Advances and Applications in Statistics
Volume 17, Number 1, 2010, Pages 41-60
Published Online: September 28, 2010
This paper is available online at http://pphmj.com/journals/adas.htm

ESTIMATING RECURRENT TIME USING NONLINEAR MIXED EFFECTS MODEL IN PATIENTS WITH COLORECTAL CANCER

NUENGRUITHAI SONTIMOON[1], MONTIP TIENSUWAN[1,*], SIMONA PANUNZI[2] and WUTTHI SUMETCHOTIMAYTHA[3]

[1]Department of Mathematics
Faculty of Science
Mahidol University
Bangkok 10400, Thailand
e-mail: scmts@mahidol.ac.th

[2]C. N. R.
IASI Biomatlab University
Cattolica del Sacro Cuore
Largo A. Gemelli 800168 Rome, Italy

[3]National Cancer Institute
Bangkok 10400, Thailand

Abstract

This study focuses on a nonlinear mixed effects model for repeated measures data to describe longitudinal changes in carcinoembryonic antigen (CEA) levels of colorectal cancer patients over time. The CEA level of colorectal cancer patients is regarded as the marker choice for monitoring. Following surgery, if the CEA levels begin to rise, then there will be a recurrence of disease. The fixed effects parameter ϕ_2 of the

2010 Mathematics Subject Classification: 62J02, 62P10.

Keywords and phrases: carcinoembryonic antigen, nonlinear mixed effects model, recurrent time, repeated measures.

*Corresponding author

Received June 2, 2010

proposed model represents the recurrent time which is between the flat and rapid increase in an exponential phase. Parameters of the proposed model are estimated by using Lindstrom and Bates (LB) and the stochastic approximation version of the standard expectation and maximization step (SAEM) algorithms. The results show that the estimates of the recurrent time (ϕ_2) by using LB and SAEM algorithms are equal to 17.4 and 22.8 months, respectively. Further, the residual sum of squares of a proposed model by using the SAEM algorithm is equal to 25.14 which is less than the residual sum of squares using the LB algorithm (88.04). Hence, the predicted curve of the proposed model with parameters estimated by the SAEM algorithm is outperforms the LB algorithm for the CEA level of colorectal cancer patients.

1. Introduction

By definition, studies of growth and decay involve repeated measurements taken on sample units which could be human or animal, subjects, plants, or cultures. Modeling data of this kind usually involves characterization of the relationship between the measured response, *y*, and the repeated measurement factor, or covariate, *x*. In many applications, the proposed systematic relationship between *y* and *x* is nonlinear with unknown parameters of interest. With the nonlinear mixed effects model there are many methods to estimate the unknown parameter vector. One method is the algorithm proposed by Lindstrom and Bates (LB) [9] to estimate the parameters of their model. Another method is the stochastic approximation version of the standard expectation and maximization step (SAEM) algorithm developed by Kuhn and Lavielle [8]. Colorectal cancer is one of the leading causes of cancer death. A rising carcinoembryonic antigen (CEA) level indicates progression or recurrence of the cancer. The CEA is a type of protein molecule that can be found in many different cells of the body, but it is typically associated with certain tumors and the developing fetus [2]. The CEA is used as a tumor marker, especially for cancers of the gastrointestinal tract such as colorectal cancer. When the CEA level is abnormally high before surgery or other treatment, it is expected to fall to a normal level after success in removing all of cancer. If the level begins to rise above 6 ng/ml, then there will be a high correlation of recurrence of the cancer. However, this does not always happen in every case as there might be other factors which can increase the CEA level, including diverticulitis, pancreatitis, hepatitis and smoking. If these other causes are excluded, then we must look for recurrence of

cancer. Local recurrence of colorectal cancer after ‘curative’ surgery is a major clinical problem. Typically, 50-70 percent of patients presented to a surgical clinic will undergo apparently curative surgery for disease and about 10-25 percent of those will develop local recurrence in either the tumor bed or bowel wall [7]. In 1987, Claudio et al. [5] made a comparative study of sixty-four consecutive patients who had undergone curative resection for colorectal carcinoma which was conducted to evaluate the roles of sequential CEA determinations and independent instrumental. The study was also a follow up in the early detection of respective recurrences. They found that CEA was the best predictor of recurrence when compared with the other two markers (tissue polypeptide antigen (TPA), colon cancer antigen detected with a monoclonal antibody (Ca19-9)). It is in accordance with 2000 update of American Society of Clinical Oncology colorectal cancer surveillance guidelines by Benson et al. [3] that follow up testing was done by protocol guidelines. Ninety-six of the 421 patients who developed recurrent disease underwent surgical resection with curative intent. For the subgroup of patients with resectable, the first test to detect recurrence was CEA, chest X-ray, colonoscopy, and other tests. The CEA was the most cost-effective approach to detect potentially metastases from colon cancer. Another study followed up patients with a specified testing strategy after curative colorectal surgery. Here, 64% of recurrences were detected first by CEA, far more than the other tests in the battery [4].

Therefore, the aim of this study is to estimate the fixed effects parameter ϕ_2 of the proposed CEA model represents the recurrent time in patients with colorectal cancer which is between the flat and rapid increase in an exponential phase. Parameters of the proposed model are estimated by using Lindstrom and Bates (LB) and the stochastic approximation version of the standard expectation and maximization step (SAEM) algorithms.

The paper is organized as follows: Section 2 describes on colorectal cancer data. Section 3 briefly introduces the models and uses nonlinear mixed effects model estimator by Lindstrom and Bates algorithm (LB), stochastic approximation version of the standard expectation and maximization step (SAEM) algorithm to estimate the parameters of the proposed model. The results including presentation of estimates of the parameters and estimates of the population and individual curves by LB and SAEM algorithms are given in Section 4. Section 5 provides conclusion and discussion.

