	5	1.581 (1SE)	3	C15	100	5	0.707 (1SE)	3
C4	20	5	3.162 (1SE)	0	C16	100	5	1.414 (1SE)	0
C5	20	5	3.162 (1SE)	1	C17	100	5	1.414 (1SE)	1
C6	20	5	3.162 (1SE)	3	C18	100	5	1.414 (1SE)	3
C7	20	20	6.325 (1SE)	0	C19	100	20	2.828 (1SE)	0
C8	20	20	6.325 (1SE)	1	C20	100	20	2.828 (1SE)	1
C9	20	20	6.325 (1SE)	3	C21	100	20	2.828 (1SE)	3
C10	20	20	12.649 (1SE)	0	C22	100	20	5.657 (1SE)	0
C11	20	20	12.649 (1SE)	1	C23	100	20	5.657 (1SE)	1
C12	20	20	12.649 (1SE)	3	C24	100	20	5.657 (1SE)	3

Sce - Scenarios; n – Total number of experimental units; cv - Coefficient of variation; SE – Standard error.

Table 2. Scenarios (C25 to C28) simulated for the 2 x 2 crossover design with 3 repeated measures, considering the difference between the treatment means (τ) equal to 1*SE* or 2*SE* and carryover effects $\lambda_1 = \lambda_2 = 4SE_\gamma$ of the time regression coefficient.

Sce	n	cv(%)	$\lambda_1 = \lambda_2$	π	τ	γ	γτ
C25	20	5	6.325 (4SE)	0.1	1.581 (1SE)	3.00	0
C26	20	20	25.298 (4SE)	0.1	6.325 (1SE)	1.00	0
C27	100	5	5.657 (4SE)	0.1	1.414 (2SE)	0.00	0
C28	100	20	11.314 (4SE)	0.1	2.828 (1SE)	1.32	0

Sce - Scenarios; n – Total number of experimental units; cv - Coefficient of variation; SE – Standard error.

Bayesian approach

The Bayesian approach was used for the analysis of the simulated data of the simple crossover study. This approach makes inferences about the unknown population parameters through the posterior distribution, which is obtained from the combination of information extracted from the sample (likelihood function) and information prior to sampling, called prior distribution of the parameters of interest (Degroot & Schervish, 2012).

It is considered that the unknown population parameters θ are random, and obtaining the posterior distribution is an application of the Bayes' theorem, thereby:

$$P(\theta|y) \propto L(\theta|y)P(\theta), \quad (5)$$

in which $P(\theta|y)$ is the posterior distribution of θ given y; $L(\theta|y)$ is the likelihood function of y and $P(\theta)$ is the prior distribution of parameter θ.

The prior distribution probabilistically represents the knowledge that exists about the population parameters (θ) before carrying out the experiment. The definition of the prior distribution is made by the researcher, specifying the distribution form and the hyperparameter values. The parameters of the prior distribution are called hyperparameters. Thus, Bayes' theorem updates the prior knowledge with sample information (Paulino, Turkman, & Murteira, 2003).

Analysis and modeling

The model considered for the data analysis was based on the model proposed by Zhang, Toubouti, and Carlin (2017), in which, considering the correlations between the three measurements taken within a period for the same treatment, the probability distribution is:

$$\begin{pmatrix} Y_{ijk1} \\ Y_{ijk2} \\ Y_{ijk3} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{ijk1} \\ \mu_{ijk2} \\ \mu_{ijk3} \end{pmatrix} , \quad \sum_t \right), \quad (6)$$

in which μ_{ijkl} is given by:

$$\mu + \lambda\, Seq_i + \pi\, Per_j + \tau\ Trat_{t[i,j]} + \gamma\ Time_l + \gamma\tau\ Time_l \times Trat_{t[i,j]}, \quad (7)$$

t is determined by the i-th sequence and the j-th period. In addition, *Trat* is the treatment to be tested; τ is the difference between the test treatment and the control; *Seq* is the sequence; *Per* is the period; and *Time* refers to the repeated measure. In this model, the subject effect was not explicitly specified, considering the marginal mixed model as described in equation (4).

