Analysis of cross-over designs with serial correlation within periods using semi-parametric mixed models

Peer-reviewed author version

MARINGWA, John; Geys, Helena; SHKEDY, Ziv; FAES, Christel; MOLENBERGHS, Geert; AERTS, Marc; Van Ammel, Karel; Teisman, Ard & BIJNENS, Luc (2008) Analysis of cross-over designs with serial correlation within periods using semi-parametric mixed models. In: STATISTICS IN MEDICINE, 27(28). p. 6009-6033.

DOI: 10.1002/sim.3363
Handle: http://hdl.handle.net/1942/9077

TUTORIAL IN BIOSTATISTICS

Analysis of Cross-Over Designs with Serial Correlation within Periods Using Semi-parametric Mixed Models

*John T. Maringwa[1], Helena Geys[2], Ziv Shkedy[1], Christel Faes[1], Geert Molenberghs[1], Marc Aerts[1], Karel Van Ammel[2], Ard Teisman[2], Luc Bijnens [2]

[1] *Center for Statistics, Hasselt University, BE3590 Diepenbeek, Belgium*

[2] *Johnson and Johnson Pharmaceutical Research and Development, Beerse, Belgium*

SUMMARY

The use of semi-parametric mixed models has proven useful in a wide variety of settings. Here, we focus on application of the methodology in the particular case of a cross-over design with relatively long sequences of repeated measurements within each treatment period and for each subject. Other than an overall measure of the difference between each one of the experimental groups and the control group, specific time point comparisons may also be of interest. To that effect, we propose the use of flexible semi-parametric mixed models, enabling the construction of simulation-based simultaneous confidence bands. The bands take into account both between- and within-subject variability, while simultaneously correcting for multiple time point comparisons. Owing to the relatively long sequences of measurements per subject, presence of serially correlated errors is anticipated and investigated. We illustrate how several formulations of semiparametric

*Correspondence to: Center for Statistics, Hasselt University, Agoralaan 1, B-3590 Diepenbeek, Belgium

E-mail: john.maringwa@uhasselt.be

mixed models can be fitted and construction of simulation-based simultaneous confidence bands using SAS PROC MIXED.

KEYWORDS: Cross-over; Penalized splines; Semi-parametric mixed models; Simultaneous confidence bands, Serial correlation.

1. INTRODUCTION

In a cross-over trial, each unit or subject receives a sequence of experimental treatments rather than a single one, in randomized order. The main advantage of a cross-over trial is that treatments are compared within subject such that the difference between treatment measurements removes any subject effect from the comparison, ordinarily greatly increasing precision. Administering the treatments in random order helps to minimize, remove, and/or estimate effects stemming from time period or from so-called carry-over treatment effect from earlier into later time periods. The theory is well established, whether for the simplest case of two treatments and two periods, or for higher-order designs [1]. In this tutorial, a particular case of a cross-over design with the salient feature of a relatively long sequence of repeated measurements within each of the treatment periods is considered. Focus is put on modeling the mean evolution using semi-parametric mixed models, accounting for correlation between observations through random effects. A considerable amount of literature with regard to repeated-measures cross-over designs already exists, although mainly focusing on two treatment and two periods designs [2, 3, 4, 5]. Dunsmore [4] uses the Bayesian growth curves of Fearn [6], with a quadratic time effect, in a two-period repeated measures cross-over design. Analyzing the same experiment as Dunsmore [4], Grender and Johnson [5] discuss a two-stage approach wherein the repeated measures across time for each subject are modeled

parametrically, also using a quadratic trend, and later analyzing the parameter estimates using multivariate methods. More recently, Putt and Chinchilli [7] analyze a two treatment and four period design using a mixed effects model that eliminates the need for preliminary testing for such nuisance factors as, for example, carry-over. Again, a parametric model assuming quadratic effects in time is employed.

There are many practical situations where determining an appropriate parametric function for the mean may not at all be easy. As pointed out by Dunsmore [4], checking for the assumption of the presumed time trend (quadratic, in that case) may be a difficult task. It is with such cases in mind that we propose modeling the mean evolution using flexible semi-parametric models, ridding of the need to specify any particular parametric form. Application of semi-parametric smoothing techniques such as, for example, penalized splines [8, 9] will prove useful. Jones and Kenward [1] consider a cross-over study with many periods and model the period effects using natural cubic splines. Related applications in longitudinal data settings, not necessarily cross-over designs, can be found in, for example [10, 11, 12, 13, 14].

Having estimated the group mean profiles using penalized splines, focus shifts to constructing confidence bands around the fitted functions. A detailed account on frequentist pointwise and simultaneous confidence bands, both in the cross-sectional and longitudinal settings, is offered in [9]. Guo [12] constructs Bayesian confidence intervals around the fitted functions in a parallel design. Lin and Zhang [15] discuss, in the context of generalized linear mixed models, both frequentist and Bayesian confidence intervals around fitted functions. Wood [16] constructs confidence intervals for generalized additive models fitted using penalized splines. Our intention is to make adaptations of the bands of Ruppert *et al.* [9] to accommodate correlation between measurements through random effects, as well as more complex models for residual covariances; specifically, models including

serial correlation and measurement error [17]. Indeed, experiments with long sequences of repeated measurements are bound to yield some form of residual dependencies, in addition to what is captured by random effects and which, at least, should be modeled parametrically.

In Section 2, a description of the experiment and data used, together with a discussion of the research questions is provided. Section 3 exemplifies use of the AUC as a summary statistic. In Section 4, focus is on semi-parametric mixed models and their application. Confidence intervals are discussed in Section 5 and Section 6 is devoted to the application of the methods discussed. A general discussion winds up the tutorial in Section 7.

2. CARDIOVASCULAR SAFETY EXPERIMENT DATA EXAMPLE

In this section, we present a case study that will be used to illustrate application of the methods discussed subsequently. The data come from a cross-over study in a cardiovascular safety experiment carried out in dogs. A balanced Latin square Williams design of four experimental groups and four periods is used (Table I). Eight female beagle dogs, weighing between 10.0 and 12.9 kg, were implanted with a device for telemetric study. The animals were orally dosed with a vehicle or a compound, the latter at low, medium, and high doses, on four successive sessions, separated by a wash-out period of at least 3 days.

Table I ABOUT HERE

Cardiovascular parameters of interest were recorded at 5 minute intervals for 6 hours, hence 72 time points per subject per period are available. There are no missing data. One of the primary objectives of the study was to assess the effect of the compound on the so-called QT, a measure of the complete electrical activity of the heart's ventricle. A drug-induced prolongation of the

ventricular repolarization and a concomitant QT prolongation is known to be associated with lethal arrhythmias. Several other parameters are measured, and for purposes of this tutorial, the response, referred to as 'TAU,' is a measure of the relaxation capacity of the heart (in milliseconds) after contraction; it is a measure of how well or badly a heart relaxes after a contraction.

Table II ABOUT HERE

Table II lists the variables used for analysis in this paper. Data are stored is the SAS data file crossdata. Using the print procedure in SAS provides output, a selection of which looks like:

dog	period	tau	group	timeh	carry
1	1	15.6260	4	0.0833	1
1	1	16.8389	4	0.1667	1
1	1	17.2002	4	0.2500	1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
1	4	19.2730	2	5.8333	1
1	4	19.3113	2	5.9167	1
1	4	19.8975	2	6.0000	1

The construction of the carry-over variable `carry` is such that for a particular period, the variable takes the value of the treatment group given in the preceding period. The excerpt from the data reveals that treatment group 1 (control) was given to dog 1 in the third period. The data in the first period are consistently allocated to one of the treatment levels, bearing in mind that period effects

are always kept in the model. Note that the carry-over effect should be defined as a class variable. Obviously, there is no carry-over effect during the first period and, consequently, there should be no corresponding level for carry-over in that period. To circumvent this problem, we allocate the same level to all observations in each first period and inclusion of period effects ensures elimination of the carry-over in the first period [1].

The question of interest here is twofold. First, an overall measure of the differences amongst the treated groups (i.e., low, medium, and high dose levels) and the control group is required. Second, there is need to detect specific sections in the time window, if any, where the treated groups significantly differ from control. Hence, in line with detecting the possible cardiovascular effects of the compound on the response, it is necessary to examine whether the effect is in the earlier or later stages of the experiment.

In current practice, summary measures, such as the area under the curve (AUC), which reduce the problem to a conventional cross-over design, are often used. An obvious shortcoming of such an approach is loss of information. Another frequently used approach is to consider a full factorial structure for time by treatment and then to construct relevant contrasts at the required time points. However, this approach may only be feasible with relatively few time points since experiments with long sequences of repeated measurements yield excessively large numbers of parameters. Also, addressing multiple comparisons with such long sequences may not be straightforward and lead to severe losses in power.

Figure 1 ABOUT HERE

Figure 1 shows the observed mean profiles in the various experimental groups for each of the treatment periods. The treatment groups do not appear to differ; however, the response seems to increase with period. Note also that a parametric form for these mean profiles might not be easily

determined, hence the need to use more flexible, semi-parametric smoothing techniques.