2. Colorectal Cancer Data

In this study, the subjects have participated in the National Cancer Institute of Thailand from July 1998 to April 2008. All of seven cancer cases have been detected clinically and confirmed histologically from biopsy tissue specimens. All subjects have undergone colorectal surgery after the colorectal cancer diagnosis and were recurrences appeared later. The follow ups started at the surgery date which is set up to be zero, and are in the range negative two to 55 months with a median of 18 months over their follow up period. The follow ups started at the surgery date and then at least six times for measurement of CEA (ng/ml) values and had a median of 10 observations over their follow up period. The recurrent times are in the range three to 38 months after the surgery date and have a median of nine months over their follow up period. Table 1 contains descriptive statistics for number of the participants, the measurement month, number of repeated CEA measurements and the recurrent time after surgery.

Table 1. Description of study participants

Colorectal cancer patients	
No. of participants	7
Measurement month (surgery month is zero)	
Median	18
Range	–2-55
No. of repeated measurements	
Median	10
Range	6-20
Recurrent time after surgery (months)	
Median	9
Range	3-38

CEA values are transformed with the logarithm to make it easier to visualize. The curves of logarithm of the CEA data of colorectal cancer patients are shown in Figure 1. The logarithm of the CEA value increases slowly exponentially for each curve. The black dots for each curve are the recurrence time (months after surgery). The logarithm of the CEA values after the recurrent time is higher than the logarithm of the CEA values before the recurrent time.

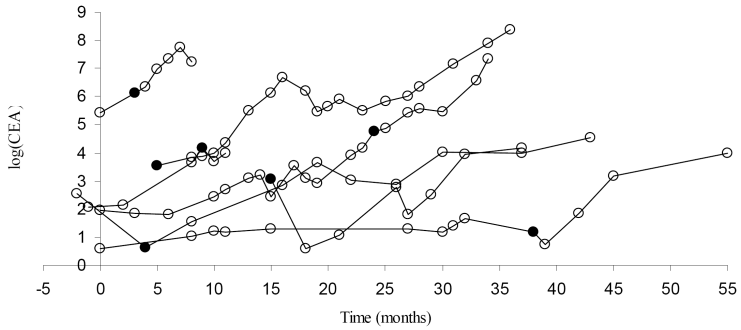


Figure 1. Log(CEA) (in nanograms per milliliter) versus time (months) of seven subjects. The black dots are the recurrent time for each subject.

3. The Nonlinear Mixed Effects Model

Let us consider the following general nonlinear mixed effects model by using the notation of Lindstrom and Bates [9], the model for the *j*th observation on the *i*th individual in the study is written as

$$y_{ij} = f(x_{ij}, \phi_i) + e_{ij}, \tag{1}$$

where

y_{ij} is the *j*th response on the *i*th individual, $i = 1, ..., m$,

x_{ij} is the vector of predictor variables for the *j*th response on the individual *i*,

f is a nonlinear function of the predictor variables and parameter vector,

ϕ_i is an $r \times 1$ vector of parameters for the *i*th individual, and $e_{ij} \sim N(0, \sigma^2 I_i)$.

Random-effects terms may be included in the parameter vector to allow the parameter value to vary from individual to individual by writing

$$\phi_i = A_i\phi + B_i b_i, \quad b_i \sim N(0, D), \tag{2}$$

where

ϕ is a $p \times 1$ vector of fixed population parameters,

b_i is a $q \times 1$ vector of random effects associated with individual *i*,

A_i is a design matrix of size $r \times p$ for the fixed effects,

B_i is a design matrix of size $r \times q$ for the random effects, and

D is a positive definite variance-covariance matrix.

Collecting the responses and errors for the *i*th individual into $n_i \times 1$ vectors. This is accomplished by letting

$$y_i = \begin{bmatrix} y_{i1} \\ \cdot \\ \cdot \\ \cdot \\ y_{in_i} \end{bmatrix}, \quad e_i = \begin{bmatrix} e_{i1} \\ \cdot \\ \cdot \\ \cdot \\ e_{in_i} \end{bmatrix} \quad \text{and} \quad f_i(\phi_i) = \begin{bmatrix} f(x_{ij}, \phi_i) \\ \cdot \\ \cdot \\ \cdot \\ f(x_{in_i}, \phi_i) \end{bmatrix}.$$

Hence, $y_i = f_i(\phi_i) + e_i$, where $e_i \sim N(0, \sigma^2 I_i)$ and I_i is an identity matrix for individual. We may summarize the data for the *i*th individual as

$$y_i = f_i(\phi_i) + e_i, \quad \phi_i = A_i\phi + B_i b_i \quad \text{and} \quad b_i \sim N(0, D).$$

From the logarithm of the CEA value curve (Figure 1), we see that trends in each subject case represent a period of slow exponential increase in the peripheral logarithm of CEA levels. Next, we try to use the exponential model for these data. The exponential model will be approached to find a recurrent time after surgery, which is the time point between the flat and rapid increase in the exponential phase.

We used the LB and SAEM algorithms to estimate the parameters of the proposed model. Let y_{ij} be logarithm of CEA value for the *j*th response on the *i*th individual. Then the nonlinear model is $y_{ij} = f(x_{ij}, \phi_i) + e_{ij}$.

The proposed CEA model is written as

$$f(x_{ij}, \phi_i) = \exp(\phi_{0i}) \exp(\phi_{1i}(x_{ij} - \phi_{2i})).$$

Since $\phi_i = A_i\phi + B_i b_i$, $\phi_{0i} = \phi_0 + b_{0i}$, $\phi_{1i} = \phi_1 + b_{1i}$ and $\phi_{2i} = \phi_2 + b_{2i}$.

Therefore, we can rewrite the CEA model as follows:

$$\begin{aligned} f(x_{ij}, \phi_i) &= \exp(\phi_0 + b_{0i}) \exp((\phi_1 + b_{1i})(x_{ij} - (\phi_2 + b_{2i}))) \\ &= \exp(\phi_0) \exp(b_{0i}) \exp((\phi_1 + b_{1i})(x_{ij} - (\phi_2 + b_{2i}))), \end{aligned} \tag{3}$$

here, $b_i = [b_{0i} \quad b_{1i} \quad b_{2i}]'$ has mean zero and variance-covariance matrix *D*,

x_{ij} is time (months) (at surgery date, $x_{ij} = 0$),

ϕ_0 is the log(CEA) level (ng/ml) at the exponential phase,

ϕ_1 is the exponential rate constant during the exponential log(CEA) phase,

ϕ_2 represents the time (months after surgery date) between the flat and rapid increase in the exponential log(CEA) phase,

b_{0i} is random and allows the exponential phase to vary among participants,

b_{1i} is random and allows the exponential rate constant to vary among participants, and

b_{2i} is random and allows the time between flat and rapid increase in the exponential log(CEA) phase to vary among participants.

