STUDYING PHYSICAL ACTIVITY DECLINE FROM ADOLESCENCE TO ADULTHOOD USING LATENT GROWTH CURVE AND RANDOM COEFFICIENT MODELS

by

Binqi Yang

BS, Remin University of China, 2003

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2007

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Binqi Yang

It was defended on

December 6th, 2006

and approved by

Thesis Advisor:
Sati Mazumdar, Ph.D.
Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Vincent C. Arena, Ph.D.
Associate Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Deborah Aaron, Ph.D.
Associate Professor
Department of Health and Physical Activity
School of Education
University of Pittsburgh

STUDYING PHYSICAL ACTIVITY DECLINE FROM ADOLESCENCE TO ADULTHOOD USING LATENT GROWTH CURVE AND RANDOM COEFFICIENT MODELS

Binqi Yang, M.S.

University of Pittsburgh, 2007

The level of physical activity is important for maintenance of good health. Research has demonstrated that virtually all individuals can benefit from physical activities which have been shown to reduce the morbidity from many chronic diseases, like cardiovascular disease and diabetes. Therefore, understanding the trend in activity level from adolescence to young adulthood is very important for public health study.

The purpose of this thesis is to describe the natural history of participation in leisure time physical activity from adolescence to young adulthood. The study data are from the University of Pittsburgh Physical Activity Study (PittPAS), which recorded physical activities of 1245 high school students over a period of 14 years. Two longitudinal growth models, a latent growth curve (LGC) model and a random coefficient model are applied to characterize the changes in activity hours per week (HRWK) as well as the effects of sex, race, and grade on these changes. Our analysis results show: Male students are more physically active and have the larger decline rate than Female students; White students are more active, and also have the larger decline rate than Black students; Students from the lower grades spend more time in physical activity and also have the larger decline rate than students in the higher grades.

Through analyzing the above physical activity, we also investigate the similarities and differences of LGC models and random coefficient models, such as both models share the same

objectives. The LGC model is a multivariate approach, while random coefficient model is a univariate one in terms of the dependent variables. Random coefficient model does not require time-structure data and allows the explanatory variable ‘time’ to take on different values for each subject. Thus, the random coefficient model has the advantage to handle large amount of missing and irregular data acquired in non-uniform time occasions. Since our study data have a large amount of missing observations and are non-uniformly acquired, random coefficient model is more appropriate in characterizing the changes of HRWK.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to express my sincere gratitude to my advisor, Professor Sati Mazumdar. This thesis would never have been possible without her constant guidance, suggestions and encouragement. I am deeply indebted to Professor Vincent C. Arena and Professor Deborah Aaron for their financial support during my study in University of Pittsburgh, and their helpful suggestions on my thesis. I would also like to thank Dr. Shui He, Dr. Fengshou Ko, and other students in our department, Jia Li, Tao Song for their help and friendship. Finally, and above all, my heartfelt gratitude goes to my parents and my fiancé, Lijie Liu, for their tremendous love, support and sacrifice over the years.

1.0 INTRODUCTION

Longitudinal data are common in the social, education and biology sciences. A wide array of statistical models is available for the analysis of longitudinal data. In recent years, methods that study a growth curve of longitudinal data have become popular. Such growth curve models provide a way to account efficiently for the dependency caused by the fact that the same subjects have been assessed repeatedly. The typical growth curve models include random coefficient model (Goldstein H, 1998) and latent growth curve model (McArdle JJ et al, 1987; Meredith W et al, 1990; Duncan TE et al, 2006). These two models are elegant in representing both collective and individual change as a function of time. They are highly similar because both approaches share the same objectives and have a similar model representation. However, they have different model assumptions, leading to the different features in processing the longitudinal data. In this thesis, we studied the similarities and differences between these two models by applying them to a set of longitudinal psychosomatic data.

1.1 RESEARCH STATEMENT

In psychosomatic research, there has been a long-standing interest in understanding how physical activity declines during adolescence (Mechelen WV et al, 2000; Kimm SYK et al, 2002). General findings in the epidemiological research of physical activity in young people are that boys are more active than girls, the amount of physical activity declines with increasing age, the decline rate is greater in girls than in boys (Mechelen WV et al, 2000), and declines are greater in black girls than in white girls (Kimm SYK et al, 2002). Studying data from the Amsterdam Longitudinal Growth and Healthy, Mechelen and his colleagues tried to describe the natural development of habitual physical activity (HPA) of young Dutch male and female individuals between the ages of 13 and 27 (Mechelen WV et al, 2000). They used ANOVA method to separately analyze the repeated measurements of total HPA, sports activity, leisure time activity and all other activity, and found that the male individuals had a significant decrease in weekly time spent on HPA. Moreover, they found that the activity changes of males and females were different regarding the three levels of intensity, i.e., moderate activity, vigorous activity, and very vigorous activity. Kimm and her colleagues did cross-sectional studies for 1213 black girls and 1166 white girls who had been followed annually from the ages of 9 or 10 years to 18 or 19 years (Kimm SYK et al, 2002). They used a validated questionnaire to measure leisure-time physical activity on the basis of metabolic equivalents (MET) of the reported activity and their MET-times per week. They used two-sample tests to examine racial differences in descriptive characteristics. Their findings showed that race was a factor, with black girls having a decline in activity twice that of white girls. Behavioral risk factors such as smoking and pregnancy also affected the decline in activity. An interesting finding was that the level of parental education instead of household income was associated with the decline of activity.

Although people have studied the physical activity during adolescence, none have captured the activity change over time. To fully understand the development of physical activity patterns throughout the life span, longitudinal change in physical activity needs to be examined. This thesis will evaluate the role that different factors, (i.e., sex, race, socioeconomic status, and grade), play in the decline of physical activity. We use latent growth curve (LGC) models and random coefficient models to analyze how the trend of physical activity changes over time or at each time point; how much it changes, and to compare the change among different groups. Through this study, we could not only provide answers to some questions such as, “Do boys do more exercise than girls as they grow up?”, “Do adolescents from wealthy families better maintain their exercise habit than those from poor families?”, “Is race relevant?”, but also we could address how the time spent in activity changes over time, and whether such changes depend on race, sex, or socio-economic status.

1.2 THESIS CONTENT AND ORGANIZATION

This thesis mainly investigates the similarities and differences between two longitudinal models: latent growth curve model and random coefficient model. These two models are both used to address the trend in physical activity from adolescence to adulthood. The study data are from University of Pittsburgh Physical Activity Study (PittPAS), which recorded physical activities of 1245 high school students at different time points over a period of 14 years. The thesis is organized as follows: Chapter 1 briefly introduces the background of this thesis work; Chapter 2 introduces basic concepts of a LGC model and a random coefficient model, and discusses similarities and differences between these two models; Chapter 3 gives the descriptive statistics of our study PittPAS data; Chapter 4 shows the similarities of a LGC model and a random coefficient model through studying the uniformly acquired data in the first four years; Chapter 5 compares the different features of these two models by analyzing the data observed in the whole time period; Finally, the characteristics of the trend of physical activity from adolescence to adulthood are presented and conclusions are drawn in the last Chapter.

