*Multivariate Behavioral Research*, 43:476–496, 2008

ISSN: 0027-3171 print/1532-7906 online
DOI: 10.1080/00273170802285941

Psychology Press
Taylor & Francis Group



# Investigating Ceiling Effects in Longitudinal Data Analysis

Lijuan Wang and Zhiyong Zhang
*University of Notre Dame*

John J. McArdle
*University of Southern California*

Timothy A. Salthouse
*University of Virginia*

Score limitation at the top of a scale is commonly termed "ceiling effect." Ceiling effects can lead to serious artifactual parameter estimates in most data analysis. This study examines the consequences of ceiling effects in longitudinal data analysis and investigates several methods of dealing with ceiling effects through Monte Carlo simulations and empirical data analyses. Data were simulated based on a latent growth curve model with $T = 5$ occasions. The proportion of the ceiling data [10%–40%] was manipulated by using different thresholds, and estimated parameters were examined for $R = 500$ replications. The results showed that ceiling effects led to incorrect model selection and biased parameter estimation (shape of the curve and magnitude of the changes) when regular growth curve models were applied. The Tobit growth curve model, instead, performed very well in dealing with ceiling effects in longitudinal data analysis. The Tobit growth curve model was then applied in an empirical cognitive aging study and the results were discussed.

Many tests and scales have been developed in psychological and educational research to measure participants' abilities, well-being, and other constructs.

Correspondence concerning this article should be addressed to Lijuan Wang, Department of Psychology, University of Notre Dame, Notre Dame, IN 46556. E-mail: lwang4@nd.edu

However, if a test is relatively easy, high-scoring participants may answer every item correctly and reach the highest possible score, or ceiling, on the test. When this happens, the true extent of the high-scoring participants cannot be correctly measured, and this phenomenon is usually termed “ceiling effects.” Uttl (2005) defined the ceiling effects as occurring when the tests or scales are relatively easy such that substantial proportions of individuals obtain either maximum or near-maximum scores and the true extent of their abilities cannot be determined.

Ceiling effects are related to, but different from, performance asymptotes. Asymptotes occur when participants’ scores cannot exceed a specific value with more information, additional practice, or retests (Miller, 1956). The asymptotic values are the greatest true values that participants can actually demonstrate. In this study, we assume ceiling effects happen before participants reach asymptotic values. The concept of ceiling effects is also distinct from the concept of semicontinuous variables (Olsen & Schafer, 2001). A semicontinuous variable combines a continuous distribution with point masses at one or more locations. For example, in alcohol usage research, the alcohol usage variable is a mixture of 0s and continuously distributed positive values, which is one of the typical semicontinuous variables. The difference between semicontinuous variables and ceiling effects is that the 0s are valid data values, not proxies for negative or missing response (Olsen & Schafer, 2001), whereas the ceiling threshold is not a valid data value and is a proxy for some larger true values.

However, the idea of ceiling effects is similar to the concept of right censoring in the framework of survival analysis. In survival analysis, right censoring is considered to be where the event is observed only if it occurs prior to some prespecified time (Klein & Moeschberger, 2005). Similarly for ceiling effects, the true score can be observed only if it is less than or equal to the ceiling threshold. The difference between these two concepts is that right censoring is related to the time until an event whereas ceiling effects are more related to measurement and testing properties.

Many researchers have observed ceiling effects in their research and recognized that there could be some problems in the results such as artifactual nonlinearity or underestimated regression parameters (e.g., Genia, 2001; Ledbetter, Smith, Vosler-Hunter, & Fischer, 1991; Murrell, Kenealy, Beaumont, & Lintern, 1999). Uttl (2005) provided a comprehensive discussion of severe ceiling effects in widely used memory tests such as the verbal paired associates and word list tests from the Wechsler Memory Scales (WMS; Wechsler, 1945, 1987, 1997), the Rey Auditory Verbal Learning Test (RAVLT; Rey, 1964), and the California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987). Among the adverse effects of low ceilings mentioned by Uttl were underestimated means and standard deviations and attenuated reliability and validity.

To deal with cross-sectional ceiling data, Muthen (1989, 1990) developed and applied the Tobit approach to analyze censored data (ceiling or floor data) for a factor analysis model in two steps. The Tobit correlation matrix is estimated in the first step, and the factor loadings and variances are estimated based on the Tobit correlation matrix in the second step. Van den Oord & Van der Ark (1997) adjusted this Tobit approach to analyze sums of Likert items with ceiling or floor effects for factor analysis and found that the results were more accurate.

With the use of repeated measures in longitudinal data, the problems of ceiling effects could be more severe in that participants improve on the tests over time and more participants reach the ceiling at later time points. However, there has not yet been much discussion of their impact, or possible solutions, with this type of data. The goal of this study was therefore to investigate the influences of ceiling effects in longitudinal data analysis and explore possible methods of dealing with ceiling effects. First, influences of ceiling effects in longitudinal data are evaluated through simulation studies. Different ceiling proportions are manipulated by setting different ceiling thresholds. The impact of longitudinal ceiling effects is examined by comparing parameter estimates from models that do not consider ceiling effects with the true parameter values. Accuracy (comparison between the mean of the empirical distribution of the parameter estimates and the true parameter value) and precision (standard deviation of the empirical distribution of the estimates and the true standard error) are used to evaluate the bias. Second, several analytical methods are discussed and their performance compared. Finally, an empirical longitudinal data set with ceiling problems is analyzed and the results discussed.



## CEILING EFFECTS IN LONGITUDINAL STUDIES

Many longitudinal studies have been conducted to simultaneously investigate the intraindividual change pattern and the interindividual differences of the intraindividual change (Baltes & Nesselroade, 1979). Because participants are repeatedly measured in the longitudinal studies, it is almost an article of faith that longitudinal data are correlated. Regular regression techniques cannot be applied in analyzing longitudinal data due to the violation of independent observations assumption. Instead, growth curve models or mixed-effects models were proposed and are widely used to analyze longitudinal data by considering the joint probability density function for the repeated measures (Fitzmaurice, Laird, & Ware, 2004; Laird & Ware, 1982; McArdle & Nesselroade, 2003; Meredith & Tisak, 1990).