We will contrast our semi-parametric analysis, presented in Section 4 with the AUC-based summary measure analysis, the topic of the next section.

3. ANALYSIS OF THE CARDIOVASCULAR SAFETY EXPERIMENT USING THE AUC AS SUMMARY STATISTIC

In this section, we discuss application of the area under the curve (AUC) as one way of summarizing data from a repeated measures cross-over design. As mentioned earlier, this may be seen as loss of information. Indeed, if the aim is to compare evolution over time across the experimental groups, such an approach is not useful. However, for an overall profile comparison, the AUC may sometimes be a viable option. As pointed out by Jones and Kenward [1], the approach makes few modeling assumptions about the joint behavior of the repeated measurements, making it robust. Also, given that in our situation the data are completely balanced, each subject provides approximately the same amount of information, a key assumption for the use of such a summary statistic [1]. Let us now focus on the model considered for the AUC summary statistic. As is usually done, to uphold the ubiquitous assumption of normally distributed errors, the model is based on a log transformation of the AUC. Let Y_{ijk} denote the log of AUC for animal i in period j, receiving experimental group k, for $i = 1, \ldots, n$, $j = 1, \ldots, p$, and $k = 1, \ldots, g$. Taking the last period ($j = 4$) and the control group ($k = 1$) as reference categories, define P_j, G_k, and C_j as indicator variables for period, treatment group, and carry-over respectively, such that, for example, $P_j = 1$ if period $= j$, and 0 otherwise, for all $j \leq p - 1$, with a similar definition for C_j and G_k and for $k = 2, 3, 4$. The model takes the form:

$$Y_{ijk} = \beta_0 + \alpha_j P_j + \tau_k G_k + \zeta_j C_j + U_i + \varepsilon_{ijk}, \tag{1}$$

where β_0 is an intercept, α_j is the effect associated with period j, τ_k is the effect associated with treatment group k, ζ_j is the carry-over effect in period j, U_i is the random intercept accounting for the correlation of observations from one subject, and ε_{ijk} is the random error term. Following [1], we do not include an interaction between period and treatment group. Such an interaction may emanate from subjects being affected by some factors other than treatment, and/or when the effect of a treatment level might depend on the current state of the subjects [19]. Needless to say that then the interpretation of results becomes difficult.

Let us give a brief description of how model (1) may be fitted using the SAS procedure MIXED. It is easy to compute the area under the curve for each subject within each treatment period using, for example, the trapezoid rule. We will denote the SAS variable, corresponding to the AUC, by `logauc`, stored together with all other relevant variables in the data set `AUCdata`. Note the presence of a single measurement per subject in each period. A PRINT procedure call in SAS produces the following output:

dog	period	group	logauc	carry
1	1	4	8.8158	1
1	2	3	8.8569	4
1	3	1	8.9353	3
1	4	2	8.8602	1
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
8	1	2	8.7573	1
8	2	1	8.7921	2

8	3	3	8.8956	1
8	4	4	8.8874	3

The model may then be fitted using the following SAS statements:

```
proc mixed data=AUCdata method=ml order=data asycov;

class period group dog carry;

model logauc = period group carry / ddfm=kr solution;

random dog;

contrast "high"    group  1 0 0 -1;

contrast "Medium"  group  0 1 0 -1;

contrast "low"     group  0 0 1 -1;

run;
```

The 'method=ml' option requests estimation to be based on maximum likelihood, rather than defaulting to REML and 'asycov' invokes printing of the estimated variance-covariance matrix for covariance parameters in the model. We make the maximum likelihood choice because some of our comparisons include models with different fixed-effects structures, precluding the use of restricted maximum likelihood (REML, [17]) as a basis for testing. Fixed effects in the model are reflected under the MODEL statement, while the RANDOM statement is used to define the dog-specific parameter U_i in (1). Assuming the group variable is sorted in descending order, the CONTRAST statements are used to compare each of the other treatment groups to the control. Running the model yields the following selected output, displaying tests for fixed effects and estimates of variance components:

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
period	3	24	37.31	<.0001
group	3	24	6.64	0.0020
carry	3	24.1	2.38	0.0950

Covariance Parameter Estimates

Cov Parm	Estimate
dog	0.004833
Residual	0.000370

We observe that period and treatment effect are significant, unlike carry-over. This is the more desirable situation. It is however important to note that the sample size ($n = 8$) under consideration is rather small, so it could introduce power issues when trying to detect carry-over effects.

4. ANALYSIS OF CROSS-OVER DESIGNS USING SEMI-PARAMETRIC MIXED MODELS

In the analysis presented in Section 3, the repeated measurements for each dog were summarized using the log of AUC as a summary statistic. However, the data fall within the realm of continuous longitudinal data and hence can be modeled by use of a linear mixed model [17]. A flexible route, situated within the framework of mixed models, is to model the mean with a semi-parametric smooth function, $f(t)$, which can be estimated, among others, with penalized splines. In Section 4.1, we give a brief review of penalized splines formulated as mixed models, while in Section 4.2 focus is on the inference problem for the study. Section 4.3 reviews penalized splines, specifically adapted to the cross-over setting.

4.1. Penalized Smoothing Splines and Mixed Models

Let Y_{ij} denote the response taken at time t_{ij} ($j = 1, \ldots, m$) on subject i ($i = 1, \ldots, n$). A penalized spline model, with a subject-specific random intercept U_i, based on a truncated line

basis, can be written as:

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + \sum_{v=1}^{K} b_v (t_{ij} - \kappa_v)_+ + U_i + \varepsilon_{ij} = f(t_{ij}) + U_i + \varepsilon_{ij}, \tag{2}$$

where $\kappa_1 < \kappa_2 < \cdots < \kappa_K$ are a set of distinct knots [18] in the range of t_{ij}, and $t_+ = \max(0, t)$, $b_v \sim N(0, \sigma_b^2)$ and $U_i \sim N(0, \sigma_U^2)$. The two sets of random effects b_v and U_i are assumed to be independent. The subject-specific random intercept accounts for the correlated nature of the observations. The truncated lines basis is simple in formulation and performs adequately in many circumstances [20] and is therefore a sensible choice. Using matrix notation [13], a stacked version of (2) becomes:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\varepsilon}. \tag{3}$$

The matrix $\boldsymbol{Z}$ contains the elements of the truncated line basis, as well as columns of ones for the random subject effect. Further, $\boldsymbol{b} = (b_1, \ldots, b_K, U_1, \ldots, U_n)'$. The correspondence between the penalized spline smoother and the optimal predictor in a mixed model framework, assuming normality for the b_k, is a key feature in fitting the models. This connection presents an opportunity for using standard mixed-models software, such as, for example, the **lme** library in S-plus or the MIXED procedure in SAS, to fit the penalized spline model. The method, with radial basis, is also routinely implemented in the SAS procedure GLIMMIX with radial basis. We give a brief description of the radial basis. For $q \geq 1$, let

$$\boldsymbol{Z}_{K_i} = \left[|x_{ij} - \kappa_v|^{2q-1} \right]_{1 \leq i \leq n,\ 1 \leq j \leq m,\ 1 \leq v \leq K}.$$

To overcome the problem of overfitting the data, the parameters in $\boldsymbol{b}_b = (b_1, \ldots, b_K)'$, coefficients for the knot points, are restricted by assuming that $\boldsymbol{b}_b$ is normally distributed with zero mean vector and variance-covariance matrix [9]

$$\text{Cov}(\boldsymbol{b}_b) = \sigma_b^2 (\boldsymbol{\Omega}_K)^{-1/2} (\boldsymbol{\Omega}_K^{-1/2})', \quad \text{where} \quad \boldsymbol{\Omega}_K = \left[|\kappa_v - \kappa_{v'}|^{2q-1} \right]_{1 \leq v, v' \leq K}.$$

Now, let $\boldsymbol{Z}_i^* = \boldsymbol{Z}_{K_i}\boldsymbol{\Omega}_K{}^{-1/2}$ and, as usual, construct the matrix $\boldsymbol{Z}$ by stacking the matrices $\boldsymbol{Z}_i^*$. This transformation then enables fitting the model using standard mixed-model software, resulting in the model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}_b, \quad \text{where} \quad \text{Cov}(\boldsymbol{b}_b) = \sigma_b^2 \boldsymbol{I}_K.$$

Note that the vector $\boldsymbol{b}_b$ here refers only to random effects for smoothing. Taking $q = 2$ results in the so-called cubic radial spline basis.

Fitting penalized splines by the linear mixed model approach has some appealing advantages, such as the automatic determination of the smoothing parameter, a unified framework for inference and ease to extend the models. In what follows, we proceed along these lines by formulating a series of hypothetical semi-parametric mixed models.