2.0 LATENT GROWTH CURVE MODEL AND ROMDOM COEFFICIENT MODEL

Under the names of growth models (McArdle JJ et al, 1987), hierarchical linear models (Li F et al, 1999), random regression (Laird NM et al, 1982), or mixed models (Goldstein H, 1995), the popularity of latent growth curve (LGC) models and random coefficient models is the result of theoretical developments and the availability of software such as SAS, Mplus, LISREL, Amos, to conduct data analyses (Llabre MM et al, 2004). The most important reason for the popularity of the longitudinal growth models is their elegance in representing both collective and individual change as a function of time. In this Chapter, we briefly introduce the LGC model, and discuss its similarities and differences with random coefficient model.

2.1 LGC MODEL AND ITS TWO-FACTOR MODEL SPECIFICATION

As one type of growth curve models, the LGC model aims to capitalize on repeated observations of subjects across time (Goldstein H, 1998). It has emerged as a versatile tool for analyzing data in repeated measures designs, especially for studying longitudinal changes (McArdle JJ et al, 1987; Meredith W et al, 1990; Duncan TE et al, 2006). Considering data change over time as an underlying latent process, the LGC model establishes a trajectory of change over time for each individual in a sample. Characteristics of each trajectory varying across individuals are then treated as latent variables. The change pattern over time can be assumed to follow a linear or nonlinear function. The LGC model provides large flexibility and potentials to allow the complex representations of growth and correlates of change as a function of time. Particularly, the flexibility not only relies on the combination of individual and group levels of analysis, but also concerns the integration of factorial structure of the repeatedly measured variables, the shape of the growth curves, residual structures, missing data on predictor variables (Duncan TE et al, 2006).

A simple LGC model consists of two-level equations, where parameters in the first level are dependent variables of the second level equation. As depicted in Figure 1, a linear straight line growth curve model can be expressed as

$$\begin{aligned} y_{it} &= \text{int}_i + \lambda_{it}\text{slp}_\text{i} + \varepsilon_{it} \\ \text{int}_i &= \mu_{\text{int}} + \alpha_1 x_{1i} + \alpha_2 x_{2i} + e_{\text{int},i} \ , \\ \text{slp}_i &= \mu_{slp} + \alpha_3 x_{1i} + \alpha_4 x_{2i} + e_{\text{slp,i}} \end{aligned} \tag{2.1}$$

where y_{it} represents the measure of each individual i, $i = 1,2,\ldots,N$, on four occasion times t, $t = 0,1,2,3$. ε_{it} is a normally distributed residual at the measurement level, and $\varepsilon_{it} \sim N(0,\sigma_\varepsilon^2)$. The initial level and linear shape for each individual subject are presented by the latent factors, int_i and slp_i, which have initial values μ_{int} and μ_{slp}, and random deviations $e_{\text{int},i}$ and $e_{\text{slp,i}}$, respectively. Time factor is introduced by constraining parameters λ_{ti} to known values [0, 1, 2, 3]. α_1, α_2, α_3 and α_4, are the effects of the time-invariant covariate x_{1i} and x_{2i}. The random residuals $e_{\text{int},i}$ and $e_{\text{slp,i}}$ are assumed to be Gaussian distributed as $\begin{pmatrix} e_{\text{int},i} \\ e_{\text{slp,i}} \end{pmatrix} \sim N(0,\sum_e)$, where $\sum_e = \begin{pmatrix} \sigma_{\text{int}}^2 & \sigma^2{}_{slp,\,\text{int}} \\ \sigma^2{}_{slp,\,\text{int}} & \sigma_{slp}^2 \end{pmatrix}$. Actually, (2.1) is known as conditional latent growth curve model (Li F et al, 1999). The individual slope and intercept values deviate from the group mean values, resulting in the random effects. Thus, the LGC model given in (2.1) combines the fixed and random effects, and is very similar to the random coefficient model.

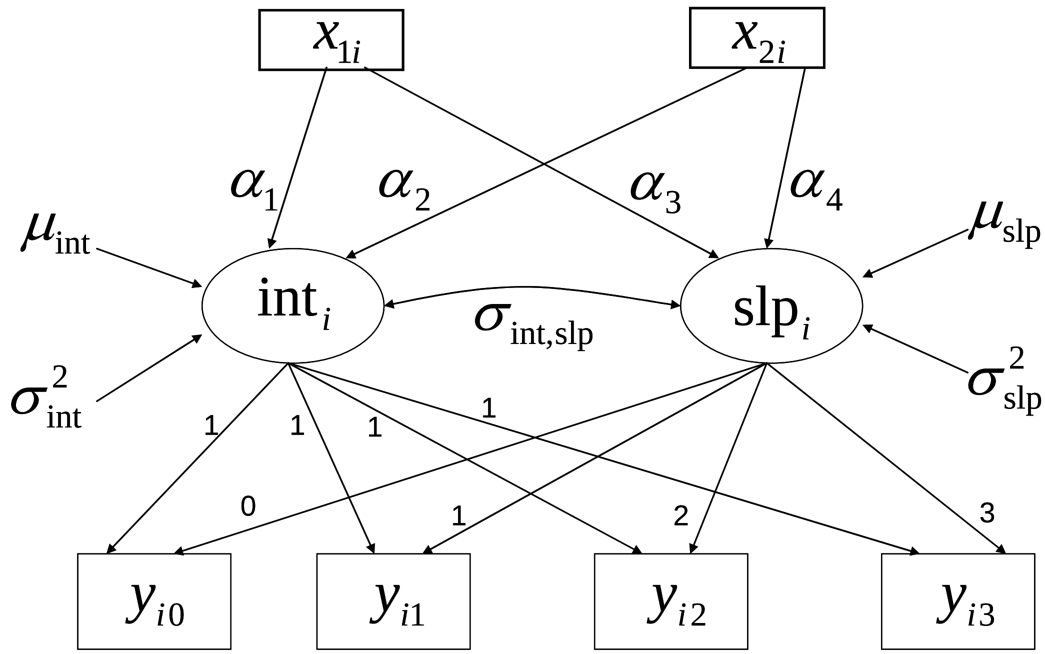


Figure 1. Conceptual latent growth curve diagram for each individual *i*

2.2 RANDOM COEFFICIENT MODEL

Random coefficient model is another popular growth curve model, which can also be applied to study longitudinal data changes. A random coefficient model shares the same objectives as that of a LGC model and have a similar model representation. In order to see this point clearly, we first recall the model specification of a simple random coefficient model. Specifically, a random coefficient model with linear growth pattern can be denoted as

$$Y_{it} = u_{\text{int}} + e_{\text{int},i} + u_{slp}Time_{it} + e_{slp,i}Time_{it} + a_1X_{1i} + a_2X_{2i} + a_3X_{1i}Time_{it} + a_4X_{2i}Time_{it} + \varepsilon_{it}, \quad (2.2)$$

where each individual $i = 1,2,\ldots,N$ is examined repeatedly at four time occasions $t = 0,1,2,3$. By simple manipulation, (2.2) can be written as

$$Y_{it} = (\ u_{\text{int}} + a_1 X_{1i} + a_2 X_{2i} + e_{\text{int},i}) + (\ u_{slp} + a_3 X_{1i} + a_4 X_{2i} + e_{slp,i})\ Time_{it} + \varepsilon_{it}\ . \tag{2.3}$$

Writing $\text{int}_i = u_{\text{int}} + a_1 X_{1i} + a_2 X_{2i} + e_{\text{int},i}$, $\text{slp}_i = u_{slp} + a_3 X_{1i} + a_4 X_{2i} + e_{sl,i}$, (2.2) can be expressed as

$$\begin{aligned} Y_{it} &= \text{int}_i + Time_{it}\text{slp}_i + \varepsilon_{it} \\ \text{int}_i &= u_{\text{int}} + a_1 X_{1i} + a_2 X_{2i} + e_{\text{int},i}\ . \\ \text{slp}_i &= u_{slp} + a_3 X_{1i} + a_4 X_{2i} + e_{slp,i} \end{aligned} \tag{2.4}$$