3.1. Lindstrom and Bates (LB) algorithm

In the above model, there are many methods to estimate the unknown parameter vector. One method is nonlinear mixed effects model algorithm as proposed by Lindstrom and Bates [9]. We also employed the LB algorithm to estimate the parameters of our proposed model. The algorithm uses a combination of least square estimators for nonlinear fixed effects models and maximum likelihood estimators for a linear mixed effects model. The linear mixed effects model is used to estimate parameters from an approximation to the marginal distribution of the complete data vector. This combination provides approximate maximum likelihood estimates for the nonlinear mixed effects model. The Lindstrom and Bates algorithm is suggested two steps of estimation schemes as follows:

Step 1. Pseudo-data (PD)

Lindstrom and Bates [9] showed that $\hat{\phi}$ is a maximum likelihood estimate relative to an approximate marginal distribution of y. As in the linear case, these estimates can be calculated as the solution to a (nonlinear) least squares problem formed by augmenting the data vector with "pseudo-data" as

$$y_i = f_i(x_i, \phi_i, b_i) + e_i,$$

where $e_i \sim N(0, \sigma^2 I_i)$. The probability density function of y_i is given by

$$f(y_i) = \frac{1}{\sqrt{2\pi}|\sigma^2 I_i|^{1/2}} \cdot \exp\left(-\frac{1}{2}(y_i - f_i(x_i, \phi_i, b_i))^T(\sigma^2 I_i)^{-1}(y_i - f_i(x_i, \phi_i, b_i))\right).$$

Since b_i has a normal distribution with mean zero and variance-covariance matrix D that is $b_i \sim N(0, D)$, hence the probability density function of b_i is given as

$$f(b_i) = \frac{1}{\sqrt{2\pi}|D|^{1/2}} \cdot \exp\left(-\frac{1}{2} b_i^T D^{-1} b_i\right).$$

Thus the joint probability density function of the response y_i and b_i is written as

$$f(y_i, b_i) = \frac{1}{\sqrt{2\pi}|\sigma^2 I_i|^{1/2}} \cdot \exp\left(-\frac{1}{2}(y_i - f_i(x_i, \phi_i, b_i))^T (\sigma^2 I_i)^{-1}(y_i - f_i(x_i, \phi_i, b_i))\right)$$
$$\times \frac{1}{\sqrt{2\pi}|D|^{1/2}} \exp\left(-\frac{1}{2} b_i^T (D)^{-1} b_i\right),$$

and the likelihood function of the response y_i is given by

$$L = \prod f(y, b)$$
$$= \left(\frac{1}{2\pi}\right)^{m/2} \left(\frac{1}{|\sigma^2 I_i|}\right)^{m/2}$$
$$\times \exp\left(-\frac{1}{2}\sum (y_i - f_i(x_i, \phi_i, b_i))^T (\sigma^2 I_i)^{-1}(y_i - f_i(x_i, \phi_i, b_i))\right)$$
$$\times \left(\frac{1}{2\pi}\right)^{m/2} \left(\frac{1}{|D|}\right)^{m/2} \exp\left(-\frac{1}{2}\sum b_i^T \hat{D}^{-1} b_i\right).$$

Then the log-likelihood function of the response y_i is obtained by

$$\log L = -\frac{m}{2}\log(2\pi) - \frac{m}{2}\log|\sigma^2 I_i|$$
$$-\frac{1}{2}\sum (y_i - f_i(x_i, \phi_i, b_i))^T (\sigma^2 I_i)^{-1}(y_i - f_i(x_i, \phi_i, b_i))$$
$$-\frac{m}{2}\log(2\pi) - \frac{m}{2}\log|D| - \frac{1}{2}\sum b_i^T \hat{D}^{-1} b_i.$$

Given the current estimate of ω, $\hat{\omega}$ (and thus $\hat{\sigma}^2$ and $\hat{D}$), minimize in ϕ and b_i, $i = 1, ..., m$. The twice negative log-likelihood is given as

$$-2\log L = \sum_{i=1}^{m} (\log|\hat{D}| + b_i^T \hat{D}^{-1} b_i + \log|\hat{\sigma}^2 I_i|$$
$$+ (y_i - f_i(x_i, \phi_i, b_i))^T (\hat{\sigma}^2 I_i)^{-1}(y_i - f_i(x_i, \phi_i, b_i))),$$

where

$\hat{D}$ is variance-covariance of random effects,

b_i is random effect,

y_i is observation, and

$f_i(x_i, \phi_i, b_i)$ is the CEA model.

Denote the resulting estimates as $\hat{b}_i$ and $\hat{\phi}_0$.

Step 2. Linear mixed effects (LME)

Lindstrom and Bates [9] defined maximum likelihood estimators for ϕ with respect to the marginal density function of *y*, that is

$$f(y) = \int f(y|b)\, f(b)\, db.$$

However, the expectation function $f_i(\phi_i)$ is nonlinear in *b*, and there is no closed-form expression for this density function and the calculation of such estimates would be very difficult. Instead, they approximated the conditional distribution of *y* for *b* near $\hat{b}$ by a multivariate normal with expectation that is linear in *b*. To accomplish this, they approximated the residual $y - f(\phi)$ near $\hat{b}$ as Taylor series expansion about $b_i = \hat{b}_i$ of $f_i(x_i, \phi, b_i)$ so that

$$y_i = f_i(x_i, \phi, b_i) \approx f_i(x_i, \phi, b_i) + Z_i(\hat{\phi}, \hat{b}_i) \cdot (b_i - \hat{b}_i) + e_i,$$