4.2. Semi-parametric Models for Mean Evolution

Let us, for illustrative purposes, consider a two-group parallel design. We are interested in investigating whether there is a difference between the two groups, that is, comparing the average profiles in the two groups. Further, we intend to investigate which specific sections of the profiles exhibit significant differences. The semi-parametric model discussed in Section 4 implies that the mean response for each treatment group can be represented by an additive model of two components, a linear component and a smooth component. Let the group-specific mean profiles for two groups A and B be denoted by $f_A(t)$ and $f_B(t)$, respectively. To test whether the group-specific mean profiles are equal, without loss of generality, we formulate the hypotheses:

$$H_0 : f_A(t) = f_B(t), \qquad H_1 : f_A(t) \neq f_B(t). \tag{4}$$

The null hypothesis obviously implies a common mean for both groups. Figure 2 illustrates, with hypothetical examples, several possible scenarios related to the evolution of the means over time.

In all examples, the mean is a sum of a linear part and a smooth part. In panel A, the two groups have the same mean, implying that the null hypothesis in (4) is enforced. Panel B allows for a pattern in which the means of the two groups differ only by a constant, while in panel C the groups are different in the linear part but the smooth component of the mean is identical. Finally, panel D reveals a pattern in which the means of the two groups have different evolutions over time and the groups are different in both the linear and smooth parts. Note that the formulation of our models implies, in certain instances, more complicated versions of the null hypothesis in (4), since testing for both fixed effects and variance components may be involved. In what follows, we formulate linear mixed models within the cross-over setting, following each of the scenarios illustrated in Figure 2.

FIGURE 2 ABOUT HERE

4.3. Formulation of the Models for the Cross-over Design

This section focuses on formulation of possible models which can be used to describe the data at hand. The model with a full factorial structure for treatment and time as in [1] is a sensible starting point. We show how one can move from this very general model to more parsimonious models, based on describing the time evolution through use of penalized smoothing splines. The formulation of the models is similar in spirit as in [14]. Using appropriately constructed matrices, all models given in this section can be represented using the matrix notation of Section 4.1. For the desired flexibility, 40 equally spaced knots, selected as quantiles of the time variable, are used [9, 18]. Following [9], the 40 knots κ_k, which are separated by approximately 9 minutes, occupy the $\left(\frac{k+1}{K+2}\right)$ th quantiles of unique values of time.

Models 1–6 in Section 4.3.1 are used to model the cross-over aspects of the experiment, as

in (3) where $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{I})$. In Section 4.4, the same set of models are considered focusing on decomposing the covariance matrix of $\boldsymbol{\varepsilon}$ into components of serial correlation and measurement error. All models are fitted using the SAS procedure MIXED, by means of maximum likelihood techniques. An illustration is given in Section 4.3.1.

4.3.1. Modeling the Cross-over Aspect of the Design

Model 1. Let $Y_{ijk\ell}$ denote the measurement on subject i, in period j, corresponding to treatment group k at time point ℓ, for $i = 1, \ldots, n$; $j = 1, \ldots, p$; $k = 1, \ldots, g$; and $\ell = 1, \ldots, m$. Define t_ℓ as an indicator variable for time, such that $t_\ell = 1$ indicates that time is ℓ and 0 otherwise, for $\ell \leq m - 1$. Consider a model with a full factorial structure for treatment group and time [1], expressed as:

$$Y_{ijk\ell} = \beta_0 + \alpha_j P_j + \tau_k G_k + \lambda_\ell t_\ell + \gamma_{kl} G_k t_\ell + \psi_{jl} P_j t_\ell + \zeta_j C_j + U_i + \varepsilon_{ijk\ell}. \tag{5}$$

The parameter λ_ℓ refers to the effect of time, $\gamma_{k\ell}$ denotes the interaction between treatment group and time, $\psi_{j\ell}$ is the interaction between period and time, and $\varepsilon_{ijk\ell}$ are random terms.

Often, $\Sigma_i = \text{Cov}(\boldsymbol{\varepsilon}_i)$ is assumed to be $\sigma_\varepsilon^2 \boldsymbol{I}$, resulting in a conditional independence model. A more general residual covariance structure, for example decomposing the vector $\boldsymbol{\varepsilon}_i$ into components of serial correlation, $\boldsymbol{\varepsilon}_{(1)i}$, and measurement error, $\boldsymbol{\varepsilon}_{(2)i}$, can be considered [17] and will be discussed in Section 4.4.

Given that for each subject, 72 measurements in each period are taken, Model 1 is bound to yield a large number of parameters, hence the need for parsimony. The following models, adjusting for possible period effects, show different possible approaches to modeling the time evolution in the experimental groups. Using the SAS procedure MIXED, the mean structure is specified with the MODEL statement while between-subject heterogeneity is taken into account by specifying a

subject-specific random intercept in the RANDOM statement. The REPEATED statement is used to specify the covariance structure for the random error; for Model 1, this is $\boldsymbol{\varepsilon}_i \sim N(0, \sigma_{\boldsymbol{\varepsilon}}^2 I)$. The complete code for the model is as follows:

```
*SAS CODE FOR MODEL 1

proc mixed data=crossuse method=ml order=data asycov;

class period group dog time carry;

model tau = period group time group*time period*time carry / ddfm=kr;

random intercept/ subject=dog solution;

repeated time/type=simple subject=dog(period);

run;
```

Model 2. In Model 2, it is assumed that the time evolution is the same in all treatment groups (see Figure 2, panel A). As such, smoothing of the time trend occurs at the highest level of the model, thereby ignoring the treatment groups. The model can be represented as:

$$Y_{ij\ell} = \alpha_j P_j + \beta_0 + \beta_1 t_{ij\ell} + \sum_{v=1}^{K} b_v (t_{ij\ell} - \kappa_v)_+ + U_i + \zeta_j C_j + \varepsilon_{ijk\ell}, \tag{6}$$

where κ_v are knots, and the coefficients b_v are common to all treatment groups, such that Var$(b_v) = \sigma_b^2$. Model 2 can be expressed in matrix notation by adopting the following notation: $\boldsymbol{Y} = (Y_{ij\ell})_{i,j,\ell}$ and $\boldsymbol{X} = (1, t_{ij\ell}, P_1, P_2, P_3, C_1, C_2, C_3)_{i,j,\ell}$. Further, define, for each subject, a smoothing matrix

$$\boldsymbol{Z}_{\boldsymbol{b}i} = ((t_{ij\ell} - \kappa_k)_+)_{1 \le \kappa \le K}\,, \quad \text{with stacked version} \quad \boldsymbol{Z}_{\boldsymbol{b}} = \begin{pmatrix} \boldsymbol{Z}_{b1} \\ \boldsymbol{Z}_{b2} \\ \vdots \\ \boldsymbol{Z}_{bn} \end{pmatrix}. \tag{7}$$

The SAS code for Model 2 is shown in the panel below. The MODEL statement is used to specify the fixed terms in (6), i.e., period, carry-over effects and the linear trend over time.

```
*SAS CODE FOR MODEL 2

proc mixed data=crossuse method=ml;

class period group dog time carry;

model tau = period timeh carry / solution ddfm=kr;

random Z1-Z40 / type=toep(1) solution;

random intercept / subject=dog solution;

repeated time / type=simple subject=dog(period);

run;
```

Since there are two variance components in the model, σ_b^2 and σ_U^2, two RANDOM statements are used. The first one is used to specify the non-parametric part $\boldsymbol{Z_b}$. Our application uses truncated lines, and therefore one needs to construct the matrix $\boldsymbol{Z_b}$, with columns $Z_1, \ldots, Z_K$, and denoted (Z1-Z40 in the application). Example SAS macros for achieving this can be found in [9] and will not be discussed here. In particular, the option 'type=toep(1)' in this RANDOM statement implies that the covariance matrix of b_v has a $K \times K$ diagonal Toeplitz structure given by

$$\begin{pmatrix} \sigma_b^2 & 0 & \ldots & 0 \\ 0 & \sigma_b^2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & \sigma_b^2 & 0 \\ 0 & \ldots & \ldots & \sigma_b^2 \end{pmatrix} = \sigma_b^2 \boldsymbol{I}_{K \times K}$$

The second RANDOM statement is used, as in Model 1, to specify the random intercept for each subject. Note that, in case additional subject-specific parameters are considered, the second RANDOM statement ought to be modified accordingly. The REPEATED statement is used to

specify the covariance structure for the random error. The option 'type=simple' implies that $\varepsilon_{ijk\ell} \sim N(0, \sigma_\varepsilon^2)$, i.e., Cov$(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \boldsymbol{I}$. The parameter estimates for the variance components in the model are given in the panel below, where `Variance`, `Intercept`, and `Time` are the parameter estimates for σ_b^2, σ_U^2, and σ_ε^2, respectively.

```
*Covariance Parameters: Model 2

Cov Parm      Subject        Estimate

Variance                       1.0889

Intercept     dog              2.1466

Time          dog(Period)      1.2473
```

Model 3. Model 3 assumes that the underlying linear trends in the treatment groups differ by a shift only. However, the same non-parametric part is fitted to all treatment groups. This model assumes the difference amongst the treatment groups, if present, does not depend on time. A penalized spline representation of the model is:

$$Y_{ijk\ell} = \alpha_j P_j + \tau_k G_k + \beta_0 + \beta_1 t_{ij\ell} + \sum_{v=1}^{K} b_v (t_{ij\ell} - \kappa_v)_+ + U_i + \zeta_j C_j + \varepsilon_{ijk\ell}. \tag{8}$$

This scenario corresponds to panel B of Figure 2. Compared with Model 2, the current model has additional fixed effects parameters, τ_k. Hence, the MODEL statement includes the fixed effect `group`. Note that the covariance structure is the same as in Model 2 and hence the two random statements remain unchanged.