Non-informative normal prior distributions were used for the model parameters, i.e., no subset of the parametric space was favored. Therefore, the following prior distributions were considered:

$$\mu \sim N(0; 0.001); \quad \lambda \sim N(0; 0.001); \quad \pi \sim N(0; 0.001);$$

$$\tau \sim N(0; 0.001); \quad \gamma \sim N(0; 0.001); \quad \gamma\tau \sim N(0; 0.001),$$

which are presented with the mean and precision hyperparameters.

A non-informative inverse Wishart prior distribution was considered for the variance-covariance matrix $\mathbf{\Sigma}_t$, so it was defined that the precision matrix $\mathbf{\Sigma}_{t\text{-}1} \sim Wishart(\boldsymbol{R}_t, \nu)$, in which $\boldsymbol{R}_t = \boldsymbol{I}\sigma^2$, with $\sigma^2 = 0.001$, $t = 1, 2$, and $\nu = dim(\mathbf{\Sigma}_t) = 3$ is the degrees of freedom. According to Li, Zhou, Zhang, and Yang (2020), the dimension of matrix $\mathbf{\Sigma}_t$ is the smallest value of degrees of freedom for which non-informative Wishart prior distribution is adequate.

To fit the model and obtain the marginal posterior distributions of the parameters μ, λ, π, τ, γ, $\gamma\tau$, and the $\mathbf{\Sigma}_t$ matrices, the *OpenBUGS* software (Lunn, Spiegelhalter, Thomas, & Best, 2009) through the *R2OpenBUGS* package interface (Sturtz, Ligges, & Gelman, 2005) was used. This package implements Monte Carlo methods via Markov chains to obtain the marginal chains. A chain with 800,000 iterations was generated, with burn-in of 775,000 and thin of 5 iterations, obtaining a final chain of 5,000 samples.

To verify the chain convergence, the following criteria were used: Dependence factor of Raftery and Lewis (1992) close to 1; stationarity of Heidelberger and Welch (1983) chain, p-value $\geqslant 0.05$; and the non-significance for the comparison of the means of the Geweke (1992) Z test by the standard Normal.

Finally, point estimates were obtained for each parameter through the posterior means, as well as the 95% Highest Probability Density (HPD) intervals. To verify the behavior of our approach for a nominal confidence

of 5%, the type I error rates were estimated for the parameters of the differences in carryover (λ) and time (γ) effects, as well as the power of these tests. The mean squared error (MSE), accuracy and precision of the estimates of parameters λ, τ and γ, in each one of the n data sets, were also estimated.

Results and discussion

Considering the 1,000 simulations performed in each scenario and the average values of the posterior estimates of the model parameters, we observed that the estimated values were close to the real values simulated for the scenarios with samples n = 20 and 100. Analyzing the averages of the limits of 95% HPD intervals for the parameters, we could note that the effect of the general mean added to the effect of treatment B (μ_B) was significant, therefore being different from zero.

Furthermore, the differences in carryover and period effects, as well as the interaction between time and treatment, were not significant, as expected, since these conditions were established in the simulations. In scenarios C3, C6, C9, C12 (n = 20) and C15, C18, C21, C24 (n = 100), in which a linear time effect corresponding to $6SE_\gamma$ was simulated, the variable time showed a significant effect. However, in scenarios C2, C5, C8, C11 (n = 20) and C14, C17, C20, C23 (n = 100), in which a time effect of $2SE_\gamma$ was simulated, the time was not significant on average.

Although a difference of 1SE has been simulated for the treatments effects, this difference was not statistically significant in any of the following scenarios: C1, C2, C3, C7, C8, C9, C13, C14, C15, C19, C20 and C21. On the other hand, when $\tau = 2SE$ was simulated, there were no significant effects for the results averages of scenarios with cv = 5% (C4, C5, C6, C16, C17 and C18); however, the differences in treatments for scenarios with cv = 20% (C10, C11, C12, C22, C23 and C24) were significant. This result may be due to the fact that the standard error depends on the variability, and if it is greater, higher standard errors and, consequently, greater simulated differences are expected.