To demonstrate the problems with ceiling effects in longitudinal data, we considered using the latent basis growth curve model, which is a relatively flexible model in the growth curve modeling framework (Meredith & Tisak, 1990). A form of the latent basis growth curve model is expressed in Equation (1). The observed score $y_{it}$ for $i$th individual at occasion $t(t = 1, 2, \ldots T)$ is regressed on the latent random-effects parameters $(b_{0i}, b_{1i})$ by some basis coefficients or shape parameters $A(t)$. Assigning $A(t)$ different values will lead to different shapes of the change pattern. Let $A(1) = 0$ and $A(T) = 1$, then $b_{0i}$ can be explained as the individual initial level at Time 1 and $b_{1i}$ can be explained as the individual change from Time 1 to Time T for person $i$ (McArdle, 2004). And $A(t)(t = 2 \ldots T-1)$ represents the free loading reflecting the *proportion* of change between two timepoints relative to the total change occurring from the first to the last timepoints. The fixed-effects parameter vector $(\beta_0, \beta_1)$ is the initial level and change estimates for the "average person" in the sample. The $\Phi$ matrix represents the interindividual variation of the latent random-effects parameter vector $(b_{0i}, b_{1i})$ and $\sigma_t^2$ estimates the residual variance of each occasion.

$$
\begin{aligned}
y_{it} &= b_{0i} + A(t)b_{1i} + e_{it} \\
A(1) &= 0, A(T) = 1 \\
b_{0i} &= \beta_0 + u_{0i} \\
b_{1i} &= \beta_1 + u_{1i} \\
\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} &\sim N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Phi = \begin{pmatrix} \phi_{00} & \phi_{01} \\ \phi_{10} & \phi_{11} \end{pmatrix} \right) \\
e_{it} &\sim N(0, \sigma_t^2)
\end{aligned} \tag{1}
$$

If there are no ceiling data in the longitudinal data, then $y_{it}$ is the true observed score. If ceiling effects exist in the longitudinal data, we can only observe the ceiling threshold for the ceiling data, which may misrepresent the true scores of those individuals.

In order to investigate the consequences of ceiling effects in longitudinal data analysis, data were simulated based on a latent growth curve model. In each simulated sample, data with $N = 200$ participants and $T = 5$ occasions were simulated based on the following true parameter values. $A(t)$ was set to be as (0,.25,.50,.75,1), which indicates a linear growth change pattern. The fixed-effects parameter vector $(\beta_0, \beta_1)$ was set to be as (7.5, 5.0), $\Phi$ matrix was set to be as $\phi_{00} = 2^2$, $\phi_{11} = 1.5^2$, $\phi_{01} = \phi_{10} = 0$. $\sigma_t^2$ was set to be as $1^2$. The correlation between initial level and change latent variables was set to be 0 to have a clearer observation of the influences of ceiling data. These population values were chosen partly based on the empirical data analysis results in a later section and modified or simplified to make the results clearer. Five hundred replications of data simulation and analysis were implemented.

The data without ceiling effects were first simulated based on the true model described earlier. New data sets were then created by adding ceiling effects.

TABLE 1
Ceiling Proportions at 5 Occasions With Different Ceiling Thresholds (CT)

| *Time* | *Range of the True Scores* | *CT = 15* | *CT = 14* | *CT = 13* |
|---|---|---|---|---|
| Occasion 1 | [−2.36, 16.38] | 0.04% | 0.18% | 0.69% |
| Occasion 2 | [−0.87, 18.76] | 0.30% | 1.04% | 3.09% |
| Occasion 3 | [−0.16, 19.98] | 1.64% | 4.42% | 10.19% |
| Occasion 4 | [0.71, 23.73] | 6.64% | 13.68% | 24.30% |
| Occasion 5 | [−0.14, 25.12] | 17.61% | 28.87% | 42.63% |

Because different proportions of ceiling data may have different magnitudes of influences on parameter estimates, three ceiling thresholds (13, 14, and 15) were used to manipulate different ceiling proportion conditions. For example, if the ceiling threshold is 13, then individual scores larger than 13 in each occasion were forced to be 13. The ceiling proportions with different ceiling thresholds are displayed in Table 1. There are larger proportions of ceiling data in later occasions than those in earlier occasions due to simulated increasing growth trajectories. The ceiling proportions of the generated data in the fifth occasion ranged from 18% to 43% based on different ceiling thresholds.

The sample mean and covariance matrix of the generated data with ceiling thresholds = 13 were compared with those true values (Table 2). Both the mean vector and the covariance matrix were underestimated in most occasions. With larger proportions of ceiling data, biases were larger. The maximum likelihood

TABLE 2
Sample Mean Vector and Covariance Matrix With Ceiling Threshold (CT) = 13
Compared With the True Values

| | *True Values Without Ceilings* | | | | | *CT = 13* | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | y1 | y2 | y3 | y4 | y5 | y1 | y2 | y3 | y4 | y5 |
| Mean vector | 7.5 | 8.8 | 10.0 | 11.3 | 12.5 | 7.5 | 8.7 | 9.9 | 10.9 | 11.7 |
| Covariance matrix | | | | | | | | | | |
| y1 | 5.0 | | | | | 4.9 | | | | |
| y2 | 4.0 | 5.2 | | | | 3.9 | 4.9 | | | |
| y3 | 4.0 | 4.3 | 5.6 | | | 3.6 | 3.8 | 4.7 | | |
| y4 | 4.0 | 4.4 | 4.8 | 6.3 | | 3.0 | 3.3 | 3.5 | 4.0 | |
| y5 | 4.0 | 4.6 | 5.1 | 5.7 | 7.2 | 2.3 | 2.6 | 2.8 | 2.8 | 3.0 |

*Note.* The values in the sample mean vector and the covariance matrix are the average values of the 500 replications.

estimation (MLE) is the most widely used method in estimating a growth curve model (Fitzmaurice et al., 2004; Verbeke & Molenberghs, 2000). MLE, especially in the structural equation modeling (SEM) framework, is based on fitting the sample mean vector and sample covariance matrix to the model. Thus, the parameter estimates could also be biased when regular MLE is applied to analyze longitudinal ceiling data and the ceiling effects are not considered. Table 3 displays the MLE results of fitting a latent basis growth curve model, which freely estimates basis coefficient parameters $A(2)$, $A(3)$, and $A(4)$ to allow us to quantify the amount of artifactual nonlinearity for the different levels of ceiling data, to the generated data without considering ceiling effects. Values in the second column are the parameter estimates for the simulated data without ceilings and are consistent with the true values. Columns 3–5 display the parameter estimates for data with different ceiling thresholds.

