

# MODELLING VARIATION OF LOWER LEG LENGTH GROWTH IN EARLY LIFE

D. Y. T. FONG

*Clinical Trials Centre, Faculty of Medicine, The University of Hong Kong,  
2/F, Block B, Nurses' Quarters, Queen Mary Hospital, 102 Pokfulam Road, Hong Kong.*

*Tel: (852) 2855-4665 Email: dytfong@hku.hk*

S. W. S. WONG

*Clinical Trials Centre, Faculty of Medicine, The University of Hong Kong,  
2/F, Block B, Nurses' Quarters, Queen Mary Hospital, 102 Pokfulam Road, Hong Kong.*

*Tel: (852)2855-4673 Email: sswong@hku.hk*

AND

J. KARLBERG

*Clinical Trials Centre, Faculty of Medicine, The University of Hong Kong,  
2/F, Block B, Nurses' Quarters, Queen Mary Hospital, 102 Pokfulam Road, Hong Kong.*

*Tel: (852)2855-4666 Email: jpekarl@hku.hk*

# MODELLING VARIATION OF LOWER LEG LENGTH GROWTH IN EARLY LIFE

## SUMMARY

We consider the estimation of sources of variation for panel data with repeated measurements. With no repeated measurements and known measurement error, models for variation decomposition were proposed earlier when there are one or more types of measurements. Estimation was performed using the EM algorithm accompanied with model augmentation that demands more computational efforts. In this article, we extend previous variation models and modify the estimation methods in order to estimate various variation components after eliminating the unknown effects of measurement error. Specifically, methods that dispense with model augmentation, and estimation of time-dependent covariates are considered. A set of lower leg length data from Chinese infants is analysed by using the proposed model. Interestingly, our results are consistent to the well-accepted three-phase (infancy-childhood-puberty) growth transition proposition for human growth. Moreover, gender effect is found to be time-varying.

# 1. INTRODUCTION

Good understanding of human growth is important to clinicians and medical researchers involved in nutrition and hormone studies. In particular, modelling the time change of mean and variation of body size measurements helps to understand the full profile of human growth. Various growth curve models were therefore proposed to capture the individual or population growth in humans.<sup>1-5</sup> One of the well-studied growth curve model is perhaps the Infancy-Childhood-Puberty (ICP) model which put forward the theory that human growth is a composition of three overlapping phases that correspond to infancy, childhood and puberty.<sup>2</sup> However, this model as well as virtually all previous growth curve models focuses on the study of mean growth profile. The modelling of biological variation during human growth, to our knowledge, has not received any scrutiny.

Intuitively, biological variation of body size generally increases with age due to the influence of increasing number of subject-specific factors during postnatal growth. It would therefore be worthwhile to study how this variation accumulates. We hypothesise that biological variation at a certain stage of a growth process is an additive sum of the past and current contributions which shall be named as transmitted and added variation respectively. On the other hand, measurement error is prone to exist in body size measurements which contribute an additional variation component. It is therefore interesting to study the decomposition of the total variation of body size measurements taken at a certain time in life into components that link to different sources.

For panel data with no repeated measurements, models for variation decomposition are available to identify the sources of variation in manufacturing processes consisting of se-

quential process stages.<sup>6–8</sup> Both ad hoc methods and the EM algorithm utilising Kalman filtering were proposed in order to estimate the various variation components when one or more types of measurements are taken.<sup>7,8</sup> Previously developed methods assumed a known variation component from measurement error and have not explicitly examined the effects of time-dependent covariates. In addition, model augmentation was required during the EM algorithm which undesirably increased its computing complexity.<sup>8</sup>

In this article, we extend previous models for variation decomposition and modify the estimation methods to identify the transmitted and added variation of body size in early life after eliminating the unknown variation due to measurement error. Moreover, estimation methods that dispense with model augmentation, and estimation of time-dependent covariates are also considered.

The next section describes a dataset that contains body size measurements in early life. This is followed by the development of an extended variation decomposition model in Section 3, where a simple estimation procedure when there are no missing values is also proposed. With the presence of missing values, an EM estimation procedure without the need of model augmentation is outlined in Section 4. Using the proposed model and methods, the set of lower leg length data is analysed in Section 5. Interestingly, the results are consistent to the previously hypothesised theory for a phase transition from infancy to childhood.<sup>2</sup> Moreover, the effect of gender on postnatal growth is found to be time-varying. Finally, some concluding remarks regarding the use of the proposed model and methods are provided in Section 6.

## 2. THE LOWER LEG LENGTH DATA

Body growth in early life is preferably measured by the lower leg length (LLL) because it is less prone to measurement error when compared with other measures such as body length.<sup>9</sup> A total of 61 new births (29 females and 32 males) delivered at a university hospital in Hong Kong SAR, PR China were randomly selected and followed for 2 years. The parents of all the infants that participated in the study signed an informed consent form before the infants were monitored. The lower right leg of each infant was measured by using a knemometer, expressed in millimeters, and 10 readings were obtained each time at birth, every 2 months during the first year, and every 3 months during the second year. Some infants did not however attend every scheduled visit or did not have all the 10 measurements recorded; only 40 infants were fully compliant to the study and 15.2% of scheduled measurements were missing. Interesting questions to be considered, after eliminating the unknown effects of measurement error are: how the biological variation was incremented in successive measurement visits, and is there any effects of gender on LLL growth?