Model 4. Here, it is assumed that the linear parts of the models differ, while the same smooth part is considered for all groups. This resembles a scenario where, relative to Model 3, are tilted at some angle, such that treatment effect is no longer constant in time. A representation of such a

model is:

$$Y_{ijk\ell} = \alpha_j P_j + \tau_k G_k + \beta_0 + (\beta_1 + \beta_{1k} G_k) t_{ik\ell} + \sum_{v=1}^{K} b_v (t_{ij\ell} - \kappa_v)_+ + U_i + \zeta_j C_j + \varepsilon_{ijk\ell}, \quad (9)$$

with Var$(b_v) = \sigma_b^2$. Panel C of Figure 2 graphically illustrates such a scenario. The MODEL statement of Model 4 includes the interaction term `group*timeh`, which allows for different slopes for different groups.

```
model tau = period group timeh group*timeh / solution ddfm=kr;
```

Similarly as in Model 3, the two random statements remain the same.

Model 5. All models considered so far assume that the same smooth component is fitted to the different treatment groups. It is possible to go one step further and fit a model with different non-parametric parts of the model in the different treatment groups, although the same smoothing parameter is used. The linear parts of the models are assumed different and, although the random effects are assumed independent from group to group, a single parameter is used to smooth the groups. A representation of such a model is:

$$Y_{ijk\ell} = \alpha_j P_j + \tau_k G_k + \beta_0 + (\beta_1 + \beta_{1k} G_k) t_{ik\ell} + \sum_{v=1}^{K} b_{kv} (t_{ik\ell} - \kappa_v)_+ + U_i + \zeta_j C_j + \varepsilon_{ijk\ell}. \quad (10)$$

Note that part of the design matrix, $\boldsymbol{Z_b}$, corresponding to smoothing, is now block-diagonal with each diagonal entry corresponding to a particular treatment group and the coefficients for the truncated lines basis, b_{kv}, are now group-specific with Var$(b_{kv}) = \sigma_b^2$. The smoothing matrix is now given by:

$$\boldsymbol{Z_b} = \begin{pmatrix} \boldsymbol{Z}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Z}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{Z}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{Z}_4 \end{pmatrix},$$

where $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_4$ are group-specific smoothing matrices each constructed by stacking the $\boldsymbol{Z}_{bi}$ as in (7).

This situation is similar to the illustration in panel D of Figure 2. To define the block-diagonal matrix for $\boldsymbol{Z}_b$, the first random statement is changed to

```
random Z1-Z40 / type=toep(1) subject=group;
```

where the 'subject=group' option implies that $\boldsymbol{Z}_b$ is a block-diagonal matrix with the number of blocks equal to the number of groups. Complete code for Model 5 is given below.

```
*SAS CODE FOR MODEL 5

proc mixed data=crossuse method=ml order=data asycov;

class period group dog time carry;

model tau = period group timeh group*timeh carry / solution ddfm=kr;

random z1-z40 / type=toep(1) subject=group;

random intercept / subject=dog solution;

repeated time / type=simple subject=dog(period);

run;
```

Model 6. A further step is to relax the assumption on the smoothing parameter and to assume that the groups can be smoothed separately but with different smoothing parameters. Hence, both the fixed effects part and the non-parametric part differ by group and four variance components corresponding to smoothing the different treatment groups are estimated. The penalized spline representation of this model and the $\boldsymbol{Z}$ matrix is the same as in Model 5, with $\text{Var}(b_{kv}) = \sigma^2_{kb}$.

The covariance matrix pertaining to smoothing is given by

$$\begin{pmatrix} \sigma_{1b}^2 \boldsymbol{I}_{K\times K} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_{2b}^2 \boldsymbol{I}_{K\times K} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_{3b}^2 \boldsymbol{I}_{K\times K} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma_{4b}^2 \boldsymbol{I}_{K\times K} \end{pmatrix}.$$

The reference panel for this model is panel D, as in the previous model, since the difference between these models cannot be seen graphically. To define a group-specific random component for $\boldsymbol{b}$, one needs to replace the option 'subject=group' with option 'group=group' in the RANDOM statement, as follows:

```
random Z1-Z40 / type=toep(1) group=group;
```

Note that, this change excepting, the specification of Models 5 and 6 in SAS is identical.

```
*SAS CODE FOR MODEL 6

proc mixed data=crossuse method=ml order=data asycov;

class period group dog time carry;

model tau = period group timeh group*timeh carry / ddfm=kr;

random Z1-Z40 / type=toep(1) group=group;

random intercept / subject=dog solution;

repeated time / type=simple subject=dog(period);

run;
```

Based on Akaike's Information Criterion (AIC) [21, 22], a comparison of models presented in this section (see also Section 6) suggests that Model 4 be chosen as the current best model. The covariance parameter estimates for Models 4, 5, and 6 are given below, where `Variance` corresponds to the smoother, `Intercept` to the random intercept and `Time` to the residual errors.

Note that, for Model 5, all groups share the same variance component for σ_b^2, and for Model 6, `group` in the `Subject` column indicates that each group has a different vector of random effects, since $\boldsymbol{Z}_b$ is a block-diagonal matrix.

```
*Covariance Parameters: Model 4

Cov Parm      Subject        Estimate

Variance                       1.1383

Intercept     dog              2.1463

Time          dog(Period)      1.1523


*Covariance Parameters: Model 5

Cov Parm      Subject        Estimate

Variance      group            0.7384

Intercept     dog              2.1428

Time          dog(Period)      1.1338


*Covariance Parameters: Model 6

Cov Parm      Subject        Group      Estimate

Variance                     group 1      0.4898

Variance                     group 2      0.7575

Variance                     group 3      2.2180

Variance                     group 4      0.3784

Intercept     dog                         2.1427

Time          dog(Period)                 1.1308
```


4.4. Modeling the Covariance Structure

Two types of covariance structures for measurements from a particular subject need to be accounted for in the analysis. First, the correlations amongst measurements across different treatment periods and, second, dependencies amongst measurements within one treatment period. Assuming that the covariances applying to one period are similar to those in other periods, the between- and within-period covariance structures are separable [1]. As such, accommodation of between-period dependencies can be achieved by introducing the subject-specific random intercepts U_i. Commonly used models for repeated measures covariance structures, such as, for example, an AR(1) process, can be used to model the remaining within-period dependencies. In particular, we consider the decomposition of the residual variance into components of serial correlation and measurement error, such that:

$$\mathsf{Cov}(\boldsymbol{\varepsilon}_i) = \sigma_\varepsilon^2 \boldsymbol{I} + \tau^2 \boldsymbol{H}_i, \tag{11}$$

where elements of $\boldsymbol{H}_i$, the serial correlation matrix, are modeled by some parametric function, particular cases of which are the exponential and Gaussian functions [17]. Let the variance of the serial process be denoted by τ^2 and the rate of decay of correlations with distance $d_{\ell\ell'}$ between time point ℓ and ℓ' by θ. Table III shows the forms of the covariance structure we consider for within-period residual covariance.

Table III ABOUT HERE

As mentioned before, owing to the length of the sequences of measurements per subject, one would expect the residuals to be serially correlated. To gain insight into this phenomenon, we fit an unstructured mean model that includes other fixed effects, like period and the necessary interactions, and assess the behavior of the residuals. Note that, at this stage, neither random

effects are included nor covariance structure is modeled. Denote the residuals at time point ℓ by $r_{\ell v}$, $v = 1, \ldots, 32$. At each of the 72 distinct time points, there are 32 observations. Figure 3 shows a plot of r_{1v} on the horizontal axis against $r_{\ell' v}$ on the vertical axis, with $\ell' = 2, 6, 10, \ldots, 62$, a selection of time points.

Figure 3 ABOUT HERE

It is clear from the residual plots that, after removing the mean structure, the residuals do not appear independent and, as expected, the dependencies tend to weaken with distance in time. As such, conditional independence models as in Section 4.3.1 may not be appropriate in this case, hence the modeling of serial correlation.

