Estimation and Inference for a Spline-Enhanced Population Pharmacokinetic Model

Lang Li,[1,*] Morton B. Brown,[2] Kyung-Hoon Lee,[3] and Suneel Gupta[4]

[1]Division of Biostatistics, Indiana University, Indianapolis, Indiana 46202, U.S.A.

[2]Department of Biostatistics, The University of Michigan, Ann Arbor, Michigan 48109, U.S.A.

[3]Department of Pharmacology, College of Medicine, Sung Kyun Kwan University,
Suwon, Kyunggi-do, 440-746, Korea

[4]ALZA Corporation, Clinical Pharmacology, Mountain View, California 94039, U.S.A.

*email: lali@iupui.edu

SUMMARY. This article is motivated by an application where subjects were dosed three times with the same drug and the drug concentration profiles appeared to be the lowest after the third dose. One possible explanation is that the pharmacokinetic (PK) parameters vary over time. Therefore, we consider population PK models with time-varying PK parameters. These time-varying PK parameters are modeled by natural cubic spline functions in the ordinary differential equations. Mean parameters, variance components, and smoothing parameters are jointly estimated by maximizing the double penalized log likelihood. Mean functions and their derivatives are obtained by the numerical solution of ordinary differential equations. The interpretation of PK parameters in the model and its flexibility are discussed. The proposed methods are illustrated by application to the data that motivated this article. The model's performance is evaluated through simulation.

KEY WORDS: Double penalized log-likelihood function; Multicompartment models; Natural cubic spline; Runge–Kutta method.

1. Introduction

The main purpose of a clinical pharmacokinetic (PK) study is to understand the pharmacokinetics of a drug, in particular, the absorption phase and the disposition phase of an orally administrated drug in the human body, where the disposition phase includes two subprocesses, elimination and distribution. The drug's pharmacokinetics is described by multicompartment models.

Two-compartment models for the amounts of drug in the gut, blood, and tissue are standard in pharmacokinetic studies. In our motivating example, there is strong evidence that a two-compartment model with time-varying rates was considered, but its parameters proved to be essentially nonidentifiable. As an alternative, we consider a one-compartment model combining the blood and tissue compartments of the two-compartment model. We show that a one-compartment model with time-varying rates, which we call the SEPK model (spline enhanced pharmacokinetic model), can mimic the behavior of a two-compartment model but with fewer parameters. The SEPK model fits our example well and all parameters can be estimated reasonably accurately.

We now provide some background. Let us assume that subject i takes an oral dose and the pharmacokinetics follow a one-compartment model (Figure 1a). At time 0, all the drug is in the gut. Let $A_{1,i}(t)$ and $A_{2,i}(t)$ represent the amount of drug in the gut and the rest of the human body at time t, respectively, with initial value $\{A_{1,i}(0), A_{2,i}(0)\} = \{1, 0\}$. The absorption rate $k_{a,i}$ and the elimination rate $k_{e,i}$ for subject i are assumed to be constant over time. The one-compartment model is

$$\frac{dA_{1,i}(t)}{dt} = -k_{a,i}A_{1,i}(t),$$
$$\frac{dA_{2,i}(t)}{dt} = k_{a,i}A_{1,i}(t) - k_{e,i}A_{2,i}(t). \tag{1}$$

In a two-compartment model (Figure 1b) following oral administration, let $A_{1,i}(t)$, $A_{2,i}(t)$, $A_{3,i}(t)$ represent the amounts of drug in the gut, blood, and tissue at time t, respectively, with initial value $\{A_{1,i}(0), A_{2,i}(0), A_{3,i}(0)\} = \{1, 0, 0\}$. Unlike the one-compartment model, the two-compartment model has both elimination and distribution phases in the drug disposition. We assume that all parameters $(k_{a,i}, k_{e,i}, k_{23,i}, k_{32,i})$ are constant over time, where $k_{23,i}$ and $k_{32,i}$ are the distribution rates for subject i. The two-compartment model is

$$\frac{dA_{1,i}(t)}{dt} = -k_{a,i}A_{1,i}(t),$$
$$\frac{dA_{2,i}(t)}{dt} = k_{a,i}A_{1,i}(t) - (k_{e,i} + k_{23,i})A_{2,i}(t) + k_{32,i}A_{3,i}(t),$$
$$\frac{dA_{3,i}(t)}{dt} = k_{23,i}A_{2,i}(t) - k_{32,i}A_{3,i}(t). \tag{2}$$

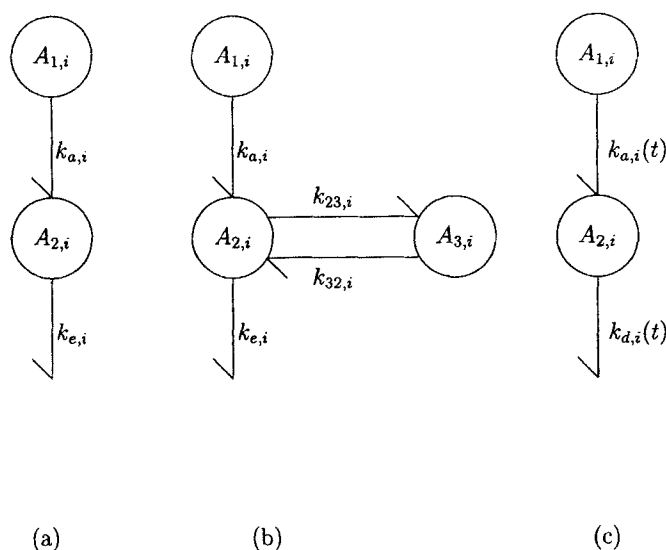(a)   (b)   (c)

Figure 1. Pharmacokinetic models. **a.** One-compartment model. **b.** Two-compartment model. **c.** Spline-enhanced PK model.

In a population PK study, we observe drug plasma concentrations with some noise for each of $i = 1, \ldots, m$ subjects. The concentration for subject i is $A_{2,i}(t)/V_i$, where V_i is the subject-specific volume of distribution. The concentration measurement is usually log transformed to render the measurement error homogeneous. Let the true log concentration be $f(\beta_i, t) = \log\{A_{2,i}(t)/V_i\}$, where β_i is a PK parameter vector for subject i. For example, $\beta_i = (V_i, k_{a,i}, k_{e,i})^T$ when the PK model is the one-compartment model (1). The classical population PK models are $y_{ij} = f(\beta_i, t_{ij}) + \epsilon_{ij}$ and $\log(\beta_i) = X_i\alpha + Z_ib_i$, where y_{ij} is the observed log plasma drug-concentration for subject i at time t_{ij} and ϵ_{ij} is the noise. In particular, X_i and Z_i are design matrices for the fixed effect α and random effect b_i, respectively, and represent baseline covariates for subject i. These matrices are time independent because $f(\beta_i, t)$ is derived from multicompartment models with a constant PK parameter assumption. This model is a nonlinear mixed effects model (NLMM) (Davidian and Giltinan, 1995). Estimation and inference procedures for the model, such as first-order linearization (Sheiner and Beal, 1980) and conditional first-order linearization (Lindstrom and Bates, 1990), assume that the subject-specific PK parameters are linearly related to the covariates.

If a drug follows a constant PK parameter assumption and the drug is administrated repeatedly (multiple doses), the concentration should superimpose each time, i.e., the overall concentration profile for multiple doses is equivalent to the summation of the profiles from the individual doses. This is referred to as the superposition principle. This principle is important in establishing the bioequivalence (BE) between a controlled-release drug formulation and an immediate-release drug formulation, as illustrated in the example below. Violation of the superposition principle means that the drug does not follow the constant PK parameter assumption.

Figure 2a is a plot of S-oxybutynin plasma concentration (on the log scale) from 40 patients in a Ditropan (ALZA Corporation) study with three doses and 8 hours per dose.

S-oxybutynin is a major active component of Ditropan. It is commonly prescribed to treat symptoms of urge incontinence, urgency, and frequency of urination arising from overactivity of the detrusor muscle. Following oral administration of S-oxybutynin, absorption from the gut is rapid, and it appears to be metabolized by cytochrome P450 3A4 in the liver and the gut wall. Based on Douchamps et al. (1988), S-oxybutynin has two-compartment kinetics. There are two orally administrated drug formulations: immediate-release product and controlled-release product. Immediate-release S-oxybutynin must be given frequently (three times a day), while the controlled-release drug has the advantage of convenience, i.e., only one dose a day. This PK study was designed to evaluate the bioequivalence (BE) between two drug formulations. In this article, we do not discuss the BE analysis but rather the implication of this BE study. Each formulation of the drug can have different dose levels. FDA guidelines grant waivers for higher dose BE studies if BE can be demonstrated at a lower dose level, provided that the drug satisfies several requirements, one of which is the superposition principle. Hence, verification of this assumption is crucial in order to save time and money during the drug development.