## 3. THE VARIATION DECOMPOSITION MODEL

Let  $N$  be the number of subjects who undergo  $T$  measurement visits. At the  $t^{\text{th}}$  visit,  $K_{it}$  measurements on subject  $i$  are obtained and denoted by  $y_{itk}$ 's ( $k = 1, 2, \dots, K_{it}$ ). These measurements  $y_{itk}$ 's are however subject to measurement error and let the corresponding values that are free from measurement error be  $z_{it}$ 's. Then the variation decomposition

model is formulated as

$$z_{it} = a_t + b_t z_{i,t-1} + \beta_t x_{it} + \delta_{it}; \quad \delta_{it} \sim \mathcal{N}(0, \sigma_{\delta_t}^2) \quad (1)$$

$$y_{itk} = z_{it} + e_{itk}; \quad e_{itk} \sim \mathcal{N}(0, \sigma_{e_t}^2) \quad (2)$$

where  $b_1 = 0$ , and  $\delta_{it}$ 's and  $e_{itk}$ 's are independent. The  $x_{it}$  represents a covariate that can be a demographic variable or an indicator of treatment group in a clinical trial. This model is consistent to the intention-to-treat principle and allows the occurrence of missing values at any measurement visits. In particular,  $K_{it} = 0$  corresponds to the situation when no measurements are taken from subject  $i$  at visit  $t$  in which case (2) is omitted. The model however assumes the missing mechanism is random. From (1) and (2), we can easily derive that

$$\text{Var}(y_{itk}) = b_t^2 \text{Var}(z_{i,t-1}) + \sigma_{\delta_t}^2 + \sigma_{e_t}^2 \quad (3)$$

$$= \begin{array}{l} \text{variation} \\ \text{transmitted} \end{array} + \begin{array}{l} \text{variation} \\ \text{added} \end{array} + \begin{array}{l} \text{variation due to} \\ \text{measurement error} \end{array} \quad (4)$$

for  $t = 1, 2, \dots, T$ . As a result, variation at a visit, other than the one at birth, can be decomposed nicely into three components: biological variation transmitted from previous visits, biological variation added since last visit, and variation due to measurement error. Particularly,  $b_t = 1$  refers to 100% variation transmission while  $b_t > 1$  and  $b_t < 1$  refer to a variation explosion and absorption respectively. It is also worth noting that modification of model (1)-(2) to accommodate more than one type of measurement and more covariates is straightforward.

Indeed, the variation decomposition model can be cast into the general growth curve model proposed by Potthoff and Roy.<sup>4</sup> However, most general growth curve models focused on examining the dependence of mean growth on time and the corresponding usual inferential procedures are not directly applicable.<sup>1,3</sup> Moreover, the particular covariance structure of  $y_{itk}$ 's from (1) and (2) was not considered during the estimation of growth curves.<sup>10</sup>

Note that the set of unknown parameters under (1) and (2) includes  $a_t$ 's,  $b_t$ 's,  $\beta_t$ 's,  $\sigma_{\delta_t}^2$ 's, and  $\sigma_{e_t}^2$ 's which can generally be estimated by brute force through direct maximisation of the likelihood function based on observed data, i.e.  $y_{itk}$ 's. On the other hand, Markov Chain Monte Carlo methods can be employed which also provide standard errors of the estimates. The computational effort and time may however be substantially high in view of the large number of parameters.<sup>11</sup> Alternatively, a more efficient method is considered in Section 4. Meanwhile, when there are no missing values and all subjects have the same number of measurements taken at each visit, i.e.  $K_{it} = K$ , we can construct a set of naive estimates in closed form. With the notation

$$\begin{aligned}
K_{\cdot p} &= \sum_{i=1}^N K_{ip}, \quad \bar{y}_{ip\cdot} = \frac{1}{K_{ip}} \sum_{k=1}^{K_{ip}} y_{ipk}, \quad \bar{y}_{\cdot p\cdot} = \frac{1}{NK_{\cdot p}} \sum_{i=1}^N \sum_{k=1}^{K_{ip}} y_{ipk}, \\
\bar{x}_{\cdot p} &= \frac{1}{N} \sum_{i=1}^N x_{ip}, \quad S_x(p, q) = \frac{1}{N} \sum_{i=1}^N x_{ip}x_{iq} - \bar{x}_{\cdot p}\bar{x}_{\cdot q}, \\
S_y(p, q) &= \frac{1}{NK} \sum_{i=1}^N \sum_{k=1}^K y_{ipk}y_{iqk} - \bar{y}_{\cdot p}\bar{y}_{\cdot q}, \quad S_{yx}(p, q) = \frac{1}{NK} \sum_{i=1}^N \sum_{k=1}^K y_{ipk}x_{iq} - \bar{y}_{\cdot p}\bar{x}_{\cdot q}, \\
\text{and } BS_y(p, q) &= \frac{1}{N} \sum_{i=1}^N \bar{y}_{ip\cdot}\bar{y}_{iq\cdot} - \bar{y}_{\cdot p}\bar{y}_{\cdot q}.
\end{aligned}$$