Models with correlated errors may be obtained by modifying the REPEATED statement of the models presented in Section 4.3.1. The functional form of the correlation is specified in the '`type=`' option. For an AR(1) process, the statement is modified to:

```
repeated time / type=ar(1) subject=dog(period);
```

The output for the covariance parameters, corresponding to this statement, is:

Cov Parm	Subject	Estimate
Variance		0.6051
Intercept	dog	2.0993
AR(1)	dog(Period)	0.6840
Residual		1.1770

From this output, `Residual` refers to λ^2, the variance, and `AR(1)` represents θ, the rate of decay of the correlations, as in Table III. The estimated residual variance function is $\widehat{\lambda^2}\widehat{\theta}^{|\ell-\ell'|} =$

$1.177 \times 0.6840^{|\ell-\ell'|}$. To fit a model with exponential-type serial correlation, the statement is altered to

```
repeated time /type=sp(exp)(timeh) local subject=dog(period);
```

Note that `time` is a class variable whilst `timeh` is continuous. The 'local' option requests a split of the residual variance into elements of serial correlation and additional measurement error. The estimated covariance parameters are as follows:

Cov Parm	Subject	Estimate
Variance		0.8778
Intercept	dog	2.0772
Variance	dog(Period)	1.0312
SP(EXP)	dog(Period)	0.3527
Residual		0.1553

Here, the `Variance` for `dog(Period)` corresponds to λ^2 and 'SP(EXP)' corresponds to θ in Table III. The estimate of the residual variance, σ_ε^2 in (11) is given under `Residual`. The estimate for the covariance function in (11) is $\widehat{\sigma}_\varepsilon^2 \boldsymbol{I} + \widehat{\lambda}^2 \widehat{\boldsymbol{H}}_i = 0.1553\boldsymbol{I} + 1.0312\widehat{\boldsymbol{H}}_i$, where $\widehat{\boldsymbol{H}}_{i\ell\ell'} = \exp(-d_{\ell\ell'}/0.3527)$. Finally, for the Gaussian type of serial correlation, the following output is obtained:

Cov Parm	Subject	Estimate
Variance		0.5097
Intercept	dog	2.1044
Variance	dog(Period)	0.7844


```
SP(GAU)        dog(Period)      0.2656

Residual                        0.3259
```

The estimate for θ is now given under `SP(GAU)`, resulting from the statement:

```
repeated time /type=sp(gau)(timeh) local subject=dog(period);
```

Similarly, the estimated residual covariance function is $\widehat{\sigma}_{\varepsilon}^2 \boldsymbol{I} + \widehat{\lambda}^2 \widehat{\boldsymbol{H}}_i = 0.3259\boldsymbol{I} + 0.7844\widehat{\boldsymbol{H}}_i$, where $\widehat{\boldsymbol{H}}_{i\ell\ell'} = \exp(-d_{\ell\ell'}/0.2656^2)$.

5. POINTWISE CONFIDENCE INTERVALS AND SIMULTANEOUS CONFIDENCE BANDS

This section focuses on the construction of confidence bands around the group-specific fitted functions. Such bands can be used to compare the different treatment groups at specific time points, if necessary. Ruppert *et al.* [9] give details for constructing such intervals or bands for smoothed functions. Our intention is to adapt their results so as to accommodate the correlation structure, as accounted for by the random intercept, and the residual covariances allowing for presence of serial correlation. Consider the penalized spline model expressed in the mixed model form, as in Section 4, re-expressed as:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}_b\boldsymbol{b}_b + \overbrace{\boldsymbol{Z}_U\boldsymbol{b}_U + \boldsymbol{\varepsilon}}^{\boldsymbol{\varepsilon}^*},$$

such that,

$$\text{Cov}(\boldsymbol{\varepsilon}_i^*) = \sigma_U^2\boldsymbol{J} + \boldsymbol{\Sigma}_i = \boldsymbol{R}_i^*, \tag{12}$$

where $\boldsymbol{Z}_b$ and $\boldsymbol{Z}_U$ are matrices corresponding to smoothing and random intercepts, respectively, and $\boldsymbol{J}$ is a matrix of ones. Construction of pointwise confidence intervals, as well as simultaneous

confidence bands, requires the covariance for the vector of contrasts between the estimated and true parameters for the fixed and random effects, such that [9]:

$$\text{Cov}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widehat{\boldsymbol{b}}_b - \boldsymbol{b}_b \end{pmatrix} \simeq \left(\boldsymbol{C}^T \widehat{\boldsymbol{R}}^{-1} \boldsymbol{C} + \widehat{\boldsymbol{B}}\right)^{-1}, \tag{13}$$

where $\boldsymbol{C}$ is a design matrix containing linear time effects and truncated line basis, $\widehat{\boldsymbol{R}}$ is the residual covariance, and $\widehat{\boldsymbol{B}}$ is a matrix constructed from variance components corresponding to smoothing. Consider a random-intercept model in a parallel design with balanced longitudinal data consisting of n subjects, each with m repeated measurements. Assuming independent residual errors, the required covariance matrix takes the form [14]:

$$\left(\boldsymbol{C}^T \widehat{\boldsymbol{R}}^{-1} \boldsymbol{C} + \widehat{\boldsymbol{B}}\right)^{-1} = \widehat{\sigma}_\varepsilon^2 \left[\sum_{i=1}^{n} \left\{ \boldsymbol{C}_i^T \left(\boldsymbol{I}_{m\times m} - \frac{\widehat{\sigma}_{b_0}^2}{\widehat{\sigma}_\varepsilon^2 + m\widehat{\sigma}_U^2} \boldsymbol{J}_{m\times m} \right) \boldsymbol{C}_i \right\} + \frac{\widehat{\sigma}_\varepsilon^2}{\widehat{\sigma}_b^2} \boldsymbol{D} \right]^{-1}, \tag{14}$$

where $\boldsymbol{D} = \text{diag}(0, 0, 1, \ldots, 1)$. The simultaneous confidence bands are based on simulations, assuming a multivariate normal distribution for the vector of contrasts between the estimated and true parameters for both fixed and random effects,

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widehat{\boldsymbol{b}}_b - \boldsymbol{b}_b \end{pmatrix} \sim \boldsymbol{N}\left\{\boldsymbol{0}, \left(\boldsymbol{C}^T \widehat{\boldsymbol{R}}^{-1} \boldsymbol{C} + \widehat{\boldsymbol{B}}\right)^{-1}\right\}. \tag{15}$$

For details on how the simulations from (15) are used to construct the bands, we refer to [9]. Note, the generality of (13) makes it applicable in different settings. For example, expression (14) follows specifically from a random-intercept model with independent errors, implying a compound-symmetry structure for $\widehat{\boldsymbol{R}}$, which is readily invertible. The following section focuses on how the preceding developments can be extended to the particular case of a cross-over design.

5.1. Adaptation of the Confidence Bands to the Cross-over Setting

Consider a random-intercept model as discussed in Section 4. Further, assume the model also includes both components of serial correlation and measurement error in the residual covariance

structure. The resulting $\widehat{\boldsymbol{R}}$ implies a simplified version of (13), but may not be straightforwardly obtained. However, the matrix can still be used in its most general form. For illustrative purposes, consider a particular dog i. Assuming an exponential type of serial correlation, the part of Cov($\boldsymbol{\varepsilon}_i^*$) in the first period corresponding to the first three observations is given by:

$$\mathbf{R}^*_{i[3]} = \begin{pmatrix} \sigma_U^2 & \sigma_U^2 & \sigma_U^2 \\ \sigma_U^2 & \sigma_U^2 & \sigma_U^2 \\ \sigma_U^2 & \sigma_U^2 & \sigma_U^2 \end{pmatrix} + \tau^2 \begin{pmatrix} 1 & \exp(-d_{12}/\theta) & \exp(-d_{13}/\theta) \\ \exp(-d_{12}/\theta) & 1 & \exp(-d_{23}/\theta) \\ \exp(-d_{13}/\theta) & \exp(-d_{23}/\theta) & 1 \end{pmatrix}$$

$$+ \begin{pmatrix} \sigma_\varepsilon^2 & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & \sigma_\varepsilon^2 \end{pmatrix}, \tag{16}$$

where τ^2 is the variance for the serial correlation part, θ is the rate of decay of the correlations as a function of $d_{\ell\ell'}$, which is the Euclidean distance between coordinates of the time variable, and σ_ε^2 is the variance of the measurement error. Therefore, $\boldsymbol{R}_i^*$ consists of matrices of the form (16) as diagonal elements, corresponding to each period, and matrices of the form $\sigma_U^2\boldsymbol{J}$ in the off diagonal entries. Now, replacing $\widehat{\boldsymbol{R}}$ with $\widehat{\boldsymbol{R}}^*$ in (13), where $\widehat{\boldsymbol{R}}^*$ is block-diagonal with diagonal elements $\widehat{\boldsymbol{R}}_i^*$, provides the expression which can used to construct either pointwise or simultaneous confidence bands. Focusing on Model 4, estimates of σ_U^2, τ^2, θ, and σ_ε^2 are as given in Section 4.4.