Here, ε_{it} is a residual at the measurement level and $\varepsilon_{it} \sim N(0,\sigma_{\varepsilon}^2)$, $Time_{it}$ is a variable, which denotes one of the measurement occasions [0, 1, 2, 3]. The initial level (intercept) and linear shape (slope) for each individual are expressed by int_i and slp_i. The effects of time-invariant covariate X_{1i} and X_{2i} on the intercept and slope are represented by coefficients a_1, a_2, a_3, a_4. And inter-individual differences between the intercept and slope are characterized by random effects $e_{\text{int},i}$ and $e_{slp,i}$, which are assumed to be normally distributed as $\begin{pmatrix} e_{\text{int},i} \\ e_{\text{slp,i}} \end{pmatrix} \sim N(0, \sum_e)$, and $\sum_e = \begin{pmatrix} \sigma_{\text{int}}^2 & \sigma_{slp,\ \text{int}}^2 \\ \sigma_{slp,\ \text{int}}^2 & \sigma_{slp}^2 \end{pmatrix}$. Therefore, $\begin{pmatrix} \text{int}_i \\ \text{slp}_\text{i} \end{pmatrix} \sim N(\boldsymbol{\mu}, \sum_e)$ with $\mathbf{u} = \begin{bmatrix} \mu_{\text{int}} + a_1 X_{1i} + a_2 X_{2i} \\ \mu_{\text{slp}} + a_3 X_{1i} + a_4 X_{2i} \end{bmatrix}$.

2.3 COMPARISIONS OF LGC MODEL AND RANDOM COEFFICIENT MODEL

Comparing (2.4) with (2.1), there are no important differences in the specification of the LGC model and random coefficient model for the linear growth curve. The major difference between these two models is the way how time is introduced. In the random coefficient model (2.4), time is introduced as a fixed explanatory variable $Time_{it}$, whereas in the LGC model (2.1), time is introduced via parameter λ_{it}, which is constrained in such a way that they represent time. The consequence of such difference is that the random coefficient model is essentially a univariate approach, with time points treated as observations of the same variables, whereas the LGC model is essentially a multivariate approach, with each time point treated as a separate variable. This distinction leads to the different properties of these two methods in the following three aspects: time estimation, residual specification and missing data handling.

2.3.1 Estimating the Growth Curve Pattern

(2.1) and (2.4) show that the LGC model and random coefficient model share the same features for the linear growth curve. In practice, the linear pattern of the growth curve is usually too restricted to fit the data. A higher-order polynomial such as quadratic or cubic pattern of the growth curve can be used to model the nonlinear change or development in the longitudinal data. It shows that the LGC model and random coefficient model also share the same features regarding the estimation of nonlinear growth curves (MacCallum RC et al, 1997).

The difference between the LGC model and random coefficient model lies on the way to process the time factor. As given in (2.1) and (2.4), the time factor is introduced differently, by setting parameter λ_{it} in the LGC model, or including time as an independent variable $Time_{it}$ in random coefficient model. LCG subjects to a restriction that even time space for all individuals between different interview occasions should be maintained. However, time factor in random coefficient model treated as an observation variable could include irregular time interval. Generally speaking, the fixed values of λ_{it}, and the values of time variable $Time_{it}$ represent the occasions at which individuals have been measured. As a univariate approach, time is given as an observation variable in the LGC model. However, it is still possible to estimate time factor parameters in the LGC models (McArdle JJ et al, 1987). In other words, instead of fixing the factor parameters λ_{ti} to known values [0, 1,…, T], we can set them to [0,1, b_2 …, b_T] , where b_2 …, b_T are to be estimated for a specific growth curve pattern. Such freedom to estimate time factor parameters instead of fixing them is quite helpful especially in the case when the measurement time is unknown.

2.3.2 Specification of Residual Correlation Structures

LGC model and random coefficient model are different in the flexibility of incorporating the alternative structures for the residuals. As a multivariate method, the LGC model allows very flexible specifications of residual covariance structures. Since y_{it} measured at time occasion [0, 1, 2, 3] is treated as a separate variable, it is easy to estimate its variances, and covariance between each dependent variable in the LGC model. Therefore, the effectiveness of a variety of reasonable error structures can be compared, and the most appropriate one can be used for the particular problem (Willett JB et al, 2004).

For random coefficient model, the available residual structures are designed to handle repeated measurer data, and usually provided by the software packages, such as SAS Proc Mixed.

2.3.3 Missing data

Comparing with LGC model, random coefficient model has more advantages to handle missing data. In LCG model, if one individual has the missing data in a subject at one occasion, all data relative to this individual will be considered as missing. That’s, observations with missing values for any variables in the analysis are omitted from the computations in the LGC model (SAS, 2006). However, if there is a missing observation in one or more occasions in random coefficient model, the data related to those occasions are omitted and other time occasions associated with this individual are still be processed. Therefore, random coefficient model is more robust in modeling the data including a large amount of missing observations.

2.4 IMPLEMENTATION OF LGC AND RANDOM COEFFICIENT MODEL

2.4.1 Random Coefficient Model Fitting

In this thesis, random coefficient model is fitted by PROC MIXED procedure in SAS version 9.0. The estimation of random coefficient model can be done either by the Maximum Likelihood (ML) method or the restricted/residual maximum likelihood (REML). The ML method is an iterative estimation approach, which produces estimates for the population parameters that maximizing the probability of observing the data given one model (Diggle P et al, 2002). By maximizing the likelihood function, ML estimation is expected to provide asymptotically efficient and consistent estimations given the relatively large sample size. The REML is another method can be used to fit the model structure. Both ML and REML iteratively optimize the parameters estimates for the model effects. Differently, REML only maximizes the likelihood of the data for the random effects. Therefore, REML leads to a restricted solution (SAS, 2006). In Proc Mixed procedure, REML is the default estimation method. Akaike's Information Criterion (AIC) is calculated to assess the goodness of model fitness. The model with the smallest AIC among all competing models is considered as the best model.

2.4.2 LGC Model Fitting

The LGC modelling approach is one of structural equation modelling (SEM) approaches, which expresses relationships among several directly observed variables (manifest variables) or unobserved hypothetical variables (latent variables). Hence, fitting growth trajectory models can be carried out using standard computer software for SEM, such as Amos, EQS, LISREL, Mplus, and SAS PROC CALIS (Li F et al, 1999). There have been a lot of literatures about comparing these software performances (Li F et al, 1999; Bollen KA et al, 2006). In this thesis, PROC CALIS (Covariance Analysis of Linear Structural Equations) routine in SAS 9.0 is used to estimate the LGC model parameters. It uses the default method maximum likelihood (ML) and generalizes the least-squares to estimate model parameters (SAS, 2006). Moreover, CALIS procedure assumes that random variables have an approximately multivariate normal distribution. CALIS routines in SAS provide several statistical tests to determine the adequacy of model fit to the data. Commonly accepted indices of fit are Chi-square test, the goodness fit index (GFI), and the root mean square error of approximation (RMSEA) (Duncan TE et al, 1999; Suhr DD, 2004). These three indices are used to assess the goodness of model fit in this Chapter.