for  $p, q = 1, 2, \dots, T$ , the naive estimates are defined as

$$\tilde{b}_t = \frac{BS_y(t, t-1)}{BS_y(t-1, t-1)} \quad (\tilde{b}_1 = b_1 = 0), \quad (5)$$

$$\tilde{\beta}_t = \frac{S_{yx}(t, t)}{S_x(t, t)}, \quad \tilde{a}_t = \bar{y}_{\cdot t} - \tilde{b}_t \bar{y}_{\cdot t-1} - \tilde{\beta}_t \bar{x}_{\cdot t}, \quad (6)$$

$$\tilde{\sigma}_{e_t}^2 = S_y(t, t) - BS_y(t, t), \quad \text{and} \quad \tilde{\sigma}_{\delta_t}^2 = BS_y(t, t) - \tilde{b}_t BS_y(t, t-1) \quad (7)$$

for  $t = 1, 2, \dots, T$ . These estimates are expressed explicitly and thus can be easily computed. Indeed, they are the maximum likelihood estimates when there are only two measurement visits. With known variation due to measurement error, a similar set of estimates was proposed which has been shown to perform reasonably well when compared with the corresponding set of maximum likelihood estimates.<sup>8,12</sup>

#### 4. PARAMETER ESTIMATION

The log-likelihood function based on  $y_{itk}$ 's can be written as

$$\begin{aligned} l_{\mathcal{O}} = & -\frac{K_{\cdot\cdot}}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^N \sum_{t=1}^T \left[ K_{it} \log \sigma_{e_t}^2 + \log \left( 1 + \frac{\sigma_{iz}^2(t|t-1)}{\sigma_{e_t}^2} K_{it} \right) \right] \\ & - \frac{1}{2} \sum_{i=1}^N \sum_{t=1}^T \left[ \sum_{k=1}^{K_{it}} \frac{(y_{itk} - \bar{y}_{it\cdot})^2}{\sigma_{e_t}^2} + \frac{(\bar{y}_{it\cdot} - z_{it|t-1})^2}{\sigma_{iz}^2(t|t-1)} K_{it} p_{it} \right] \end{aligned} \quad (8)$$

where  $K_{\cdot\cdot} = \sum_{i=1}^N \sum_{t=1}^T K_{it}$  is the total number of measurements,  $z_{ip|q} = E[z_{ip}|y_{itk}; k = 1, 2, \dots, K_{it}, t = 1, 2, \dots, q]$ ,  $\sigma_{iz}^2(p|q) = Var[z_{ip}|y_{itk}; k = 1, 2, \dots, K_{it}, t = 1, 2, \dots, q]$ , and  $cov_{iz}(p_1, p_2|q) = Cov[z_{ip_1}, z_{ip_2}|y_{itk}; k = 1, 2, \dots, K_{it}, t = 1, 2, \dots, q]$  for  $p, q, p_1, p_2 = 1, 2, \dots, T$ . It can be maximised by using EM-algorithm with the E-step facilitated by a



Kalman filter recursion. With a set of initial estimates, e.g. the naive estimates, the EM recursion continues to iterate until the increase in  $l_O$  is smaller than a pre-specified tolerance. The corresponding procedures are described in Appendix I. Note that  $cov_{iz}(t, t-1|t)$ 's are explicitly evaluated. Otherwise, it would be computed by augmenting (1) into  $(z_{it}, z_{i,t-1})^T = (a_t, 0)^T + \begin{bmatrix} b_t & 0 \\ 1 & 0 \end{bmatrix} (z_{i,t-1}, z_{i,t-2})^T + (\beta_t, 0)^T x_{it} + (\delta_{it}, 0)^T$  and the Kalman filter recursion would be derived in a bivariate setting that made computation less efficient especially during the estimation of standard errors (see the followings).

Note however that allowing the covariate effect to differ from time epoch may result in over-fitting for datasets of only moderate size. It is therefore desirable to represent  $\beta_t$ 's in terms of a small number of underlying parameters. The corresponding inference requires a modification of the M-step of the EM recursion and may slightly increase the computing time. The E-step however remains virtually the same. Details of the modification are described in Appendix II.