5.2. SAS Macro for Calculating Simultaneous Confidence Bands

Implementing the confidence bands involves use of (15) and (16). For this purpose, required output data sets are kept using the ODS OUTPUT statement, as in Section 4.3.1. Calculation of the confidence intervals/bands is implemented in a SAS macro, the usage of which is briefly explained. The macro essentially uses covariance parameter estimates obtained in Section 4.4 as input to

compute (16). Generation of vectors from a multivariate normal distribution, see (15), is done using a separate SAS macro, included within the macro, by the call

```
%inc(<location of the file containing the MVN macro>)
```

The general call of the macro, which calculates pointwise and simultaneous bands has the form

```
%macro CBmacro(data=,model=,resp=,time=,subject=,group=,
period=,carry=,K=,sctype=,nsim=);
```

The data for analysis is specified under `data`. Our macro fits Models 2–6 and choice of the required model is given under MODEL. The option of either exponential or Gaussian serial correlation is specified in `sctype`, which takes values `exp` or `gau`. The response and the 'time' variable are passed on via `resp` and `time`, respectively. One has to also provide the number of knot points `K`, and the required number of simulations (`nsim`) for construction of simultaneous confidence bands. Finally, it is imperative to indicate from the data set to be analyzed, which variables represent the subject of analysis, the treatment groups, carry-over effect and the treatment period.

The macro CBmacro, available from the authors, outputs panels with tests for fixed effects and covariance parameter estimates. Further, a data set containing group-specific fitted values together with 95% pointwise and simultaneous confidence bands is created by the macro. As an example, a call for Model 3 takes the form

```
%macro CBmacro(data=crossuse,model=3,resp=tau,time=time,subject=dog,
group=group, period=period,carry=carry,K=40,sctype=exp,nsim=10000);
```

and the predicted values and confidence intervals may be obtained from the output data set `myout` exemplified in the following panel, which may be useful in producing plots like Figure 4.

TIME	PRED	GROUP	SE	L95	U95	MALPHA	LS95	US95
0.08333	19.7630	1	0.54270	18.6993	20.8267	2.4179	18.4508	21.0753
0.16667	19.8702	1	0.53545	18.8207	20.9197	2.4179	18.5755	21.1649
0.25000	19.9633	1	0.53304	18.9186	21.0081	2.4179	18.6744	21.2522
0.33333	20.0632	1	0.53252	19.0194	21.1069	2.4179	18.7755	21.3508
0.41667	20.1696	1	0.53297	19.1250	21.2142	2.4179	18.8809	21.4583
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.

6. RESULTS

This section focuses on the application of the methodology discussed in the previous sections. Emphasis will be on the semi-parametric mixed model approach. Within the semi-parametric models, a comparison of models assuming independent residual errors with models assuming some form of residual correlation structure will be undertaken. In addition, we briefly discuss the issue of carry-over where, loosely, the observed response in one period could be the result of the effect of the previously allocated treatment. Regardless of the wash-out period used in the study, carry-over effects are not unexpected. For models that do not require preliminary testing for carry-over, we refer to [7]. The approach we take is to select one of the semi-parametric models with independent errors, based on Akaike's Information Criterion (AIC) [21, 22]. The selected model will then be improved by modeling the covariance structure and all further inferences will be based on the model

selected based on comparing covariance structures. The results from such a model are compared with the summary statistics analysis.

6.1. Model Fitting, Selection, and Hypotheses Testing

Let us now focus on fitting and selection of the models discussed in Section 4.3.1. For each of the models, the independent errors structure, a commonly used approach in practice, is assumed and results given in Table IV.

Table IV: ABOUT HERE

The exploratory comparison of the models with independent residual errors appears to indicate Model 4, with differing linear effects by group and the same non-parametric component is a plausible starting point. A formal likelihood ratio test (LRT) can be performed to see if indeed there is need to move from Model 3 to Model 4. Such a test, based on a χ^2_3, and a LRT statistic of $29.6 = 6922.9 - 6893.3$ yields a highly significant result ($p = 0.0001$). Hence, a model with different linear effects by group fits better. Note, the test between Model 3 and 4 is based on fixed effects only; no variance components are involved.

Different ways of modeling the residual covariance are applied and the results in Table IV indicate a substantial improvement in the fit of Model 4 upon modeling of the residual covariance structure. In particular, the model with exponential-type serial correlation appears to fit better than other models. However, the group-by-time interaction is now insignificant, implying Model 3. Note that, the group-by-time interaction is the characteristic separating Models 3 and 4. Thus a formal test between both models would be based on χ^2_3, with a LRT statistic of $3.2 = 5521.6 - 5518.4$ and $p = 0.3618$, corroborating that a constant difference in time suffices in this situation. Henceforth, Model 3, with exponential serial correlation, becomes our chosen model and any further inferences

will be based upon this model. The selection of Model 3 already suggests presence of treatment effect. However, a formal test may be required, and that translates to testing the chosen model against Model 2 (the null model). The hypothesis of interest then becomes, following (6 and (8): $H_0 : \tau_k = 0, k = 2, 3, 4$. The null model was fitted under the various covariance structures, such that appropriate comparisons of fixed effects against any of the (chosen) alternative models may be effectuated. In this case, a test between Models 3 and 2, both under the exponential serial correlation would be appropriate. Again, there are no variance components equated to zero in this test and therefore it is based on χ^2_3. The LRT statistic is 22.5, which is significant ($p < 0.0001$), hence groups differ.

We present the fixed-effects parameter estimates for Model 3, p-values from the associated t-tests, and variance components in Table V. For the sake of comparison, we have included parameter estimates from the model with independent residual errors.

Table V ABOUT HERE

Focusing on the model with serial correlation in Table V, it can be observed that only the medium dose group differs from the control group ($p = 0.0014$). Let us focus on comparing this model with the model assuming independent errors. While parameter estimates do not change much, it is the standard errors that substantially change, rendering some previously significant effects insignificant, such as, for example, the difference between the low and high doses. This highlights the problem of underestimating variability, often ignored when models such as the conditional independence model are applied in practice.

Table V also gives results from the AUC analysis. Although the time dimension is lost in this analysis, an overall comparison amongst the doses can be salvaged. Note the parameter estimates for the AUC are not directly comparable to those from the semi-parametric mixed models since they are

on the log scale. Similar to the conclusion made above, the results indicate a significant difference between the medium dose and the control group, albeit with weakened evidence ($p = 0.0129$). Hence, the animals receiving the medium dose group tend to have higher values of the measure of relaxation capacity of the heart than the control group, and the difference is constant over time.

A common issue with cross-over designs is carry-over [1] and [19]. To investigate how carry-over may have influenced our results, we re-consider Model 3 with serial correlation and, in addition, introduce carry-over effects. The results for parameters of interest from this model are also given in Table V. Although carry-over (parameter estimates not given) appears significant ($p = 0.0390$), no major changes are obtained in parameter estimates, their standard errors, or the conclusions reached previously.

6.2. Confidence Intervals and/or Bands

The final model emerging as the 'best' explicitly suggests that treatment effect is constant over time, implying that for this situation time point comparisons are redundant. We proceed to construct the confidence bands around the fitted profiles. This endeavor requires estimation of the variance components pertaining to smoothing and random intercept, as well as the parameters associated with the exponential serial correlation. Note that smoothing is performed at the highest level of the model, and hence, one variance component is estimated for all four treatment groups. Following the discussion in Section 5, using the estimates in Table V, one can construct the pointwise as well as simultaneous confidence bands. Constructing simultaneous confidence bands involves estimating a value $m_{(1-\alpha)}$, which would replace $Z_{(1-\frac{\alpha}{2})}$ usually used in constructing confidence intervals under the normal distribution assumption [9].

Figure 4 ABOUT HERE

Five independent simulations of 10,000 draws each, using results discussed in Section 5 are performed. We obtain $m_{0.95} \approx 2.4587$, 2.4603, 2.4593, 2.4669 and 2.4128. The minimum of these values, i.e., 2.4128, can be taken as the estimate of $m_{0.95}$, implying that the simultaneous confidence bands are about $2.4128/1.96 = 1.23$ times wider than their pointwise counterparts. Figure 4 show the group-specific fitted profiles and the corresponding 95% confidence bands. The model appears to fit well. The simultaneous confidence bands constructed enable one to make joint statements about the profiles' evolution in time. In case the model featured treatment effect changing over time, such intervals could be used, for example, to compare each of the active treatment groups to the control and each time point.

7. DISCUSSION

We have exemplified the flexibility of penalized smoothing-splines methodology, within the linear mixed model framework, to the context of cross-over designs. We have illustrated that one can formulate different possible scenarios, showing how the different treatment groups could possibly differ. Such an approach then enables one to select the model deemed 'best' according to some criterion. Amongst the models considered, an interesting comparison is between Model 4, with an overall smooth part, and Model 5, having independent random effects by group, with the same variance component. Although in Model 5 the random effects responsible for smoothing are independent and vary from group to group, only a single smoothing parameter is estimated as in Model 4. The two models are not the same. However, from a parametric point of view, they contain the same number of parameters. It is not clear how formal tests between such models may be performed, hence the use of information criteria, such as AIC.