The Chi-square test indicates the amount of difference between expected and observed covariance matrices. A chi-square value close to zero indicates little difference between the expected and observed covariance matrices. When one model with the constraints is nested with the model without the constraints, the Chi-square difference test can be used to compare these two models. If the constraints lead to a significant deterioration of the model fit, we can conclude such constraints are not valid.

The GFI represents the discrepancy function adjusted for sample size. The GFI should be between 0 and 1, with a larger value indicating better model fit. The acceptable model fit is indicated by a GFI value of 0.90 or greater (Hu LT et al, 1999). More information about the GFI can be found in the SAS online help (SAS, 2006).

The RMSEA is related to residuals. It represents the discrepancy per degree of freedom for the model. The RMSEA values range from 0 to 1 with a smaller RMSEA value indicating better model fit. Better model fit is indicated by an RMSEA value of 0.06 or less.

Beside the above three indices, we also use Akaike's Information Criterion (AIC) to justify our LGC model fitting. This is the same as random coefficient model. The model yields the smallest value of AIC is considered the best.

3.0 DESCRIPTION OF THE MATERIAL

3.1 STUDY DATA

A longitudinal study, University of Pittsburgh Physical Activity Study (PittPAS), was conducted to examine changes in physical activity of adolescents recruited from a single suburban school district. The study began in Year 1990 and included two phases: phase 1 conducted from Year 1990 ~ Year 1993 and phase 2 conducted from Year 2000 ~ Year 2003. A total of 1245 individuals were recruited into the study at Year 1990. 1171 of these individuals had been followed until 1993 in phase 1 study. To better describe the study data, we denote the interviews happened during Year 1990 ~ Year 1993 as the first, second, third and fourth interviews for each individual. Information had been recorded about their time spent in some typical sport activity for each of four years. The activities included aerobics, band/drill team, baseball, basketball, bicycling, bowling, cheerleading, dance class, football, garden/yard work, and tennis. Moreover, the recorded data also included some summaries of physical activity information, such as the number of activity taken per week, the total time spent per week, the time spent per week in team activities and individual activities, and the activity intensity, which was characterized into light, moderate and vigorous classes. Phase 1 of the study ended in 1993. Phase 2 began in 1999, and the individuals completed their fifth interviews during the period between 24-MAY-1999 and 23-FEB-2003 (denoted as R1), of which the median interview date is 25-January-2001, and their sixth visits between 18-JUN-2001 and 27-SEP-2004 (denoted as R2), of which the median interview date is 17-March-2003. Due to the difficulty in tracking the individuals, these two interview periods overlapped, resulting in non-regular observation time.

During the study periods, students' information had been obtained by a trained interviewer. Several questionnaire forms were used as the standard assessment procedures. Table 1 and Table 2 present the notations of continuous and categorical data, which are used through the thesis. Table 1 lists the notations of the continuous data such as age and total time spent in physical activity (HRWK). Moreover, we denote HRWK90, HRWK91, HRWK92 and HRWK93 as HRWK acquired in the first, second, third, and fourth interviews, respectively. HRWKR1 and HRWKR2 are defined as HRWK values acquired in the fifth and sixth interview, respectively. Table 2 presents the notations of categorical variables of each individual. We define HARDEX90-HARDEX93, EASYEX90-EASYEX93, and TV90-TV93 as the values of HARDEX, EASYEX, and TV from Year 1990 ~ Year 1993.

Due to long time intervals between two study phases, many participants could not be tracked and followed, thus resulting in many missing data. As presented in Table 3, the maximum drop-out rate reaches 20%. The study shows that the White participants are more likely to stay in the study compared to the Black participants; individuals from the "High SES" families are more likely to stay; individuals from the "Middle SES" families have a nearly constant rate to stay, while students from the "Low SES" families have the highest dropping rates.

Table 1. Continuous data

Descriptions	Variable Name
Age when attend the study, calculated using birthday	AGE
Number of activity students reported	NUMACT
Total time spent per week in physical activity	HRWK
Time spent per week in team physical activity	THRWK
Time spent per week in individual activity	IHRWK
Time spent per week in vigorous activity	VHRWK
Time spent per week in light moderate activity	LMHRWK

Table 2. Categorical data

Descriptions	Variable
Grade when students enter this study	GRADE
Sex	SEX
Race	RACE
Social and economics status	SES
Days of hard exercise in past two weeks	HARDEX
Days of easy exercise in past two weeks	EASYEX
Hours of TV/Computer/Video per day	TV

Table 3. Demographic status (%) of the whole population through years

		Adolescence				Young Adulthood	
Visit time		1st	2nd	3rd	4th	5th	6th
Year		1990	1991	1992	1993	R1	R2
Students participated (N)		1171	1088	957	809	828	678
SEX	Male	51.6	52.6	52.5	52.4	47.6	45.8
	Female	48.4	47.4	47.5	47.6	52.4	54.3
RACE	White	75.8	76.3	78.4	81.8	82.8	85.9
	Black	24.2	23.8	21.6	18.2	17.2	14.1
SES	High SES	27.5	28.0	29.9	32.3	32.9	36.4
	Mid SES	53.5	53.3	52.3	52.5	52.7	50.0
	Low SES	19.0	18.7	17.9	15.2	14.5	13.6
GRADE	Grade 7	33.1	33.0	33.9	33.6	31.8	30.8
	Grade 8	31.3	33.2	32.4	31.2	31.6	31.4
	Grade 9	35.5	33.8	33.8	35.2	36.6	37.8

3.2 DESCRIPTIVES OF CONTINUOUS VARIABLES

Table 4 lists the detailed characteristics of each activity. It shows that at baseline students usually attended four activities (NUMACT), and spent more than eighteen hours per week in all activities (HRWK), including ten hours in individual activity (IHRWK), eight hours in team activity (THRWK); twelve hours in vigorous activity (VHRWK), and six hours in light-moderate activity (LMHRWK). Over the course of the study, there is an obvious decline in the NUMACT and HRWK. At their sixth interview, individuals only had three activities averagely, and HRWK has declined to 5.4 hours, including only 4.6 hour of IHRWK, and 0.8 hours of THRWK, 1.8 hours of VHRWK, and 3.6 hours of LMHRWK.

Table 4. Means and standard deviations (in parenthesis) of the continuous variables

	Adolescence				Young Adulthood	
Visit time	1st	2nd	3rd	4th	5th	6th
Year	1990	1991	1992	1993	R1	R2
Students participated (N)	1171	1088	957	809	828	682
NUMACT	4.54 (2.42)	4.20 (2.51)	3.22 (2.26)	2.71 (2.07)	3.39 (2.18)	3.09 (1.96)
HRWK	18.36 (21.40)	15.97 (15.31)	12.61 (13.13)	10.42 (11.24)	6.87 (8.75)	5.52 (6.06)
IHRWK	10.32 (15.14)	7.92 (10.66)	5.19 (6.97)	4.04 (6.06)	5.81 (7.81)	4.65 (5.15)
THRWK	8.04 (11.39)	8.05 (9.72)	7.42 (10.03)	6.39 (8.44)	1.06 (2.24)	0.87 (2.22)
VHRWK	12.32 (16.83)	10.23 (11.07)	7.95 (10.00)	6.51 (8.72)	2.43 (3.97)	1.88 (3.06)
LMHRWK	6.05 (9.39)	5.74 (7.89)	4.66 (6.25)	3.92 (5.39)	4.44 (5.94)	3.64 (4.39)

3.3 DESCRIPTIVES OF CATEGORICAL VARIABLES

Table 5 presents the percentage of time spent in each activity within two weeks. During the transition from adolescence to adulthood, we can see that more participants spent little time in HARDEX and EASYEX.