We focus on the immediate-release drug here. In the study, it was administrated three times at 8-hour intervals, and 25 plasma concentrations were sampled in a 2-day time period for each subject. The sampling time points are 0.5, 1, 1.5, 2, 3, 5, 8, 8.5, 9, 9.5, 10, 11, 13, 16, 16.5, 17, 17.5, 18, 19, 21, 24, 27, 30, 36, and 48 hours after the initial dose. If the superposition principle holds, we expect the three peak concentrations ($Cmax$) to be nondecreasing, and similarly for the $AUC_{(0,8)}$, where $AUC_{(0,8)}$ represents the area under the plasma concentration curve from time 0 to 8 hours after each dose. We performed an initial noncompartment analysis (PK analysis without a model assumption). The subject–dose-specific $\{AUC_{(0,8),i1}, AUC_{(0,8),i2}, AUC_{(0,8),i3}\}$ are calculated using the trapezoidal rule (Rowland and Tozer, 1995, p. 469), where the final subscript denotes dosing interval. The subject–dose-specific $\{Cmax_{i1}, Cmax_{i2}, Cmax_{i3}\}$ are the maximum concentrations within each dosing interval. Based on Figure 2b and 2c, the $AUC_{(0,8)}$ and $Cmax$ for dose 3 appear to be lower than those of dose 1 and 2. The paired Wilcoxon signed-rank one-sided test suggests that there is strong evidence against the null hypotheses of $AUC_{(0,8)}$ for dose 3 being equal to or higher than those of dose 1 or 2. The p-values are 0.003 and 0.00001, respectively. The average AUC ratio between dose 3 and 1 is 0.92 with standard error 0.03, and the ratio is 0.83 with standard error 0.03 between dose 3 and 2. $Cmax$ follows the same pattern, and the p-values based on the paired Wilcoxon signed-rank one-sided test are 0.00001 and 0.00001, respectively. The average $Cmax$ ratio between dose 3 and 1 is 0.81 with standard error 0.05, and the ratio is 0.75 with standard error 0.06 between dose 3 and 2. Hence, neither the superposition principle nor the constant-PK parameter assumption hold for this drug. In another PK study, Sathyan, Chancellor, and Gupta (2001) presented a similar finding.

An immediate therapeutic consequence of time-dependent PK is its effects on the $Cmax$ and AUC because both of them are important criteria to evaluate the drug's efficacy. For example, the $Cmax$ of S-oxybutynin can be much lower after several days or weeks of repeated oral administration
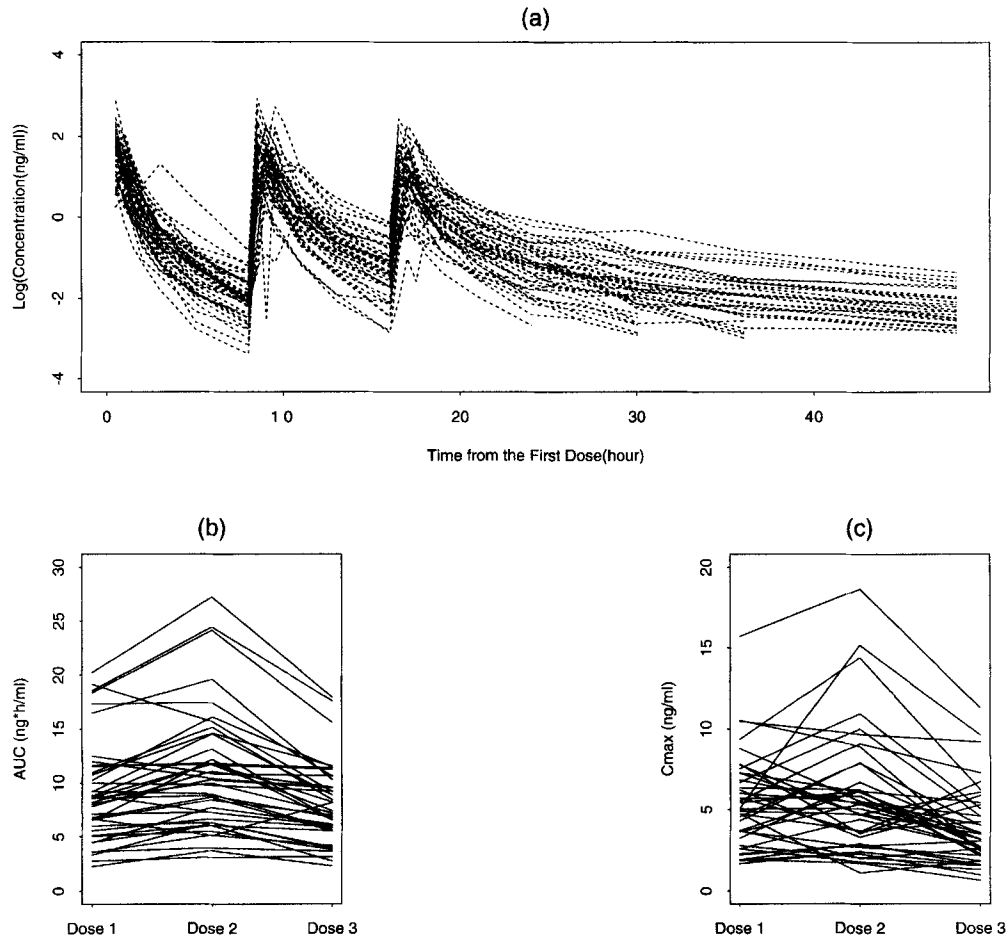
Figure 2. Observed oxybutynin concentration, AUC and $Cmax$. **a.** Observed oxybutynin concentration. **b.** Subject–dose-specific $AUC_{(0,8)}$. **c.** Subject–dose-specific $Cmax$.

than that of the first dose, based on the results of this study. There are two ways in which we can try to maintain the $Cmax$ at the target level. One way is to modify the dose levels for the multiple doses, and the other is to change the drug's kinetics. However, without a proper model, it is not possible to understand the time-dependent kinetics, thus, neither approach is feasible. The multicompartment models are not proper models as they are based on a constant-PK parameter assumption. Thus, we need a modified model.

In order to search for the evidence of time-dependent PK parameters during drug absorption and disposition phases, let the model be (Figure 1c)

$$\frac{dA_{1,i}(t)}{dt} = -k_{a,i}(t)A_{1,i}(t),$$

$$\frac{dA_{2,i}(t)}{dt} = k_{a,i}(t)A_{1,i}(t) - k_{d,i}(t)A_{2,i}(t), \qquad (3)$$

where $k_{a,i}(t)$ is an absorption process function and $k_{d,i}(t)$ is a disposition process function for subject i. Unlike multicompartment models, we use two arbitrary functions instead of a few parameters to describe the absorption and the disposition phases of the drug. Practically, the disposition process

continues all the way through the PK study, while the absorption process usually finishes very quickly and the plasma samples later on in the study are not informative with respect to $k_{a,i}(t)$. Hence, $k_{a,i}(t)$ can only be modeled as a step function with a few steps, and $k_{d,i}(t)$ can be modeled as a natural cubic spline function. We call this model a spline-enhanced PK model. An alternative way is to model some PK parameters in the two-compartment model (2) as arbitrary functions of time, e.g., $\{k_{a,i}(t), k_{e,i}(t), k_{23,i}(t)\}$. Apparently it has a better interpretation than (3) does because it has both elimination and distribution processes while model (3) cannot distinguish them in $k_{d,i}$. However, they are not statistically identifiable, and we experienced extremely high collinearity during the estimation. One reason is that both $k_{e,i}(t)$ and $k_{23,i}(t)$ tend to explain the same drug disposition process if either one of them is allowed to be an arbitrary function of time. Hence, we focus on model (3) in this article.

There are statistical, numerical, and pharmacokinetic challenges in fitting and interpreting this model. In Section 2, model specification, estimation, and inference are developed; it allows subject-specific PK parameters to be time-dependent and the nonparametric functions are incorporated directly into the differential equations. Models (1) and (2) have

constant PK parameter assumptions, while (3) does not. Although (3) is more flexible than (1) and (2), it is necessary to clarify the connections among model (1), (2), and (3) in order to establish the interpretation of the parameters in model (3). Section 3 discusses the interpretation and performance of model (3) when the true models are (1) and (2). Data analysis is presented in Section 4, simulation results are described in Section 5, and conclusions are presented in Section 6.

2. Statistical Method

2.1 *Model Specification*

We choose the natural cubic spline to model the nonparametric function because the smoothing parameter can be treated as a variance component and can be jointly estimated with the mean parameters. Zhang et al. (1998) and Lin and Zhang (1999) successfully implemented this approach in the linear additive mixed models and the generalized linear additive mixed models (GAMM), respectively.

Let us start with ordinary differential equations (3) for subject i and rewrite them as

$$\frac{d\mathbf{A}_i(t)}{dt} = \mathbf{G}\{\boldsymbol{\phi}_i(t)\}\mathbf{A}_i(t), \tag{4}$$

where $\mathbf{A}_i(t) = \{A_{1,i}(t), A_{2,i}(t)\}^{\mathrm{T}}$, $A_{2,i}(t)$ is the amount of drug in the blood at time t and the true plasma drug concentration is $A_{2,i}(t)/V_i$. $\mathbf{G}$ is a 2×2 matrix and $\boldsymbol{\phi}_i(t) = \{k_{a,i}(t), k_{d,i}(t)\}^{\mathrm{T}}$ is a subject-specific time-dependent PK vector.