$$b_i \sim N(0, D) \quad \text{and} \quad e_i \sim N(0, \sigma^2 I_i),$$

where $Z_i(\hat{\phi}, \hat{b}_i)$ is the $n_i \times k$ matrix of derivatives of $f_i(x_i, \phi, b_i)$ with respect to b_i. The expectation of y_i is given as

$$\begin{aligned} E(y_i) &= E(f_i(x_i, \phi, b_i) + Z_i(\hat{\phi}, \hat{b}_i) E(b_i - \hat{b}_i) + e_i) \\ &= E(f_i(x_i, \phi, b_i)) + Z_i(\hat{\phi}, \hat{b}_i) E(b_i - \hat{b}_i) + E(e_i) \\ &= f_i(x_i, \phi, b_i) - Z_i(\hat{\phi}, \hat{b}_i)\hat{b}_i, \end{aligned}$$

and the variance of y_i is given by

$$\begin{aligned} \mathrm{Var}(y_i) &= \mathrm{Var}(f_i(x_i, \phi, b_i) + Z_i(\hat{\phi}, \hat{b}_i) E(b_i - \hat{b}_i) + e_i) \\ &= \mathrm{Var}(f_i(x_i, \phi, b_i)) + Z_i(\hat{\phi}, \hat{b}_i)\mathrm{Var}(b_i - \hat{b}_i) Z_i^T(\hat{\phi}, \hat{b}_i) + \mathrm{Var}(e_i) \\ &= Z_i(\hat{\phi}, \hat{b}_i)\mathrm{Var}(b_i) Z_i^T(\hat{\phi}, \hat{b}_i) + \sigma^2 I_i \\ &= Z_i(\hat{\phi}, \hat{b}_i) D Z_i^T(\hat{\phi}, \hat{b}_i) + \sigma^2 I_i. \end{aligned}$$

Then, the response variables, y_i are distributed as a normal distribution with

$$y_i \sim N(f_i(x_i, \phi, b_i) - Z_i(\hat{\phi}, \hat{b}_i)\hat{b}_i, Z_i(\hat{\phi}, \hat{b}_i)DZ_i^T(\hat{\phi}, \hat{b}_i) + \sigma^2 I_i).$$

Therefore, the marginal distribution of y_i is written as

$$f(y_i) = \frac{1}{\sqrt{2\pi}|V_i|^{1/2}} \exp\Big(-\frac{1}{2}(y_i - f_i(x_i, \phi, \hat{b}_i) + Z_i(\hat{\phi}, \hat{b}_i)\hat{b}_i)^T \times V_i^{-1}(y_i - f_i(x_i, \phi, \hat{b}_i) + Z_i(\hat{\phi}, \hat{b}_i)\hat{b}_i)\Big),$$

where $V_i = Z_i(\hat{\phi}, \hat{b}_i)DZ_i^T(\hat{\phi}, \hat{b}_i) + \sigma^2 I_i$ and $Z_i(\hat{\phi}, \hat{b}_i)$ is the $n_i \times k$ matrix of derivatives of $f_i(x_i, \phi, b_i)$ with respect to b_i.

The likelihood function of y_i is written as

$$L = \prod f(y_i) = \left(\frac{1}{2\pi}\right)^{m/2}\left(\frac{1}{|V_i|}\right)^{m/2} \exp\Big(-\frac{1}{2}\sum(y_i - f_i(x_i, \phi, \hat{b}_i) + Z_i(\hat{\phi}, \hat{b}_i)\hat{b}_i)^T \times V_i^{-1}(y_i - f_i(x_i, \phi, \hat{b}_i) + Z_i(\hat{\phi}, \hat{b}_i)\hat{b}_i)\Big).$$

The log-likelihood function of y_i is obtained by

$$\log L = -\frac{m}{2}\log(2\pi) - \frac{m}{2}\log(|V_i|) - \frac{1}{2}\sum(y_i - f_i(x_i, \phi, \hat{b}_i) + Z_i(\hat{\phi}, \hat{b}_i)\hat{b}_i)^T \times V_i^{-1}(y_i - f_i(x_i, \phi, \hat{b}_i) + Z_i(\hat{\phi}, \hat{b}_i)\hat{b}_i).$$

Estimate ϕ and ω as the values $\hat{\phi}$ and $\hat{\omega}$ that give twice the negative approximate marginal normal log-likelihood is obtained as

$$-2\log L = \sum(\log|V_i| + (y_i - f_i(x_i, \phi_i, \hat{b}_i) + Z_i(\hat{\phi}, \hat{b}_i)\hat{b}_i)^T \times V_i^{-1}(y_i - f_i(x_i, \phi_i, \hat{b}_i) + Z_i(\hat{\phi}, \hat{b}_i)\hat{b}_i)).$$

3.2. Stochastic approximation version of the standard expectation and maximization step (SAEM) algorithm

The SAEM algorithm which is developed by Kuhn and Lavielle [8] provides general idea of the algorithm in order to replace the expectation phase of the EM algorithm, that is, the calculation of the conditional expectation of the likelihood of

the complete data. By using a stochastic approximation, this expectation can be analytically calculated in the case of nonlinear mixed effects models.

At iteration $[k]$, let $Q_k(\theta)^{[k]}$ be the expectation function of the complete likelihood conditional on the observations y and the vector of parameters θ estimated at iteration $[k-1]$,

$$Q_k(\theta)^{[k]} = E(\log p(y, \phi; \theta \mid y, \theta^{[k-1]})).$$

The key idea is to recycle variates generated from the previous iterations of the EM. Therefore, instead of approximating $Q_k(\theta)^{[k]}$ by the complete likelihood, i.e., $\log p(y, \phi^{[k]}; \theta)$, it is replaced by the following stochastic approximation:

$$Q_k(\theta)^{[k]} = Q(\theta)^{[k-1]} + \gamma_k(\log p(y, \phi^{[k]}; \theta) - Q(\theta)^{[k-1]}),$$

where ϕ is simulated according to the conditional distribution $p(\cdot \mid y, \theta^{[k-1]})$. The SAEM is consisted in replacing the usual E step of EM by a stochastic procedure. At iteration k, SAEM consists in three steps as follows:

1. Simulation step. Draw $\phi^{[k]}$ from $f(y_{ij}; \theta_k)$ by Metropolis-Hasting process which consists of

1.1 generate u from $U(0, 1)$,

1.2 draw $\tilde{\phi}_{i,p}$ using the normal distribution and draw $\phi_{i,p-1}$ from OLS,

1.3 find $f(y, \tilde{\phi}_{i,p}; \theta_k)$ and $f(y, \phi_{i,p-1}; \theta_k)$ by substituting $\tilde{\phi}_{i,p}$ and $\phi_{i,p-1}$, respectively,

1.4 $\alpha(\phi_{i,p-1}, \tilde{\phi}_{i,p}) = \min\left(1, \dfrac{f(y \mid \tilde{\phi}_{i,p}; \theta_k)}{f(y \mid \phi_{i,p-1}; \theta_k)}\right)$,

1.5 if $u \le \alpha(\phi_{i,p-1}, \tilde{\phi}_{i,p})$, then $\phi_{i,p} = \tilde{\phi}_{i,p}$ else $\phi_{i,p} = \phi_{i,p-1}$.

2. Stochastic approximation step. Update $Q_k(\theta)$ according to

$$Q_k(\theta)^{[k]} = Q(\theta)^{[k-1]} + \gamma_k(\log p(y, \phi^{[k]}; \theta) - Q(\theta)^{[k-1]}),$$

where γ_k is a decreasing sequence of positive numbers. It has been set by MONOLIX, where $\gamma_k = 1$ for $1 \le k \le K$ and $\gamma_k = \dfrac{1}{k-K}$ for $k \ge K+1$ (the default value is $K = 300$). The stochastic approximation step consists in updating

the sufficient statistics of the complete model as follows:

$$s_{1,i,k} = s_{1,i,k-1} + \gamma_k(\phi_i^{[k]} - s_{1,i,k-1}), \quad i = 1, 2, ..., m,$$

$$s_{2,k} = s_{2,k-1} + \gamma_k\left(\sum_{i=1}^{m} \phi_i^{[k]}\phi_i^{[k]^T} - s_{2,k-1}\right),$$

$$s_{3,k} = s_{3,k-1} + \gamma_k\left(\sum_{i,\,j} (y_{ij} - f(x_{ij}, \phi_i^{[k]}))^2 - s_{3,k-1}\right).$$

3. Maximization step. Update θ_k according to $\theta_k = \text{Arg}\max_\theta Q_k(\theta)$. The θ_{k+1} is obtained in the maximization step as follows:

$$\phi^{[k+1]} = \left(\sum_{i=1}^{m} C_i^{/} D_k^{-1} C_i\right)^{-1} \sum_{i=1}^{m} C_i^{/} D_k^{-1} S_{1,i,k},$$

$$D^{[k+1]} = \frac{1}{m}\left(S_{2,k} - \sum_{i=1}^{m} (C_i\phi_{k+1}) S_{1,i,k}^{/} - \sum_{i=1}^{m} S_{1,i,k}(C_i\phi_{k+1})^{/} + \sum_{i=1}^{m} (C_i\phi_{k+1})(C_i\phi_{k+1})^{/}\right),$$

$$\sigma^{2^{[k+1]}} = \sqrt{\frac{S_{3,k}}{N_{\text{tot}}}}.$$

We used the software MONOLIX developed by Kuhn and Lavielle [8] which implements SAEM algorithm for maximum likelihood estimator in nonlinear mixed effects models. This MATLAB software is available at http://software.monolix.org.

4. Results

The full model with all the parameters of (3) is fitted to the data. The estimates of the fixed effects parameters (the ϕ's) as well as their asymptotic z values (z value = estimate/standard error (SE)) and estimates the error variance, variance-covariance matrix for random effects (D) and also the residual sum of squares by the LB and SAEM algorithms are given in Table 2. Initial values for parameters (parameters of vector θ consist of ϕ_0, ϕ_1, ϕ_2, D and σ^2) must be specified for the first step. As in standard nonlinear estimation, poor starting values can cause poor results. Afterwards we use an ordinary least square (OLS) method to find the initial values.

Table 2. Estimates of the fixed effects, SE and approximate estimates of the error variance, variance-covariance matrix for the random effects, and the residual sum of squares by the LB and SAEM algorithms

	LB			SAEM		
Parameters	Estimates	SE	z	Estimates	SE	z
ϕ_0	1.1325	0.5138	2.204	1.26	0.43	2.93
ϕ_1	0.0283	0.0062	4.564	0.0325	0.0046	7.06
ϕ_2	17.489	8.584	2.037	22.8	10	2.28
σ^2	0.831			0.585		
D_{11}	0.9044			0.483		
D_{22}	0.0007			0.285		
D_{33}	0.1868			0.293		
Residual sum of squares $\sum_{ij}(y_{ij} - f(x_{ij}, \phi_i))^2$			88.04			25.1396

The residual sum of squares of the CEA model which is estimated from parameters by the LB and the SAEM algorithms are 88.04 and 25.14, respectively.

The prediction curves based on the estimated population mean fixed effects from the fitted CEA model are shown in Figure 2. The diamond and the square lines represent the estimated population mean fixed effects from the LB and SAEM algorithms, respectively.