Tables 3 and 4 show the estimates of type I error rates for the differences in carryover effects ($H_O : \lambda = 0$) and time effects ($H_O : \gamma = 0$). The test power for time is given in the scenarios in which a linear time effect of $r \cdot SE_\gamma$, with r = 2 or 6 was simulated.

Table 3. Type I error rates of the tests of differences in carryover effects (λ) and time effects (γ) in the scenarios in which these effects were assumed null, and the power of γ test for the scenarios in which a linear time effect of $r \cdot SE_\gamma$, with r = 2 or 6 was simulated; results for scenarios with 20 subjects, with τ = 1SE (C1, C2, C3, C7, C8, C9) and τ = 2SE (C4, C5, C6, C10, C11, C12), in which $\lambda_1 = \lambda_2 = 0$ was assumed.

Sce (cv=5%)	Par	τ	Tipe I error	Power	Sce (cv=20%)	Par	τ	Tipe I error	Power
C1	λ	1.581	0.122	-	C7	λ	6.325	0.132	-
	γ	1.581	0.085	-		γ	6.325	0.080	-
C2	λ	1.581	0.114	-	C8	λ	6.325	0.128	-
	γ	1.581	-	0.423		γ	6.325	-	0.390
C3	λ	1.581	0.104	-	C9	λ	6.325	0.142	-
	γ	1.581	-	0.995		γ	6.325	-	0.991
C4	λ	3.162	0.116	-	C10	λ	12.65	0.130	-
	γ	3.162	0.090	-		γ	12.649	0.068	-
C5	λ	3.162	0.133	-	C11	λ	12.649	0.136	-
	γ	3.162	-	0.477		γ	12.649	-	0.399
C6	λ	3.162	0.116	-	C12	λ	12.649	0.134	-
	γ	3.162	-	0.994		γ	12.649	-	0.989

Note: ' -' measure not calculated.

The type I error rates of the differences in carryover effects for the scenarios with cv = 5 and 20% were similar (Table 3). Considering the exact confidence interval (99%) for the proportion $p_\alpha = 0.05$, $CI_{99\%}(p_\alpha)$ = [0.034; 0.068], all type I error rates for λ were statistically greater than the 5% significance level, since the values were greater than the interval upper limit. Therefore, the test for the difference in carryover effects was liberal for scenarios with 20 subjects.

The hypothesis test of time effect was liberal, except for scenario C10, in which the type I error rate was included in the interval $CI_{99\%}(p_\alpha)$, indicating that the test was exact.

From these results, it is noteworthy that the type I error rates of the difference tests in carryover effects with samples of 20 subjects were all greater than 10%, which was the level of significance suggested by Grizzle (1965) to perform the carryover effect test. If the null hypothesis that the difference in carryover effects is equal to zero is rejected, only the data from the first period should be used in the analysis (Jones & Kenward, 2014).

Although the results presented here are theoretical, based on data simulation with repeated measures, there are applied studies, such as the one performed by Yin, Paoletti, Sargent, and Mandrekar (2017), in which the authors studied the toxicity time-trend of five doses of a drug using a sample size of 36 subjects. In this study, the type I error rate of the test of time effects as repeated measures was less than 13%.

Furthermore, for scenarios C2, C5, C8 and C11, in which a time effect of $2SE_\gamma = 1$ was considered, the power values were less than 50%, indicating that in the 1,000 simulations performed in these scenarios, the test rejected the null hypothesis (H_0: $\gamma = 0$) in a little more than 500 simulations, which is an unexpected result. On the other hand, for scenarios C3, C6, C9 and C12, in which $\gamma = SE_\gamma = 3$, the power values were approximately 99%, since a greater difference between the values of γ was considered.

For $n = 100$ samples, the test for λ was exact in scenarios C14, C20, C21, C22, C23 and C24, according to the interval $CI_{99\%}(p_\alpha)$ (Table 4). Contrasting the results with scenarios of $n = 20$ samples, the scenarios with samples of 100 subjects presented all type I error rates for the carryover hypothesis test lower than 10%, which was suggested by Grizzle (1965).