Because the measurement error of the drug concentration often has a constant coefficient of variation, a log transformation can make the variance homogeneous. Davidian and Carroll (1987) provided detailed discussion of nonconstant variance. Let

$$\boldsymbol{\beta}_i(t) = \left\{\boldsymbol{\phi}_i^{\mathrm{T}}(t), V_i\right\}^{\mathrm{T}} = \{k_{a,i}(t), k_{d,i}(t), V_i\}^{\mathrm{T}},$$

and let

$$f\{\boldsymbol{\beta}_i(t), t\} = \log[A_{2,i}\{\boldsymbol{\phi}_i(t), t\}/V_i]$$

(we use $A_{2,i}\{\boldsymbol{\phi}_i(t), (t)\}$ and $A_{2,i}(t)$ interchangeably for the convenience, and they represent the same function). Although V_i can also be treated as $V_i(t)$ in general, there is no scientific reason to believe that it is time dependent. Hence, the spline-enhanced PK model (SEPK) is

$$y_{ij} = f\{\boldsymbol{\beta}_i(t_{ij}), t_{ij}\} + \epsilon_{ij}, \tag{5}$$
$$\log\{\boldsymbol{\beta}_i(t_{ij})\} = \mathbf{X}_{ij}\boldsymbol{\alpha} + \mathbf{S}_{ij}g(t_{ij}) + \mathbf{Z}_{ij}\mathbf{b}_i,$$
$$= \mathbf{X}_{ij}\boldsymbol{\alpha} + \mathbf{N}_{ij}\mathbf{g} + \mathbf{Z}_{ij}\mathbf{b}_i, \tag{6}$$

where y_{ij} is the observed log plasma drug concentration for subject i at time t_{ij}. The first-stage model (5) describes the relationship between subject–time-specific PK parameters $\boldsymbol{\beta}_i(t_{ij})$ and log drug concentration. The function $f\{\boldsymbol{\beta}_i(t_{ij}), t_{ij}\}$ can be evaluated through a numerical solution of (4) (Section 2.3). The second-stage model (6) describes a semiparametric relationship between subject–time-specific parameters and time-dependent covariates. It may be expressed more explicitly for (3) as

$$\log(V_i) = \mathbf{X}_{1,i}\boldsymbol{\alpha}_1 + \mathbf{Z}_{1,i}\mathbf{b}_{1i},$$
$$\log\{k_{a,i}(t_{ij})\} = \mathbf{X}_{2,ij}\boldsymbol{\alpha}_2 + \mathbf{Z}_{2,ij}\mathbf{b}_{2i},$$

$$\log\{k_{d,i}(t_{ij})\} = \mathbf{X}_{3,ij}\boldsymbol{\alpha}_3 + g(t_{ij}) + \mathbf{Z}_{3,ij}\mathbf{b}_{3i},$$
$$= \mathbf{X}_{3,ij}\boldsymbol{\alpha}_3 + \mathbf{N}_{3,ij}\mathbf{g} + \mathbf{Z}_{3,ij}\mathbf{b}_{3i},$$

where $\mathbf{X}_{ij} = \mathrm{diag}(\mathbf{X}_{1,i}, \mathbf{X}_{2,ij}, \mathbf{X}_{3,ij})$ is a $3 \times p$-dimensional time-dependent design matrix for the p-dimensional fixed effect vector $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^{\mathrm{T}}, \boldsymbol{\alpha}_2^{\mathrm{T}}, \boldsymbol{\alpha}_3^{\mathrm{T}})^{\mathrm{T}}$, $\mathbf{Z}_{ij} = \mathrm{diag}(\mathbf{Z}_{1,i}, \mathbf{Z}_{2,ij}, \mathbf{Z}_{3,ij})$ is a $3 \times q_i$-dimensional time-dependent design matrix for the q_i-dimensional random effect vector $\mathbf{b}_i = (\mathbf{b}_{1i}^{\mathrm{T}}, \mathbf{b}_{2i}^{\mathrm{T}}, \mathbf{b}_{3i}^{\mathrm{T}})^{\mathrm{T}}$, ϵ_{ij} is i.i.d. $N(0, \sigma_0^2)$, and $\mathbf{b}_i$ is i.i.d. $N(\mathbf{0}, \mathbf{D}_i)$. $\mathbf{S}_{ij} = (0, 0, 1)^{\mathrm{T}}$ in (6) is an indicator matrix for the spline function $g(t_{ij})$. Let l be the total number of distinct sampling time points t_{ij}s, $\mathbf{g}$ be an $l \times 1$ vector of the function g at those points, and $\mathbf{N}_{ij} = (\mathbf{0}, \mathbf{0}, \mathbf{N}_{3,ij}^{\mathrm{T}})^{\mathrm{T}}$ be a $3 \times l$ indicator matrix for $\mathbf{g}$. Note that we use only one spline function in model (6). More spline functions can be similarly defined. When $t_k \leq t < t_{k+1}$, $g(t)$ has the following expression:

$$g(t) = \frac{(t - t_k)g_{k+1} + (t_{k+1} - t)g_k}{t_{k+1} - t_k}$$
$$- \frac{(t - t_k)(t_{k+1} - t)}{6}$$
$$\times \left\{\left(1 + \frac{t - t_k}{t_{k+1} - t_k}\right)g_{k+1}'' + \left(1 + \frac{t_{k+1} - t}{t_{k+1} - t_k}\right)g_k''\right\},$$
$$g_k = g(t_k),$$
$$g_k'' = g''(t_k), \qquad k = 1, \ldots, l - 1. \tag{7}$$

We can rewrite the models as

$$\mathbf{Y} = \mathbf{f}\{\boldsymbol{\beta}(\mathbf{t}), \mathbf{t}\} + \boldsymbol{\epsilon}, \tag{8}$$
$$\log\{\boldsymbol{\beta}(\mathbf{t})\} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{N}\mathbf{g} + \mathbf{Z}\mathbf{b}, \tag{9}$$

where $\mathbf{Y}$ is an n-dimensional vector, $\boldsymbol{\beta}$ is a $3n$-dimensional vector, $\mathbf{X} = (\mathbf{X}_{11}^{\mathrm{T}}, \ldots, \mathbf{X}_{1J_1}^{\mathrm{T}}, \ldots, \mathbf{X}_{m1}^{\mathrm{T}}, \ldots, \mathbf{X}_{mJ_m}^{\mathrm{T}})^{\mathrm{T}}$ is a $3n \times p$ matrix, $\mathbf{N} = (\mathbf{N}_{11}^{\mathrm{T}}, \ldots, \mathbf{N}_{1J_1}^{\mathrm{T}}, \ldots, \mathbf{N}_{m1}^{\mathrm{T}}, \ldots, \mathbf{N}_{mJ_m}^{\mathrm{T}})^{\mathrm{T}}$ is a $3n \times l$ indicator matrix for the spline function $\mathbf{g}$, $\mathbf{Z}_i = (\mathbf{Z}_{i1}^{\mathrm{T}}, \ldots, \mathbf{Z}_{iJ_i}^{\mathrm{T}})^{\mathrm{T}}$, $\mathbf{Z} = \mathrm{diag}(\mathbf{Z}_1, \ldots, \mathbf{Z}_m)$ is a $3n \times q$ matrix, $q = \Sigma_1^m q_i$, and m is the total number of subjects.

The double penalized log likelihood for $\mathbf{Y}$ and $\mathbf{b}$ is

$$l_{pp}(\boldsymbol{\alpha}, \mathbf{g}, \boldsymbol{\theta} \mid \mathbf{Y}, \mathbf{b})$$
$$= -\frac{n}{2}\log\left(\sigma_0^2\right) - \frac{1}{2}\log|\mathbf{D}|$$
$$- \frac{1}{2}\left[\mathbf{b}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{b} + \lambda \int_{T_1}^{T_2}\left\{g^{(2)}(t)\right\}^2 dt \right.$$
$$\left. + \frac{1}{\sigma_0^2}\{\mathbf{Y} - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{g}, \mathbf{b})\}^{\mathrm{T}}\{\mathbf{Y} - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{g}, \mathbf{b})\}\right],$$
$$= -\frac{n}{2}\log\left(\sigma_0^2\right) - \frac{1}{2}\log|\mathbf{D}|$$
$$- \frac{1}{2}\left[\mathbf{b}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{b} + \lambda\mathbf{g}^{\mathrm{T}}\mathbf{K}\mathbf{g} \right.$$
$$\left. + \frac{1}{\sigma_0^2}\{\mathbf{Y} - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{g}, \mathbf{b})\}^{\mathrm{T}}\{\mathbf{Y} - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{g}, \mathbf{b})\}\right],$$

$$(10)$$

where $\lambda \geq 0$ is a smoothing parameter controlling the balance between the goodness of fit and the roughness of the estimated $g(t)$. T_1 and T_2 specify the range of t, and $\mathbf{K}$ is the nonnegative

definite smoothing matrix defined in equation (2.3) of Green and Silverman (1994). $\boldsymbol{\theta}$ consists of σ_0^2 and the unique elements of $\mathbf{D} = \mathrm{diag}(D_1, \ldots, D_m)$. We use $\mathbf{f}\{\boldsymbol{\beta}(\mathbf{t}), \mathbf{t}\}$ and $\mathbf{f}(\boldsymbol{\alpha}, \mathbf{g}, \mathbf{b})$ interchangeably. Although they have different functional forms, they represent the same true log concentration.

