Published in final edited form as:

Stat Methods Med Res. 2017 February ; 26(1): 21–42. doi:10.1177/0962280214537255.

A Multiphase Non-Linear Mixed Effects Model: An Application to Spirometry after Lung Transplantation

Jeevanantham Rajeswaran* and

Department of Quantitative Health Sciences, and Heart and Vascular Institute, Cleveland Clinic

Eugene H. Blackstone

Heart and Vascular Institute and Department of Quantitative Health Sciences, Cleveland Clinic

Abstract

In medical sciences, we often encounter longitudinal temporal relationships that are non-linear in nature. The influence of risk factors may also change across longitudinal follow-up. A system of multiphase non-linear mixed effects model is presented to model temporal patterns of longitudinal continuous measurements, with temporal decomposition to identify the phases and risk factors within each phase. Application of this model is illustrated using spirometry data after lung transplantation using readily available statistical software. This application illustrates the usefulness of our flexible model when dealing with complex non-linear patterns and time varying coefficients.

Keywords

spirometry data; non-linear model; mixed-effects model; temporal decomposition; time varying coefficient; additive regression

1 Introduction

In medical science, data collected repeatedly over time are often encountered. For example, after aortic valve repair, investigators may be interested in determining long-term durability of the repaired valve. Its durability might be determined by investigating hemodynamics of the valve, such as aortic peak or mean gradient, and degree of aortic valve regurgitation (Banbury et al [1]). The important feature of longitudinal data is that while, as with cross-sectional data, the observations between subjects are assumed to be uncorrelated, the observations from a single subject tend to be correlated. Hence any statistical analyses or model that involves longitudinal data should take such correlation into consideration in order to draw valid inferences. In general, the objectives of a longitudinal data analysis can be either population based inferences (marginal), or subject-specific inferences or transitional (Diggle et al. [2]).

When the subjects are sampled from a large population, one can model the natural heterogeneity due to unmeasured factors by introducing random effects into the model. In

*Cleveland Clinic, JJNB-01, 9500 Euclid Avenue, Cleveland, OH 44195, USA, rajeswj@ccf.org.

other words, the natural heterogeneity across subjects is represented by a probability distribution. In modeling of longitudinal response data, the correlation among the observations within a subject is accounted for by using the unobservable random effects. Hence, this type of model is also called a *latent variable* model. Laird and Ware [3] introduced linear random effects models, more appropriately called *mixed effect models* in a longitudinal setup. The extended model that accommodates non-normal responses is often called a *generalized linear mixed effects model*.

Suppose there are m subjects. The generalized linear or non-linear random effects model can be written as follows:

$$g(\mu_i(\mathbf{b}_i)) = f(\beta, \mathbf{b}_i; \mathbf{X}_i, \mathbf{Z}_i), \quad (1)$$

where $\mu_i(\mathbf{b}_i) = E(\mathbf{Y}_i | \mathbf{b}_i)$ and g is a link function. The random effects $\mathbf{b}_i (i = 1, \dots, m)$ are mutually independent with a common multivariate distribution $F_b(\mathbf{G})$, often a multivariate normal. Conditional on the random effects $\mathbf{b}_i$, the longitudinal responses are assumed to be independently, identically distributed, often as a distribution from the exponential family. Note that, since $E_b\{E(\mathbf{Y}_i | \mathbf{b}_i)\} = E(\mathbf{Y}_i)$, one can obtain estimates of the marginal expectation of $\mathbf{Y}$ from conditional models. In fact, Lee and Nelder [4] argue that the conditional model is the basic model and any conditional model leads to a specific marginal model. Further, in practice, longitudinal data are often highly unbalanced in the sense that each subject has a different number of longitudinal responses observed at non-fixed time points. In this scenario, a mixed effect modeling approach is preferred over a marginal modeling approach.

Laird and Ware [3] used the EM algorithm for estimation in linear mixed effect models. For linear mixed effect models for unbalanced data, Jennrich and Schluchter [5] discussed the implementation of Newton-Raphson and Fisher's scoring algorithms for computing maximum likelihood estimates and a generalized EM algorithm to estimate restricted and unrestricted maximum likelihood estimates. For generalized linear mixed-effects models, McCulloch [6] provides some details on using an extension of the EM algorithm for parameter estimation.

Because of the need to have a flexible modeling approach to model either unbalanced or balanced longitudinal data, particularly, in temporal relationships in biological and pharmacokinetics data, use of non-linear mixed effect models have been increasing. For example, Wu and Ding [7] proposed a nonlinear temporal trend of viral load after antiretroviral therapy. Most of the literature focuses on the case of continuous longitudinal data, because of computational difficulties in modeling categorical data. Sheiner and Beal [8] proposed a non-linear one compartment model to describe plasma concentration over time. A thorough review of non-linear models and estimation methods can be found, for example, in Davidian and Giltinan [9] or Vonesh and Chichilli [10]. Parameter estimation in nonlinear random effects models can be carried out by maximum likelihood estimation, which can be implemented either using exact or approximation methods. See, for example, Pinheiro and Bates [11] Vonesh [12], and Davidian and Giltinan [9, ch 6.] for further details.

In Section 2, we introduce a multiphase mixed effects model and discuss the motivation and objectives of this approach. We discuss the estimation strategy in section 3. In Section 4, we illustrates the application of this model using spirometry data after lung transplantation. In section 5, we perform a simulation study to verify the model estimation, and brief concluding remarks are given in Section 6.

2 A Multiphase Non-Linear Mixed Effects Model

Let Y_{ij} be the j^{th} ($j = 1, \dots, n_j$) continuous response for subject i ($i = 1, \dots, m$). We now consider the following non-linear mixed effect model,

$$y_i = \mathbf{X}_{io}\beta_o + \text{l\ddot{o}g} \left(\sum_{l=1}^L \mu_l(\mathbf{X}_{il}; \beta_l, b_{il}) T_l(\mathbf{t}_i, \boldsymbol{\Theta}_l) \right) + \varepsilon_i, \quad (2)$$

where, $\text{l\ddot{o}g}(\cdot)$ is a $n_j \times 1$ vector with $\log(\sum_{l=1}^L \mu_l(\mathbf{X}_{il}; \beta_l, b_{il}) T_l(t_{ij}, \boldsymbol{\Theta}_l))$ as its j^{th} element; $\mathbf{t}_i$ is a $n_j \times 1$ vector of time points at which longitudinal response vector $\mathbf{y}_i$ is observed or measured for subject i ; $\mathbf{X}_{i\beta}$ are matrices of design matrix of fixed effects that are not necessarily equivalent, and β_{β} are the corresponding fixed-effect regression parameters. Since the functions $\mu_l(\mathbf{X}_{il}, \beta_l, b_{il})$ and $T_l(\mathbf{t}_i, \boldsymbol{\Theta}_l)$ (which are described below) are positive, we have used logarithmic transformation of the additive components. The subject-specific random effect vector $\mathbf{b}_i = (b_{i1}, \dots, b_{iL})^T$ is assumed to follow a L -dimensional multivariate normal distribution, $N(\mathbf{0}, \mathbf{G})$. Positive (or negative) associations between the random effects essentially quantify how well the responses in one time phase are associated with other time phases. It is further assumed that $\varepsilon_i \sim_{iid} N(0, \sigma_{\varepsilon}^2 \mathbf{I})$. That is, we assume conditional independence.