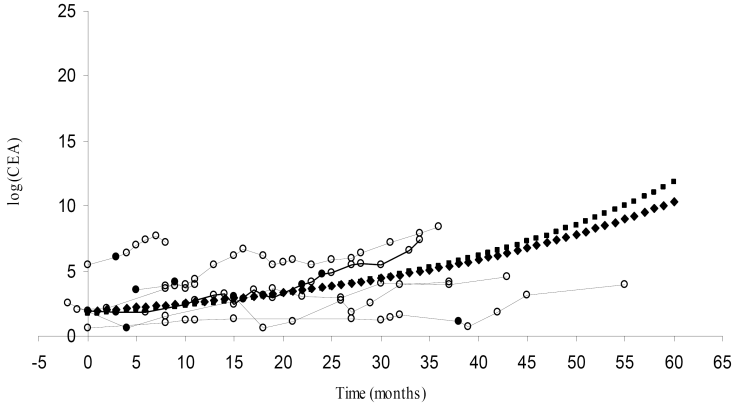
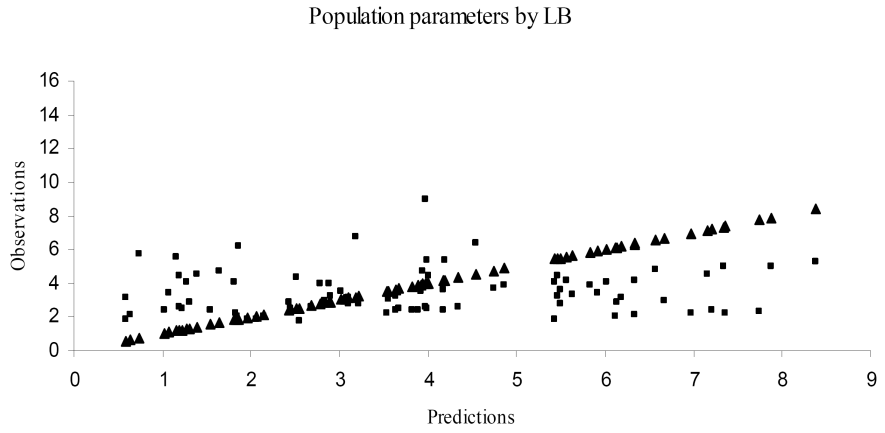
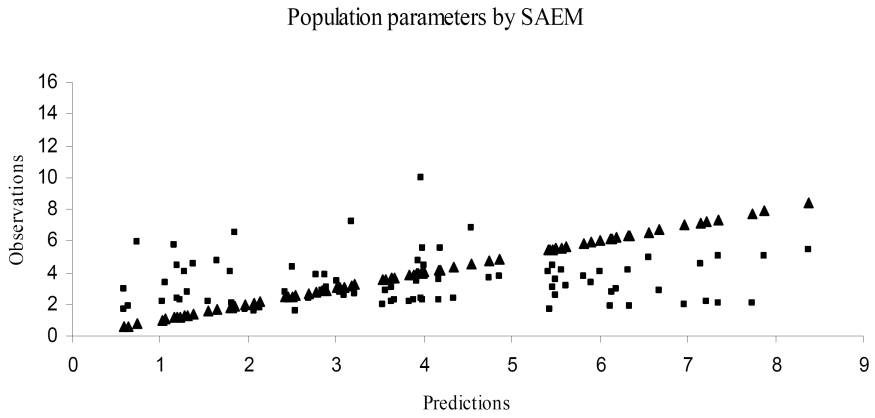


Figure 2. The log(CEA) value (in ng/ml) of seven subjects, population observation and predictive of log(CEA), where open circles represent observed, diamonds and squares represent population predictions of log(CEA) by the LB and SAEM algorithms, respectively.

The goodness of fit of the observations with predictions for population prediction of the CEA model by the LB and SAEM algorithms are displayed in Figure 3.



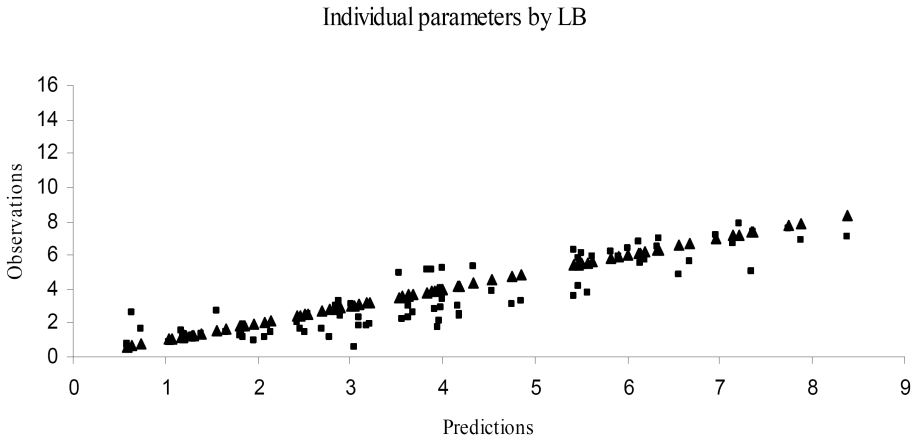
(a)



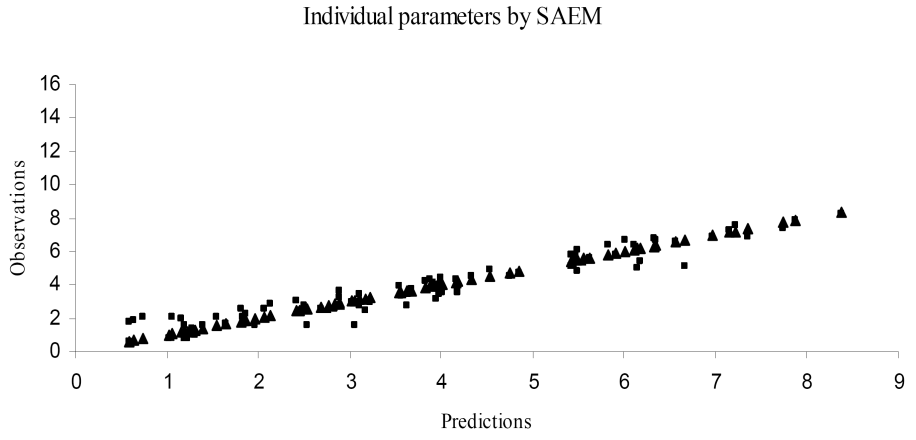
(b)

Figure 3. Observations versus prediction for all subjects. Predicted values (squares) are based on the estimated population mean fixed effects from the fitted CEA model $(b_i = 0)$; (a) by the LB and (b) by the SAEM algorithms, respectively. The triangles represent observations.

The goodness of fit of the observations with predictions for individual prediction of the CEA model by the LB and SAEM algorithms are displayed in Figure 4.