Table 5. Status (%) of categorical variables

		Adolescence				Adulthood	
Visit time		1st	2nd	3rd	4th	5th	6th
Year		1990	1991	1992	1993	R1	R2
HARDEX	None	9.40	8.64	10.95	11.37	27.93	29.91
	1 to 2 days	14.46	14.15	18.14	20.07	15.96	13.78
	3 to 5 days	26.99	26.84	27.53	29.87	19.71	21.85
	6 to 8 days	20.08	21.23	16.68	15.36	15.96	17.60
	9 or more days	25.14	29.14	26.69	23.34	20.44	16.86
EASYEX	None	11.14	11.2	13.67	**	14.99	16.28
	1 to 2 days	18.59	19.49	20.46	**	13.06	13.93
	3 to 5 days	25.63	24.82	26.30	**	20.56	24.49
	6 to 8 days	14.99	14.89	13.67	**	13.30	12.46
	9 or more days	29.65	29.60	25.89	**	38.09	32.84
TV	None	1.83	1.56	2.71	3.02	3.38	**
	1 hour or less	14.49	17.37	23.28	23.46	25.12	**
	2 to 3 hours	42.63	47.89	46.97	49.33	48.19	**
	4 to 5 hours	19.65	19.85	17.85	15.24	14.25	**
	6 or more hours	21.40	13.33	9.19	8.95	9.06	**

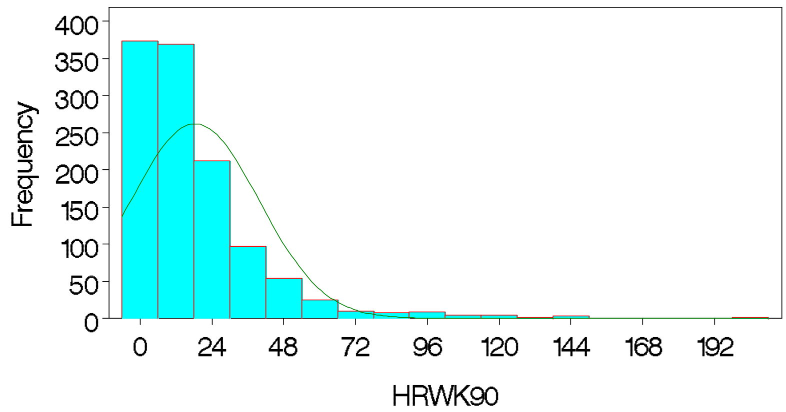
** indicates that the corresponding data are not available.

3.4 DATA TRANSFORMATION

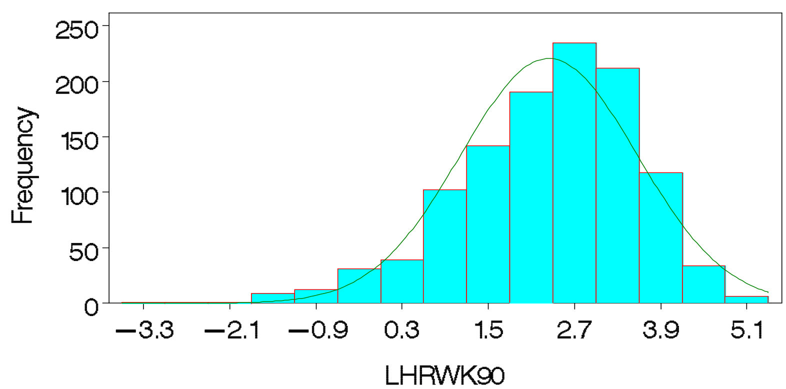
In this thesis, we study the change in physical activity through the observations of HRWK. As normality is the major assumption of both LGC model and random coefficient model, we need to examine the normality of HRWK. Figure 2(a) demonstrates the HRWK distribution at the first interview. It shows that HRWK is not normally distributed. To improve the normality of HRWK, we apply log transform after comparing several transformations including the cubic, square, square-root, log, reciprocal root functions. The log transform could eliminate the serious “skewness” problem and shrink extreme high values. As the log function cannot be directly applied to the zero-value HRWK, we can either exclude zero-value HRWK or assign a small value, for example 0.0005, to the zero-value HRWK. We denote the log values of HRWK after excluding zero-value observations as LHRWK (i.e., LHRWK=log(HRWK)), and define the log values of HRWK after introducing the small number 0.0005 as rLHRWK (i.e., rLHRWK=log(HRWK+0.0005)). Figure 2(b) and Figure 2(c) depict the distributions of LHRWK and rLHRWK, respectively. Figure 2(b) shows that LHRWK has approximate normal distribution. Thus, we label zero-value HRWK data missing when we compute LHRWK values for the tradeoff of normality. As demonstrated in Figure 2(b), we can preserve zero-value HRWK observations when calculating rLHRWK, while we still have the skewness problem. Figure 3 ~ Figure 7 depict the distribution of HRWK, LHRWK and rHRWK in the 2^{nd}-6^{th} interviews. However, to label zero-value HRWK observations missing results in many missing data. Table 6 lists the number of zero-value HRWK and its percentage among the total number of participants (given in Table 4) in each interview.

Table 6. Number of zero-value HRWK in each interview

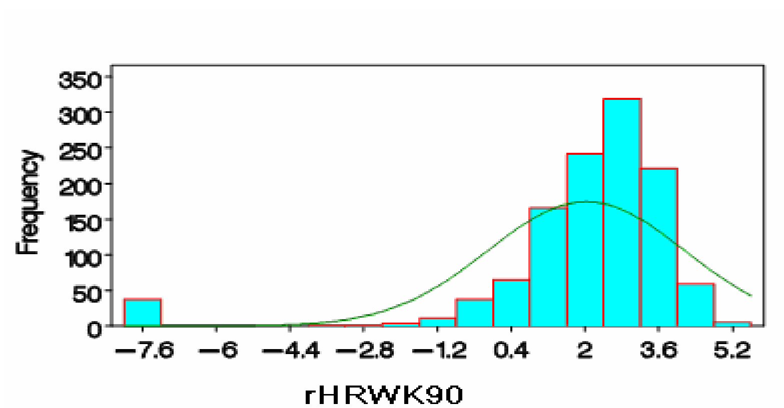
Visit time	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	6^{th}
Number of zero-value HRWK	38	55	95	99	37	34
Percentage (%)	3.24	5.06	9.93	12.24	4.47	4.99