Standard errors of the estimates may be obtained by approximating the observed information matrix from a hybrid accelerator.<sup>13</sup> However, in view of our large set of unknown parameters, we suggest to use the more straightforward parametric bootstrap techniques.<sup>14</sup> Specifically, the maximum likelihood estimates, say  $\hat{\theta}$ , are treated as if they were the true values of the parameters, and  $M$  sets of data are sampled from model (1)-(2). By repeating the estimation algorithm for each set of sampled data,  $M$  sets of maximum likelihood estimates are obtained and the corresponding standard deviations are then used to approximate the standard errors of  $\hat{\theta}$ . Diagnostics for model adequacy can be performed by checking if the generalised conditional residuals, i.e. the summand within the last square bracket of (8),

are approximately independent and distributed as  $\chi^2$  with  $K_{it}$  degrees of freedom.

Maximum likelihood estimates utilise all available data while naive estimates are simple to implement. A small size simulation study was performed to assess the performance of the maximum likelihood estimates with respect to the amount of missing values. A total of 200 samples were generated from model (1)-(2) by using the estimates obtained from analysing the LLL data in the next section as the hypothesised values. The corresponding estimates were examined in terms of their biases and nominal properties when compared with a Normal distribution. The general results are that when the amount of missing values varied, the biases were consistently small (less than 5%) while the skewnesses and kurtoses consistently well agreed with 0 and 3 respectively. The simulation study also compared maximum likelihood estimates with the corresponding naive estimates. Naive estimates were slightly more biased and skewed than maximum likelihood estimates when there were less than 10% of missing values in samples of size 61. With more missing values, differences between the two sets of estimates became larger.

## 5. DECOMPOSITION OF LOWER LEG LENGTH VARIATION

We now decompose the LLL variation of the 61 Chinese infants as described in Section 2. The only covariate considered was gender, and  $x_{it}$  was taken as 1 if subject  $i$  is a male and 0 otherwise. Although gender does not vary with time, its effect was however treated as time-varying. Hence, the corresponding inferential procedure was the same as if a time-dependent covariate was considered. To compute the naive estimates, we used 46 infants who had at least 5 repeated LLL measurements, i.e.  $K = 5$ , taken at all visits. The maximum likelihood

estimates that utilised all data were then obtained from the EM recursion in Appendix I with the naive estimates served as the initial estimates. The EM iteration was stopped when the increase in  $l_O$  was less than  $10^{-5}$ . Standard errors of the resulting estimates were obtained by parametric bootstrap of size 10,000. All analyses were performed using the Statistical Analysis System (SAS) version 6.12 in a Pentium III 500Hz machine.

The observed log-likelihood was maximised at  $-14261.9$  which improved the initial value from the naive estimates by 96.0. Plot of the empirical quantiles from generalised conditional residuals against the  $\chi^2(10)$  quantiles did not reveal any severe model departures. Besides, the bias, skewness and kurtosis were obtained from parametric bootstrap for each maximum likelihood estimate which were reasonably close to 0, 0, and 3 respectively. The Normal distribution was therefore used for inferences on the estimates.

Table I summarises the maximum likelihood estimates of the variation components, and Figure 1 depicts the corresponding values. Standard errors of these estimates were less than 30% of the corresponding estimates. Both the variation due to measurement error and biological variation added between successive visits were fairly stable over time. The cumulative biologically added variation since birth is also shown for comparison with the total biological variation. Note that the former was consistently larger than the later since not all added variation was transmitted to later infant growth. Specifically, the cumulative biologically added variation fairly agreed with the total biological variation during the first year but started to take off at the end of one year (Figure 1). In other words, the biologically added variation was almost all transmitted during the first year of infant growth but most of it was being absorbed afterwards. This happens to match nicely with the ICP theory that

there is an abrupt change of growth velocity during the second half of the first year of life when there is a transition from the infancy phase to the childhood phase.<sup>15</sup> Particularly, the deceleration of total biological variation is consistent to the observation that the variation of growth velocity becomes smaller after the onset of childhood.<sup>15</sup>

Apart from studying the variation contribution from different sources, we can also examine the effects of covariates after accounting for measurement error. Figures 2 and 3 display the mean LLL and the estimated effects of gender with the corresponding two-sided 95% confidence band respectively for infant growth during the first two years after birth. Gender did not appear to have an effect on the mean LLL at birth nor six months after birth. Male infants however tended to have a longer LLL between two to four months after birth. Indeed, the likelihood ratio test statistics for a constant gender effect over time was  $28.36 > 18.31$  ( $= \chi_{10}^2(0.95)$ ). Therefore, male Chinese infants grew faster at the first few months while female Chinese infants caught up at six months after birth. The underlying biological rationale would be of interest to the paediatricians.

## 6. CONCLUDING REMARKS

For panel data with repeated measurements, we proposed the variation decomposition model for the study of variation decomposition in the presence of covariates and unknown measurement error. Estimation methods were developed and previous model augmentation during maximum likelihood estimation is no longer required. Clinical trials often have more than one repeated measurement visit in addition to baseline in order to explore treatment effectiveness over time. While our model studies the variation decomposition, it can also be employed to

determine the treatment effect, in the form of a covariate, at different time epochs. In the sequel, study of sample size calculation procedures are worthwhile accompany the analysis of the model. Moreover, multivariate extension of the model to accommodate multiple types of body measurements can also be proceeded unambitiously.<sup>8</sup>

The additive assumption of variation components for LLL seems appropriate as no evidence was found that this was violated. Interestingly, a change of variation transmission was observed that parallels with the transition from infancy to childhood under the proposition of the three phases of human growth under the ICP model.<sup>2</sup> A longer observation period (up to 20 years) would perhaps be desirable to unveil, if it exists, the second transition from childhood to puberty in variation transmission. Another interesting result is the significant effect of gender at the first few months after birth but not at later life. This poses a question to the biologist for further studies.