For the shape parameters $A(t)$, the true shape is linear (true $A(t) = 0$, .25, .50, .75, 1), but the estimated curve was biased to be nonlinear when ceiling effects existed. For example, when ceiling threshold was set to be 13, the estimated $A(t)$ was equal to (0,.29, .57, .82, 1). For the fixed-effects parameters, the estimate of the initial level parameter $\beta_0$ from the data with ceilings was consistent with the true values because the ceiling proportion in the first occasion is relatively small (less than 1%; see in Table 1). However, the estimate of the change parameter $\beta_1$ was substantially underestimated. The true value of the covariance $\phi_{01}$ of the latent random-effects level and slope parameters is 0, which indicates that the individual initial level is not correlated with the individual change over

TABLE 3
Regular Maximum Likelihood Estimation (MLE) Estimates of the Parameters From Fitting a Latent Basis Growth Curve Model to the Simulated Data With or Without Ceilings

| *Parameters and True Values* | *No Ceilings* | *CT = 15* | *CT = 14* | *CT = 13* |
|---|---|---|---|---|
| $A(2) = 0.25$ | 0.25 (0.02) | 0.26 (0.02) | 0.28 (0.02) | 0.29 (0.02) |
| $A(3) = 0.50$ | 0.50 (0.02) | 0.53 (0.02) | 0.55 (0.02) | 0.57 (0.02) |
| $A(4) = 0.75$ | 0.75 (0.02) | 0.78 (0.02) | 0.80 (0.02) | 0.82 (0.02) |
| $\beta_0 = 7.50$ | 7.50 (0.16) | 7.49 (0.16) | 7.50 (0.16) | 7.50 (0.16) |
| $\beta_1 = 5.00$ | 5.01 (0.14) | 4.75 (0.13) | 4.52 (0.12) | 4.16 (0.12) |
| $\phi_{00} = 4.00$ | 4.00 (0.46) | 4.27 (0.49) | 4.40 (0.49) | 4.51 (0.48) |
| $\phi_{01} = 0.00$ | −0.01 (0.30) | −0.67 (0.31) | −1.13 (0.33) | −1.71 (0.34) |
| $\phi_{11} = 2.25$ | 2.23 (0.39) | 1.80 (0.33) | 1.72 (0.31) | 1.77 (0.31) |
| $\sigma^2 = 1$ | 0.99 (0.06) | 0.96 (0.06) | 0.92 (0.05) | 0.86 (0.05) |
| $\mathrm{Corr}(b_{0i}, b_{1i}) = 0$ | 0.00 (0.10) | −0.24 (0.10) | −0.41 (0.09) | −0.60 (0.07) |

*Note.* The values outside the parentheses are the means of the estimates from the 500 replications and the values inside the parentheses are the standard deviations of the estimates from the 500 replications.

TABLE 4
Results of the Model Selections for the Simulated Data With and Without Ceilings

| | $\chi^2$ *Difference (Linear vs. Quadratic)* | | | |
|---|---|---|---|---|
| *Conditions* | *Mean* | *Minimum* | *Maximum* | *Correct Selection Proportion* |
| No ceiling | 3.95 | 0.07 | 15.82 | 95.4% |
| CT = 15 | 12.73 | 1.05 | 39.20 | 34.2% |
| CT = 14 | 25.04 | 4.77 | 55.34 | 2.2% |
| CT = 13 | 45.88 | 14.39 | 82.92 | 0% |

*Note.* Difference in degrees of freedom = 4. $\chi^2(0.95, 4) = 9.49$; $\chi^2(0.99, 4) = 13.28$.



time. However, the estimated value was significantly less than 0 in three ceiling threshold conditions, which incorrectly implies that the initial level is negatively correlated with the change over time. In other words, when an individual has a higher initial level, the individual tends to have a smaller slope. Certainly, this bias could make the interpretation of the results very misleading.

To further evaluate the artifactual nonlinearity consequences from the model selection perspective, two comparative models were fitted on the generated data to examine whether the true model can be correctly selected or not. The true model is a linear growth curve model and the alternative model is a quadratic growth curve model. Likelihood ratio tests were used to select a better fitted model from these two nested models. Table 4 displays the model selection results from the 500 replications with mean chi-square difference (chi-square value of the linear model minus chi-square value of the quadratic model), minimum and maximum chi-square difference, and the correct model selection proportion. For the data without ceilings, the likelihood ratio test can successfully select the true model with a proportion of 95.4%. However, when there are ceiling data in the longitudinal data, the correct selection proportion rapidly decreased to almost 0%. For example, for the data with ceiling threshold = 13, among the 500 replications, the minimum chi-square difference was 14.39. With the degree of freedom difference of 4, we always concluded that the quadratic model fitted the data relatively better than the linear model for all 500 replications.

## COMPARISON OF METHODS IN DEALING WITH LONGITUDINAL CEILING DATA

In this section, we investigate three possible methods to deal with longitudinal data including deleting ceiling cases, considering ceiling cases as missing data, and Tobit growth curve modeling.

### Listwise Deleting

Listwise deleting is a simple and easily implemented method. If a case has ceiling data, then the whole case is deleted. Intuitively, it can avoid the ceiling effects because the new data set no longer has any observations at ceiling. However, an obvious disadvantage of this method is the loss of data and shrinkage of sample size. Furthermore, in longitudinal studies, ceiling effects always happen more frequently for high-scoring participants than for low-scoring participants. By deleting ceiling cases, one will lose more high-scoring participants than low-scoring participants. To numerically evaluate the influences of the listwise deleting method, the generated ceiling data in the previous section were analyzed again by deleting the cases with ceiling data. The results are provided online at http://www.cdhrm.org/upload/Ceiling_Tables.pdf. Overall, the estimates were still biased. Compared with the results in Table 3, the estimates of shape parameters $A(t)$ and the covariance parameters $\phi_{01}$ were less biased. However, the estimates of the fixed initial level parameter $\beta_0$ were more biased due to deleting some high-scoring participants.

### Considering Ceiling Data as Missing Data

Another possible way to deal with ceiling data is to consider ceiling data as missing data. A datum can be considered missing when it reaches the ceiling threshold. It is widely known that there are three kinds of missing data patterns: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR; Fitzmaurice et al., 2004; Little & Rubin, 1987; Rubin, 1976). When we treat ceiling data as missing data, the missing data pattern is neither MCAR nor MAR and belongs to the MNAR pattern because the incompleteness at the current occasion is related to the ceiling data at the current occasion (only the observed data at the ceiling threshold are considered missing data). If we apply regular methods dealing with missing data such as the expectation-maximization (EM) algorithm or multiple imputations to analyze ceiling data as missing data, the parameter estimation may still have problems. To illustrate, we analyzed the generated ceiling data using the EM algorithm implemented in the Mplus program. The results are provided online at http://www.cdhrm.org/upload/Ceiling_Tables.pdf. Clearly, the parameter estimates were still biased. The results were similar to the results in Table 3, although the estimates of the covariance parameter $\phi_{01}$ improved a little bit.