2.2 Estimation of Semiparametric Function and Variance Components

Following Green and Silverman(1994), Zhang et al. (1998), and Lin and Zhang (1999), it is convenient to reparameterize the spline functions using mixed model representation. This provides a foundation for the joint estimation procedure for the variance and smoothing parameters. Let $\mathbf{g} = \mathbf{T}\boldsymbol{\delta} + \mathbf{B}\mathbf{a}$, where $\mathbf{T}$ is an $l \times 2$ matrix, $\mathbf{B}$ is the $l \times (l-2)$ matrix, $\mathbf{B} = \mathbf{L}(\mathbf{L}^{\mathrm{T}}\mathbf{L})^{-1}$, $\mathbf{K} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$, $\mathbf{L}^{\mathrm{T}}\mathbf{T} = \mathbf{0}$, and $\mathbf{g}^{\mathrm{T}}\mathbf{K}\mathbf{g} = \mathbf{a}^{\mathrm{T}}\mathbf{a}$. Both $\mathbf{T}$ and $\mathbf{B}$ are full rank matrices. $\mathbf{T}$ represents basis vectors for the constant and linear terms in the cubic spline $\mathbf{g}$, while $\mathbf{B}$ represents the higher order terms; they are orthogonal to each other. In addition, because the penalty term is the second derivative of the spline function, the constant and linear terms disappear and it penalizes only the higher order terms. Therefore, l_{pp} can be rewritten as

$$
\begin{aligned}
l_{pp}(\boldsymbol{\gamma}, \boldsymbol{\zeta}, \mathbf{a} \mid \mathbf{Y}, \mathbf{b}) \\
= -\frac{n}{2}\log\left(\sigma_0^2\right) - \frac{1}{2}\log|\mathbf{D}| \\
- \frac{1}{2}\left[\mathbf{b}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{b} + \lambda\mathbf{a}^{\mathrm{T}}\mathbf{a}\right. \\
\left. + \frac{1}{\sigma_0^2}\{\mathbf{Y} - (\boldsymbol{\gamma}, \mathbf{a}, \mathbf{b})\}^{\mathrm{T}}\{\mathbf{Y} - \mathbf{f}(\boldsymbol{\gamma}, \mathbf{a}, \mathbf{b})\}\right], \quad (11)
\end{aligned}
$$

where $\boldsymbol{\gamma} = (\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\delta}^{\mathrm{T}})^{\mathrm{T}}$ represents the fixed effect, the cubic term $\mathbf{a}$ is treated as a random effect, and $\boldsymbol{\zeta} = (\boldsymbol{\theta}^{\mathrm{T}}, \lambda)^{\mathrm{T}}$ represents the variance components. We use $\mathbf{f}\{\boldsymbol{\beta}(\mathbf{t}), \mathbf{t}\}$, $\mathbf{f}(\boldsymbol{\alpha}, \mathbf{g}, \mathbf{b})$, and $\mathbf{f}(\boldsymbol{\gamma}, \mathbf{a}, \mathbf{b})$ interchangeably. Although they have different functional forms, they represent the same true log concentration.

The main problem is how to estimate $\boldsymbol{\gamma}$, $\mathbf{a}$, and $\mathbf{b}$. Maximizing $l_p = \log\{\int \exp(l_{pp})d\mathbf{b}\}$ is our goal (O'Sullivan, Yandell, and Raynor, 1986), but this involves a high-dimensional intractable integral. Thus, we compromise by maximizing l_{pp} instead. Maximizing l_{pp} is equivalent to maximizing a second-order Laplace approximation of l_p (Lin and Zhang, 1999).

Given the variance and smoothing parameters, l_{pp} can be maximized by iteratively fitting a linear mixed model with the design matrix representing the tangent plane of the nonlinear function at the current estimate $\boldsymbol{\gamma}^*, \mathbf{a}^*, \mathbf{b}^*$. When $\mathbf{g}$ is not included in the model, the model reduces to an ordinary NLMM and the algorithm is the pseudodata step of Lindstrom and Bates (1990), although they did not express the problem in a linearization form. Specifically, we repeatedly fit the linear mixed model,

$$
\mathbf{W} = \tilde{\mathbf{X}}\boldsymbol{\gamma} + \tilde{\mathbf{C}}\mathbf{a} + \tilde{\mathbf{Z}}\mathbf{b} + \boldsymbol{\epsilon}, \quad (12)
$$

where $\tilde{\mathbf{X}} = \mathbf{F}(\mathbf{X}, \mathbf{NT})\mid_{\boldsymbol{\gamma}^*, \mathbf{a}^*, \mathbf{b}^*}$, $\tilde{\mathbf{C}} = \mathbf{FNB}\mid_{\boldsymbol{\gamma}^*, \mathbf{a}^*, \mathbf{b}^*}$, $\tilde{\mathbf{Z}} = \mathbf{FZ}\mid_{\boldsymbol{\gamma}^*, \mathbf{a}^*, \mathbf{b}^*}$, $\mathbf{W} = (\mathbf{Y} - \mathbf{f} + \tilde{\mathbf{X}}\boldsymbol{\gamma} + \tilde{\mathbf{C}}\mathbf{a} + \tilde{\mathbf{Z}}\mathbf{b})\mid_{\boldsymbol{\gamma}^*, \mathbf{a}^*, \mathbf{b}^*}$, and $\mathbf{F} = \partial\mathbf{f}/\partial\boldsymbol{\beta} \times \partial\boldsymbol{\beta}/\partial(\log\boldsymbol{\beta})$.

Let $\mathbf{M}_1 = (\tilde{\mathbf{X}}, \tilde{\mathbf{C}}, \tilde{\mathbf{Z}})$ at convergence, $\mathbf{H} = \{\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_1 + \mathrm{diag}(\mathbf{0}, \sigma_0^2\lambda\mathbf{I}, \sigma_0^2\mathbf{D}^{-1})\}^{-1}$, and $\mathbf{H}_0 = \mathbf{M}_1^{\mathrm{T}}\mathbf{M}_1$. The covariance

matrix of the best linear unbiased predictor (BLUP) is $\mathrm{cov}(\hat{\boldsymbol{\gamma}}^{\mathrm{T}}, \hat{\mathbf{a}}^{\mathrm{T}}, \hat{\mathbf{b}}^{\mathrm{T}}) = \mathbf{H}^{-1}\mathbf{H}_0\mathbf{H}^{-1}$.

Hastie and Tibshirani (1990) proposed several estimates of the degrees of freedom (d.f.) for linear additive models that include splines. Here we extend one of them to the SEPK model. Let $\boldsymbol{\Omega} = \tilde{\mathbf{Z}}\mathbf{D}\tilde{\mathbf{Z}}^{\mathrm{T}}/\sigma_0^2 + \mathbf{I}$. After some derivation (Appendix 1), the d.f. can be expressed as the trace of the following matrix:

$$
\begin{aligned}
& \begin{pmatrix} \tilde{\mathbf{X}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\mathbf{X}} & \tilde{\mathbf{X}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\mathbf{C}} \\ \tilde{\mathbf{C}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\mathbf{X}} & \tilde{\mathbf{C}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\mathbf{C}} + \sigma_0^2\lambda\mathbf{I} \end{pmatrix}^{-1} \\
& \times \begin{pmatrix} \tilde{\mathbf{X}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\mathbf{X}} & \tilde{\mathbf{X}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\mathbf{C}} \\ \tilde{\mathbf{C}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\mathbf{X}} & \tilde{\mathbf{C}}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\tilde{\mathbf{C}} \end{pmatrix}. \quad (13)
\end{aligned}
$$

In order to estimate the smoothing parameters and variance components, we use restrictive maximum likelihood (REML). Wahba (1985) initially developed REML to choose a smoothing parameter for a natural cubic spline. The key feature of this approach is to treat the smoothing parameters as variance components and regard the penalty functions as normal densities of random effects. The mixed model representation (12) fulfills these conditions. Harville (1974) gave a Bayesian interpretation of REML. Following Wolfinger (1993) and assuming a flat prior for $\boldsymbol{\gamma}$, we obtain

$$
\begin{aligned}
l_r(\boldsymbol{\zeta}) &= \log\int e^{l_{pp}}d\boldsymbol{\gamma}d\mathbf{a}d\mathbf{b} \\
&\approx -\frac{1}{2}\left\{\log|\mathbf{V}| + \log\left|\tilde{\mathbf{X}}^{\mathrm{T}}\mathbf{V}^{-1}\tilde{\mathbf{X}}\right|\right. \\
&\left. + (\mathbf{W} - \tilde{\mathbf{X}}\hat{\boldsymbol{\gamma}})^{\mathrm{T}}\mathbf{V}^{-1}(\mathbf{W} - \tilde{\mathbf{X}}\hat{\boldsymbol{\gamma}})\right\}, \quad (14)
\end{aligned}
$$

where $\mathbf{V} = \tilde{\mathbf{C}}\mathbf{A}\tilde{\mathbf{C}}^{\mathrm{T}} + \tilde{\mathbf{Z}}\mathbf{D}\tilde{\mathbf{Z}}^{\mathrm{T}} + \sigma_0^2\mathbf{I}$. Here again we use the second-order Laplace approximation expanded not only around $\hat{\mathbf{b}}$ but also around $\hat{\boldsymbol{\gamma}}$ and $\hat{\mathbf{a}}$. The scores and the elements of the information matrix for $\boldsymbol{\zeta} = (\boldsymbol{\theta}^{\mathrm{T}}, \boldsymbol{\lambda}^{\mathrm{T}})^{\mathrm{T}}$ are given by

$$
\begin{aligned}
u_{\zeta_j} &= -\frac{1}{2}\left\{\mathrm{tr}\left(\mathbf{P}\frac{\partial\mathbf{V}}{\partial\zeta_j}\right)\right. \\
&\left. - (\mathbf{W} - \tilde{\mathbf{X}}\hat{\boldsymbol{\gamma}})^{\mathrm{T}}\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\zeta_j}\mathbf{V}^{-1}(\mathbf{W} - \tilde{\mathbf{X}}\hat{\boldsymbol{\gamma}})\right\} \\
&= 0, \quad (15)
\end{aligned}
$$

$$
I_{\zeta_j\zeta_k} = -\frac{1}{2}\mathrm{tr}\left(\mathbf{P}\frac{\partial\mathbf{V}}{\partial\zeta_j}\mathbf{P}\frac{\partial\mathbf{V}}{\partial\zeta_k}\right), \quad (16)
$$

where $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\tilde{\mathbf{X}}(\tilde{\mathbf{X}}^{\mathrm{T}}\mathbf{V}^{-1}\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^{\mathrm{T}}\mathbf{V}^{-1}$.