(a) Distribution of HRWK in Year 1990

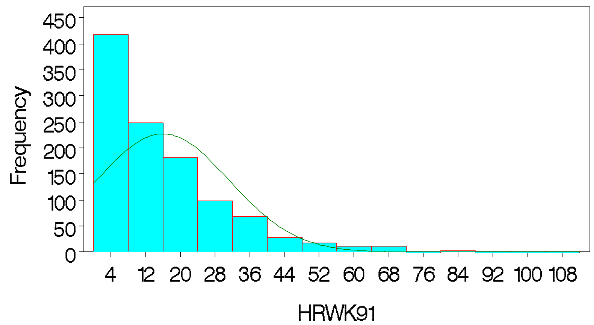


(b) Distribution of LHRWK in Year 1990

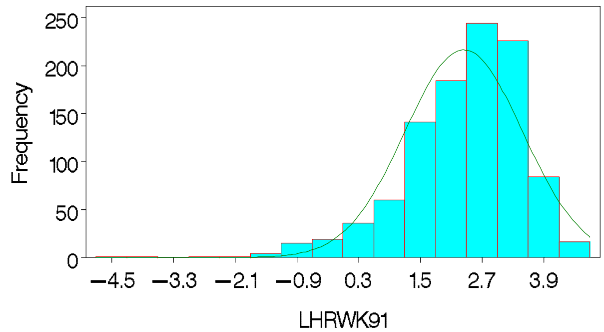


(c) Distribution of rLHRWK in Year 1990

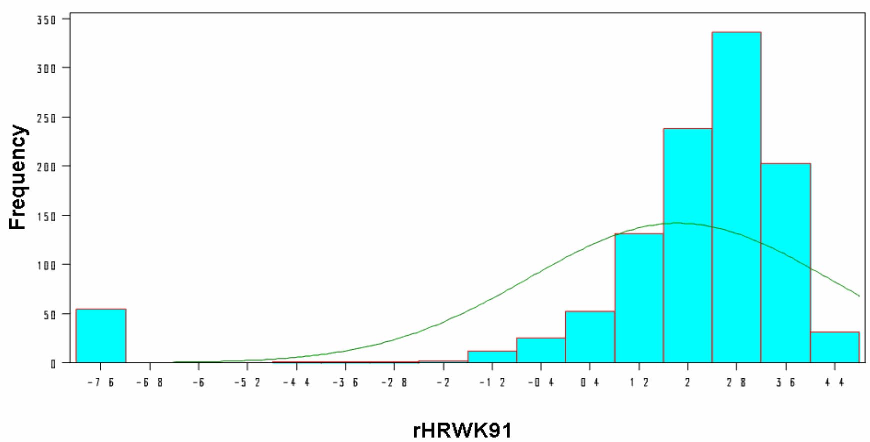
Figure 2. Transformation of HRWK in the 1st interview



(a) Distribution of HRWK in Year 1991

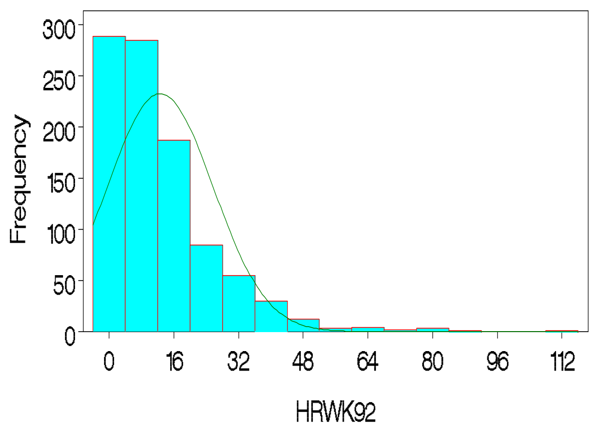


(b) Distribution of LHRWK in Year 1991

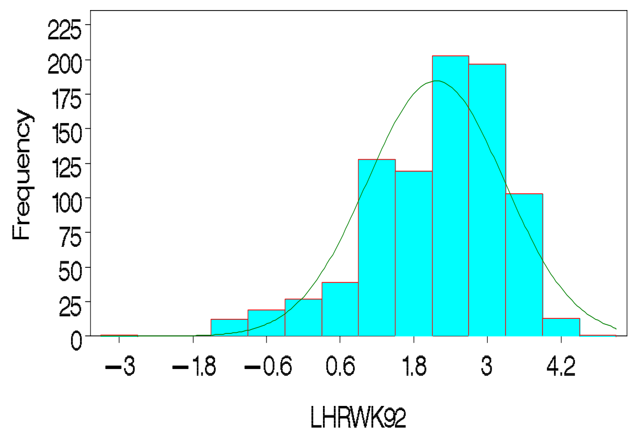


(c) Distribution of rLHRWK in Year 1991

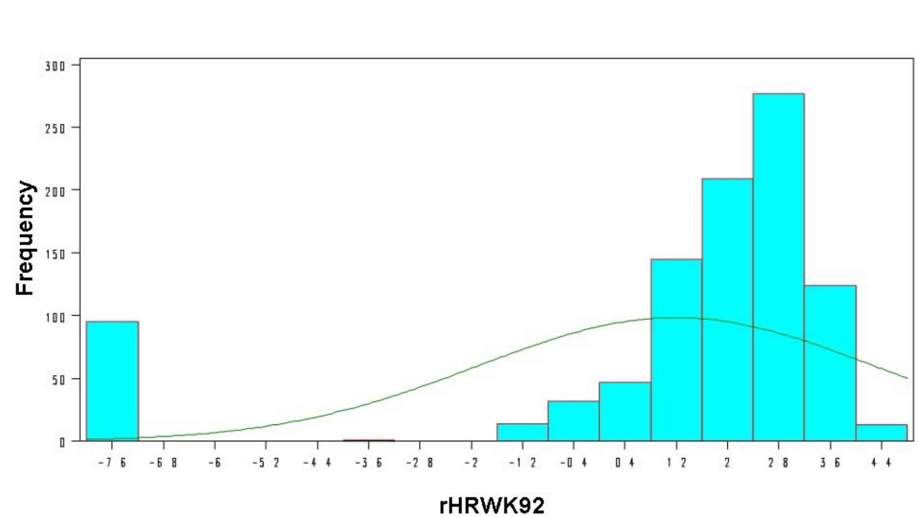
Figure 3. Transformation of HRWK in 2nd interview