$T_l(t, \boldsymbol{\Theta}_l)$ is a function of time that depends only on time t and a shaping parameter vector $\boldsymbol{\Theta}_l$, and this function can be any of the forms or transformations of (4) given below. The model (2) has two main fixed components:

1. An overall model (namely, $\mathbf{X}_{io}\beta_o$) that does not depend on time;
2. A series of log-linear mixed effect models $\mu_l(\mathbf{X}_{il}, \beta_l, b_{il}) = \exp\{\mathbf{X}_{il}\beta_l + b_{il}\}$ ($l = 1, \dots, L$) that are modulated by the time functions $T_l(t, \boldsymbol{\Theta}_l)$. Hence, the effects of covariates $\mathbf{X}_{i\beta}$ are time varying.

Note that model (2) can be written as

$$\text{e\ddot{x}p}\{E(\mathbf{y}_i | \mathbf{b}_i)\} = \text{e\ddot{x}p}\{\mathbf{X}_{io}\beta_o\} \times \sum_{l=1}^L \mu_l(\mathbf{X}_{il}; \beta_l, b_{il}) T_l(t, \boldsymbol{\Theta}_l), \quad (3)$$

where $\text{e\ddot{x}p}\{\cdot\}$ is a vector with n_j elements. Hence, multiple overlapping time phases of outcome are additive in the conditional expectation domain, with each phase individually shaped by a function of time $T_l(t, \boldsymbol{\Theta}_l)$ and scaled by a function of concomitant information

$\mu(\mathbf{X}_{ik}; \beta_i, b_{ii})$. Note that we can use as many phases as the data warrants, but in our data analysis experience, at most 2 phases – early and late – are usually adequate and rarely 3 phases – an early, constant, and a late.

Following the spirit of Blackstone and colleagues' [13] multiphase parametric hazard modeling strategy to model time-to-event data, the motivation for the nonlinear mixed effect model (2) is two fold:

1. In medical sciences, the temporal trend of a longitudinal biological response of interest plays an important role. It essentially describes how a patient or a group of patients behave after an intervention and most of the post-intervention patient management protocols are based on these temporal trends. Hence, it is important to model the temporal trend of a longitudinal response accurately. It is also known that most of the biological responses (such as, pharmacokinetics, lab results from blood samples, biomarkers [14], [15], [16]) are nonlinear in trend. The additive flexible linear and non-linear components in (2) can handle most non-linear and linear trends.
2. It is known that the magnitude of the effect of a risk factor on a longitudinal response can change with time. Especially, after a medical intervention (for e.g., surgery), different risk factors can influence the longitudinal response at different time segments. For example, after a mitral valve repair, the type of mitral valve ring and experience of the surgeon may influence the early measurements of ejection fraction, say, measured within one month of surgery. However, patient comorbidities, such as diabetes and history of smoking may be two of the risk factors that influence late ejection fraction, say, one year after the repair. Older age, on the other hand, may be a predictor for ejection fraction regardless of the time. Hence, although the effect of some factors on a longitudinal response may diminish and become negligible or start increasing and become noticeable after some time, some other factors' effect may stay the same regardless of the time. In other words, there is an interaction effect between some factors and time on the longitudinal response. The model (2) can accommodate this scenario.

It is noted that our model is similar to, but more flexible than, a model proposed by Wu [17]. Wu [17] proposed a focused multiphase exponential decay model to fit a specific shape of temporal trend of virus load in an AIDS study; our model is more flexible, in the sense that it can handle more complex shapes and more than two phases. Model (2) is similar to an additive hazard regression model proposed by Blackstone et al. [13]. The effect of covariates change with time in model (2), similar to time-varying coefficient models. Fahrmeir et al. [18] used a Bayesian version of penalized splines in additive regression to model non-linear effect of covariate and time trends. Many have studied time-varying coefficient models through a non-parametric approach (Wu and Yu [19], Wu and Zhang [20]). We now briefly describe the time function and the 3 phases in the following sub sections.

2.1 Multi-Phase time function: $T(t, \Theta)$

We describe the time function and some commonly used special cases of the function.

2.1.1 Mathematical Formulation—The generic equation of the time for the multiphase model was originally described as a model of cumulative mortality by Hazelrig and colleagues [21] and then used in a multiphase hazard model to fit time-to-event data by Blackstone and colleagues [13]. We transform these equations into the conditional expectation domain. The family of equations is given as:

$$G(t, \Theta) = \frac{|\nu| - \nu}{2|\nu|} + \frac{\nu}{|\nu|} \left[1 + \phi(m) \left(\frac{|m| - m}{2|m|} + \frac{|\nu|t}{\rho(t_{1/2})} \right)^{-1/\nu} \right]^{-1/m}, \quad (4)$$

where $m > 0$ and/or $\nu > 0$, $\phi(m) = m$ if $m > 0$, and $\phi(m) = -1$ if $m < 0$. $\Theta \equiv (m, \nu, t_{1/2})$ is the shaping parameter, $\rho(t_{1/2})$ is a function of $t_{1/2}$, m and ν . By defining $t_{1/2}$ by the equation $G(t_{1/2}) = 1/2$, one can deduce the function $\rho(t_{1/2})$ in terms of $m, \nu, t_{1/2}$. In survival terminology, Blackstone and colleagues [13] called $t_{1/2}$ as the *half life time*. Natural constraints on G are that $G(0, \Theta) = 0$ and $G(t, \Theta) \rightarrow 1$ as $t \rightarrow \infty$.

The first derivative of G with respect to t is

$$g(t, \Theta) = \left[1 + \phi(m) \left(\frac{|m| - m}{2|m|} + \frac{|\nu|t}{\rho(t_{1/2})} \right)^{-1/\nu} \right]^{-\frac{1}{m}-1} \left(\frac{|m| - m}{2|m|} + \frac{|\nu|t}{\rho(t_{1/2})} \right)^{-\frac{1}{\nu}-1} \frac{\phi(m)}{\rho(t_{1/2})m},$$

which is non-negative. Thus, G is non-decreasing.

Note that, when $m < 0$ and $\nu < 0$ $G(0, \Theta) > 0$, and this violates the constraints. Thus, this generic formulation (4) does not exist for $m < 0$ and $\nu < 0$. Hence, the generic formulation (4) simplifies into three cases, depending on the signs of m and ν :

Case 1: $m > 0$ and $\nu > 0$:

$$G(t, \Theta) = \left[1 + m \left(\frac{\nu t}{\rho(t_{1/2})} \right)^{-1/\nu} \right]^{-1/m},$$

and the limiting case as $m \rightarrow 0^+$ is

$$G(t, \Theta) = \exp \left[- \left(\frac{\nu t}{\rho(t_{1/2})} \right)^{-1/\nu} \right],$$

Case 2: $m < 0$ and $\nu > 0$:

$$G(t, \Theta) = \left[1 - \left(1 + \frac{\nu t}{\rho(t_{1/2})} \right)^{-1/\nu} \right]^{-1/m},$$

and the limiting case as $\nu \rightarrow 0^+$ is

$$G(t, \Theta) = \left[1 - \exp \left(-\frac{t}{\rho(t_{1/2})} \right) \right]^{-1/m},$$

Case 3: $m > 0$ and $\nu < 0$:

$$G(t, \Theta) = 1 - \left[1 + m \left(\frac{-\nu t}{\rho(t_{1/2})} \right)^{-1/\nu} \right]^{-1/m},$$

and the limiting case as $m \rightarrow 0^+$ is

$$G(t, \Theta) = 1 - \exp \left[- \left(\frac{-\nu t}{\rho(t_{1/2})} \right)^{-1/\nu} \right],$$

2.2 Model components

We briefly describe three commonly occurring phases in the following sub sections.