Table 4. Type I error rates of the tests of differences in carryover effects (λ) and time effects (γ) in the scenarios in which these effects were assumed null, and the power of the γ test for the scenarios in which a linear time effect of $w \, . \, SE$, with $w = 2$ or 6 was simulated; results for scenarios with 100 subjects, with $\tau = 1SE$ (C13, C14, C15, C19, C20, C21) and $\tau = 2SE$ (C16, C17, C18, C22, C23, C24), in which $\lambda_1 = \lambda_2 = 0$ was assumed.

Sce (cv=5%)	Par	τ	Tipe I error	Power	Sce (cv=20%)	Par	τ	Tipe I error	Power
C13	λ	0.707	0.089	-	C19	λ	2.828	0.075	-
	γ	0.707	0.058	-		γ	2.828	0.052	-
C14	λ	0.707	0.066	-	C20	λ	2.828	0.054	
	γ	0.707	-	0.307		γ	2.828	-	0.346
C15	λ	0.707	0.078	-	C21	λ	2.828	0.057	-
	γ	0.707	-	0.995		γ	2.828	-	0.994
C16	λ	1.414	0.085	-	C22	λ	5.657	0.065	-
	γ	1.414	0.049	-		γ	5.657	0.054	-
C17	λ	1.414	0.071	-	C23	λ	5.657	0.065	-
	γ	1.414	-	0.340		γ	5.657	-	0.330
C18	λ	1.414	0.077	-	C24	λ	5.657	0.061	-
	γ	1.414	-	0.987		γ	5.657	-	0.988

Note: ' -' measure not calculated.

Considering the $CI_{99\%}(p_\alpha)$, the hypothesis test of time effect was exact in all scenarios in which $\gamma = 0$ was simulated (C13, C16, C19 and C22). Li et al. (2020) reported that the type I error rates for the test of time effect were close to the 5% significance level in an applied study in which a simulation was performed with a 4 × 4 crossover design with repeated measures using sample sizes greater than or equal to 100 experimental units.

Similarly, to the scenarios with $n = 20$, the test power for the time effect was greater than 98% in the scenarios C15, C18, C21 and C24, in which $\gamma = 6SE_\gamma$ was simulated. These results show that the sample size did not influence the test power for γ when there was a time effect; nevertheless, an increase in power was observed when the time effect was greater.

It can be highlighted that, regarding type I error and power of the difference tests in carryover and time effects, better results were obtained in the scenarios with samples of 100 experimental units and $\gamma = 6SE\gamma$, highlighting the scenarios with $cv = 20\%$ (C21 and C24).

Additionally, Figure 1 presents the results of the mean squared errors, accuracies and amplitudes of the 95% HPD intervals (from which the precision was verified) of the parameters estimates of the differences in carryover and treatments effects, and time effects, in each scenario.

In scenarios with coefficient of variation of 5%, lower mean squared errors and amplitudes of the 95% HPD intervals were obtained for the estimates of λ and τ parameters compared to the scenarios with coefficients of variation of 20%. For the time effect, these measures were similar when the same number of samples was considered, even with different coefficients of variation. In all scenarios, high accuracies (above 85%) were obtained.

Furthermore, for the scenarios with samples $n = 100$ subjects and $cv = 5\%$, lower mean squared errors were obtained for the parameters of the differences in carryover effects, differences in treatment effects, and time effects. In addition, for these scenarios, smaller amplitudes of the 95% HPD intervals were obtained compared to the scenarios with samples of 20 subjects, which indicates greater precision of the estimates.