Particular attention has also been given to models including serial correlation, wherein time-

honored functions, such as exponential and Gaussian, have been considered. Indeed, with relatively long sequences of repeated measurements within a period, residual correlations are expected. As we have seen, ignoring such correlations can possibly lead to misleading results. It is worth mentioning that the flexible models considered here for the mean can be considered to model the serial correlation as well, something that has received limited attention thus far in the literature.

Often, researchers require comparisons of treatment groups at specific time points. This could possibly be done by fitting a full factorial structure in time and compare groups using appropriate contrasts. However, the large number of time points involved here makes such an approach prohibitive. An attractive alternative is the use of confidence intervals and/or bands, constructed around the fitted profiles. Once a suitable model has been selected, confidence bands can then be constructed around the fitted functions. Focus has been on the adaptation of the confidence intervals and bands of Ruppert *et al.* [9] for application in the specific cross-over design situation. Using the confidence bands, one is able to identify specific sections where the bands do not overlap, indicating significant differences. Other than overcoming the disadvantages of the full factorial structure approach mentioned above, the problem of multiple comparison is also elegantly solved here. Note that, as mentioned before, the model we have focused on does not warrantee use of confidence bands for time point comparisons, since treatment effect is constant over time.

In summary, we have demonstrated the use of penalized splines to data from a repeated measures cross-over design. Particular attention was on comparing models with different mean structures as well as various covariance structures. Further, we touched upon the construction of simulation-based simultaneous confidence bands for the fitted profiles.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support from the IAP research network no. P5/24 of the Belgian Government (Belgian Science Policy) and from the Institute for the Promotion of Innovation by Science and Technology (IWT) in Flanders, Belgium.

REFERENCES

1. Jones B, Kenward MG. *Design and Analysis of Cross-Over Trials*. Chapman & Hall/CRC, 2003.
2. Wallenstein S, Fisher AC. The analysis of the two-period repeated measurements cross-over design with application to clinical trials. *Biometrics* 1977; **333**: 251–269.
3. Patel HI, Hearne EM. An application of multivariate analysis to the two-period repeated measures crossover design with application to clinical trials. *Communications in Statistics - Theory and Methods.* 1980; **9**: 1919–1929.
4. Dunsmore IR. Growth curves in two-period change over models. *Applied Statistics* 1981; **30**: 223–229.
5. Grender JM, Johnson WD. Fitting multivariate polynomial growth curves in two period crossover designs. *Statistics in Medicine* 1994; **13**: 359–365.
6. Fearn T. A Bayesian approach to growth curves. *Biometrika* 1975; **62**: 89–100.
7. Putt M, Chinchilli VM. A mixed effects model for the analysis of repeated measures cross-over studies. *Statistics in Medicine* 1999; **18**: 3037–3059.
8. Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science* 1996; **11**: 89–121.
9. Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge University Press, 2003.
10. Zeger SL, Diggle PJ. Semiparametric models for longitudinal data with application to cd4 cell numbers in hiv seroconverters. *Biometrics* 1994; **50**: 689–699.
11. Verbyla AP, Cullis BR, Kenward MG, Welman SJ. The analysis of designed experiments and longitudinal data by using smoothing splines. *Applied Statistics* 1999; **48**:269–311.
12. Guo W. Functional mixed effects models. *Biometrics* 2002; **58**: 121–128.
13. Durbán M, Harezlak J, Wand MP, Carroll RJ. Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine* 2005; **24**: 1153–1168.
14. Maringwa JT, Geys H, Shkedy Z, Faes C, Molenberghs G, Aerts M, Van Ammel K, Teisman A, Bijnens L. Application of Semi-parametric Mixed Models and Simultaneous Confidence Bands in a Cardiovascular Safety Experiment with Longitudinal Data. In: *Proceedings of the 21st International Workshop on Statistical Modelling*

2006, Hinde J, Einbeck J, Newell J. (eds). Galway, 2006; 346–353.

15. Lin X, Zhang D. Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society B* 1999; **61**: 384–400.
16. Wood S. On confidence intervals for GAMs based on penalized regression splines. *Tech. Report*. Dept. Statistics, University of Glasgow, 2004.
17. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal data.* New York: Springer, 2000.
18. Ruppert D. Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics* 2002; **11**: 735–757.
19. Senn S. *Crossover Trials in Medical Research*. John Wiley & Sons, 1993.
20. Ngo L, Wand MP. Smoothing with mixed model software. *Journal of Statistical Software* 2004; **9**:1–56.
21. Akaike H. Maximum likelihood identification to Gaussian autoregressive moving average models. *Biometrika* 1973; **60**: 255–265.
22. Burnham KP, Anderson DR. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer, 2002.

Table I. Williams design for a cross-over study with four dose groups, control ($C \equiv 1$), low ($L \equiv 2$), medium ($M \equiv 3$), and high ($H \equiv 4$). The design is replicated twice, resulting in a total of 8 animals being used.

	Period			
Subject	1	2	3	4
1	H	M	C	L
2	M	L	H	C
3	L	C	M	H
4	C	H	L	M

Table II. Meaning of the variables used in the SAS programs.

Variable	Explanation
`tau`	The response variable
`logauc`	Logarithm of the area under curve for Tau
`group`	Class variable of experimental group
`period`	Class variable indicating treatment period
`dog`	Class variable for subject indicator
`timeh`	Time in hours (continuous variable)
`time`	Duplicate of timeh for use as a class variable
`Carry`	Class variable for carry-over effect
`Z1-Z40`	Columns of $\boldsymbol{Z}$ matrix for smoothing

Table III. Various covariance structures for modeling residual covariance within periods.

	Serial correlation	
AR(1)	Gaussian (GSC)	Exponential (ESC)
$\tau^2\theta^{d_{\ell\ell'}}$	$\tau^2\exp\left(-d_{\ell\ell'}/\theta^2\right)$	$\tau^2\exp\left(-d_{\ell\ell'}/\theta\right)$

Table IV. Minus twice loglikelihood values and AIC values for models in Section 4.3.1, assuming independent residual errors. For the null model (Model 2) as well as the model with the smallest AIC (Model 4), different covariance structures are considered.

Within-period cov.	Crit.	Mod. 1	Mod. 2	Mod. 3	Mod. 4	Mod. 5	Mod. 6
Independent	-2 loglik	9025.3	7128.7	6922.9	6893.3	6941.8	6938.3
	AIC	10041.3	7150.7	6950.9	$\boxed{6927.3}$	6975.8	6978.3
AR(1)	-2 loglik		5590.7	5557.1	5552.5		
	AIC		5614.7	5587.1	5588.5		
Gaussian ser. corr.	-2 loglik		5650.4	5600.2	5593.9		
	AIC		5670.4	5632.2	5631.9		
Exp. ser. corr.	-2 loglik		5544.1	5521.6	5518.4		
	AIC		5570.1	$\boxed{5553.6}$	5556.4		

Table V. Fixed-effects parameter estimates, standard errors (s.e) and p-values for associated t-tests corresponding to Model 3 with independent errors within periods, exponential serial correlation (ESC) and Model 3 with exponential serial correlation plus carry-over effects. Also included are estimates, standard errors and p-values for the AUC analysis. The last period and the control group are taken as reference categories.

		Model 3 (Independent errors)		Model 3 (ESC)		Model 3 (ESC+carry-over)		AUC	
Effect	Par.	Est. (s.e.)	p	Est. (s.e.)	p	Est. (s.e.)	p	Est. (s.e.)	p
Fixed Effects:									
Intercept	β_0	20.679(0.567)	0.0001	20.767(0.587)	0.0001	20.654(0.596)	0.0001	8.930(0.027)	0.0001
Period 1	α_1	-2.122(0.079)	0.0001	-2.267(0.179)	0.0001	-2.082(0.216)	0.0001	-0.103(0.012)	0.0001
Period 2	α_2	-1.693(0.063)	0.0001	-1.646(0.179)	0.0001	-1.649(0.171)	0.0001	-0.080(0.010)	0.0001
Period 3	α_3	-0.569(0.063)	0.0001	-0.550(0.179)	0.0027	-0.551(0.171)	0.0017	-0.026(0.010)	0.0138
High	τ_4	0.232(0.066)	0.0012	0.164(0.179)	0.3594	0.197(0.179)	0.2746	0.010(0.010)	0.3539
Medium	τ_3	0.758(0.066)	0.0001	0.585(0.179)	0.0014	0.729(0.179)	0.0001	0.036(0.010)	0.0015
Low	τ_2	-0.136(0.066)	0.0383	-0.119(0.179)	0.5052	-0.10(0.179)	0.5555	-0.005(0.010)	0.6274
Time	β_1	1.206(1.184)	0.3212	1.353(1.064)	0.2236	1.350(1.069)	0.2260		
Variance components:									
$\mathrm{Var}(b_{vk})$	σ_b^2	1.151(0.526)		0.888(0.546)		0.903(0.500)			
$\mathrm{Var}(U_i)$	σ_U^2	2.134(1.069)		2.088(1.060)		2.056(1.053)		0.005(0.003)	
Exponential serial correlation:									
$\mathrm{Var}(\varepsilon_{(1)ijkl})$	τ^2			1.078(0.072)		1.144(0.073)			
Rate of decay	θ			0.386(0.043)		0.431(0.040)			
Measurement error variance:									
$\mathrm{Var}(\varepsilon_{(2)ijkl})$	σ_ε^2	1.129(0.034)		0.163(0.023)		0.172(0.024)			