2.3 Numerical Methods

For the estimation of the nonparametric function, we not only have to know the mean profile $f\{\boldsymbol{\beta}(t), t\}$ but also the first derivative $\partial f\{\boldsymbol{\beta}(t), t\}/\partial\boldsymbol{\beta}$. (In this section, we use $\boldsymbol{\beta}(t) = \{\phi(t), V\} = \{k_a(t), k_d(t), V\}$ to represent any particular subject-specific parameter vector $\boldsymbol{\beta}_i(t)$ for convenience and so is $\mathbf{A}$ for $\mathbf{A}_i$). Because $f\{\boldsymbol{\beta}(t), t\}$ is a log transformation of $A_2\{\phi(t), t\}/V$ or $A_2(t)/V$, where $A_2(t)$ is the amount of drug in the blood and tissue, it is enough to show how to estimate $\mathbf{A}(t)$ and $\mathbf{A}_\phi(t) = \partial\mathbf{A}(t)/\partial\phi$. Take the derivative with

respect to ϕ on both sides of (4),

$$\frac{d\mathbf{A}_\phi(t)}{dt} = \mathbf{G}_\phi\{\phi(t), t\}\mathbf{A}(t) + \mathbf{G}\{\phi(t), t\}\mathbf{A}_\phi(t), \qquad (17)$$

where $\mathbf{G}_\phi = \partial\mathbf{G}/\partial\phi$. Hence, both $\mathbf{A}$ and $\mathbf{A}_\phi$ can be solved jointly with (4) and (17). The initial value, $\mathbf{A}_\phi(0)$, is always zero because changing PK parameters, ϕ, will not affect the initial value of $\mathbf{A}(0)$.

The numerical solution (18) of the differential equation is a fifth-order Runge–Kutta formula (Press et al., 1992, p. 711),

$$\mathbf{p}_1 = h\mathbf{G}\{\phi(t_n), t_n\}\mathbf{A}_n,$$
$$\mathbf{p}_2 = h\mathbf{G}\{\phi(t_n + l_2 h), (t_n + l_2 h)\}$$
$$\times (\mathbf{A}_n + w_{21}\mathbf{p}_1),$$
$$\mathbf{p}_3 = h\mathbf{G}\{\phi(t_n + l_3 h), (t_n + l_3 h)\}$$
$$\times (\mathbf{A}_n + w_{31}\mathbf{p}_1 + w_{32}\mathbf{p}_2),$$
$$\mathbf{p}_4 = h\mathbf{G}\{\phi(t_n + l_4 h), (t_n + l_4 h)\}$$
$$\times (\mathbf{A}_n + w_{41}\mathbf{p}_1 + w_{42}\mathbf{p}_2 + w_{43}\mathbf{p}_3),$$
$$\mathbf{p}_5 = h\mathbf{G}\{\phi(t_n + l_5 h), (t_n + l_5 h)\}$$
$$\times (\mathbf{A}_n + w_{51}\mathbf{p}_1 + w_{52}\mathbf{p}_2 + w_{53}\mathbf{p}_3 + w_{54}\mathbf{p}_4),$$
$$\mathbf{p}_6 = h\mathbf{G}\{\phi(t_n + l_6 h), (t_n + l_6 h)\}$$
$$\times (\mathbf{A}_n + w_{61}\mathbf{p}_1 + w_{62}\mathbf{p}_2 + w_{63}\mathbf{p}_3 + w_{64}\mathbf{p}_4 + w_{65}\mathbf{p}_5),$$
$$\mathbf{A}_{n+1} = \mathbf{A}_n + v_1\mathbf{p}_1 + v_2\mathbf{p}_2 + v_3\mathbf{p}_3 + v_4\mathbf{p}_4 + v_5\mathbf{p}_5$$
$$+ v_6\mathbf{p}_6 + \mathbf{O}(h^6),$$
$$\mathbf{A}^*_{n+1} = \mathbf{A}_n + v_1^*\mathbf{p}_1 + v_2^*\mathbf{p}_2 + v_3^*\mathbf{p}_3 + v_4^*\mathbf{p}_4 + v_5^*\mathbf{p}_5$$
$$+ v_6^*\mathbf{p}_6 + \mathbf{O}(h^5), \qquad (18)$$

where $\mathbf{A}^*_{n+1}$ is a fourth-order formula and $\{l_s, w_{s_1 s_2}, v_s, v_s^*\}_{s,s_1,s_2=1}^6$ are the known coefficients. Let the error estimate be $\delta = \max|\mathbf{A}_{n+1} - \mathbf{A}^*_{n+1}|$, which is of order h^5. If we take a step h_1 and produce an error δ_1, the step h_0 that would have given some other value δ_0 is readily estimated as $h_0 = h_1|\delta_0/\delta_1|^{0.2}$. Henceforth, we let δ_0 denote the desired accuracy. We have implemented the numerical solution of the differential equation in Fortran 77, and the functions are modifications of those proposed by Press et al. (1992).

3. Interpretation and Flexibility of SEPK

The interpretation of the SEPK is simple when the true model for the data is one compartment since $k_{a,i}(t)$ represents $k_{a,i}$ and $k_{d,i}(t)$ represents $k_{e,i}$. However, the interpretation of SEPK is not so obvious when the true model for the data is two compartment. Let us reformulate the two-compartment model (2) as follows:

$$A_{2,i}(t) = \frac{k_{a,i}(k_{32,i} - k_{a,i})e^{-k_{a,i}\times t}}{(a_{1,i} - k_{a,i})(a_{2,i} - k_{a,i})}$$
$$+ \frac{k_{a,i}(k_{32,i} - a_{1,i})e^{-a_{1,i}\times t}}{(a_{2,i} - a_{1,i})(k_{a,i} - a_{1,i})}$$
$$+ \frac{k_{a,i}(k_{32,i} - a_{2,i})e^{-a_{2,i}\times t}}{(a_{1,i} - a_{2,i})(k_{a,i} - a_{2,i})},$$
$$a_{1,i} = \frac{1}{2}\left[k_{23,i} + k_{32,i} + k_{e,i}\right.$$

$$\left. + \{(k_{23,i} + k_{21,i} + k_{e,i})^2 - 4k_{e,i}k_{32,i}\}^{1/2}\right],$$

$$a_{2,i} = \frac{1}{2}\left[k_{23,i} + k_{32,i} + k_{e,i}\right.$$

$$\left. - \{(k_{23,i} + k_{32,i} + k_{e,i})^2 - 4k_{e,i}k_{32,i}\}^{1/2}\right], \qquad (19)$$

where $A_{2,i}$ is the analytic solution of the two-compartment model (2) and $(a_{1,i}, a_{2,i})$ are two phases of drug disposition. We call $a_{1,i}$ and $a_{2,i}$ the fast phase and slow phase of the drug disposition, respectively. $(k_{a,i}, k_{e,i}, k_{23,i}, k_{32,i})$ and $(k_{a,i}, k_{32,i}, a_{1,i}, a_{2,i})$ are two sets of equivalent parameterizations. Let the true model be

$$y_{ij} = \log\left\{\frac{(k_{a,i}, k_{32,i}, a_{1,i}, a_{2,i}, t)}{V_i}\right\}$$
$$+ e_{ij},$$
$$\log(k_{a,i}, k_{32,i}, a_{1,i}, a_{2,i}) = \log(k_a, k_{32}, a_1, a_2)$$
$$= (\alpha_1, \alpha_2, \alpha_3, \alpha_4),$$
$$\log(V_i) = \log(V \times b_i^*) = \alpha_0 + b_i, \qquad (20)$$

where V_i is the subject-specific volume of distribution and $V = e^{\alpha_0}$ is the mean. $b_i = \log(b_i^*) \sim N(0, 0.3)$, $e_{ij} \sim N(0, 0.2)$, $i = 1, \ldots, 20$, $j = 1, \ldots, 10$, and $t = 0.1, 0.3, 0.5, 1, 2, 3, 4, 6, 8$, and 12 hours. Let $(V, k_a, k_{32}, a_1, a_2) = (0.2, 4, 0.1, 1.43, 0.06)$ be a set of PK parameters to be estimated. In the simulation model (20), we treat only $\log(V_i)$ as a mixed effect because of the simplicity of the computation. Although more mixed effects can be specified for all the PK parameters, it does not make a difference because we are only interested in the connection between a two-compartment model and an SEPK. Let the differential equations be (3). The SEPK is

$$y_{ij} = \log[A_2\{k_{a,i}, k_{d,i}(t_{ij}), t_{ij}\}/V_i] + e_{ij},$$
$$\log\{k_{a,i}, k_{d,i}(t_{ij})\} = (\alpha_1, \mathbf{N}_{ij}\mathbf{g}),$$
$$\log(V_i) = \alpha_0 + b_i,$$

where $k_{a,i}$ is treated as a constant over time and $k_{d,i}(t)$ models the drug disposition process. $\mathbf{N}_{ij}$ is a 10×1-dimensional indicator matrix for the spline vector $\mathbf{g}$. $k_{d,i}(t)$ can be visualized as a blending of $a_{1,i}$ and $a_{2,i}$. During the early phase of drug disposition, $k_{d,i}(t)$ is dominated by the fast phase of drug disposition, $a_{1,i}$, while it is mainly $a_{2,i}$ at the end of drug disposition, as our simulation demonstrates.