2.2.1 Overall model— $\mathbf{X}_o \beta_o$ is a model that identifies the risk factors that are related to the subject-specific mean response in an overall fashion and do not involve time t .

2.2.2 Early phase—The time function $T_f(t, \Theta) = g(t, \Theta)$, the first derivative of $G(t, \Theta)$ is the most commonly used function in this phase. Under this scenario, the scaling parametric function $\mu_f(\mathbf{X}_f, \beta_f, b_{if})$ is related to the area beneath the function $g(t, \Theta)$ for subject i . This was the reason that we made the concomitant function $\mu_f(\mathbf{X}_f, \beta_f, b_{if})$ to be positive. Various shapes of g for different values of m and ν are given in Figure 1. In Figure 1, while we vary m and ν , we keep the $t_{1/2}$, the time at which the area under curve is $1/2$ fixed at 3 months.

Note that, for the early peaking function, by changing $t_{1/2}$, one can change the peaking point.

2.2.3 Constant phase—This phase is time independent. Hence, it will only have concomitant information, $\mu_f(\mathbf{X}_f, \beta_f, b_{if})$. Note that, when $\nu = -1$ and $m = 0$, $h(t, \Theta)$ described below changes to a constant phase (Figure 2).

2.2.4 Late phase—The most commonly used function for the late phase is

$$T_l(t, \Theta_l) = h(t, \Theta_l) = \frac{g(t, \Theta_l)}{1 - G(t, \Theta_l)} \quad (5)$$

Hazellrig and colleagues [21] used this transformation as a hazard function and their motivation for this transformation is that, suppose $G(t, \Theta)$ is a *CDF* and $g(t, \Theta)$ the corresponding *pdf* then $h(t, \Theta)$ is a *hazard function*. Note that, since $h(t, \Theta)$ is of the form

$\frac{1}{t_{1/2}} f\left(\frac{t}{t_{1/2}}\right)$. For identifiability, we fix the intercept in $\mu(\mathbf{X}_i, \beta_i, b_{il})$ to 0. Further note that, $h(t, \Theta)$ becomes a constant phase when $m = 0$ and $\nu = -1$. Four different shapes of $h(t, \Theta)$ for different values of m and ν are given in Figure 2.

3 Estimation

Estimation of fixed effects parameters and parameters of the variance covariance matrices is obtained by the method of maximum likelihood estimation. Let $\beta = (\beta_0, \beta_1, \dots, \beta_L)$, $\Theta = (\Theta_1, \dots, \Theta_L)$ and $\theta = (\beta, \Theta, \sigma_\varepsilon^2, \mathbf{G})$, then the i^{th} subject's contribution to the likelihood is given by

$$\begin{aligned} \mathcal{L}_i(\theta | \mathbf{y}_i) &= \int_{\mathcal{R}^L} f(\mathbf{y}_i, \mathbf{b}_i | \theta) d\mathbf{b}_i \\ &= \int_{\mathcal{R}^L} \prod_{j=1}^{n_i} f_{y|\mathbf{b}}(y_{ij} | \mathbf{b}_i; \beta, \sigma_\varepsilon^2, \Theta) f_{\mathbf{b}}(\mathbf{b}_i; \mathbf{G}) d\mathbf{b}_i \\ &= \int_{\mathcal{R}^L} (2\pi)^{(n_i+L)/2} \sigma_\varepsilon^{-n_i} |\mathbf{G}|^{-1/2} \exp\left\{-[\mathbf{y}_i - \tilde{f}(\beta, \Theta, \mathbf{b}_i)]^\top [\mathbf{y}_i - \tilde{f}(\beta, \Theta, \mathbf{b}_i)] / 2\sigma_\varepsilon^2 - \mathbf{b}_i^\top \mathbf{G}^{-1} \mathbf{b}_i / 2\right\} d\mathbf{b}_i, \end{aligned}$$

where $\tilde{f}(\beta, \Theta, \mathbf{b}_i) = \mathbf{X}_{io}\beta_o + \log\left(\sum_{l=1}^L \mu_l(\mathbf{X}_{il}; \beta_l, b_{il}) T_l(t_i, \Theta_l)\right)$ and $f_{y|\mathbf{b}}$ is the conditional density of the longitudinal response, and $f_{\mathbf{b}}$ is the density of the random effects. Then the marginal likelihood is given by

$$\mathcal{L}(\theta, \mathbf{y}) = \prod_{i=1}^m \mathcal{L}_i(\theta, \mathbf{y}_i), \quad (6)$$

Equation (6) is simply the marginal distribution of $\mathbf{Y}$ obtained by integrating the joint distribution of $\mathbf{Y}$ and $\mathbf{b}$ with respect to $\mathbf{b}$. For non-linear models, where the random-effects $\mathbf{b}_i$ enters the model non-linearly, the integral in the marginal likelihood does not have a closed form. That is, except for some special cases, the integral in (6) does not have a closed form. Hence, first some numerical methods, such as, for example, numerical integration or Monte-Carlo integration technique, may have to be implemented to evaluate the integral before maximizing the marginal likelihood, again using some numerical methods such as the Newton-Raphson method. We use Laplace approximation to evaluate the integral in (6). Laplace approximation is essentially a second-order Taylor-series approximation to the integrand in (6) with respect to some estimate of random effects $\mathbf{b}$, usually an empirical bayes estimate of $\mathbf{b}$ (Pinheiro and Bates [11]). Note that while Wolfinger [22] expanded the

integral around both $\hat{\mathbf{b}}$ and $\hat{\beta}$, Pinheiro and Bates [11] expanded only around $\hat{\mathbf{b}}$. We have implemented this estimation process using PROC NLMIXED (SAS, SAS Institute, Cary NC). The implementation of this SAS procedure is based on Pinheiro and Bates [11].

4 Data Analysis

Continuous monitoring to assess lung function after lung transplantation is an important part of patient management. The advantages in using double lung transplantation over single lung transplantation is debated (Sundaresan et al. [23]). Three hundred and seventy-nine patients who underwent lung transplant at Cleveland Clinic between 1990 and 2005 are considered in this data analysis. Average age of the study cohort is 49 years, 51% were male and 44% had double lung transplantation. A total of 6372 evaluations of postoperative forced vital capacity (FVC) were obtained over a period of 7 years; mean follow-up was 1.5 years with 5% of the evaluation were obtained after 5.2 years. The main objective of this analysis is to explicitly model the temporal trend of FVC and to assess the temporal difference in FVC between single lung transplantation and double lung transplantation patients. We further, for illustration purpose, perform a limited multivariable analysis to assess the following covariate effect on FVC; recipient age at the time of lung transplantation, gender, indication for transplantation (chronic obstructive pulmonary disease (COPD) vs other), and preoperative FVC. Note that, because the FVC is a measure of lung function and depends on the patient size, the raw measures of FVC are adjusted (normalized) using National Health Examination Survey (NHANES) equations (Hankinson et al. [24]).