(a) Distribution of HRWK in Year 1992

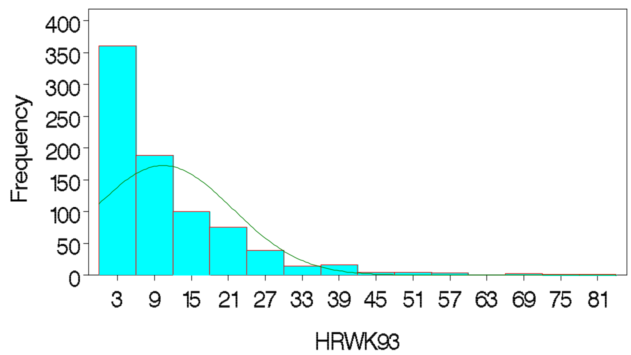


(b) Distribution of LHRWK in Year 1992

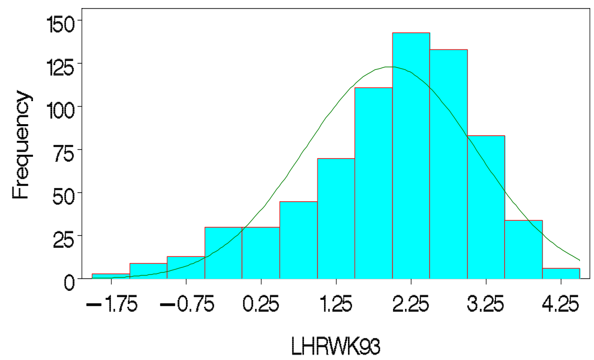


(c) Distribution of rLHRWK in Year 1992

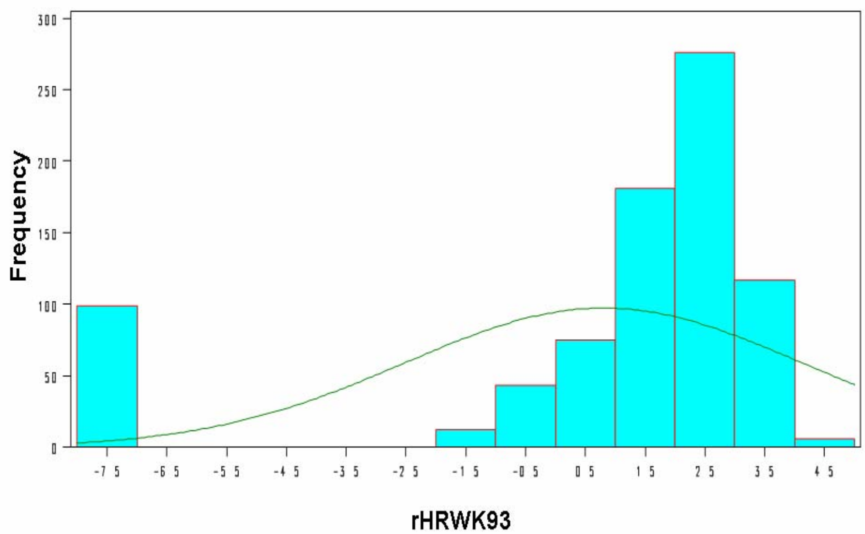
Figure 4. Transformation of HRWK in the 3rd interview



(a) Distribution of HRWK in Year 1993

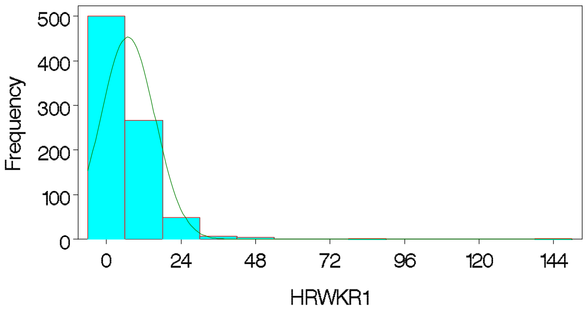


(b) Distribution of LHRWK in Year 1993

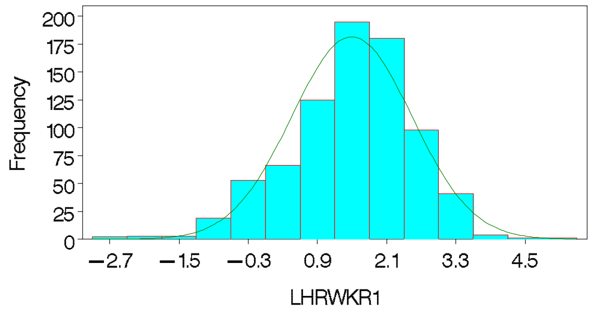


(c) Distribution of rLHRWK in Year 1993

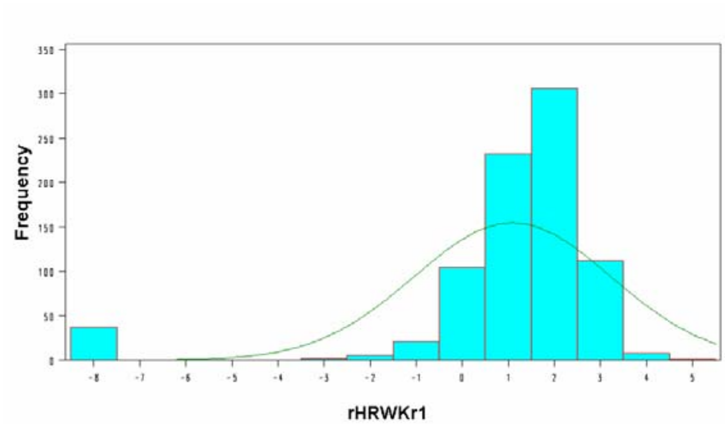
Figure 5. Transformation of HRWK in the 4th interview.



(a) Distribution of HRWK during Year R1

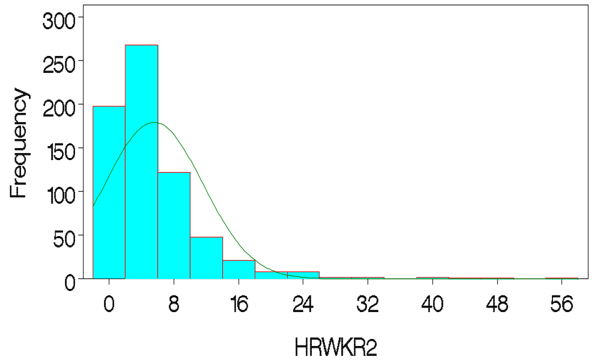


(b) Distribution of LHRWK during Year R1

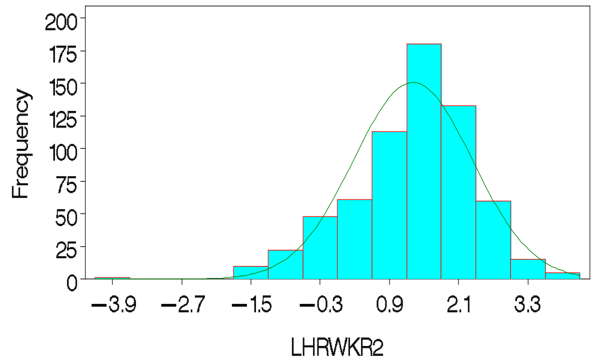


(c) Distribution of rLHRWK during Year R1

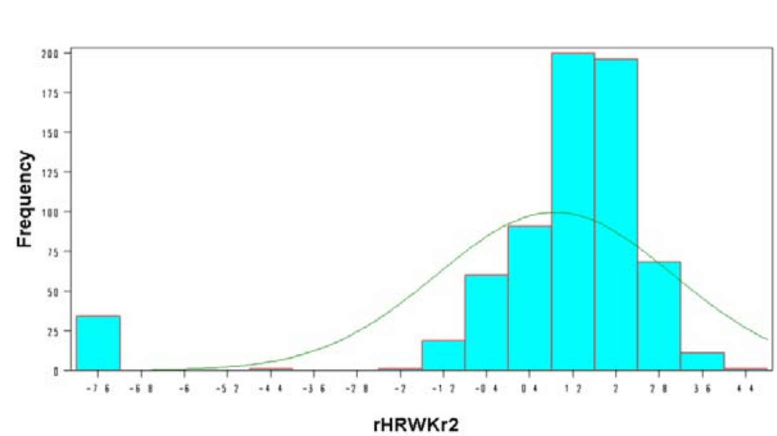
Figure 6. Transformation of HRWK in the 5^{th} interview



(a) Distribution of HRWK during Year R2



(b) Distribution of LHRWK during Year R2



(c) Distribution of rLHRWK during Year R2

Figure 7. Transformation of HRWK in the 6th interview

3.5 PHYSICAL ACTIVITY DECLINE ACROSS TIME

In this thesis, we study the change of HRWK, which represents the change in physical activity across time. The mean of HRWK depicted in Figure 8 shows that physical activity declines across time. Such decline is almost linear with respective to time during the first four interview occasions, as shown in Figure 9. Figure 10 and Figure 11 demonstrate the change of mean values of rLHRWK and LHRWK, respectively. The trend of rLHRWK is hard to be observed. Especially, the mean value of rLHRWK in the fourth interview (Year 1993) is much smaller than that of R1. It is mainly because there are more zeros value HRWK in the fourth interview (Year 1993) than the fifth interview at R1 as listed in Table 6. By labeling the zero-value HRWK as missing, we can largely improve the linearity of the mean LHRWK as depicted in Figure 11.