In the simulation, we ran 500 replications. In each replication, both models are fitted. In order to compare the performance of PK parameter estimates, the root mean square error (RMSE) was calculated. We obtained $\hat{V}^{2PK} = 0.2(0.02)$, where 0.02 is RMSE, the subscript 2PK refers to the two-compartment model, and $\hat{V}^{SEPK} = 0.19(0.02)$. Hence, $\hat{V}^{2PK}$ and $\hat{V}^{SEPK}$ are close, as are

$$\hat{k}_a^{2PK} = 4.03(0.44)$$

and

$$\hat{k}_a^{SEPK} = 3.74(0.38).$$

Although

$$\left\{\hat{a}_1^{\text{SEPK}}, \hat{a}_2^{\text{SEPK}}\right\} = \{1.55(0.23), 0.06(0.04)\}$$

have larger RMSEs than

$$\left\{\hat{a}_1^{\text{2PK}}, \hat{a}_2^{\text{2PK}}\right\} = \left\{\hat{k}_d(t_1), \hat{k}_d(t_{10})\right\}$$

$$= \{1.44(0.12), 0.06(0.02)\},$$

they are acceptable. Thus, SEPK is flexible in estimation because we do not have to specify the number of the compartments and it provides robust PK parameter estimates even if SEPK is a misspecified model. Particularly, the absorption rate estimate based on SEPK is very comparable with that based on the true two-compartment model, and the early and later phases of $k_d(t)$ can be used to estimate the fast and slow phases of the disposition process (a_1 and a_2), respectively. The simulation establishes a connection between the SEPK and the two-compartment model. A similar simulation could be carried out to compare a three-compartment model and the SEPK, and we expect similar results.

4. Data Analysis

In the Ditropan study introduced in Section 1, there was evidence that the constant PK assumption was violated. We thus apply the SEPK model to these data.

We use a piecewise natural cubic spline to model the disposition phase. It is composed of three spline functions, $g1(t)$, $g2(t)$, and $g3(t)$. The discontinuity points among them represent the gaps between the slow disposition phase of the prior dose and the fast disposition phase of next dose. Also, we choose a step function to model the absorption phase. Let $A_{2,i}(k_{a,i}(t), k_{d,i}(t), t)$ be the amount of drug in the blood and tissue. Let the differential equations be (3). Then the SEPK is as follows:

$$y_{ij} = \log[A_{2,i}\{(k_{a,i}(t_{ij}), k_{d,i}(t_{ij}), (t_{ij})\}/V_i] + \epsilon_{ij},$$

$$\log(V_i) = \log(V \times b_{1i}^*) = \alpha_0 + b_{1i},$$

$$\log\{k_{a,i}(t_{ij})\} = b_{2i} + I_{\{t_{ij}<8\}}\log(k_{a_1})$$
$$+ I_{\{8<t_{ij}<16\}}\log(k_{a_2})$$
$$+ I_{\{16<t_{ij}\}}\log(k_{a_3}),$$
$$= b_{2i} + I_{\{t_{ij}<8\}}\alpha_1 + I_{\{8<t_{ij}<16\}}\alpha_2$$
$$+ I_{\{16<t_{ij}\}}\alpha_3,$$

$$\log\{k_{d,i}(t_{ij})\} = b_{3i} + I_{\{t_{ij}<8\}}\{g1(t_{ij}) + b_{4i}(8 - t_{ij})\}$$
$$+ I_{\{8<t_{ij}<16\}}\{g2(t_{ij} - 8) + b_{4i}(16 - t_{ij})\}$$
$$+ I_{\{16<t_{ij}<24\}}\{g3(t_{ij} - 16) + b_{4i}(24 - t_{ij})\}$$
$$+ I_{\{24<t_{ij}\}}\{g3(t_{ij} - 16) + b_{4i} \times 8\}, \qquad (21)$$

where

$$\epsilon_{ij} \sim \text{i.i.d. } N(0, \sigma_0^2),$$

$$(b_{1i}, b_{2i}, b_{3i}, b_{4i}) \sim \text{i.i.d. } N(\mathbf{0}, \mathbf{D}_0),$$

$\mathbf{D}_0 = \{\sigma_{s_1 s_2}\}$ is a 4×4 general covariance matrix, and $g1(t)$, $g2(t)$, and $g3(t)$ are defined as (7). (Here $g1(t_{ij})$ represents the spline function at time t_{ij} in dose 1, while $g(t_{ij})$ represents the general spline function at time t_{ij} in model (6)). Because only the last dose has plasma samples 8 hours after oral administration, the model for the ith subject's $\log\{k_{d,i}(t_{ij})\}$ in (21) allows three doses to share the same random intercept b_{3i} and random slope b_{4i} within 8 hours

after each dose. The fifth term in $\log\{k_{d,i}(t_{ij})\}$ ensures that it is continuous. More general models, such as each dose having its own random intercept, slope, or spline function within each subject's $\log\{k_{d,i}(t_{ij})\}$, were also tried. However, due to the high collinearity among those parameters, the estimation procedures did not converge. Because model (21) has enough random structures to describe within- and between-subject variation, we consider it first. Figure 3a shows the fitting of the model. Figure 3b and Table 1 display the estimates of the drug absorption and disposition processes. It appears that the absorption rate is decreasing and the early phase of the drug disposition process for dose 2 is lower than the other two, although all the later phases are comparable. The within-subject variance estimate is $\hat{\sigma}_0^2 = 0.12$, the between-subject variance due to the $\log(V_i)$ is estimated as 0.21, the variance of the $\log(k_{a,i})$ is estimated as 0.05, and the estimate of the variances of the random intercept and random slope for $\log\{k_d(t_{ij})\}$ are 0.02 and 0.001, respectively.

We performed some Wald tests and approximate F-tests (Hastie and Tibshirani, 1990) to test the time-varying behavior of drug absorption and disposition. The degrees of freedom for an approximate F-test is calculated based on (13). Our first null hypothesis is $k_{a_1} = k_{a_2} = k_{a_3}$. The Wald statistic is $W = 5.95$ and $p = 0.051$ with d.f. $= 2$, and the F-statistic is $F = 3.83$ and $p = 0.022$ with d.f. $= (2, 956.33)$. Both suggest that the absorption process is time dependent. Our second null hypothesis is $g1(t_j) = g2(t_j) = g3(t_j)$, where $t_j = 0.5, 1.0, 1.5, 2.0, 3.0, 5.0$, and 7.0 are distinct sampling time points after each dose. Here $W = 14.54$ and $p = 0.41$ with d.f. $= 14$, and $F = 1.78$ and $p = 0.067$ with d.f. $= (8.98, 956.33)$. Although both tests do not provide enough evidence of difference among three disposition processes for three doses during the first 8 hours after oral administration, we do observe some difference among the early phases of drug disposition processes in Figure 3b. We thus remodeled the disposition functions as

$$g1(t) = g3(t) = g2(t) + \log(1 + \Delta) \times (1 - t/8), \qquad 0 < t < 8. \qquad (22)$$

In model (22), doses 1 and 3 have different early phase drug disposition than dose 2, which is denoted by Δ. Its effect on the disposition process exponentially diminishes after 8 hours. We obtained $W = 5.69$ and $p = 0.017$ with d.f. $= 1$ and $F = 7.06$ and $p = 0.008$ with d.f. $= (1, 964.21)$. Although both tests show evidence that the early phases of drug disposition process are time dependent among three doses, we should be aware that this test is based on the finding from model (21), so there is likely an inflation of the type I error. However, in this exploratory data analysis, we are searching for possible reasons for time-dependent PK behavior; therefore, we think the test based on model (22) is informative. It at least provides some evidence for a time-dependent pattern of $k_d(t)$, which can be further compared with the other features of the study, e.g., blood pressures or heart rates. Theoretically, high k_d leads to low AUC and $Cmax$, and high k_a leads to high AUC and $Cmax$. In this PK study, dose 1 has higher k_a and higher early phase of k_d than does dose 2. Dose 2 has higher AUC and $Cmax$ than dose 1, probably because the effect of k_d is greater than that of k_a. Similarly, dose 2's AUC and $Cmax$ are higher than those of dose 3 because dose 3 has higher k_d and lower k_a.