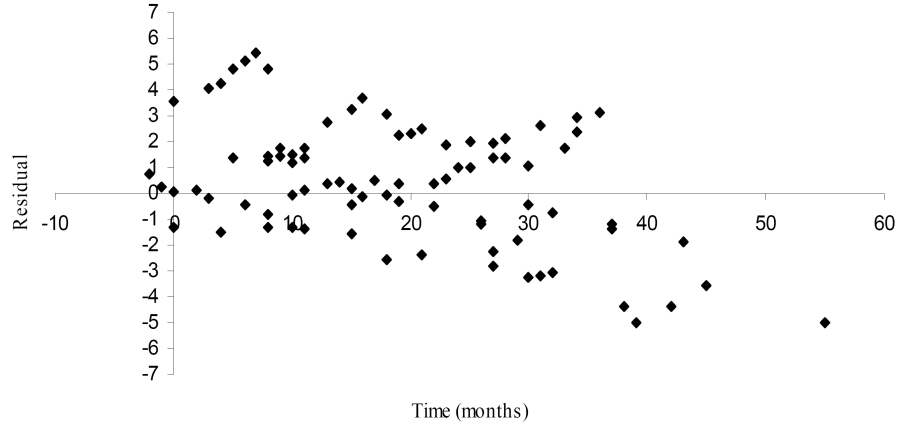
(a)



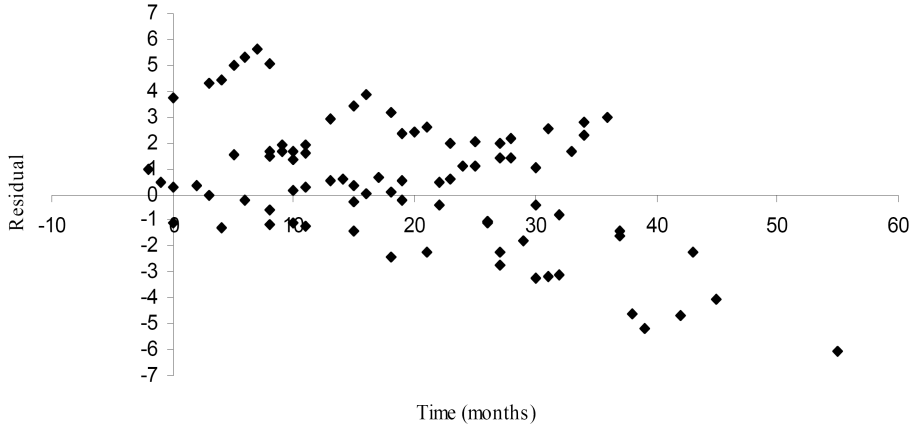
(b)

Figure 4. Observations versus prediction for all subjects. Predicted values (squares) are based on the estimated individual random effects from the fitted CEA model; (a) by the LB and (b) by the SAEM algorithms, respectively. The triangles represent observations.

The goodness of fit of the quality of the CEA model with time plots for residuals population parameters by the LB and SAEM algorithms are displayed in Figure 5.



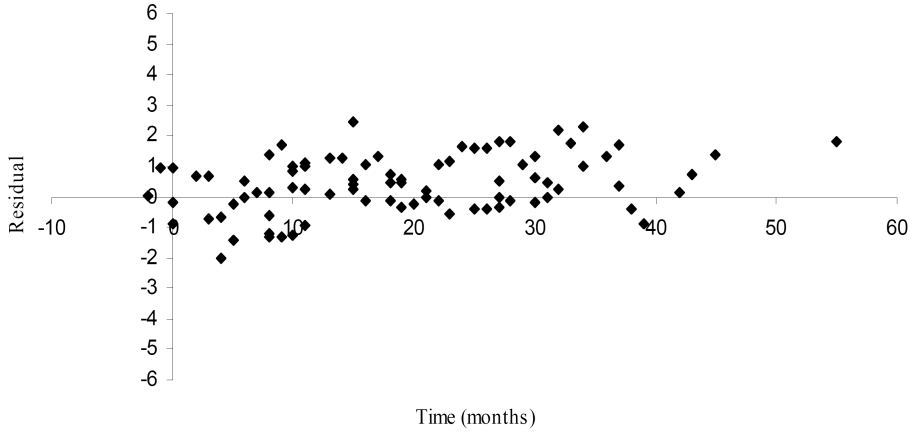
(a)



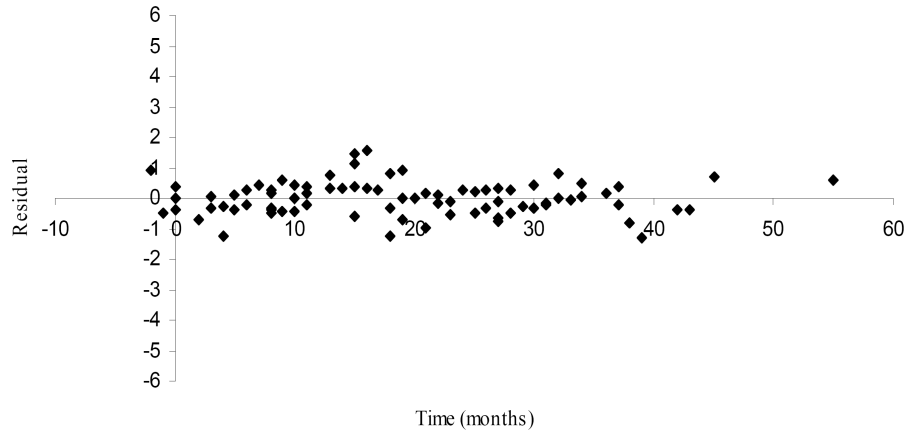
(b)

Figure 5. Residual plots of the CEA population parameters model versus time (in months) (a) by the LB algorithm (b) by the SAEM algorithm.

The goodness of fit of the quality of the CEA model with time plots for residuals individual parameters by the LB and SAEM algorithms are displayed in Figure 6.



(a)



(b)

Figure 6. Residual plots of the CEA individual parameters model versus time (in months) (a) by the LB algorithm, (b) by the SAEM algorithm.