### Tobit Growth Curve Model/Hierarchical Tobit Model

Here we propose to use Tobit growth curve model to analyze longitudinal ceiling data. Tobit regression (also called censored regression) was first proposed by

Tobin (1958) to analyze limited dependent variables in econometrics. Tobit regression model can be modified to analyze cross-sectional ceiling data. Tobit regression for the dependent variable with ceilings at one timepoint is given in Equation (2):

$$
\begin{aligned}
y_i^* &= \beta' X_i + e_i \\
y_i &= y_i^*, \text{ for } y_i^* < c \\
y_i &= c, \text{ for } y_i^* >= c \\
e_i &\sim N(0, \sigma^2)
\end{aligned}
\tag{2}
$$

In this model, the observed dependent variable $y_i (i = 1, \ldots, n)$ satisfies $y_i = \min(y_i^*, c)$, where the true dependent variable $y_i^*$ is a latent variable generated by $y_i^* = \beta' X_i + e_i$. The observed $y_i$ does not satisfy the model when $y_i^* >= c$ because the observed $y_i$ is constrained to be equal to the ceiling threshold. The residuals $e_i$ is assumed to be independent and have a normal distribution of $N(0, \sigma^2)$ conditional on the covariates $X_i$. The log-likelihood function for this Tobit model is

$$
\begin{aligned}
l(\beta, \sigma) &= \sum_{i=1}^{n} \ln[g(y_i | X_i, \beta, \sigma)] \\
&= \sum_{i=1}^{n} \ln[I(y_i < c) f((y_i - \beta' X_i)/\sigma)/\sigma \\
&\quad + I(y_i = c)(1 - F((c - \beta' X_i)/\sigma))] \\
&= \sum_{i=1}^{n} [I(y_i < c)(-1/2(y_i - \beta' X_i)^2/\sigma^2 - \ln(\sigma) - \ln(\sqrt{2\pi})) \\
&\quad + I(y_i = c) \ln(1 - F((c - \beta' X_i)/\sigma))]
\end{aligned}
$$

where I(.) is an indicator function with I(true) = 1 and I(false) = 0, f(.) is the standard normal density function, and F(.) is the cumulative normal distribution function. When the observed data $y_i$ are less than the ceiling threshold, the likelihood function of this data point comes from the normal density function $f((y_i - \beta' X_i)/\sigma)/\sigma$ based on the regression equation. When the observed data $y_i$ is equal to the ceiling threshold, the likelihood function comes from the survival function of $y_i$ (1-cumulative normal distribution function) because the information we have about this observed data point is that the true score is larger than or equal to ceiling threshold. If there are no ceilings in the data, then the log-likelihood function is reduced to

$$l(\beta, \sigma) = \sum_{i=1}^{n} \ln[g(y_i | X_i, \beta, \sigma)] = \sum_{i=1}^{n} \ln[f((y_i - \beta' X_i)/\sigma)/\sigma],$$

which is the log-likelihood function for the regular regression analysis. The log-likelihood values can be iteratively maximized and the parameter estimates can be obtained easily. This model can be used to deal with ceiling effects at one timepoint and this model has been extended to analyze censored data in factor analysis (Muthen, 1989).

The Tobit model can be extended to the hierarchical Tobit model to analyze longitudinal data or nested data (Cowles, Carlin, & Connett, 1996; Hajivassiliou, 1994; Kyriazidou, 1997). Expressed in terms of growth curve modeling and dealing with longitudinal ceiling data, the hierarchical Tobit model can be expressed by replacing the growth function in the Equation (1) with

$$\begin{aligned} y_{it}^* &= b_{0i} + A(t)b_{1i} + e_{it} \\ y_{it} &= y_{it}^*, \text{ for } y_{it}^* < c, \\ y_{it} &= c, \text{ for } y_{it}^* >= c \end{aligned} \tag{3}$$

where $y_{it}^*$ is the true score of $i$th participants at $t$th occasion, $y_{it}$ is the observed score, and c is the ceiling threshold. The path diagram of this model is portrayed in Figure 1. In Figure 1, the latent initial level and change parameters, the repeated true scores, and the measurement errors are represented by circles. The repeated observed scores are represented by squares. The constants including the ceiling threshold are represented by triangles (McArdle, 2005). The $r_{1it}$ and $r_{2it}$ are two time-varying random-effects parameters expressed by two indicator functions: $r_{1it} = I(y_{it}^* < C)$, $r_{2it} = I(y_{it}^* >= C)$, where $r_{1it} + r_{2it} = 1$ for a given $i$ and $t$. Here we name this model the *Tobit growth curve model*.

For the Tobit growth curve model, the conditional likelihood of the model can be specified as follows: If $y_{it}^* < c$ or $y_{it} < c$, then $y_{it} \sim N(b_{0i} + A(t)b_{1i}, \sigma_t^2)$; if $y_{it}^* \geq c$ or $y_{it} = c$, then $y_{it} \sim N(y_{it}^*, \sigma_t^2)$ & $y_{it}^* \geq c$. That is,

$$\begin{aligned} L_1 &= L(A(t), \sigma_t | b_{0i}, b_{1i}) = \prod_{i,t} g(y_{it} | A(t), \sigma_t, b_{0i}, b_{1i}) \\ &= \prod_{y_{it} < c} (f((y_{it} - b_{0i} - A(t)b_{1i})/\sigma_t)/\sigma_t) \prod_{y_{it} = c} (1 - F((c - b_{0i} - A(t)b_{1i})/\sigma_t)). \end{aligned}$$

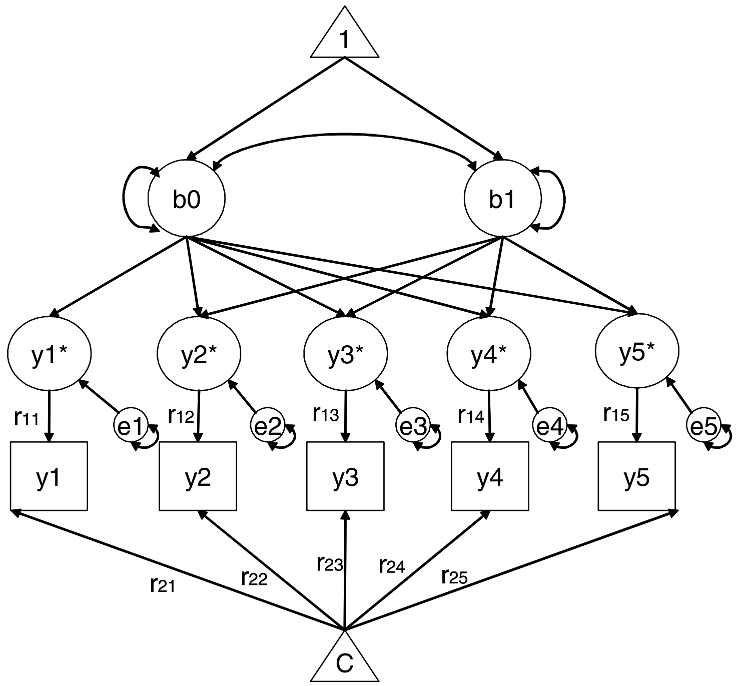



FIGURE 1 Path diagram of a Tobit growth curve model.