We first focus on the temporal trend of the longitudinal outcome and then on baseline covariates that may influence the trend. Thus, we first consider model (2) without covariates and determine the number of phases L and estimate the shaping parameter vector Θ for each phase. With only time, the model (2) can be written as,

$$\begin{aligned} y_i &= \log \sum_{l=1}^L \mu_l(\beta_{0l}, b_{il}) T_l(t_i, \Theta_l) + \varepsilon_i \\ &= \log \left(\sum_{l=1}^L \exp(\beta_{0l}, b_{il}) T_l(t_i, \Theta_l) \right) + \varepsilon_i, \end{aligned} \quad (7)$$

where β_{0l} is phase-specific intercept (fixed effect) and b_{il} is patient-specific random intercept for phase l .

The data suggested a bi-phase model for the temporal trend of FVC. Estimates of the shaping parameters are given in the Table 1.

It can be noted here that the estimated covariance between subject-specific random effects for each phases, b_{1i} and b_{2i} appears to be different from zero with a moderate correlation of 0.33. This suggests that the late postoperative values of FVC are positively influenced by the early post-op values of FVC. Based on the estimates in Table 1, the estimated multiphase temporal trend equation for FVC can simplified as follows:

$$E(\widehat{FVC}_i^Z | \mathbf{b}_i) = \log \left[\exp\{0.35 + b_{1i}\} \frac{0.85t^{0.23}}{\exp\{0.69t^{1.23}\}} + \exp\{0 + b_{2i}\} \frac{2.56\{\exp(0.96t) - 1\}^{-1}}{\{\exp(0.96t)\}^{2.67} \{\exp(0.96t) - 1\}^{-2.67} - 1} \right],$$

(8)

where $FVC_i^Z = \frac{FVC_{i-60}}{20}$, and $\mathbf{b}_i^\top = [b_{1i}, b_{2i}]$, the patient-specific random effects for the phases, are assumed to be a bivariate normal. The multiphase model yielded an early peaking and a late plateau phase (Figure 3). The early phase is from equation $\mathcal{T}(t, \Theta) = g(t, \Theta)$ with limiting case 3 and the late phase is from equation $\mathcal{T}(t, \Theta) = h(t, \Theta)$ with limiting case 2.

Figure 4 shows the estimated patient-specific profiles of the 150 patients and the average of the patient-specific profiles of all 379 patients. It can be noted from the patient specific profiles, by introducing phase-specific random effects, the non-linear trends can have different starting points with different non-linear shapes. This is similar to a linear mixed-effect model with random intercept and random slope, however, with more flexible non-linear shapes as the data warrants.

The temporal trend of postoperative FVC of the study cohort is shown in Figure 5. The temporal trend for the “average patient” or “typical patient” is obtained by setting $(b_{1i}, b_{2i})^\top = [0, 0]^\top$. Average profile is obtained by averaging the patient-specific profiles. Binned averages are obtained by using a bin smoother as follows: Based on the time of the measurement, we first partitioned the response values into a number of disjoint groups and taking mean of the response of interest, in this case, mean of FVC. Note that, when binning, the possibility of a given bin having multiple observations from the same patient was not taken into the consideration. Hence, binned average is used here as a crude verification of the model fit.

It is noted from Figure 5 that the temporal trend of the average patient and the trend based on the average profile appear to be similar early and diverge later. The average profile appears to follow the bin average closely.

Remark: Identifying phases and estimation of the temporal trend is a data driven process. That is, based on the binned smoothers, we first get an overview of the shape of the temporal non-linear (if any) trend of the postoperative FVC. We then, for this data, started with 3 phases: early ($g(t, \Theta)$), constant and late ($h(t, \Theta)$).

1. We started with $m = 1$ and $v = 1$ for the early and late phase, with a small $t_{1/2} = 0.1$ for the early and a large $t_{1/2} = 5$ for the late phase, and estimated the intercept of the constant phase. A non-significant very small intercept value for the constant phase is observed for these data. Based on this observation, we have removed the constant phase from the model.

2. Now, keeping $m = 1$ and $\nu = 1$ for the late phase, we have tried 3 possible combination of starting values for m and ν $((1,1), (-1,1), (-1,1))$ for the early phase and using PROC NLMIXED and observed the convergence and likelihood estimates under these 3 scenarios. Based on the convergence and likelihood values (larger ones), it is noted that $m = 0$ and $\nu < 0$ provide a best fit for the early phase.
3. Now, keeping $m = 0$, $\nu = -1$ and $t_{1/2} = 0.1$ for the early phase, we have tried 3 possible combination of starting values, as described above, for the late phase. Based on the convergence and likelihood values (larger ones), it is noted that $m < 0$ and $\nu = 0$ provide a best fit for the late phase.
4. We now using $m = 0$, $\nu = -1$ and $t_{1/2} = 0.1$ for early phase and $m = -1$, $\nu = 0$ and $t_{1/2} = 5$ for late phase, we have obtained the final estimates given in Table 1.

4.1 Model selection

We compare the model (8) with the following two simpler alternate models:

Alternate Model 1—The distribution of the random effect vector $\mathbf{b}_i^T = [b_{1i}, b_{2i}]$ is that of two independent normal variates. That is, the variance covariance matrix is a diagonal matrix.

$$Var(\mathbf{b}) = \mathbf{G} = \begin{pmatrix} \sigma_{b_1}^2 & 0 \\ 0 & \sigma_{b_2}^2 \end{pmatrix}.$$

This alternative model effectively assumes that the early postoperative FVC values have no influence on the late postoperative FVC values.

Alternate Model 2—Instead of having subject-specific random-effects for each phases as in (7), we consider a model where a single random effect enters the model linearly as follows:

$$\mathbf{y}_i = \log\left(\sum_{l=1}^L \exp\{\beta_{0l}\} T_l(t, \boldsymbol{\Theta}_l)\right) + b_i + \varepsilon_i.$$

Note that, alternative model 2 suggests a compound symmetry correlation structure for the continuous longitudinal response.

We use Akaike's information criteria (Akaike [25]) (AICc - corrected version) to compare the 3 models. The estimated AICc for the 3 models are given in Table 2. Based on the AICc values, model (8) which has subject-specific random effects for each phase, is better than alternate model 1 where the random effects are assumed independent Normal variates, and clearly better than the alternate model 2 which has one common random effect that enters the model linearly.