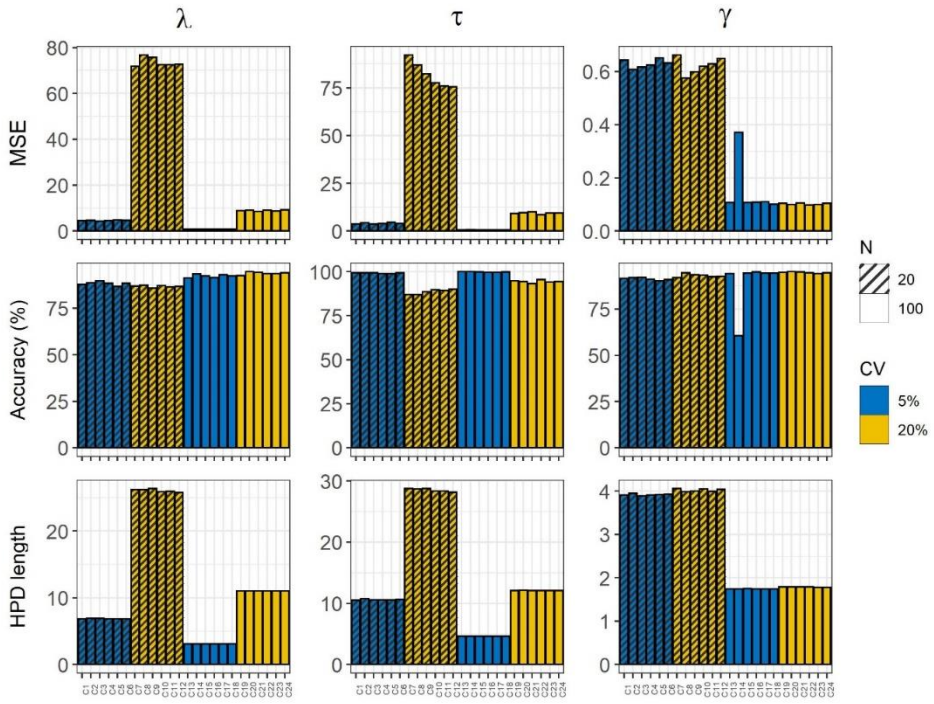

Figure 1. Mean squared errors (MSE), accuracies (%) and precision (amplitude of HPD intervals) of differences estimates in carryover effects (λ), differences in treatments (τ) and time effects (γ), from the 2 x 2 crossover model with 3 repeated measures, for scenarios n = 20, with τ = 1SE (C1, C2, C3, C7, C8, C9) and τ = 2SE (C4, C5, C6, C10, C11, C12), and scenarios with n = 100, with τ = 1SE (C13, C14, C15, C19, C20, C21) and τ = 2SE (C16, C17, C18 , C22, C23, C24).

However, satisfactory results were obtained for the estimates in the analysis for samples of 20 subjects. Highlighting one of the advantages of crossover designs, a smaller number of subjects is required compared to other designs to obtain a high test power and high precision of estimates (Jones & Kenward, 2014; Larsen, Juul, Kreilgaard, Kristensen, & Simonsson, 2018).

In all scenarios with samples of 100 subjects, low type I error rates were obtained for all parameters. The power results for the time effect test were similar for scenarios with 20 subjects and 100 subjects. In addition, Carriere and Huang (2000) emphasize that crossover designs provide greater test power, due to the less variability between subjects.

Simulated data analyses for scenarios with $\lambda_1 = \lambda_2 = 4SE$

This section presents the mean values of the parameters estimates of the 2 × 2 crossover model with 3 repeated measures, in which the carryover effects of each treatment were simulated with 4*SE* of the treatments means, $\lambda_1 = \lambda_2 = 6.325$; $\lambda_1 = \lambda_2 = 25.29$; $\lambda_1 = \lambda_2 = 5.657$; $\lambda_1 = \lambda_2 = 11.314$, for scenarios C25, C26, C27 and C28, respectively. Note that, in the simulation, the difference between the carryover effects is equal to zero ($\lambda = \lambda_2 - \lambda_1 = 0$).

The mean values of the parameters estimates were close to the simulated real values, except for the parameter of difference in period effects (π), which was significant in all scenarios (Table 5). According to Li et al. (2020), when a significant period effect is found, the model interpretation may be problematic; however, this does not preclude the possibility of performing the test for treatment effects.