The random-effects parameters $(b_{0i}, b_{1i})$ are usually assumed to have a multivariate distribution $\text{MVN}\left(\binom{\beta_0}{\beta_1}, \Phi\right)$. Then the likelihood for the random-effects parameters can be written as $L_2 = \prod_{i=1}^{n} MVN(\boldsymbol{\beta}, \boldsymbol{\Phi})$. Therefore, the full likelihood function is $L = L_1 \times L_2$.

In estimating the Tobit growth curve model, both MLE and Bayesian method (Austin, 2002; Chib, 1992; Cowles et al., 1996) can be used in estimating the parameters. In terms of programs and software, some structural equation modeling software such as Mplus and LISREL can be used with censored outcome specifications and SAS NLMIXED can also be used to estimate the parameters with specified likelihood function. Because the focus of this study is not estimation methods but analytical models, we used only Bayesian estimation to estimate the parameters in this article. Interested readers can find Mplus codes and simulation results from MLE method at http://www.cdhrm.org/upload/ML_tobit_results.pdf

To use Bayesian method, we need to specify prior distributions for the parameters. The prior distributions were specified as follows:

$$A(t) \leftarrow N(0, 1.0 + 6E)$$

$$\beta_0 \leftarrow N(0, 1.0 + 6E)$$

$$\beta_1 \leftarrow N(0, 1.0 + 6E)$$

$$\sigma^2 \leftarrow IGamma(0.001, 0.001)$$

$$\Phi \leftarrow IWishart\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, 2\right).$$

All of the prior distributions are noninformative priors (Congdon, 2001, 2003). The Bayesian estimation procedure was conducted in a free program of WinBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2003) and the WinBUGS code is contained in Appendix A (interested readers can also find the WinBUGS code for a regular growth curve model in Appendix B).

The results from fitting the Bayesian Tobit growth curve model on the generated ceiling data are displayed in Table 5. The results show that all of the parameters in the growth curve models can be recovered well by the Tobit growth curve modeling approach regardless of different ceiling proportions in the data. With larger ceiling proportions, the standard errors of the estimates became a little bit larger than the true standard errors.

TABLE 5
Results of Fitting the Tobit Growth Curve Model to the Simulated Data With Ceilings

| *Parameters and True Values* | *No Ceiling* | *CT = 15* | *CT = 14* | *CT = 13* |
|---|---|---|---|---|
| $A(2) = 0.25$ | 0.25 (0.02) | 0.25 (0.02) | 0.25 (0.02) | 0.25 (0.02) |
| $A(3) = 0.50$ | 0.50 (0.02) | 0.50 (0.02) | 0.50 (0.02) | 0.50 (0.02) |
| $A(4) = 0.75$ | 0.75 (0.02) | 0.75 (0.02) | 0.75 (0.02) | 0.75 (0.02) |
| $\beta_0 = 7.50$ | 7.50 (0.16) | 7.50 (0.16) | 7.50 (0.16) | 7.50 (0.16) |
| $\beta_1 = 5.00$ | 5.01 (0.14) | 5.00 (0.14) | 5.00 (0.15) | 5.00 (0.16) |
| $\phi_{00} = 4.00$ | 4.00 (0.46) | 4.01 (0.47) | 4.01 (0.47) | 4.01 (0.47) |
| $\phi_{01} = 0.00$ | -0.01 (0.30) | 0.04 (0.33) | 0.04 (0.34) | 0.07 (0.36) |
| $\phi_{11} = 2.25$ | 2.23 (0.39) | 2.17 (0.43) | 2.17 (0.45) | 2.16 (0.47) |
| $\sigma^2 = 1$ | 0.99 (0.06) | 1.01 (0.06) | 1.01 (0.06) | 1.01 (0.06) |
| $\text{Corr}(b_{0i}, b_{1i}) = 0$ | 0.00 (0.10) | 0.02 (0.11) | 0.02 (0.12) | 0.03 (0.13) |

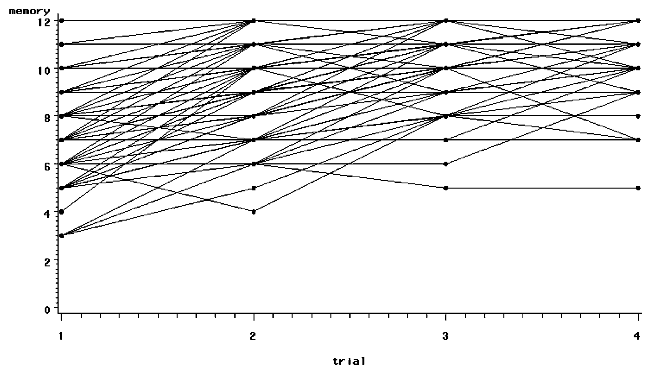
*Note.* The values outside the parentheses are the means of the estimates from the 500 replications and the values inside the parentheses are the standard deviations of the estimates from the 500 replications.

## AN EMPIRICAL STUDY

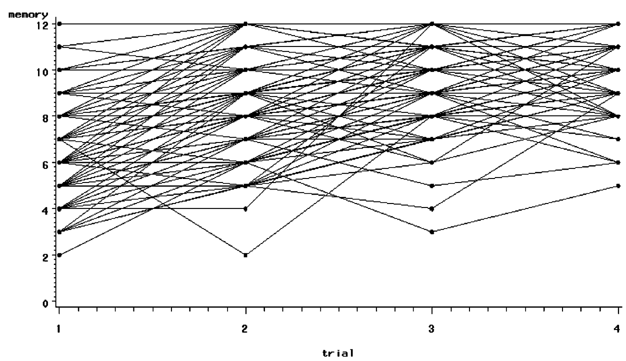
In this section, an empirical data set was used to demonstrate the influences of ceiling effects and the application of the Bayesian Tobit growth curve model in real longitudinal data analysis. The empirical data examined here were collected by Salthouse and his colleagues (Salthouse, 2004). Participants ($N = 608$) aged 19 to 97 were measured on the Wechsler Memory Scale III Word Lists subtest across three separate sessions. On each session a list of 12 unrelated words was presented to the participants followed immediately by an attempt to recall as many of the words as possible. This procedure was repeated four times with the same words in the same order. In this study, only the Session 1 data were used. Figure 2 displays the individuals' growth curves over the four trials across three groups in the first session. Because of the wide age range, we divided the total sample into three age groups (younger adult group: 19–39, $N = 135$; middle-age adult group: 40–59, $N = 236$; and older adult group: 60–97, $N = 237$) as in Zhang, Davis, Salthouse, and Tucker-Drob (2007) and compared the results among these three groups.