5. Conclusion and Discussion

We proposed a nonlinear mixed effects model to describe the longitudinal CEA data. The CEA data are obtained from colorectal cancer patients. Hence, the proposed model is called the *CEA model*. The model itself is ideal for such analyses because the random effects may provide an adequate variance-covariance matrix to explain the nonindependence among repeated CEA measurements for each subject. The data are not necessarily nicely balanced in terms of numbers of observations and intervals among subjects. Further, the average curves can be estimated from the fixed effects model and the differences in the individuals' progression are estimated by using the random effects model. We believe that the model-fitting approach is not only a methodologically but also a biologically valuable interpretation.

Because of the nonlinearity and the complexity of the nonlinear mixed effects model, implementation of parameter estimation in the model should be done very carefully. Divergence and false convergence problems sometimes occur in fitting the model since many parameters in the model need to be estimated. To deal with these problems, it is suggested to try different initial values; in this study, we therefore used the ordinary least square (OLS) method to find the initial values.

The estimates of the fixed effects parameters (the ϕ's) as well as their asymptotic z values (z value = estimate/standard error) and estimates the error variance and variance-covariance matrix for random effects (D) by the LB and SAEM algorithms are shown in Table 2. All of the fixed effects parameter estimates for the CEA model are significantly different from zero based on the z statistics. The parameter ϕ_2 represents the recurrent time (months after surgery date), which is between the flat and the rapidly increasing in the exponential phase. The recurrent times of the CEA model are equal to 17.4 and 22.8 months by using the LB and the SAEM algorithms, respectively. This is in line with recommended colorectal cancer surveillance guidelines by the American Society of Clinical Oncology [1] that postoperative serum CEA testing may be performed every 2 to 3 months in colorectal cancer patients in stage II or III disease for two or more years after diagnosis. The majority of recurrences in patients who have undergone a complete resection of a colorectal cancer will occur within 5 years, and usually within 3 years of surgery. The residual sum of squares by using the LB and SAEM algorithms for the CEA model are also illustrated in Table 2. The residual sum of squares of the CEA model by using the SAEM algorithm is equal to 25.14, which is less than by using the LB algorithm

(88.04). From these results, we can assume that the predicted curve of the CEA model, which is estimated by the SAEM algorithm, fits the logarithm of CEA of colorectal cancer patients better than as estimated by the LB algorithm. Moreover, the SAEM algorithm obviously takes the computational time much less than the LB algorithm.

The SAEM algorithm is implemented in specialized software for the phenotypic analysis of nonlinear mixed effects models called '*MONOLIX*', which can be freely downloaded from the website: http://software.monolix.org. The software program is based on a thorough statistical theory [6, 8] and several statistical developments are ongoing.

Since the diseases can be recurrent in colorectal cancer patients after surgery, the key question of 'How long does it take, after surgery, for the recurrence to appear?' is an important consideration for physicians. The results from this study hence can provide basic information as a guideline of curing colorectal cancer patients. Our study found out that the SAEM algorithm estimated the recurrent time of colorectal cancer patients as equal to 22.8 months. Therefore, the period of 22.8 months or 2 years approximately after surgery is the opportunity for disease recurrence in patients with colorectal cancer. However, it is suggested that more complete data from a larger number of patients and from CEA measurements at varied different periods of time can confirm a significant window of opportunity in detecting the recurrence of the CEA levels.

For a comparison of these two algorithms we could use a simulation study to see which algorithm is perform better in more general cases of nonlinear mixed effects colorectal cancer model. Finally, it is essential for future research on improving algorithms for cancer recurrence to collect data from other cases involving different body organs. Moreover, we can apply these algorithms to the problem of nonlinear mixed effects models in the medical, agriculture and industry fields.

Acknowledgements

We would like to thank the Commission on Higher Education Staff Development Project, Ministry of Education, Thailand, the Thailand Research Fund and Mahidol University for financial support.

References

[1] R. C. Bast, Jr., P. Ravdin, D. F. Hayes, S. Bates, H. Fritsche, J. M. Jessup, N. Kemeny, G. Y. Locker, R. G. Mennel and R. Mark, 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology, J. Clinical Oncology 15 (2001), 1865-1878.

[2] S. Benchimol, A. Fuks, S. Jothy, N. Beauchemin, K. Shirota and C. P. Stanners, Carcinoembryonic antigen, a human tumor marker, functions as an intercellular adhesion molecule, Cell 57 (1989), 327-334.

[3] Al. B. Benson, C. E. Desch, P. J. Flynn, C. Krause, C. L. Loprinzi, B. D. Minsky, N. J. Petrelli, D. G. Pfister, T. J. Smith and M. R. Somerfield, 2000 update of American society of clinical oncology colorectal cancer surveillance guidelines, J. Clinical Oncology 18 (2000), 3586-3588.

[4] A. Castells, X. Bessa, M. Daniels, C. Ascaso, A. M. Lacy, J. C. Garcia-Valdecasas, L. Gargallo, F. Novell, E. Astudillo, X. Filella and J. M. Piqué, Value of postoperative surveillance after radical surgery for colorectal cancer, Diseases of the Colon & Rectum 41 (1998), 714-724.

[5] F. Claudio, M. T. Silvia, R. Stefano, M. Giorgio, C. Gaetano, P. Sergio and B. Ugo, Follow-up of colorectal cancer resected for cure an experience with CEA, TPA, Ca19-9 analysis and second-look surgery, Diseases of the Colon & Rectum 30 (1987), 273-277.

[6] B. Delyon, M. Lavielle and E. Moulines, Convergence of a stochastic approximation version of the EM algorithm, Ann. Statist. 27 (1999), 94-128.

[7] J. G. Docherty, J. R. McGregor, P. J. O'Dwyer and D. J. Galloway, Local recurrence of colorectal cancer: the problem, mechanisms management and adjuvant therapy, British J. Surgery 81(7) (1994), 1082.

[8] E. Kuhn and M. Lavielle, Maximum likelihood estimation in nonlinear mixed effects models, Comput. Statist. Data Anal. 49 (2005), 1020-1038.

[9] M. J. Lindstrom and D. M. Bates, Nonlinear mixed effects model for repeated measures data, Biometrics 46 (1990), 673-678.