The π estimates were close to the simulated values for the carryover effect of the treatment received in the first period, λ_1 for the *AB* sequence and λ_2 for the *BA* sequence (Table 5). These results may be due to the fact that the simulated values for the carryover effects ($\lambda_1 = \lambda_2$), different from zero but equal to each other, were captured by the difference in periods effects (π), since the carryover effects were in the second period only. According to Jones and Kenward (2014), carryover effects $\lambda_1 = \lambda_2 = 0$ cannot be tested when there are period effects in the model. In this sense, it can be inferred that carryover effects are mixed or confused with period effects.

The differences in carryover effects (λ) were not significant in all scenarios. The effect of the general mean added to treatment B was significant in all scenarios, keeping the pattern observed in the simulations. As $\lambda_1 = \lambda_2$, the estimates of the differences in treatment effects are not biased, being close to the simulated values and not significant in all scenarios. In addition, significant time effects were obtained in scenarios C25 and

C28, in which linear time effects of $\gamma = 6SE_\gamma$ were considered. For the interaction between time and treatment, the effects were not significant.

Table 5. Means of descriptive statistics of posterior point and interval estimates and minimum and maximum values of the convergence criteria of the parameters of the 2 x 2 crossover model with 3 repeated measures, for the scenarios C25 (*cv* = 5%) and C26 (*cv* = 20%), in which $\lambda_1 = \lambda_2 = 4SE$, with samples of 20 subjects, and for the scenarios C27 (*cv* = 5%) and C28 (*cv* = 20%), with samples of 100 subjects.

Sce	Par	RV	mean	LL	UL	Sce	Par	RV	mean	LL	UL
C25	μ_B	265.42	260.94	256.24	265.38	C27	μ_B	265.59	265.28	263.31	267.21
	λ	0.00	0.38	-3.10	3.83		λ	0.00	0.10	-1.47	1.65
	π	0.10	6.78	3.29	10.34		π	0.10	5.82	4.27	7.39
	τ	1.58	2.73	-2.51	8.09		τ	1.41	1.61	-0.72	3.93
	γ	3.00	3.32	1.93	4.73		γ	0.00	0.07	-0.57	0.68
	$\gamma\tau$	0.00	-0.30	-2.22	1.68		$\gamma\tau$	0.00	-0.06	-0.93	0.83
C26	μ_B	260.68	247.95	232.09	262.61	C28	μ_B	264.17	260.78	254.88	266.54
	λ	0.00	5.54	-7.15	18.52		λ	0.00	0.85	-4.70	6.30
	π	0.10	30.04	17.15	43.14		π	0.10	12.36	6.82	17.75
	τ	6.32	13.01	-0.91	27.54		τ	2.83	3.87	-1.99	10.09
	γ	1.00	1.28	-0.20	2.76		γ	1.32	1.36	0.73	1.98
	$\gamma\tau$	0.00	-0.27	-2.30	1.75		$\gamma\tau$	0.00	-0.04	-0.96	0.83

Note: Sce - scenario, Par - parameter, RV - real value, LL - HPD lower limit of 95% credibility, UL - HPD upper limit of 95% credibility.

Similar values of type I error rates and power of the tests of differences in carryover (λ) and time (γ) effects were obtained, as well as the measures of mean squared error, accuracy and precision of the estimates of differences in carryover effects, differences in treatments and time effects of the model, considering the scenarios simulated with $\lambda_1 = \lambda_2 = 0$.

As noted in the literature, carryover effects affect the process of making inferences about the treatments when they are not equal ($\lambda_1 \neq \lambda_2$), as there may be biased estimates of the other model parameters.

Conclusion

Based on the results of the present study, it is possible to conclude that under the Bayesian approach, the model presented good performance in relation to accuracy, mean squared error and precision of estimates with a sample size of 20 subjects. The results of these measurements were even better with a sample size of 100 subjects. Furthermore, with larger samples, exact tests of the differences in carryover effects and time effects were obtained. However, the test of time effect proved to be powerful even with small samples.