Table 6 displays the descriptive statistics (means and standard deviations) and different ceiling proportions across four trials and three age groups. Forty-four percent of the younger adults reached ceiling in the fourth trial, and only 16% of the older adults reached ceiling in the fourth trial. Therefore, different age groups have substantially different ceiling proportions. For group means, the younger adult group obtained higher scores than the middle-age adult group and the older adult group across all trials. For the group variation (standard deviation), the older adult group had more variation than the middle-age adult group and the younger adult group.

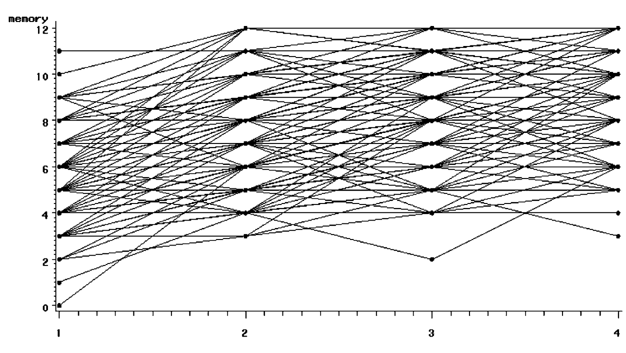
Both the regular growth curve model (Equation 1) and the Bayesian Tobit growth curve model (Equation 3) were fitted to the data for each age group. The main results of the analysis are given in Table 7. From the regular growth curve analysis (ignoring the ceiling effects), the younger age group showed the least average change ($\beta_1$) among three groups. The correlations between the latent initial levels and the latent change were significantly negative for the younger adult group and the middle-age adult group but statistically uncorrelated for the older adult group. When Tobit growth curve model was used, some of the results became very different. The younger age group showed the most average change among these three groups. The correlations between the latent initial levels and the latent change were not significantly different from 0 for all three groups, which means that the latent initial level and change variables were two statistically uncorrelated constructs. Intuitively, the results from the Tobit growth model are more reasonable, based on the similar comparison in the previous section.




(a) Younger age group




(b) Middle-age group




(c) Older age group

FIGURE 2 Individual growth curves of Wechsler Memory Scale III Word Lists subtest scores.

TABLE 6
Descriptive Statistics of the Memory Scores Across 4 Trials and 3 Age Groups

| *Age Group* | *Ceiling Proportion* | | | | *M(SD)* | | | |
|---|---|---|---|---|---|---|---|---|
| | *A1* | *A2* | *A3* | *A4* | *A1* | *A2* | *A3* | *A4* |
| 19–39 | 1% | 14% | 36% | 44% | 7.31 (1.81) | 9.59 (1.79) | 10.64 (1.47) | 10.89 (1.36) |
| 40–59 | 0% | 7% | 23% | 31% | 6.58 (1.79) | 8.96 (1.85) | 10.11 (1.65) | 10.44 (1.47) |
| 60–97 | 0% | 3% | 7% | 16% | 5.64 (1.80) | 7.81 (2.02) | 8.82 (2.02) | 9.46 (1.96) |

*Note.* The values inside the parentheses are the standard deviations of each cell. A1: first trial; A2: second trial; A3: third trial; A4: fourth trial.

TABLE 7
Results of the Tobit Growth Curve Model Fitted to the Empirical Grouped Data

| *Parameter* | *Younger 19 to 39 Estimate (SE)* | *Middle-Age 40 to 59 Estimate (SE)* | *Older 60 to 97 Estimate (SE)* |
|---|---|---|---|
| Regular growth curve analysis | | | |
| A(2) | 0.64 (0.03) | 0.61 (0.02) | 0.57 (0.02) |
| A(3) | 0.94 (0.02) | 0.91 (0.02) | 0.84 (0.02) |
| $\beta_0$ | 7.31 (0.16) | 6.59 (0.12) | 5.64 (0.12) |
| $\beta_1$ | 3.57 (0.14) | 3.87 (0.11) | 3.81 (0.11) |
| $\phi_{00}$ | 2.90 (0.56) | 3.08 (0.46) | 2.63 (0.38) |
| $\phi_{01}$ | −1.53 (0.52) | −1.68 (0.44) | −0.46 (0.35) |
| $\phi_{11}$ | 1.73 (0.55) | 1.87 (0.48) | 1.55 (0.42) |
| Tobit growth curve model | | | |
| A(2) | 0.57 (0.03) | 0.57 (0.02) | 0.55 (0.02) |
| A(3) | 0.90 (0.04) | 0.88 (0.03) | 0.82 (0.02) |
| $\beta_0$ | 7.32 (0.16) | 6.60 (0.12) | 5.64 (0.11) |
| $\beta_1$ | 4.30 (0.20) | 4.25 (0.13) | 3.96 (0.13) |
| $\phi_{00}$ | 2.61 (0.54) | 2.87 (0.48) | 2.42 (0.39) |
| $\phi_{01}$ | −0.10 (0.55) | −0.89 (−0.50) | 0.06 (0.40) |
| $\phi_{11}$ | 1.49 (0.69) | 1.70 (0.59) | 1.41 (0.54) |

## DISCUSSION

In this study, influences of ceiling effects in longitudinal data analysis were investigated through both Monte Carlo simulation studies and an empirical longitudinal study. From the simulation studies, we found that ceiling effects in longitudinal data resulted in incorrect model selection and biased parameter estimation and thus wrong interpretation of parameters and models. The magnitude of the biases was positively related to the proportion of the ceiling data in the data set. From the empirical data analysis, it was further found that when the ceiling effects were ignored, this could lead to some very misleading results, especially when comparing multiple groups that had different proportions of ceiling data. For example, when ceiling effects were ignored, the younger adult group showed the least change scores compared with the other two groups. This result could lead to the misleading conclusion that the younger adults have lower learning ability than older adults. Furthermore, when ceiling effects were ignored, negative correlations were found between latent initial levels and change parameters for younger adult group and middle-age adult group, which was also misleading for understanding the two statistically uncorrelated latent constructs (Jones et al., 2005).