4.2 Goodness of Fit

Information criteria used in the previous section are only a tool for comparing different models. That is, given more than one possible model, using information criteria, one can choose a better model. However, they do not test or provide insight into how well a given model fits the data. Vonesh et al. [26] proposed a measure of concordance between predicted and observed response as a measure of goodness-of-fit for response function. This measure is an extension of concordance correlation proposed by Lin [27] and is given by,

$$r_c = 1 - \frac{\sum_{i=1}^m (\mathbf{y}_i - \hat{\mathbf{y}}_i)^\top (\mathbf{y}_i - \hat{\mathbf{y}}_i)}{\sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}}\mathbf{1}_i)^\top (\mathbf{y}_i - \bar{\mathbf{y}}\mathbf{1}_i) + \sum_{i=1}^n (\hat{\mathbf{y}}_i - \hat{\mathbf{y}}\mathbf{1}_i)^\top (\hat{\mathbf{y}}_i - \hat{\mathbf{y}}\mathbf{1}_i) + N(\bar{\mathbf{y}} - \hat{\mathbf{y}})^2}, \quad (9)$$

where $\bar{\mathbf{y}}$ is the grand mean of the observed y_{ij} , $\hat{\mathbf{y}}$ is the grand mean of the predicted $\hat{y}_{ij} = E(\hat{Y}_{ij} | \mathbf{b}_i)$, $\mathbf{1}_i$ is the $n_i \times 1$ unit vector of 1's and $N = \sum n_i$ is the total number of observations. Vonesh [26] pointed out the following advantages in using r_c as the concordance correlation coefficient between observed and predicted values: it directly measures the level of agreement between observed and predicted; unlike R^2 , we do not need a null model for reference, line of identity serves as the point of reference; r_c ranges, $-1 \leq r_c \leq 1$, with perfect fit at 1 and lack of fit at 0.

The estimated concordance correlation based on the temporal model (8) is 0.93 which suggests a very good fit. This conclusion is affirmed by the close (to line of identity) fit in patient-specific observed profiles versus the predicted profiles plot given in Figure 6.

4.3 Double versus Single Lung Transplantation

The main purpose of the proposed multiphase non-linear mixed effect model is to assess the effect of single versus double lung transplantation on the postoperative lung function and to determine if the effect is time-varying. Estimated patient-specific profiles based on the model (8) for the patients who received single lung versus the double lung is shown in Figure 7.

It can be noted from the patient-specific profiles that there is an initial increase in FVC in both groups, and thereafter, in general, the FVC profiles of double lung recipients appear to stay constants. On contrast, FVC profiles of the single lung recipients appear to be decreasing. This suggests the effect of double versus single on the post transplant FVC may be time varying. We now can quantify the phase-specific effect by simply forcing-in the binary variable (double versus single lung transplantation) in the both phases of the temporal model (8). The estimates are given in Table 3.

From Table 3, it is noted, that single versus double lung transplantation does not have any significant effect on the early values of the FVC; however, patients with double lung transplantation have a higher late postoperative FVC. The average of profiles stratified by single versus double lung transplantation is given in Figure 8.

It can be seen from Figures 7 and 8 that the difference in FVC between double versus single lung transplantation increases with time. This suggests that the postoperative lung function in double lung recipients remained relatively constant after 2 years, lung function in single lung recipients keeps deteriorating. That is, the effect coefficient is time-varying and is easily estimated for each time phase and depicted using our multiphase decomposition model.

4.4 Covariate effect on the temporal change

For simplicity, we have considered the following 5 variables in the multivariate analyses: Age, Gender, single vs. double lung transplantation, preoperative FVC and reason for lung transplantation: COPD (vs others). The main objective of this exercise is to illustrate implementation of our model to identify overall and phase-specific variables that are associated with post transplant FVC. We performed multivariable analysis to determine which variables influence the early phase, which influence the late phase, and which influence both phases at an equal magnitude (at least approximately), so that these variables can be considered as overall common covariates. Because there is no built-in variable selection algorithm in PROC NLMIXED, we have performed the following ad-hoc forward selection strategy in the data analysis to identify the phase-specific and overall covariates: we first force in a variable in each phase to see if it is significant in that phase. If it is, we keep that variable in that phase. On the other hand, if that variable is significant in both phases and the parameter estimates are at least approximately equal, we keep it in the overall phase. We repeat this process until we have considered all the variables in the model.

Phase-specific and overall effects of the selected covariates are given in Table 4. Younger age associated with higher postoperative FVC, gender and COPD disease are not. Preoperative FVC appeared to have a positive effect on the postoperative FVC, and the magnitude of the effect appears to be higher on the early postoperative FVC values than on the late FVC values. Further note, after adjusting for age, gender, pre-operative FVC, having single lung transplant is associated with higher early FVC values and having double lung transplant is associated with higher late FVC values. That is, magnitude and direction of the effect changed with time.

5 A Simulation Study

We now perform a simulation study to assess the estimates of shaping parameters and of regression coefficients of the overall and phase-specific covariates. The objectives of this focused simulation study are two-fold. As described in the motivation of this multi-phase temporal model, first, we would like to assess if we can generate the true underlying phases of the the temporal model without considering the covariates in the model. That is, in this step, we would like to assess the shaping parameters of the model without the covariates in the model. We then, in the second step, would like to assess the shaping parameters and the regression coefficients of the covariates in the model. It should be noted that in this simulation study, we are not focusing on model building, rather, given a model, we would like to assess how well the parameters are estimated.

5.1 Simulation model

We generate a continuous longitudinal response for 250 subjects at the following 13 time points over a 5-year time period: 1-day, 2-day, 3-day, 1-week, 2-week, 1-month, 3-month, 6-month and 5 time points at year 1 to year 5. We have generated 500 simulated datasets with sample size of 400.

Phases of the temporal trend—For the temporal trend, we assume a bi-phase model with $T(t, \Theta) = g(t, \Theta)$, with $\Theta = (m = 0.5, \nu = 0.5, t_{1/2} = 0.25)$, as the early phase (Case 1 in Section 2.1.1) and $T(t, \Theta) = h(t, \Theta)$, with $\Theta = (m = 0, \nu = -0.75, t_{1/2} = 6)$, as the late phase (limiting Case 3 in Section 2.1.1). Equation $g(t, \Theta)$ with the selected values for the shaping parameters shows an early peaking function and $h(t, \Theta)$ with the selected values gives a late increasing function (Figure 9).

Multi-phase multivariable model—We then simulate the following 2-phase temporal model with 4 covariates as follows:

$$y_i = 0.75x_{4i} + \log\{\exp(-1 + 1x_{1i} - 0.5x_{2i} + b_{1i})g(t_i, \Theta) + \exp(0.5x_{1i} + 0.5x_{3i} + b_{2i})h(t_i, \Theta)\} + \varepsilon_i, i=1, \dots, m. \quad (10)$$

The values of the shaping parameters for the functions $g(t, \Theta)$ and $h(t, \Theta)$ are assigned as described above (Figure 9). For simplicity, we generate 4 binary covariates as follows: $x_1 \sim \text{Binary}(0.6)$, $x_2 \sim \text{Binary}(0.5)$, $x_3 \sim \text{Binary}(0.3)$, and $x_4 \sim \text{Binary}(0.4)$; x_4 is an overall covariate that positively influences the response y regardless of the time; while x_1 influences early values of y positively, it is negatively associated with late values of y ; x_2 is negatively associated with only early values of y ; and x_3 is positively associated with only late values of y . The random effect vector $\mathbf{b} = [b_1, b_2]^\top$ and the measurement error ε_i are distributed as follows,

$$\mathbf{b}_i \sim \mathcal{N}_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix} \right), \text{ and } \varepsilon_i \sim \mathcal{N}_{\text{iid}}(0, 0.25).$$

Positive $\text{Cov}(b_1, b_2) = 0.5$, implies that early values of y and late values of y are positively associated.