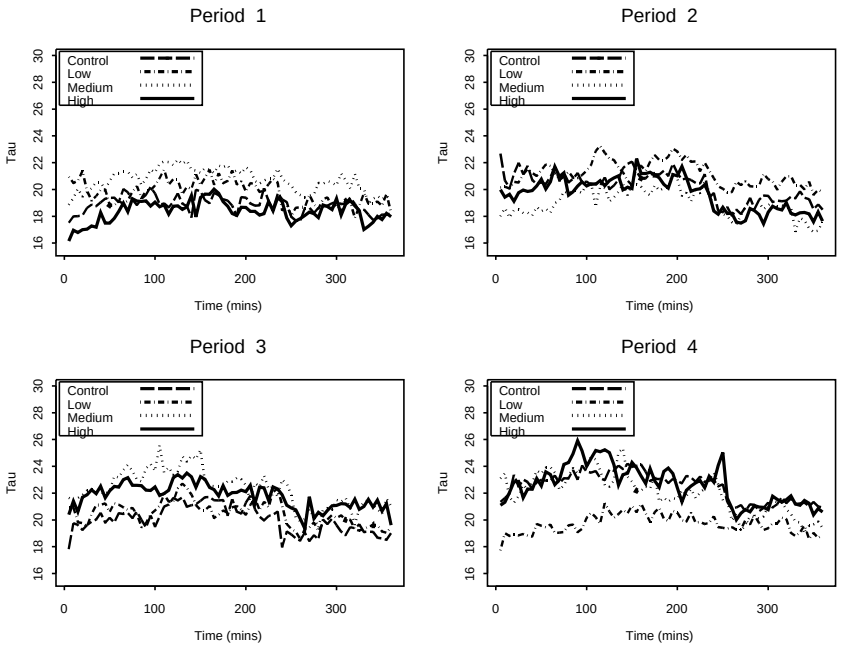

Figure 1. Observed mean profiles for each period and experimental group.

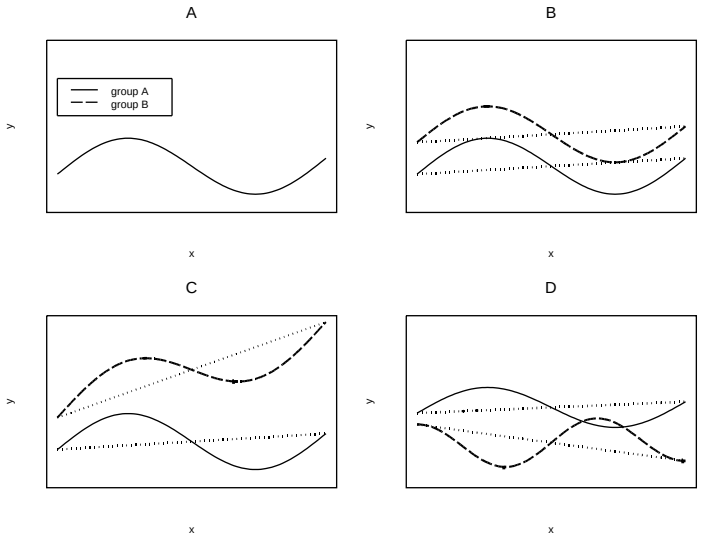

Figure 2. Hypothetical examples of the semi-parametric models showing the linear and non-parametric parts of the model. The models illustrate how the group-specific mean functions could possibly evolve over time.

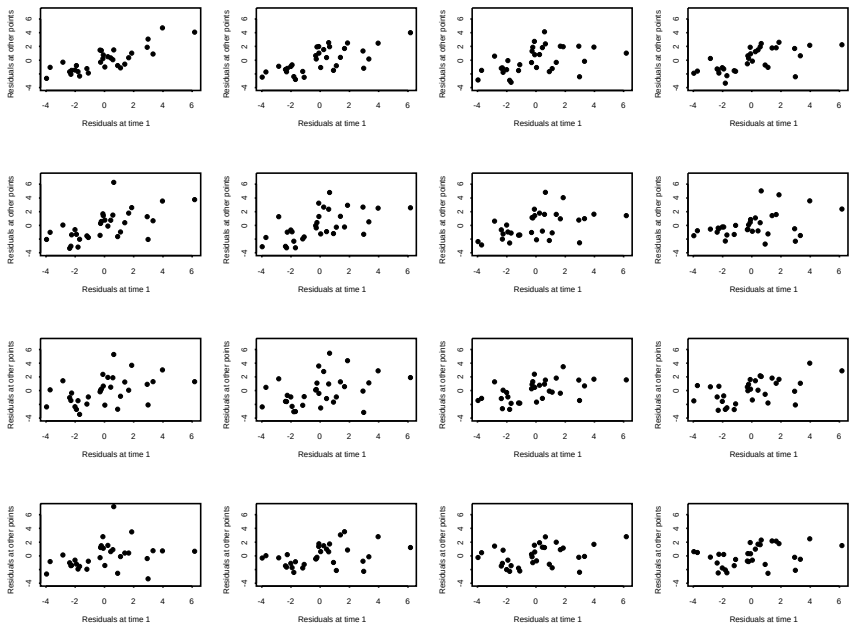

Figure 3. Residuals at first time point plotted against residuals from other 16 selected time points (2, 6, 10,. . . , 62).

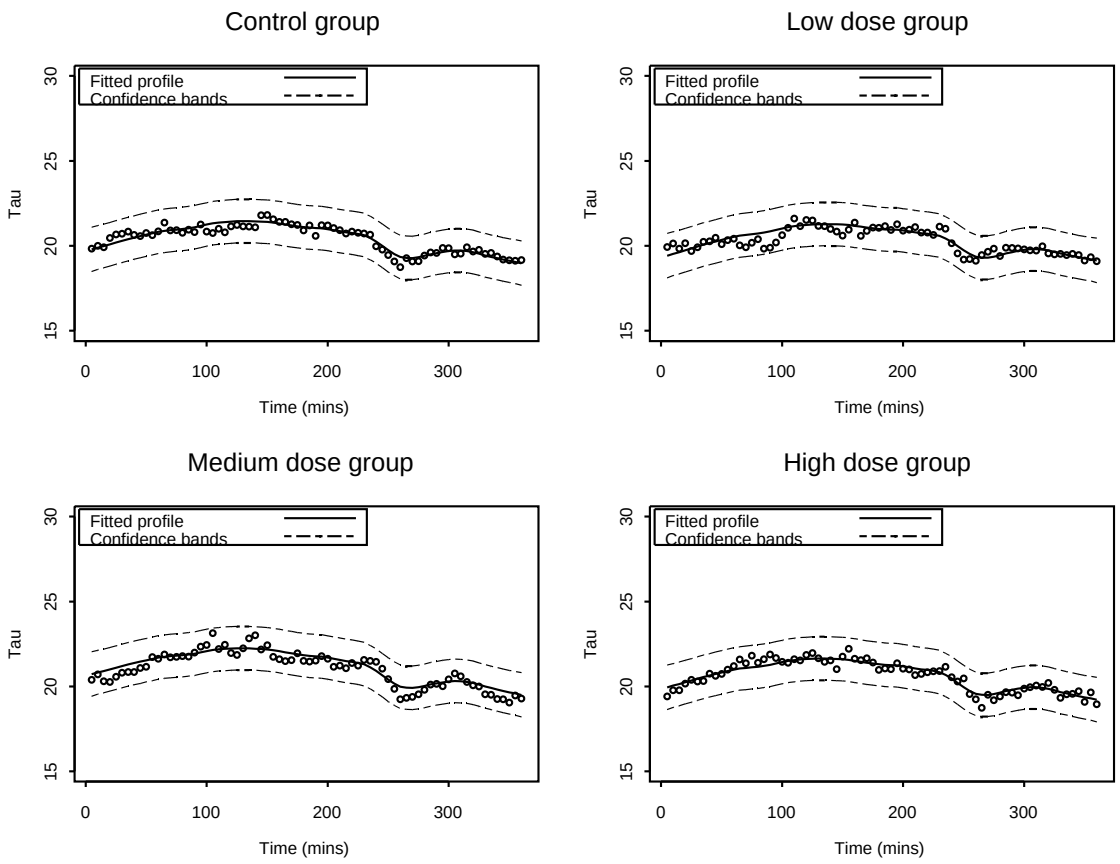

Figure 4. Group-specific fitted profiles together with the corresponding 95% simultaneous confidence bands around them. The points are observed mean values at each time point.