It is worthy to note the stagewise or previously called residual analysis approach for inference of covariates that are correlated with both previous and current growth measurements and confounded with the associated inferences.<sup>16</sup> The stagewise approach removes the covariate effects prior to the time when the previous growth measurement was taken, and is more epidemiologically sound. It was however demonstrated that the stagewise approach is less powerful and may have a relatively high bias when compared with the usual analysis approach as in ours.<sup>17,18</sup> Further consideration on the use of the two approaches is therefore deemed necessary.

## APPENDIX I

The E-step amounts to compute  $z_{it|T}$ ,  $\sigma_{iz}^2(t|T)$ , and  $cov_{iz}(t, t-1|T)$  which can be obtained from the following Kalman predictors and filters. For  $i = 1, 2, \dots, N$ , we take  $z_{i0|0} = \sigma_{iz}^2(0|0) = 0$  and compute

$$\begin{aligned}
z_{it|t-1} &= a_t + b_t z_{i,t-1|t-1} + \beta_t x_{it} \\
\sigma_{iz}^2(t|t-1) &= b_t^2 \sigma_{iz}^2(t-1|t-1) + \sigma_{\delta_t}^2 \\
p_{it} &= \left[ 1 - \frac{K_{it}}{\sigma_{e_t}^2} \left( \frac{1}{\sigma_{iz}^2(t|t-1)} + \frac{K_{it}}{\sigma_{e_t}^2} \right)^{-1} \right] \frac{\sigma_{iz}^2(t|t-1)}{\sigma_{e_t}^2} \\
z_{it|t} &= z_{it|t-1} + K_{it} p_{it} (\bar{y}_{it} - z_{it|t-1}) \\
\sigma_{iz}^2(t|t) &= (1 - K_{it} p_{it}) \sigma_{iz}^2(t|t-1) \\
\text{and } cov_{iz}(t, t-1|t) &= (1 - K_{it} p_{it}) b_t \sigma_{iz}^2(t-1|t-1)
\end{aligned}$$

for  $t = 1, 2, \dots, T$ . Then the smoothers are computed from

$$\begin{aligned}
r_{it} &= \frac{\sigma_{iz}^2(t|t)}{\sigma_{iz}^2(t+1|t)} b_{t+1} \\
z_{it|T} &= z_{it|t} + r_{it} (z_{i,t+1|T} - z_{i,t+1|t}) \\
\sigma_{iz}^2(t|T) &= \sigma_{iz}^2(t|t) - r_{it}^2 [\sigma_{iz}^2(t+1|t) - \sigma_{iz}^2(t+1|T)] \\
\text{and } cov_{iz}(t, t-1|T) &= \frac{\sigma_{iz}^2(t|T)}{\sigma_{iz}^2(t|t)} cov_{iz}(t, t-1|t)
\end{aligned}$$

backwardly from  $t = T - 1$  to 1.

Given the current iterated estimates, the M-step is the maximisation of the likelihood

function based on  $y_{itk}$ 's and  $z_{it}$ 's, i.e.

$$l_C = -\frac{1}{2} \sum_{i=1}^N \sum_{t=1}^T \sum_{k=1}^{K_{it}} \left( \log \sigma_{e_t}^2 + \frac{e_{itk}^2}{\sigma_{e_t}^2} \right) - \frac{1}{2} \sum_{i=1}^N \sum_{t=1}^T \left( \log \sigma_{\delta_t}^2 + \frac{\delta_{it}^2}{\sigma_{\delta_t}^2} \right) + \text{a constant},$$

with respect to the unknown parameters. With the notation in (3)-(3), and  $\bar{z}_{\cdot p} = \frac{1}{N} \sum_{i=1}^N z_{ip}$ ,

$S_z(p, q) = \frac{1}{N} \sum_{i=1}^N z_{ip} z_{iq} - \bar{z}_{\cdot p} \bar{z}_{\cdot q}$ , and  $S_{zx}(p, q) = \frac{1}{N} \sum_{i=1}^N z_{ip} x_{iq} - \bar{z}_{\cdot p} \bar{x}_{\cdot q}$  for  $p, q = 1, 2, \dots, T$ ,

it can be shown that  $l_C$  reaches maximum when

$$\begin{aligned} \hat{\beta}_1 &= \frac{S_{zx}(t, t)}{S_x(t, t)}, & \hat{\beta}_t &= \frac{S_z(t-1, t)S_{zx}(t-1, t) - S_{zx}(t, t)S_z(t-1, t-1)}{S_{zx}(t-1, t)^2 - S_x(t, t)S_z(t-1, t-1)}, \\ \text{and } \hat{b}_t &= \frac{S_z(t-1, t) - \hat{\beta}_t S_{zx}(t-1, t)}{S_z(t-1, t-1)} \end{aligned}$$