To minimize type I error rates and obtain more accurate and precise estimates, it is recommended to use a minimum sample size of 100 subjects in 2 × 2 crossover experiments with repeated measures.

Considering carryover effects different from zero for each treatment does not interfere in the analysis of the differences in the treatments effects. In addition, there are no losses in accuracy and precision of the estimates of the model parameters λ, τ and γ. However, special care must be taken with regard to the interpretation of the estimates of period differences, which capture the values of carryover effects present in each treatment.

Furthermore, if there are period effects in the model and the carryover effects are different from zero and equal to each other, it is impracticable to infer about the estimates of period and carryover effects for each treatment, due to the confusion between these two important effects.

In future studies we intend to consider different structures of covariances and random subject effects, as well as the use of the Bayes factor to make inferences about the significance of the model parameters.

References

Carriere, K., & Huang R. (2000). Crossover designs for two-treatment clinical trials. *Journal of Statistical Planning and Inference*, *87*, 125-134. DOI: https://doi.org/10.1016/s0378-3758(99)00185-8

Chow, S. C., & Liu, J. P. (2009). *Design and Analysis of Bioavailability and Bioequivalence Studies* (3rd ed.). London, UK: Chapman and Hall/CRC.

Degroot, M. H., & Schervish, M. J. (2012). *Probability and statistics* (4th ed.). Boston, MA: Addison-Wesley Longman.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (Research Paper, 148). Federal Reserve Bank of Minneapolis, Research Department. DOI: https://doi.org/10.21034/SR.148

Gomes, R. P., & Dantas, F. F. O. (2017). Influência do tamanho da massa muscular ativa na hipotensão pós-exercício em idosas hipertensas. *Revista UNI-RN*, *16-17*, 228-238. Retrieved from http://revistas.unirn.edu.br/index.php/revistaunirn/article/view/393

Gómez, S., Torres, V., García, Y., & Navarro, J.A. (2012). Procedimientos estadísticos más utilizados en el análisis de medidas repetidas en el tiempo en el sector agropecuario. *Revista Cubana de Ciencia Agrícola*, *46*, 1-7. Retrieved from http://www.redalyc.org/articulo.oa?id=193024313001

Grizzle, J. E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics, 21*(2), 467-480. DOI: https://doi.org/10.2307/2528104

Heidelberger, P., & Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, *31*(6), 1109-1144. DOI: https://doi.org/10.1287/OPRE.31.6.1109

Hermans, L., Nijs, J., Calders, P., De Clerck, L., Moorkens, G., Hans, G., ... Meeus, M. (2018). Influence of morphine and naloxone on pain modulation in rheumatoid arthritis, chronic fatigue syndrome/fibromyalgia, and controls: a double-blind, randomized, placebo-controlled, cross-over study. *Pain Practice*, *18*(4), 418-430. DOI: https://doi.org/10.1111/PAPR.12613

Jones, B., & Kenward, M. G. (2014). *Design and analysis of cross-over trials* (2nd ed.). London, UK: Chapman and Hall/CRC.

Krogh, H. B., Storebø, O. J., Faltinsen, E., Todorovac, A., Ydedahl-Jensen, E., Magnusson, F. L., ... Simonsen, E. (2019). Methodological advantages and disadvantages of parallel and crossover randomised clinical trials on methylphenidate for attention deficit hyperactivity disorder: a systematic review and meta-analyses. *BMJ open*, *9*(3), e026478. DOI: https://doi.org/10.1136/bmjopen-2018-026478

Larsen, M. S., Juul, R. V., Kreilgaard, M., Kristensen, A. T., & Simonsson, U. S. (2018). Impact of trial design on the estimation of drug potency and power in clinical trials of haemophilia with inhibitors. *European Journal of Pharmaceutical Sciences*, *123*, 531-538. DOI: https://doi.org/10.1016/J.EJPS.2018.07.056