(a)

Time from the First Dose(hour)

(b)

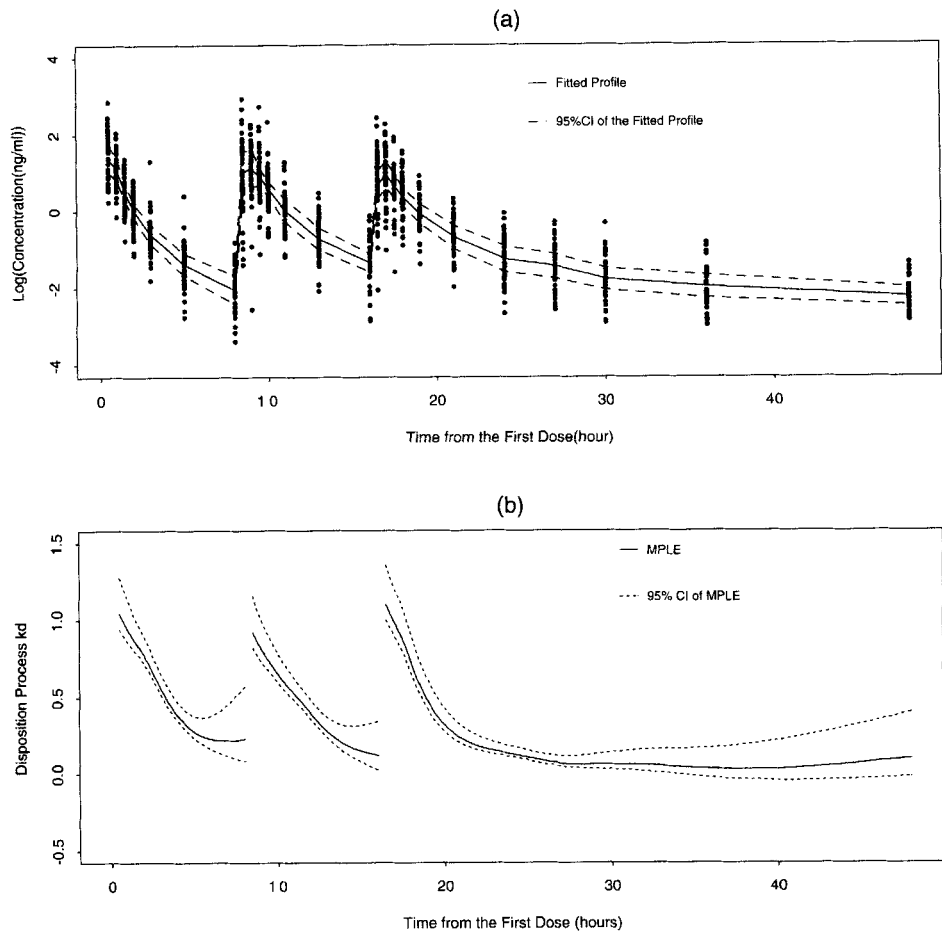Time from the First Dose (hours)

Figure 3. Observed oxybutynin concentration, model fitting, and estimation. **a.** Spline-enhanced PK model fitting; the solid line is the fitted log concentration $(\hat{f})$ and the dashed lines are the 95% pointwise confidence interval. **b.** Spline estimates of $k_d(t) = \exp\{I_{\{t<8\}}g1(t) + I_{\{8<t<16\}}g2(t) + I_{\{16<t\}}g3(t)\}$ for the oxybutynin data; MPLE represents the maximized penalized likelihood estimate; the solid line is the MPLE, $\hat{k}_d$, and the dashed lines are the 95% pointwise confidence interval.

In order to compare the performance of SEPK with that of a multicompartment model, we carried out an approximate F-test between SEPK (22) and a two-compartment model given by

$$y_{ij} = \log\{A_{2,i}(k_{a,i}, k_{32,i}, a_{1,i}, a_{2,i}, t_{ij})/V_i\} + e_{ij},$$

$$\log(V_i) = \alpha_0 + b_{1i},$$

$$\log(k_{a,i}) = \alpha_1 + b_{2i},$$

$$\log(a_{1,i}) = \alpha_2 + b_{3i},$$

$$\log(a_{2,i}, k_{32,i}) = (\alpha_3, \alpha_4). \tag{23}$$

The F-test is $F = 2.94$ and $p = 0.011$ with d.f. $= (5.22, 964.21)$. It is clear that SEPK (22) outperforms (23).

5. Simulation

In order to verify the validity of p-values of the tests for H_0: $k_{a_1} = k_{a_2} = k_{a_3}$, data were generated based on model (21) under H_0 with parameter values from Table 2, except $k_a = 4.1$. Each of the 500 replicated data sets had 40 subjects and 25 observations per subject. The 500 Wald statistics

Table 1

Estimates of fixed effects for the Ditropan data

Parameter	Model (21)	Model (22)
V	0.20 ± 0.02	0.21 ± 0.02
k_{a_1}	6.41 ± 1.41	6.12 ± 1.23
k_{a_2}	2.55 ± 0.61	2.43 ± 0.49
k_{a_3}	1.42 ± 0.23	1.49 ± 0.19
Δ		0.19 ± 0.08
σ_0^2	0.12	0.13
σ_{11}	0.20	0.19
σ_{22}	0.05	0.05
σ_{33}	0.02	0.02
σ_{44}	0.001	0.001
σ_{12}	0.08	0.09
σ_{13}	0.04	0.03
σ_{14}	0.002	0.002
σ_{23}	0.03	0.03
σ_{24}	0.002	0.002
σ_{34}	0.004	0.004

Table 2
Estimates of fixed effects by simulation

Parameter	True	MPLE	Empirical SE (MPLE)	Estimated SE (MPLE)
V	0.20	0.22	0.02	0.01
ka_1	5.50	5.97	1.32	1.23
ka_2	2.65	2.79	0.49	0.47
ka_3	1.54	1.55	0.18	0.17
Δ	0.20	0.20	0.08	0.07

were ranked after simulation, and the empirical p-value was calculated as the percentage of those simulated Wald statistics larger than the one based on the S-oxybutynin data. The empirical p-value for the approximate F-test was obtained similarly. The empirical p-value for the Wald test is 0.035, which is smaller than the one based on chi-square distribution (0.051). The empirical p-value for the approximate F-test is 0.037, which is larger than the one based on the F-distribution (0.022).

Similarly, 500 replications were produced based on the average of $g1(t), g2(t), g3(t)$ and other parameters in Table 2 to test H_0: $g1(t_j) = g2(t_j) = g3(t_j)$, where $t_j = 0.5, 1.0, 1.5, 2.0,$ 3.0, 5.0, and 8.0 are distinct sampling time points after each dose. The empirical p-value for the Wald test is 0.31 (0.35 based on the chi-square distribution). The empirical p-value for the approximate F-test is 0.22 (0.067 based on the F distribution). The same 500 replications are also used to test H_0: $\Delta = 0$ in the reduced model (22). The empirical p-value for the Wald test is 0.009 (0.017 based on the chi-square distribution). The empirical p-value for the approximate F-test is 0.011 (0.008 based on the F-distribution).

According to these simulation results, the Wald tests tend to be conservative and the approximate F-tests tend to be liberal. When the null hypothesis is to test some fixed parameters (e.g., $k_{a_1} = k_{a_2} = k_{a_3}$ or $\Delta = 0$), both tests show that the significance level (p-values) based on the distribution are comparable with the empirical ones. However, when the null hypothesis is to test the nonparametric functions, e.g., $g1(t_j) = g2(t_j) = g3(t_j)$, the Wald test gives more reliable
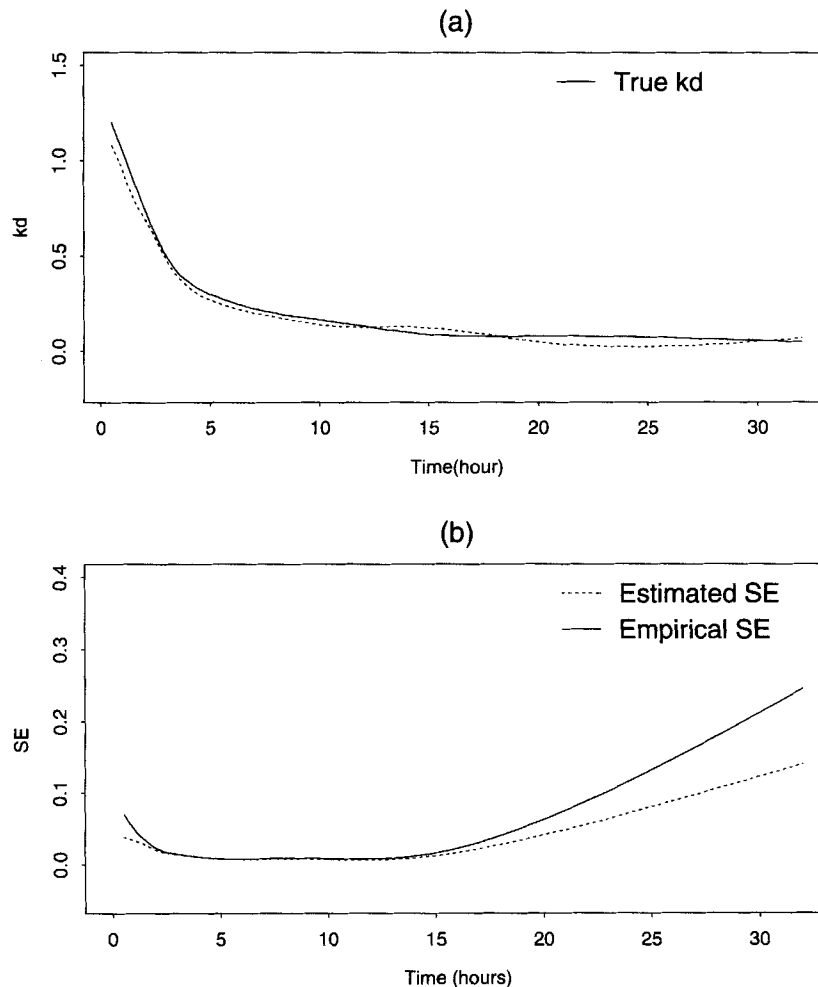
(a)

(b)

Figure 4. Spline estimates of $k_d(t) = \exp\{I_{\{t<8\}}g1(t) + I_{\{8<t<16\}}g2(t) + I_{\{16<t\}}g3(t)\}$ by simulation. **a.** MPLE represents the maximized penalized likelihood estimate; the solid line is the true $k_d(t)$ and the dashed line is the MPLE, $\hat{k}_d(t)$. **b.** The solid line is the empirical SEs of $\hat{k}_d(t)$ and the dashed line is the estimated SE of $\hat{k}_d(t)$.

p-value than the approximate F-test does. The simulation result of the approximate F-test between SEPK (22) and a two-compartment model shows the same pattern: the empirical p-value 0.044 is much large than the one (0.011) based on the F-distribution, although both p-values suggest that SEPK provides a better fit than a two-compartment model does.