for  $t = 2, 3, \dots, T$ , as well as

$$\begin{aligned} \hat{a}_t &= \bar{z}_{\cdot t} - \hat{b}_t \bar{z}_{\cdot t-1} - \hat{\beta}_t \bar{x}_{\cdot t}, \\ \hat{\sigma}_{\delta_t}^2 &= S_z(t, t) - \hat{b}_t^2 S_z(t-1, t-1) + \hat{\beta}_t^2 S_x(t, t) - 2\hat{\beta}_t S_{zx}(t, t) \\ \text{and } \hat{\sigma}_{e_t}^2 &= \frac{1}{K_{\cdot t}} \left[ \sum_{i=1}^N \sum_{k=1}^{K_{it}} y_{itk}^2 - 2 \sum_{i=1}^N \sum_{k=1}^{K_{it}} y_{itk} z_{it} + \sum_{i=1}^N K_{it} z_{it}^2 \right] \end{aligned}$$

for  $t = 1, 2, \dots, T$  where  $K_{\cdot t} = \sum_{i=1}^N K_{it}$  is the total number of observations at stage  $t$ . The

next iterated estimates are then computed from the above after replacing  $z_{it}$ ,  $z_{it}^2$ , and  $z_{it} z_{i, t-1}$

by

$$z_{it|T}, \quad \sigma_{iz}^2(t|T) + z_{it|T}^2, \quad \text{and} \quad cov_{iz}(t, t-1|T) + z_{it|T} z_{i, t-1|T}$$

respectively.

## APPENDIX II

When  $\beta_t = f_t(\theta)$  for a certain time-dependent function  $f_t$  and vector of underlying parameters  $\theta$ , the M-step amounts to solve

$$\sum_{t=1}^T \frac{1}{\sigma_{\delta_t}^2} \frac{\partial f_t(\theta)}{\partial \theta} [S_{zx}(t, t) - b_t S_{zx}(t-1, t) - f_t(\theta) S_x(t, t)] = 0,$$

$$\sigma_{\delta_t}^2 = S_z(t, t) - b_t^2 S_z(t-1, t-1) + f_t(\theta)^2 S_x(t, t) - 2f_t(\theta) S_{zx}(t, t); \quad t = 1, 2, \dots, T,$$

$$\text{and } S_z(t-1, t) - b_t S_z(t-1, t-1) - f_t(\theta) S_{zx}(t-1, t) = 0; \quad t = 2, 3, \dots, T$$

for  $\hat{\theta}$ ,  $\hat{\sigma}_{\delta_t}^2$ 's and  $\hat{b}_t$ 's. The other estimates are explicitly obtained as in Appendix I by replacing  $\hat{\beta}_t = f_t(\hat{\theta})$ . Numerical methods, such as Newton Raphson or Simple iteration, may however be needed to solve for  $\hat{\theta}$ 's and  $\hat{\sigma}_{\delta_t}$ 's.

When  $\beta_t$ 's =  $\beta$ , the M-step becomes

$$\beta = \left\{ \sum_{t=1}^T \frac{S_{zx}(t, t)}{\sigma_{\delta_t}^2} - \sum_{t=2}^T \frac{S_{zx}(t-1, t) S_z(t-1, t)}{\sigma_{\delta_t}^2 S_z(t-1, t-1)} \right\} / \left\{ \sum_{t=1}^T \frac{S_x(t, t)}{\sigma_{\delta_t}^2} - \sum_{t=2}^T \frac{S_{zx}(t-1, t)^2}{\sigma_{\delta_t}^2 S_z(t-1, t-1)} \right\}, \quad (9)$$

$$\sigma_{\delta_t}^2 = S_z(t, t) - \frac{[S_z(t-1, t) - \beta S_{zx}(t-1, t)]^2}{S_z(t-1, t-1)} I(t \neq 1) + \beta^2 S_x(t, t) - 2\beta S_{zx}(t, t), \quad (10)$$

$$\text{and } b_1 = 0, \quad b_t = \frac{S_z(t-1, t) - \beta S_{zx}(t-1, t)}{S_z(t-1, t-1)}; \quad t = 2, 3, \dots, T$$

where  $I(\cdot)$  is the usual indicator function. Simple iterations for (9) and (10) can be used to obtain  $\hat{\beta}$  and  $\hat{\sigma}_{\delta_t}^2$  while other estimates are computed as in Appendix I. Naive estimate for  $\beta$  can be taken as  $\tilde{\beta} = \frac{1}{T} \sum_{t=1}^T \frac{S_{yx}(t, t)}{S_x(t, t)}$ .

When  $\beta_t$ 's = 0, the EM recursion in Appendix I proceeds with  $\beta_t$ 's replaced by 0.