Li, B., Zhou, Z., Zhang, L., & Yang, Y. (2020). Testing treatment-by-period interaction in four- period crossover trials. *Pharmaceutical statistics*, *19*(2), 145-163. DOI: https://doi.org/10.1002/PST.1975

Li, D., & Sivaganesan, S. (2016). An objective Bayesian analysis of a crossover design via model selection and model averaging. *Statistics in medicine*, *35*(25), 4509-4527. DOI: https://doi.org/10.1002/SIM.7015

Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D., & Schabenberger, O. (2006). *SAS for Mixed Models* (2nd ed., p. 814). SAS Institute: Cary. ISBN: 9781590475003

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in medicine*, *28*(25), 3049-3067. DOI: https://doi.org/10.1002/SIM.3680

Madeyski, L. (2017). *Reproducer: Reproduce statistical analyses and meta-analyses* (R package version 0.5.1). Retrieved from http://CRAN.R-project.org/package=reproducer

McLoughlin, R., Berthon, B. S., Rogers, G. B., Baines, K. J., Leong, L. E., Gibson, P. G., ... Wood, L. G. (2019). Soluble fibre s upplementation with and without a probiotic in adults with asthma: A 7-day randomised, double blind, three way cross-over trial. *EBioMedicine*, *46*, 473-485. DOI: https://doi.org/10.1016/j.ebiom.2019.07.048

Niazi, M., Hashempur, M. H., Taghizadeh, M., Heydari, M., & Shariat, A. (2017). Efficacy of topical Rose (Rosa damascena Mill.) oil for migraine headache: A randomized double-blinded placebo-controlled cross-over trial. *Complementary therapies in medicine*, *34*, 35-41. DOI: https://doi.org/10.1016/j.ctim.2017.07.009

Paulino, D. C., Turkman, M. A. A., & Murteira, B. (2003). *Estatística Bayesiana*. Lisboa, PT: Fundação Calouste Gulbenkian.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, AU: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Raftery, A. E., & Lewis, S. (1992). Comment: One long run with diagnostics: implementation strategies for markov chain monte carlo. *Statistical Science*, *7*(4), 493-497. DOI: https://doi.org/10.1214/ss/1177011143

Samsonsen, C., Myklebust, H., Strindler, T., Brathen, G., Helde, G., & Brodtkorb, E. (2018). The seizure precipitating effect of alcohol: A prospective observational cross-over study. *Epilepsy research*, *143*, 82-89. DOI: https://doi.org/10.1016/j.eplepsyres.2018.04.007

Senn, S. S. (2002). *Cross-over Trials in Clinical Research* (2nd ed.). Chichester, West Sussex: John Wiley and Sons.

Senn, S. J. (2003). Crossover Design. In S. C. Chow (Ed.)., *Encyclopedia of Biopharmaceutical Statistics* (p. 255-262). London, KY: Taylor and Francis and CRC. Retrieved from https://archive.org/details/encyclopediaofbi0002unse_g6g9

Seoane, J. (2014). ¿Modelos mixtos (lineales)? Una introducción para el usuario temeroso. *Etologuía: boletín de la Sociedad Espanõla de Etología*, 24, 15-37. ISSN 1135-6588

West, B. T., Welch, K. B., & Galecki, A. T. (2022). *Linear mixed models: a practical guide using statistical software* (3nd ed.). New York, NY: Chapman and Hall/CRC. DOI: https://doi.org/10.1201/9781003181064

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software*, *12*(3), 1-16. DOI: https://doi.org/10.18637/jss.v012.i03

Yin, J., Paoletti, X., Sargent, D. J., & Mandrekar, S. J. (2017). Repeated measures dose-finding design with time-trend detection in the presence of correlated toxicity data. *Clinical Trials*, *14*(6), 611-620. DOI: https://doi.org/10.1177/1740774517723829

Zhang, Q., Toubouti, Y., & Carlin, B. P. (2017). Design and analysis of Bayesian adaptive crossover trials for evaluating contact lens safety and efficacy. *Statistical methods in medical research*, *26*(3), 1216-1236. DOI: https://doi.org/10.1177/0962280215572272