5.2 Simulation results: Temporal model

We first fitted the simulated data set to estimate the shaping parameters. Note that in this step, we fitted a model without covariates in the model.

The distribution of the estimates of the shaping parameters from the 500 simulated dataset are given in Figure 10. It can be seen that the estimated values are very close to the true values. Note that, when one uses $h(t, \Theta)$ as a time function the parameter $t_{1/2}$ acts as a scaler.

Hence, in our simulated model, we have $(m, \nu, t_{1/2})$ as shaping parameters in the early phase and, after setting $m = 0$, only ν as the shaping parameter in the late phase.

It is evident from Figure 10 that even though the data were simulated using the multi-phase model with 4 covariates (10), shaping parameter estimates based on the temporal model without covariates correctly estimates the underlying shaping parameters. The average of the estimates and the true value are nearly equal. The implication is that these parameters can be fixed during covariate selection and this would improve the computational speed.

5.3 Simulation results: Temporal model with covariates

We now assess the performance of the multi-phase non-linear mixed effects model based on simulated data using the following summary measures: suppose there are B simulated datasets and β is the true value and $\hat{\beta}_i$ is the estimate from the i^{th} simulated dataset;

$Bias = \beta - \bar{\hat{\beta}}$, where $\bar{\hat{\beta}} = \sum_{i=1}^B \hat{\beta}_i / B$; Average within standard error,

$AvgSE = \sum_{i=1}^B SE(\hat{\beta}_i) / B$; Empirical standard error,

$EmpSE = \sqrt{[1/(B-1)] \sum_{i=1}^B (\hat{\beta}_i - \bar{\hat{\beta}})^2}$; 95% coverage probability, CP which is defined as the fraction of times the true value lies within $(\hat{\beta}_i - 1.96SE(\hat{\beta}_i), \hat{\beta}_i + 1.96SE(\hat{\beta}_i))$.

The summary measures of the shaping and regression coefficients of the covariates based on the 500 simulated data with sample size 400, are given in Table 5.

All the estimated parameters are close to the true values of the parameters and in some cases the average of the estimates are almost equal to the true values. This suggests good performance of our non-linear mixed effect model in estimating the parameters.

5.4 Varying Sample Size

In this simulation scenario, we compare the influence of varying sample size on the parameter estimation of multiphase model (8).

It can be noted from Table 6, that in general, as sample size decreases there appears to be some increase in bias and standard error of the parameter estimates. This is especially apparent in the estimation of association parameters of the random effects, $\sigma_{b_2}^1, \sigma_{b_2}^0$ and $Cov(b_1, b_2)$. This feature is more apparent when the sample size is 100. One can speculate that because the number of parameters to be estimated for our model is relatively high (there are 15 parameters), as the sample size decreases, efficiency may decrease. That is, the asymptotic properties of the estimates are less likely to hold.

6 Conclusion

This article discusses analysis of continuous longitudinal data using a temporal decomposition mixed effects model. The generic family of equations (Hazelrig and colleagues [21]), originally used as hazard functions for modeling time to event data, provides a flexible system of mathematical functions for the fitting of non-linear temporal pattern. Once the model has been identified, fitting of it is relatively straightforward.

We have explicitly modeled the non-linear temporal pattern. One of the main advantages of our modeling is that we have applied our model using readily accessible standard software (see digital supplement for a sample SAS code). In our data analysis experience, we encounter two-phase model the most frequently, and very rarely, a three-phase model. We have never encountered a model with four- or more phases. However, if one were to encounter a model with four or more phases, one has to pay more attention to the estimation process which involves a four or more dimensional numerical integration, which may be computationally expensive or may not even be implemented using readily available software.

As demonstrated in data analysis, an advantage of formal parametric decomposition of temporal pattern and identifying time-specific risk factors is targeted patient management based on short- and long-term risk factors. For example, because it is known that single lung transplantation patients appear to have deteriorating lung function approximately 2 years after surgery, treating physicians can advise such patient to have more periodic follow-ups, say, 2 years after the surgery. That is the effect size of single versus double lung transplantation on post transplant lung function (FVC) is time-varying. An advantage over the existing non-parametric time-varying coefficient modeling approaches is that our parametric modeling approach can handle a large number of covariates, as is nearly always the case in observational studies. It should be noted that currently, our model does not handle time dependent covariates.

The main limitation of performing multivariable analysis using model (2) is that we have to use an ad-hoc method for variable selection. Variable selection in longitudinal models, particularly in nonlinear mixed-effects model is a possible area of future research. Recently, there has been some studies on variable selection in longitudinal models using cross-validation and machine learning methods. For example, Cantoni et al. [28] proposed a cross-validation technique as a variable selection tool in a marginal longitudinal model.

Model (2) can be simply extended to accommodate non-normal longitudinal responses by changing the conditional distribution in (2). We have used Laplacian approximation to obtain the marginal likelihood. Laplacian approximation to the integration works very well when we have normally distributed continuous longitudinal responses, however, it's effectiveness in applying to categorical responses is questionable [29] and [30]. In this case, for example, one has to rely on classical numerical integration methods such as Gaussian-Quadrature.

The model (2) can be extended to a multivariate setup to handle more than one longitudinal responses ([31], [32]). The lung transplantation study that we have considered in this paper actually has two longitudinal measurements: forced expiratory volume in 1 second (FEV1) and forced vital capacity (FVC). Both measurements are makers for lung function. One may wish to estimate the evolution of association between these two longitudinal measurements [33], by modeling jointly one can increase efficiency of parameter estimation ([34], [35], [36]).

Acknowledgement

The first author would like to thank Dr. Mark Schluchter, Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, for his valuable comments. This research was partially supported by a grant from National Institutes of Health - NATIONAL HEART, LUNG, AND BLOOD INSTITUTE, Grant Number: 1R01HL103552.

This research was partially supported by a grant from National Institutes of Health - NATIONAL HEART, LUNG, AND BLOOD INSTITUTE, Grant Number: 1R01HL103552.