From our simulations, even 18% of the participants reaching ceiling at one occasion (ceiling threshold = 15) could lead to some problems in longitudinal data analysis. Therefore, it is important to detect potential ceiling data before doing longitudinal data analysis. To detect ceiling data, a longitudinal plot of the data could help us to visually check if there is a substantial proportion of participants who obtained maximum scores. Frequency table of the maximum scores across occasions is quantitatively helpful. For growth curve modeling, researchers could also try to use both the regular growth curve method and the Tobit growth curve model to analyze the data. If the percentage of the participants reaching ceiling at one occasion is larger than 20% or there are some important discrepancies between the parameter estimates from two methods, researchers should be cautious about the influences of ceiling effects in the data.

Three possible methods were investigated to deal with longitudinal ceiling data in the growth curve modeling framework. The simulation results showed that Tobit growth curve models can be used to fit longitudinal ceiling data and recover the true parameters very well. This was because Tobit growth curve models made best use of ceiling data information, which refers to the information that true scores are larger than or equal to the ceiling threshold for the ceiling data. First, when ceiling effects were ignored, there were biases in the parameter estimates because the ceiling data were incorrectly treated as true ability scores in data analysis. Second, when ceiling data cases were deleted from the data analysis or considered missing data, the ceiling data information was not used at all. Finally, Tobit growth curve models effectively used all available information in analyzing the data. Although ceiling data may not reflect the true extent of ability, there is some important though limited information we can use to estimate the parameters.

Although the Tobit growth curve model performed well in both the simulation study and empirical study, we want to emphasize the important assumptions of this model. First, the distribution of the true scores at each occasion is assumed to have a specified distribution form such as a normal distribution in the model. If the distribution specified in the model is different from the distribution of the true scores from the empirical variable, the model could fail. For example, if the distribution in the model is normal and the distribution of the true scores from the empirical variable such as response time variable is log-normal or Weibull, then the estimates could be biased because the limited information from the ceiling data will be incorrectly used in the parameter estimation. In this case, we need to modify the normal Tobit growth curve model to a nonnormal Tobit growth curve model to analyze nonnormal response variables. For example, we can assume the response time variable has a Weibull distribution. Then in Equation (3), we only need to change the normal density function and the normal cumulative function to Weibull density function and Weibull cumulative function. Second, the ceiling threshold is assumed to be known in the model. This assumption is often easy to obtain when the ceiling effects happen in the measurement testing situations because researchers know the maximum scores of the tests and can set the maximum scores as ceiling thresholds.

Ceiling effects were the main focus of this study. However, floor effects can be dealt with in very similar ways. Floor effects occur when the tests are too difficult, such that some low-scoring participants cannot answer any item correctly and the scores stay on the floor, which is similar to the concept of left censoring. For example, floor effects were also observed in the Paced Visual Serial Addition Task after several repeated sessions (Feinstein, Brown, & Ron, 1994).The Tobit model and the Tobit growth curve model can also deal with floor effects in a similar way only with a little modification.

In summary, ignoring ceiling effects leads to biased parameter estimates, wrong interpretation of the relations among the random-effects parameters, and incorrect model selection. It is therefore important for researchers to pay attention to the ceiling effects in longitudinal data analysis and find ways to deal with them. When developing a new test or measurement scale, one should try to discover the true extent of participants' potentials to avoid ceiling effects. However, when analyzing existing data with ceiling effects, one needs to detect them and employ alternative methods to deal with ceiling effects. Our results suggest an alternative method to a regular growth curve model for analyzing longitudinal ceiling data. One of the strengths of the approach is the accuracy and flexibility. This permits accurate modeling of intraindividual difference and interindividual differences for longitudinal ceiling data.

## REFERENCES

Austin, P. C. (2002). Bayesian extensions of the Tobit model for analyzing measures of health status. *Medical Decision Making, 22,* 152–162.

Baltes, P. B., & Nesselroade, J. R. (1979). History and rationales for longitudinal research. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 1–39). New York: Academic.

Chib, S. (1992). Bayes inference in the Tobit censored regression model. *Journal of Econometrics, 51,* 79–99.

Congdon, P. (2001). *Bayesian statistical modelling*. New York: Wiley.

Congdon, P. (2003). *Applied Bayesian modelling*. New York: Wiley.

Cowles, M. K., Carlin, B. P., & Connett, J. E. (1996). Bayesian Tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness. *Journal of American Statistician Association, 91,* 86–98.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test: Adult version*. San Antonio, TX: Psychological Corp.

Feinstein, A., Brown, R., & Ron, M. (1994). Effects of practice of serial tests of attention in healthy subjects. *Journal of Clinical and Experimental Neuropsychology, 16,* 436–447.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004) *Applied longitudinal analysis.* New York: Wiley.

Genia, V. (2001). Evaluation of the Spiritual Well-Being Scale in a sample of college students. *The International Journal for the Psychology of Religion, 11*(1), 25–33.

Hajivassiliou, V. A. (1994). A simulation estimation analysis of the external debt crises of developing countries.*Journal of Applied Econometrics, 9,* 109–131.

Jones, R. N., Rosenberg, A. L., Morris, J. N., Allaire, J. C., McCoy, K. J. M., Marsiske, M., et al. (2005). A growth curve model of learning acquisition among cognitively normal older adults. *Experimental Aging Research, 31,* 291–312.

Klein, J. P., & Moeschberger, M. L. (2005). *Survival analysis: Techniques for censored and truncated data* (2nd ed.). New York: Springer.

Kyriazidou, E. (1997). Estimation of panel data sample selection model. *Econometrics, 65,* 1335–1364.

Laird, N., & Ware, J. (1982). A random-effects model for longitudinal data. *Biometrics, 38,* 963–974.

Ledbetter, M. F., Smith, L. A., Vosler-Hunter, W. L., & Fischer, J. D. (1991). An evaluation of the research and usefulness of the Spiritual Well-Being Scale. *Journal of Psychology and Theology, 19,* 49–55.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: Wiley.

McArdle, J. J. (2005). The development of the RAM rules for latent variable structural equation modeling. In A. Maydeau-Olivares & J. J. McArdle (Eds.), *Contemporary Psychometrics: A Festschrift for Roderick P. McDonald* (pp. 225–274). Mahwah, NJ: Erlbaum.