## ACKNOWLEDGEMENTS

The authors would like to thank Professor L. Low and Dr Selene Tam from the Department of Paediatrics and Clinical Trials Centre of The University of Hong Kong for their kind permission to use the dataset. Comments from the referees are very helpful in improving the manuscript and the financial support from the Committee on Research and Conference Grants is also gratefully acknowledged.

## REFERENCES

1. Carter, E.M. and Hubert, J. J. ‘A Growth-Curve Model Approach to Multivariate Quantal Bioassay’, *Biometrics*, **40**, 699–706 (1984).
2. Karlberg, J. ‘On the Modelling of Human Growth’, *Statistics in Medicine*, **6**, 185–192 (1987).
3. Lee, J. C. ‘Tests and Model Selection for the General Growth Curve Model’, *Biometrics*, **47**, 147–159 (1991).
4. Potthoff, R. F. and Roy, S. N. ‘A generalized multivariate analysis of variance model useful especially for growth curve problems’, *Biometrika*, **51**, 313–326 (1964).
5. Preece, M. A. and Baines, M. J. ‘A new family of mathematical models describing the human growth curve’, *Annals of Human Biology*, **5**, 1–24 (1978).
6. Lawless, J., MacKay, R. J. and Robinson, J. A. ‘Analysis of variation transmission in manufacturing processes - Part I’, *Journal of Quality Technology*, **31**, 131–142 (1999).

7. Agrawal, P., Lawless, J. F. and MacKay, R. 'Analysis of variation transmission for manufacturing processes - Part II', *Journal of Quality Technology*, **31**, 143–154 (1999).
8. Fong, D. Y. T. and Lawless, J. F. 'The Analysis of Process Variation Transmission with Multivariate Measurements', *Statistica Sinica*, **8**, 151–164 (1998).
9. Mitchell, L. C., Low, L. C. K. and Karlberg, J. 'Correlates of Lower Leg Length Growth and Weight Gain in Early Life', *Hong Kong Journal of Paediatrics*, **1**, 137–140 (1996).
10. Lee, J. C. 'Prediction and estimation of growth curves with special covariance structures', *Journal of the American Statistical Association*, **83**, 432–440 (1988).
11. Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. *Markov Chain Monte Carlo in Practice*, Chapman and Hall, New York, 1996.
12. Fong, D. Y. T. 'An Empirical Study of Two Classes of Estimators for Process Variation Transmission', *International Journal of Reliability, Quality and Safety Engineering*, **6**, 289–300 (1999).
13. Jamshidian, M. and Jennrich, R. I. 'Acceleration of the EM algorithm by using quasi-Newton methods', *Journal of Royal Statistical Society Series B*, **59**, 569–587 (1997).
14. Efron, B. and Tibsirani, R. J. *An Introduction to the Bootstrap*, Chapman and Hall, New York, 1993.
15. Karlberg, J., Engstrom, I., Karlberg, P. and Fryer, J. G. 'Analysis of linear growth using a mathematical model. I. From birth to three years', *Acta-Paediatr-Scand.*, **76(3)**, 478–488 (1987).

16. Esrey, S. A., Casella, G. and Habicht, J. 'The use of residuals for longitudinal data analysis: The example of child growth', *American Journal of Epidemiology*, **131**, 365–372 (1990).
17. Alley, W. M. 'A Note on Stagewise Regression', *The American Statistician*, **41(2)**, 132–134 (1987).
18. Smith, E. P. and Rose, K. A. 'Trend Detection in the Presence of Covariates: Stagewise versus Multiple Regression', *Environmetrics*, **2(2)**, 163–168 (1991).

Table I. Estimated variation components for the lower leg length growth data.

Age (in months)	Variation Components				Total variation
	Measurement error	Biologically added	Cumulative biologically added	Total biological	
At birth	7.20	12.36	12.36	12.36	19.56
2	7.10	11.66	24.02	22.99	30.09
4	6.61	9.90	33.92	31.75	38.36
6	6.15	6.92	40.84	34.13	40.28
8	7.70	6.00	46.84	46.10	53.80
10	7.82	9.13	55.97	45.03	52.85
12	5.65	14.71	70.68	70.27	75.92
15	5.63	10.03	80.71	53.08	58.71
18	5.84	15.50	96.21	73.27	79.11
21	7.12	11.94	108.15	71.83	78.95
24	5.61	14.72	122.87	76.78	82.39

Figure 1. Plots of total variation (—), total biological variation (—+—), cumulative biological variation (—○—), and variation due to measurement error (—●—) against age.