In order to evaluate the performance of SEPK, the 500 sets of estimates generated by model (22) are used to calculate mean and empirical and estimated standard errors. Table 2 and Figure 4a and 4b present the results.

6. Conclusion and Discussion

The spline-enhanced population PK models are able to quantify the drug concentration over time and describe time-dependent behavior for the drug absorption and disposition processes. The Wald tests and approximate F-tests verify that both absorption and disposition processes are changing over time and that SEPK performs better than a two-compartment model that is based on a constant PK parameter assumption. The simulation demonstrates that SEPK is almost as good as a two-compartment PK model in estimating absorption and disposition rates when the true model has two compartments, and it establishes an interpretation for SEPK's parameters.

We are asked frequently why we model the whole disposition process as an arbitrary function of time $k_d(t)$ instead of modeling a subprocess as a spline. For example, letting $k_e(t)$ be an arbitrary function in a multicompartment model can definitely make the model easy to interpret. We did try this idea initially; however, $k_e(t)$ is highly correlated with the other intercompartment PK parameters because the flexibility of $k_e(t)$ tends to explain the same thing as the other parameters. That is the reason why we cannnot distinguish the elimination process from the distribution process during the drug disposition through the SEPK. One possible way to solve this problem is to put some constraints on the elimination process and distribution processes when we model them. But how to restrict them needs more scientific input, and it is a potential research area in the future.

On the other hand, modeling k_a as a spline function will not be as easy as that for k_d. We know that k_a is usually much larger than k_d. As a result, the absorption phase is almost finished after 4–5 half-lives (the time at which half of the drug has been absorbed). It is difficult to estimate k_a after that, and the estimation of that part of the spline will be extremely unstable. Hence, k_a is defined as a step function in our model.

The biases of PK parameter estimates in the SEPK are not large (between 5% and 10%). This may be a problem for small sample sizes. Lin and Zhang (1999) discuss the bias correction for the generalized linear additive mixed model, and their methods can be extended to our situation.

Although S-oxybutynin has been used in the clinical setting for over 10 years, its time-dependent kinetics are not well understood. Potentially, AUC and $Cmax$ may be greatly decreased following multiple-dose administration over a period of time. Based on our SEPK analysis, one major reason is the decreased absorption rate, which is possibly related to the saturability in the transport mechanism for passage across the gastrointestinal membranes. There is some mild evidence for a time-dependent disposition process, too. If this is due to the elimination process, the time-dependent kinetics are probably associated with the saturability in metabolism during the drug's first pass through the gut wall and the liver or the induction of metabolism of coadministrated drugs producing a drug interaction (autoinduction). Additional PK studies need to be done in order to verify these results. If this is confirmed by additional PK studies, it may be desirable to reformulate the drug in order to alter this pattern of absorption and elimination. In clinical practice, the individual dose level can be calibrated based on the estimated subject-specific PK parameters and SEPK through continuous monitoring and simulation in order to keep the drug at the target $Cmax$ or AUC.

RÉSUMÉ

Cet article tire sa motivation d'une étude où des sujets, après avoir reçu trois doses identiques d'un même produit, montraient des concentrations sanguine apparemment plus basses après la troisième dose. Une explication possible était la variation au cours du temps des paramètres pharmacocinétiques (PK). Cette variation a été modélisée par des fonctions splines cubiques naturelles introduites dans des équations différentielles ordinaires. Les valeurs moyennes des paramètres, les composantes de la variance et les paramètres de lissage ont été estimés conjointement par la maximisation d'une double log-vraisemblance pénalisée. Les fonctions moyennes et leurs dérivées ont été obtenues via la résolution numérique des équations différentielles. L'interprétation des paramètres du modèle et de leur flexibilité est discutée. Les méthodes proposées sont illustrées par leur application au cas ayant motivé leur étude. Les performances des modèles sont évaluées par des simulations.

REFERENCES

Davidian, M. and Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association* **82**, 1079–1091.

Davidian, M. and Giltinan, D. G. (1995). *Nonlinear Models for Repeated Measurement Data.* London: Chapman and Hall.

Douchamps, J., Derenne, F., Stockis, A., Gangji, D., Juvent, M., and Herchuelz, A. (1988). The pharmacokinectics of oxybutynin in man. *European Journal of Clinical Pharmacology* **35**, 515–520.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Model.* London: Chapman and Hall.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* **61**, 383–385.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Model.* London: Chapman and Hall.

Lin, X. and Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society, Series B* **61**, 381–400.

Lindstrom, J. M. and Bates, M. D. (1990). Nonlinear mixed effects models for repeated measurement data. *Biometrics* **46**, 673–687.

O'Sullivan, F., Yandell, B. S., and Raynor, W. J. (1986). Automatic smoothing of regression functions in generalized linear models. *Journal of the American Statistical Association* **81**, 96–103.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (1992). *Numerical Recipes in Fortran 77.* London: Cambridge University Press.

Rowland, M. and Tozer, T. N. (1995). *Clinical Pharmacokinetics: Concepts and Principles,* 3rd edition. Baltimore: Williams and Wilkins.

Sathyan, G., Chancellor, M., and Gupta, S. (2001). Effect of OROS controlled-release delivery on the pharmacokinetics and pharmacodynamics of oxybutynin chloride. *British Journal of Clinical Pharmacology* **52**, 409–417.

Sheiner, L. B. and Beal, S. L. (1980). Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis–Menten model: Routine clinical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics* **8**, 553–571.

Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Annals of Statistics* **13**, 1378–1402.

Wolfinger, R. D. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika* **80**, 791–795.

Zhang, D., Lin, X., Raz, J., and Sowers, M. (1998). Semiparametric stochastic mixed model for longitudinal data. *Journal of the American Statistical Association* **93**, 710–719.

APPENDIX 1

At convergence, the solution of (11) is equivalent to the maximization of a second-order Laplace approximation of l_p. Hence, we calculate the d.f. based on (12). Write

$$\mathbf{W} = \mathbf{M}_2 \boldsymbol{\alpha} + \mathbf{e}, \tag{24}$$

where $\mathbf{M}_2 = (\tilde{\mathbf{X}}, \tilde{\mathbf{C}})$, $\boldsymbol{\alpha} = (\boldsymbol{\gamma}, \mathbf{a})$, $\mathrm{cov}(\mathbf{a}) = \boldsymbol{\Lambda}$, and $\mathrm{cov}(\mathbf{e}) = \sigma_0^2 \boldsymbol{\Omega} = \tilde{\mathbf{Z}} \mathbf{D} \tilde{\mathbf{Z}}^{\mathrm{T}}/\sigma_0^2 + \mathbf{I} = \mathbf{Q}^{\mathrm{T}} \mathbf{Q}$. $\mathbf{Q}^{\mathrm{T}} \mathbf{Q}$ is a Q-Q decomposition of $\boldsymbol{\Omega}$. If we take a linear transformation of $\mathbf{W}$, we have

$$\left(\mathbf{Q}^{\mathrm{T}}\right)^{-1} \mathbf{W} = \left(\mathbf{Q}^{\mathrm{T}}\right)^{-1} \mathbf{M}_2 \boldsymbol{\alpha} + \left(\mathbf{Q}^{\mathrm{T}}\right)^{-1} \mathbf{e}, \tag{25}$$

where $\mathrm{cov}\{(\mathbf{Q}^{\mathrm{T}})^{-1}\mathbf{e}\} = \sigma_0^2 \mathbf{I}$. Hastie and Tibshirani (1990) proposed several estimators for d.f. for the linear additive model with independent observations. We extend one of them, which is based on the trace of smoothing matrix $\mathbf{S}$, given by

$$\mathrm{tr}(\mathbf{S}) = \mathrm{tr}\left[\left\{ \left(\mathbf{Q}^{\mathrm{T}}\right)^{-1} \mathbf{M}_2 \mathbf{M}_2^{\mathrm{T}} \mathbf{Q}^{-1} \left(\mathbf{Q}^{\mathrm{T}}\right)^{-1} \mathbf{M}_2 \right.\right.$$
$$\left.\left. + \sigma_0^2 \mathrm{diag}(\mathbf{0}, \boldsymbol{\Lambda}) \right\}^{-1} \mathbf{M}_2^{\mathrm{T}} \mathbf{Q}^{-1} \right]$$
$$= \mathrm{tr}\left[\left\{ \mathbf{M}_2^{\mathrm{T}} \mathbf{V}^{-1} \mathbf{M}_2 + \sigma_0^2 \mathrm{diag}(\mathbf{0}, \boldsymbol{\Lambda}) \right\}^{-1} \mathbf{M}_2^{\mathrm{T}} \mathbf{V}^{-1} \mathbf{M}_2 \right]. \tag{26}$$

Expression (26) can be more explicitly written as (13).