McArdle, J. J. (2004). Latent growth curve analysis using structural equation modeling techniques. In D. M. Teti (Ed.), *The handbook of research methods in developmental psychology* (pp. 340–466). New York: Blackwell.

McArdle, J. J., & Nesselroade, J. R. (2003). Growth curve analyses in contemporary psychological research. In J. Schinka & W. Velicer (Eds.), *Comprehensive handbook of psychology: Vol. 2. Research methods in psychology* (pp. 447–480). New York: Pergamon.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55,* 107–122.

Miller, G. A. (1956). The magical number of seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Murrell, R. C., Kenealy, P. M., Beaumont, J. G., & Lintern, T. C. (1999). Assessing Quality of Life (QoL) in persons with severe neurological disability associated with multiple sclerosis (MS): The psychometric evaluation of two QoL measures. *British Journal of Health Psychology, 4,* 349–362.

Muthen, B. (1989). TOBIT factor analysis. *British Journal of Mathematical and Statistical Psychology, 42,* 241–250.

Muthen, B. (1990). Moments of the censored and truncated normal distribution. *British Journal of Mathematical and Statistical Psychology, 42,* 241–250.

Olsen, M. K., and Schafer, J. L. (2001), A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association, 96,* 730–745.

Rey, A. (1964). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63,* 581–592.

Salthouse, T. A. (2004). Localizing age-related individual differences in a hierarchical structure. *Intelligence, 32,* 541–561.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2003). *WinBUGS manual (Version 1.4.1).* MRC Biostatistics Unit, Cambridge University, UK. http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometria, 26*(1), 24–36.

Uttl, B. (2005). Measurements of individual differences: Lessons from memory assessment in research and clinical practice. *Psychological Science, 16*(6), 460–467.

Van den Oord, E. J. C. G., & Van der Ark, L. A. (1997). A note on the use of Tobit approach for tests scores with floor or ceiling effects. *British Journal of Mathematical and Statistical Psychology, 50,* 351–364.

Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* New York: Springer.

Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology, 19,* 87–95.

Wechsler, D. (1987). *Wechsler Memory Scale-Revised*. San Antonio, TX: Psychological Corp.

Wechsler, D. (1997). *Wechsler Memory Scale-III.* San Antonio, TX: Psychological Corp.

Zhang, Z., Davis, H. P., Salthouse, T. A., & Tucker-Drob, E. A. (2007). Correlates of individual, and age-related, differences in short-term learning. *Learning and Individual Differences, 17*(3), 231–240.



# APPENDIX A

## WinBUGS Scripts for a Tobit Growth Curve Model

```
model;
{
      for( i in 1 : nsubj ) {
      for( j in 1 : ntime ) {
           lower.lim[i, j] <- cutoff  * temp[i,j]
                       + lowerLIM  * (1-temp[i,j])
         y[i , j] ~ dnorm(mu[i, j], tauy)I(lower.lim[i,j], )
}
}

for (i in 1:nsubj)
```

```
      {
       mu[i , 1] <- b[i,1]
       mu[i , 2] <- b[i,1]+A2*b[i,2]
       mu[i , 3] <- b[i,1]+A3*b[i,2]
       mu[i , 4] <- b[i,1]+A4*b[i,2]
       mu[i , 5] <- b[i,1]+b[i,2]
       b[i,1:2] ~ dmnorm(mub[1:2], taub[1:2,1:2])
        }

  A2~dnorm(0,1.0E-6)
  A3~dnorm(0,1.0E-6)
  A4~dnorm(0,1.0E-6)

   tauy ~ dgamma(0.001,0.001)

  mub[1]~dnorm(0,1.0E-6)
  mub[2]~dnorm(0,1.0E-6)

  taub[1:2, 1:2] ~ dwish(R[1:2, 1:2], 2)

  sigma2b[1:2, 1:2] <- inverse(taub[1:2, 1:2])
  sigma2y <- 1 / tauy
  pho<-sigma2b[1,2]/sqrt(sigma2b[1,1]*sigma2b[2,2])
}

list( nsubj=200, ntime=5, cutoff=13,lowerLIM=-100, R=structure
(.Data=c(1,0,0,1),.Dim=c(2,2)),
y = structure(.Data = c( 6.79, 7.71, 11.3, 9.54, 8.24, 6.74,
7.67, 10.53, 9.6, 12.33, 6.85, 7.38, 10.25, NA, 9.38, 8.77,
12.29, 12.49, NA, 11.45, 7.25, 10.84, 8.17, 8.33, 9.82, 10.25,
9.51, 8.73, NA, 9.45, 9.17, 9.16, 10.9,.....)),
temp=structure(.Data = c( 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0,
....)))
```

## APPENDIX B

### WinBUGS Scripts for a Regular Growth Curve Model

```
model;
{
      for( i in 1 : nsubj ) {
      for( j in 1 : ntime ) {
           y[i , j] ~ dnorm(mu[i, j], tauy)
}
}
```

```
for (i in 1:nsubj)
      {
       mu[i , 1] <- b[i,1]
       mu[i , 2] <- b[i,1]+A2*b[i,2]
       mu[i , 3] <- b[i,1]+A3*b[i,2]
       mu[i , 4] <- b[i,1]+A4*b[i,2]
       mu[i , 5] <- b[i,1]+b[i,2]
       b[i,1:2] ~ dmnorm(mub[1:2], taub[1:2,1:2])
        }

  A2~dnorm(0,1.0E-6)
  A3~dnorm(0,1.0E-6)
  A4~dnorm(0,1.0E-6)

   tauy ~ dgamma(0.001,0.001)

  mub[1]~dnorm(0,1.0E-6)
  mub[2]~dnorm(0,1.0E-6)

  taub[1:2, 1:2] ~ dwish(R[1:2, 1:2], 2)

  sigma2b[1:2, 1:2] <- inverse(taub[1:2, 1:2])
  sigma2y <- 1 / tauy
  pho<-sigma2b[1,2]/sqrt(sigma2b[1,1]*sigma2b[2,2])
}

list( nsubj=200, ntime=5, R=structure(.Data=c(1,0,0,1),
.Dim=c(2,2)),
 y = structure(.Data = c( 6.79, 7.71, 11.3, 9.54, 8.24,
6.74, 7.67, 10.53, 9.6, 12.33, 6.85, 7.38, 10.25, 13,
9.38, 8.77, 12.29, 12.49, 13, 11.45, 7.25, 10.84, 8.17,
8.33, 9.82, 10.25, 9.51, 8.73, 13, 9.45, 9.17, 9.16, 10.9,
.....)))
```