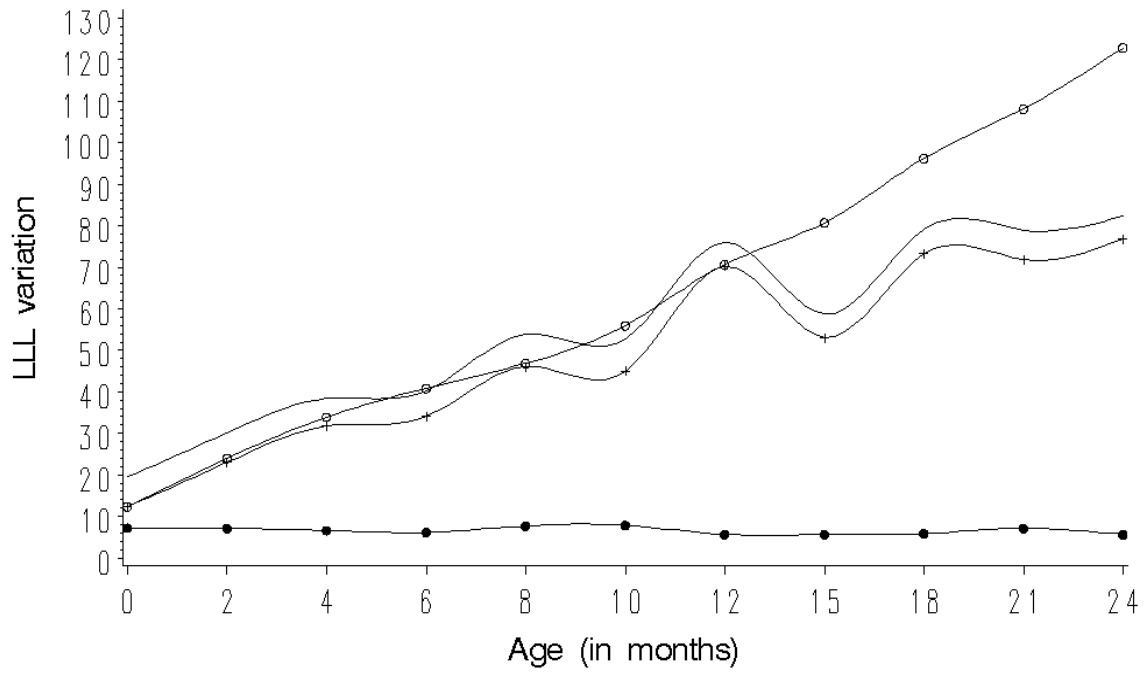


Figure 2. Mean lower leg length profile of the 61 Chinese infants.

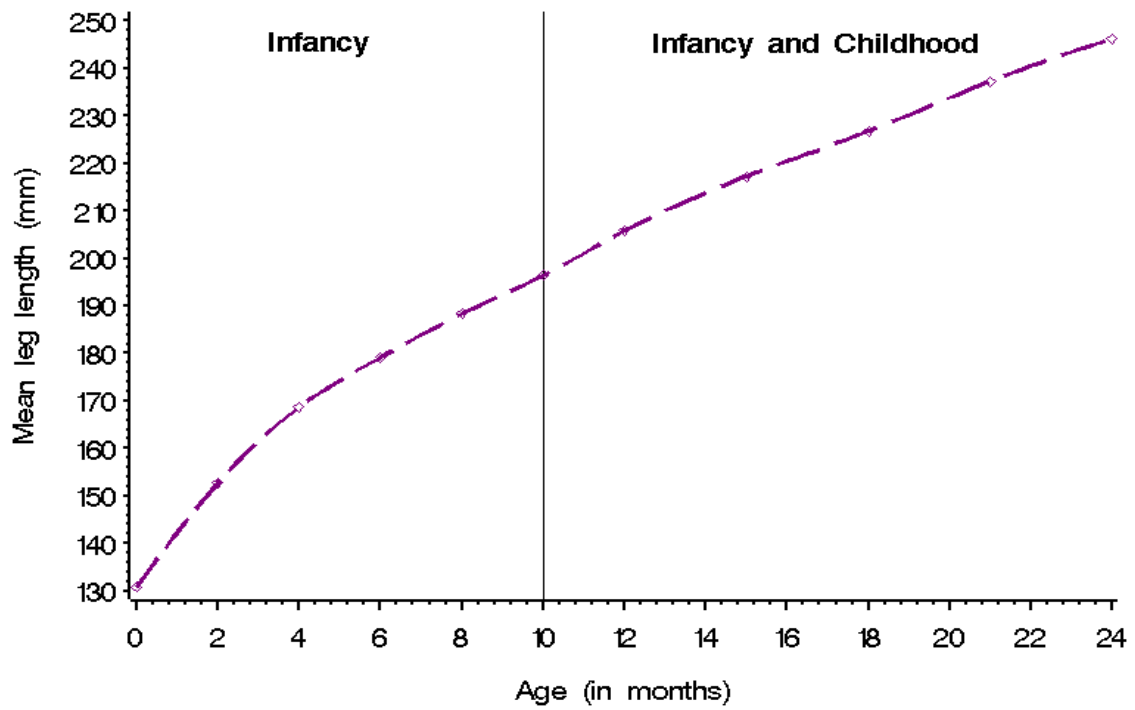


Figure 3. Estimated gender effects and the corresponding two-sided 95% confidence band.