References

1. Banbury MK, Cosgrove DM, Thomas JD, Blackstone EH, Rajeswaran J, Okies JE, Frater RM. Hemodynamic stability during 17 Years of the Carpender-Edwards Aortic Pericardial Bioprosthesis. *Annals of Thoracic Surgery*. 2002; 73:1460–1465. [PubMed: 12025820]
2. Diggle, PJ.; Heagerty, PJ.; Liang, K.; Zeger, SL. *Analysis of Longitudinal Data*. 2nd edition. New York: Oxford University press; 2002.
3. Laird NM, Ware JH. Random-Effects Models for Longitudinal Data. *Biometrics*. 1982; 38:963–974. [PubMed: 7168798]
4. Lee Y, Nelder JA. Conditional and Marginal Models: Another View. *Statistical Science*. 2004; 19:219–238.
5. Jennrich RI, Schluchter MD. Unbalanced Repeated Measures Models with Structured Covariate Matrices. *Biometrics*. 1986; 42:805–820. [PubMed: 3814725]
6. McCulloch CE. Maximum likelihood algorithm for generalized linear mixed models. *Journal of American Statistical Association*. 1997; 92:162–170.
7. Wu H, Ding A. Population HIV-1 Dynamics in Vivo: Applicable Models and Inferential tools for Virological Data from AIDS Clinical Trials. *Biometrics*. 1999; 55:410–418. [PubMed: 11318194]
8. Sheiner LB, Beal SL. Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten Model: Routine clinical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics*. 1980; 8:553–571. [PubMed: 7229908]
9. Davidian, M.; Giltinan, DM. *Nonlinear Models for Repeated Measurement Data*. New York: Chapman & Hall; 1995.
10. Vonesh, EF.; Chinchilli, VG. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. London: Chapman and Hall; 1997.
11. Pinheiro JC, Bates DM. Approximations to the Log-Likelihood Function in a Nonlinear Mixed-Effects Model. *Journal of Computational and Graphical Statistics*. 1995; 4:12–35.
12. Vonesh EF. A Note on the use of Laplace's Approximation for Nonlinear Mixed-Effects Models. *Biometrika*. 1996; 83:447–452.
13. Blackstone EH, Naftel DC, Turner ME Jr. The decomposition of time-varying hazard into phases, each incorporating a separate stream of concomitant information. *Journal of American Statistical Association*. 1986; 81:615–624.
14. Gelman A, Bois F, Jiang LM. Physiology Pharmacokinetic Analysis Using Population Modeling and Informative Prior Distributions. *Journal of American Statistical Association*. 1996; 91:1400–1412.
15. Muller P, Rosner GL. A Bayesian Population Model with Hierarchical Mixture Prior Applied to Blood count Data. *Journal of American Statistical Association*. 1997; 92:1279–1292.
16. Mikulich SK, Zerbe GO, Jones RH, Crowley TJ. Comparing Linear and Nonlinear mixed model approaches to Cosinor Analysis. *Statistics in Medicine*. 2003; 22:3195–3211. [PubMed: 14518023]
17. Wu L. A Joint Model for Nonlinear Mixed-Effects Models with Censoring and Covariates Measured with Error, With Application to AIDS Studies. *Journal of American Statistical Association*. 2002; 97:955–964.
18. Fahrmeir L, Kneib T, Lang S. Penalized Structured Additive Regression For Space time data: A Bayesian Perspective. *Statistica Sinica*. 2004; 14:731–761.

19. Wu CO, Yu KF. Nonparametric Varying-Coefficient Models for the Analyses of Longitudinal Data. *International Statistical Review*. 2000; 3:373–393.
20. Wu, H.; Zhang, JT. *Nonparametric Regression Methods for Longitudinal Data Analysis*. John Wiley & Sons; 2006.
21. Hazelrig JB, Turner ME Jr, Blackstone EH. Parametric Survival Analysis Combining Longitudinal and Cross-sectional Censored and Interval-Censored Data with Concomitant information. *Biometrics*. 1982; 39:1–5.
22. Wolfinger RD. Laplace approximation for non-linear mixed models. *Biomatrika*. 1993; 80:791–795.
23. Sundaresan RS, Shiraishi Y, Trulock EP, et al. Single or bilateral lung transplantation for emphysema? *Journal of Thoracic and Cardiovascular Surgery*. 1996; 112:1485–1494. [PubMed: 8975840]
24. Hankinson JL, Odencrantz JR, Fedan KB. Spirometry reference values from a sample of the general U.S. population. *American Journal of Respiratory Critical Care Medicine*. 1999; 159:179–187. [PubMed: 9872837]
25. Akaike H. Information Theory as an Extension of the Maximum Likelihood Principle. *Second International Symposium on Information Theory*. 1973:267–281.
26. Vonesh EF, Chincill VM, Pu K. Goodness-of-Fit in Generalized Nonlinear Mixed-Effects Models. *Biometrics*. 1996; 52:572–587. [PubMed: 10766504]
27. Lin LI. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*. 1989; 45:255–268. [PubMed: 2720055]
28. Cantoni E, Field C, Flemming JM, Ronchetti E. Longitudinal variable selection by cross-validation in the case of many covariates. *Statistics in Medicine*. 2007; 26:919–930. [PubMed: 16625521]
29. Pinheiro JC, Chao EC. Efficient Laplacian and adaptive Gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics*. 2006; 15:58–81.
30. Harry, Joe. Accuracy of Laplace approximation for discrete response mixed models. *Computational and Data Analysis*. 2008; 52:5066–5074.
31. Fieuws S, Verbeke G, Mollenbergs G. Random-effects models for multivariate repeated measures. *Statistical Methods in Medical Research*. 2007; 16:387–397. [PubMed: 17656450]
32. Verbeke G, Fieuws S, Mollenbergs G, Davidian M. The analysis of multivariate longitudinal data: A review. *Statistical Methods in Medical Research*. 2014; 23:42–59. [PubMed: 22523185]
33. Fieuws S, Verbeke G. Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. *Statistics in Medicine*. 2004; 23:3093–3104. [PubMed: 15449333]
34. Gueorguieva R. A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical Modelling*. 2001; 1(3):177–193.
35. Gueorguieva R, Sanacora G. Joint analysis of repeatedly observed continuous and ordinal measures of disease severity. *Statistics in Medicine*. 2006; 25:1307–1322. (2006). [PubMed: 16217846]
36. Fitzmaurice, G.; Davidian, M.; Verbeke, G.; Mollenbergs, G. *Longitudinal Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC Taylor & Francis Group; 2009. p. 317–391.

Figure 1.

Three cases of shapes of $g(t, \Theta)$: in all cases $t_{1/2} = 3$ months. Case I: $\nu = 0.5$ and $m = 0.5$, we have an early peaking function; Case II: $\nu = 0.5$ and $m = -1$, the function starts at a finite point and decreases; Case III: $\nu = -1.5$ and $m = 0.5$, we have a decreasing function starting at infinite.

Figure 2.

Four cases of shapes of $h(t, \Theta)$: in all cases $t_{1/2} = 6$ years. Case I: $\nu = 1.5$ and $m = 1.5$, we have an early decreasing function starting at infinite; Case II: $\nu = 0.5$ and $m = 0.5$, we have a late peaking function; Case III: $\nu = -0.75$ and $m = 0$, we have a late increasing function; and Case IV: $\nu = -1$ and $m = 0$, the function is a constant phase;

Figure 3.
Decomposition of the temporal trend of FVC.

Figure 4.

Thin lines depict estimated patient-specific profiles of 150 selected patients and the solid thick line depicts the average of the patient-specific profiles.

Figure 5. Temporal trend of FVC. Grey line is for “average patient”, black line is for average profile and the black closed circle symbols depict average actual data obtained by binning smoother.

Figure 6.

Goodness of Fit: Observed profiles versus Predicted profiles based on the temporal trend model (8).

Figure 7. Estimated patient-specific profiles. Solid (blue) lines depict the patient-specific profiles of the double lung recipients and dashed (red) lines depict the patient-specific profiles of the single lung recipients.

Figure 8.

Temporal trend of FVC stratified by Double vs. Single lung transplantation. Solid line depict the average of the patient-specific profiles for double lung recipients based on the parametric estimates in 3 and the dash line depict the average of the patient-specific profiles for single lung recipients based on the parametric estimates in 3. Symbols depict binned averages.

Figure 9.
True shapes of the phases for temporal trend in the simulated model.

Figure 10.

Distribution of the estimates of the shaping parameters. The averages of the shaping parameters of the early phase $g(t, \Theta)$ and the late phase $h(t, \Theta)$ based on the true values are depicted in solid (black) line and of the estimated values are depicted in dash-dash (grey) line.

Table 1

Estimates of shaping parameter of the temporal trend of FVC. Note that, in the early phase $m = 0$ means the limiting case of $T(t, \Theta) = g(t, \Theta)$ when $\nu < 0$ and $m \rightarrow 0+$ as described in Case 3 and in the late phase $\nu = 0$ means the limiting case of $T(t, \Theta) = h(t, \Theta)$ when $m > 0$ and $\nu \rightarrow 0+$ as described in Case 2 in Section 2.1.1. Further note that, $\hat{\beta}_{02}$ is set to 0 when using $h(t, \Theta)$ as $T(t, \Theta)$

Parameter	Estimate $\pm$ SE	P
<i>Early peaking phase</i>		
$\hat{\beta}_{01}$	0.35 ± 0.060	< 0.0001
$\hat{\nu}$	-0.81 ± 0.0083	< 0.0001
m	0	-
$\hat{t}_{1/2}$	1.01	-
<i>Late increasing phase</i>		
$\hat{\beta}_{02}$	0	-
$\hat{\nu}$	0	-
m	-0.37 ± 0.015	< 0.0001
$\hat{t}_{1/2}$	1.54	
$\widehat{Var}(\varepsilon)$	0.12 ± 0.0024	< 0.0001
$\widehat{Var}(\mathbf{b}) = \hat{\mathbf{G}}$	$\begin{pmatrix} 0.62 & \\ 0.33 & 1.59 \end{pmatrix}$	

Table 2

Model selection: AICc for 3 different models for temporal trend of FVC. is the difference in the AICc between the main model and the alternate models.

Model	AICc	
Main model (8)	6813	0
Alternate Model 1	6834	21
Alternate Model 2	8852	2039

Table 3

Effect of double versus single lung transplantation on postoperative FVC.

Parameter	Estimate $\pm$ SE	P
<i>Early peaking phase</i>		
Double	-0.16 ± 0.092	0.08
<i>Late plateau phase</i>		
Double	0.77 ± 0.13	<0.0001

Table 4

Patient-specific risk factors for higher post-op FVC. (LTx - lung transplantation)

Parameter	Estimate $\pm$ SE	P
Overall		
Age ^a	-0.60 $\pm$ 0.17	0.0005
Female	-0.098 $\pm$ 0.073	0.17
COPD (vs other reason)	0.11 $\pm$ 0.088	0.20
Early peaking phase		
Double LTx (vs Single LTx)	-0.21 $\pm$ 0.11	0.04
Preoperative FVC ^b	1.24 $\pm$ 0.16	<0.0001
Late plateau phase		
Double LTx (vs Single LTx)	0.73 $\pm$ 0.13	<0.0001
Preoperative FVC ^b	0.78 $\pm$ 0.26	0.003

^aAge scaled by 50^bFVC scaled by 60

Table 5
Summary measures of the parameter estimates based on the 500 simulated data with sample size 400.

Parameter	True value	Bias	AvgSE	EmpSE	CP
<i>Overall</i>					
x_4	0.75	0.0064	0.09777	0.09980	0.936
<i>Early peaking phase</i>					
β_0	-1	0.0012	0.1047	0.00961	0.968
x_1	1	0.0002	0.1070	0.1059	0.956
x_2	-0.5	-0.0077	0.0982	0.0939	0.960
<i>Shaping parameters</i>					
v	0.5	0.0004	0.0082	0.0082	0.956
m	0.5	-0.0009	0.0124	0.0124	0.966
$t_{1/2}$	0.25	0.0000	0.0026	0.0027	0.930
<i>Late increasing phase</i>					
$t_{1/2}$	6	-0.0230	0.5742	0.5690	0.956
x_1	-0.5	-0.0044	0.1454	0.1430	0.946
x_3	0.5	-0.0059	0.1471	0.1448	0.944
<i>Shaping parameters</i>					
v	-0.75	-0.0001	0.0018	0.0017	0.968
<i>Variances</i>					
σ_ε^2	0.25	0.0000	0.0053	0.0053	0.954
$\sigma_{b_1}^2$	1	0.0164	0.0762	0.0767	0.924
$\sigma_{b_2}^2$	2	0.0173	0.1436	0.1388	0.948
$Cov(b_1, b_2)$	0.5	-0.0012	0.0792	0.0789	0.944

Table 6

Summary measures of the parameter estimates based on the 500 simulated data with varying sample size.

Parameter	True value	Sample Size									
		n=400					n=200				
		Bias	AvgSE	EmpSE	CP	Bias	AvgSE	EmpSE	CP	Bias	AvgSE
Overall											
x_4	0.75	0.0064	0.09777	0.09980	0.936	0.0032	0.1381	0.1355	0.954	0.0102	0.1942
Early peaking phase											
β_0	-1	0.0012	0.1047	0.00961	0.968	0.0050	0.1476	0.01512	0.944	-0.0087	0.2074
x_1	1	0.0002	0.1070	0.1059	0.956	-0.0057	0.1513	0.1551	0.950	0.0012	0.2128
x_2	-0.5	-0.0077	0.0982	0.0939	0.960	-0.0038	0.1388	0.1515	0.926	-0.0033	0.1947
Shaping parameters											
ν	0.5	0.0004	0.0082	0.0082	0.956	0.0002	0.0116	0.0116	0.954	-0.0019	0.0166
m	0.5	-0.0009	0.0124	0.0124	0.966	-0.0007	0.0175	0.0772	0.960	0.0017	0.0249
$t_{1/2}$	0.25	0.0000	0.0026	0.0027	0.930	0.0000	0.0037	0.0037	0.952	-0.0008	0.0053
Late increasing phase											
$t_{1/2}$	6	-0.0230	0.5742	0.5690	0.956	-0.0325	0.8160	0.8379	0.928	-0.0589	1.1452
x_1	-0.5	-0.0044	0.1454	0.1430	0.946	-0.0041	0.2061	0.2049	0.960	-0.0002	0.2885
x_3	0.5	-0.0059	0.1471	0.1448	0.944	-0.0033	0.2079	0.2146	0.946	0.0090	0.2912
Shaping parameters											
ν	-0.75	-0.0001	0.0018	0.0017	0.968	0.0000	0.0026	0.0026	0.944	0.0001	0.0036
Variances											
σ_ε^2	0.25	0.0000	0.0053	0.0053	0.954	0.0003	0.0075	0.0078	0.930	0.0003	0.0106
$\sigma_{b_1}^2$	1	0.0164	0.0762	0.0767	0.924	0.0264	0.1068	0.1045	0.924	0.0519	0.1475
$\sigma_{b_2}^2$	2	0.0173	0.1436	0.1388	0.948	0.0220	0.2028	0.2022	0.946	0.0854	0.2784
$Cov(b_1, b_2)$	0.5	-0.0012	0.0792	0.0789	0.944	0.0061	0.1116	0.1085	0.946	0.0154	0.1539