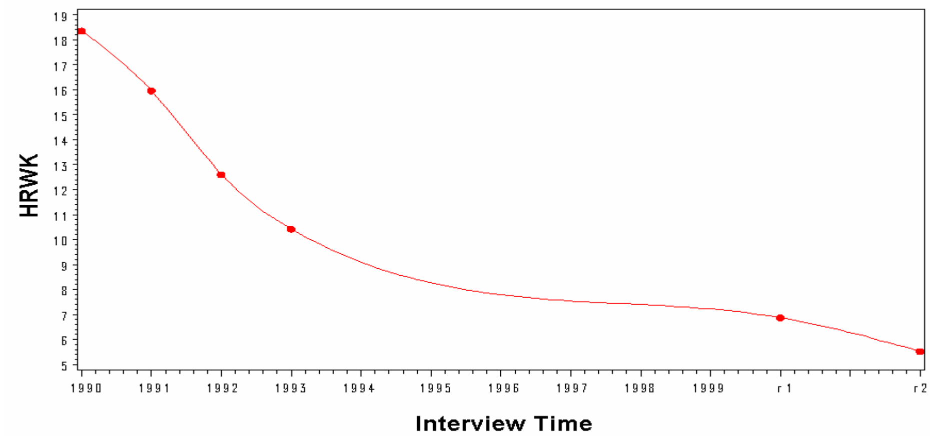


Figure 8. Mean of HRWK through years

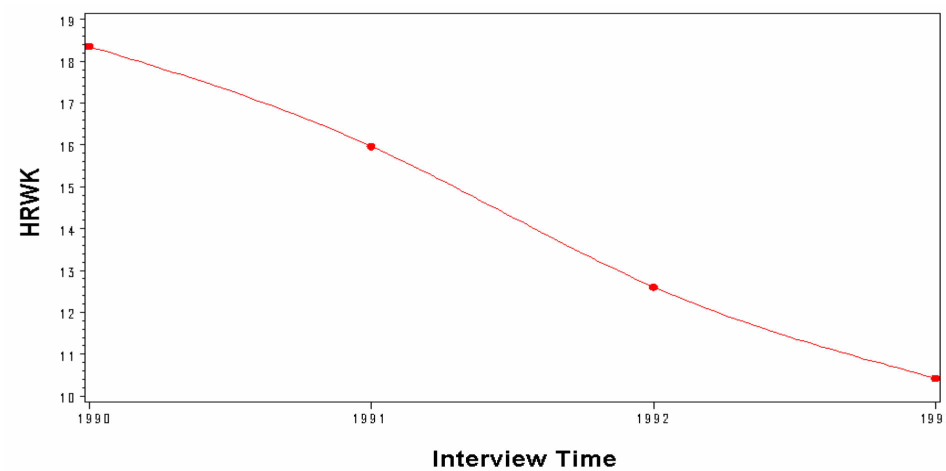


Figure 9. Mean of HRWK at first four years

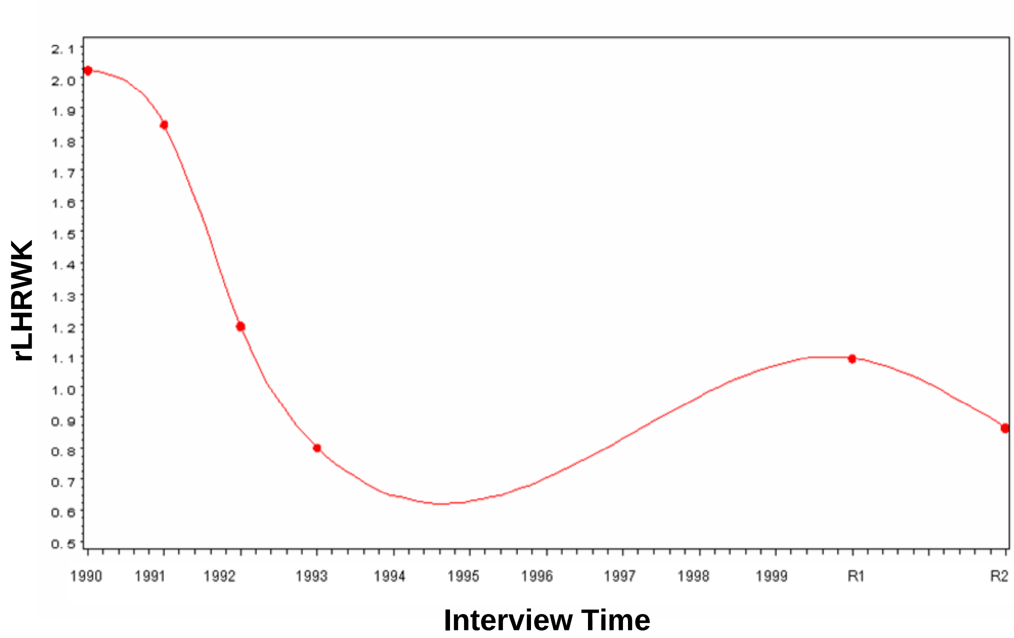


Figure 10. Mean of rLHRWK through years

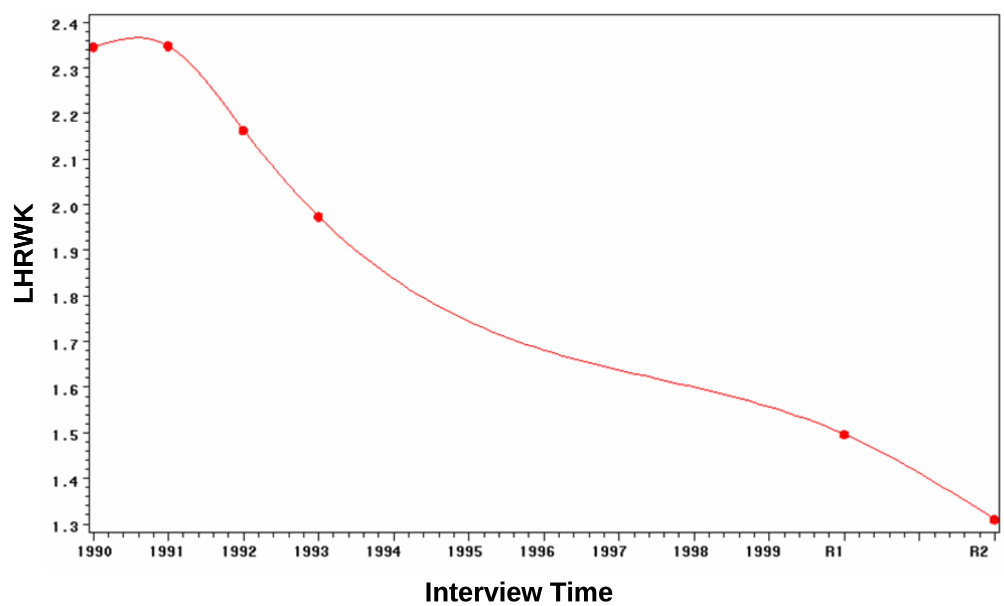
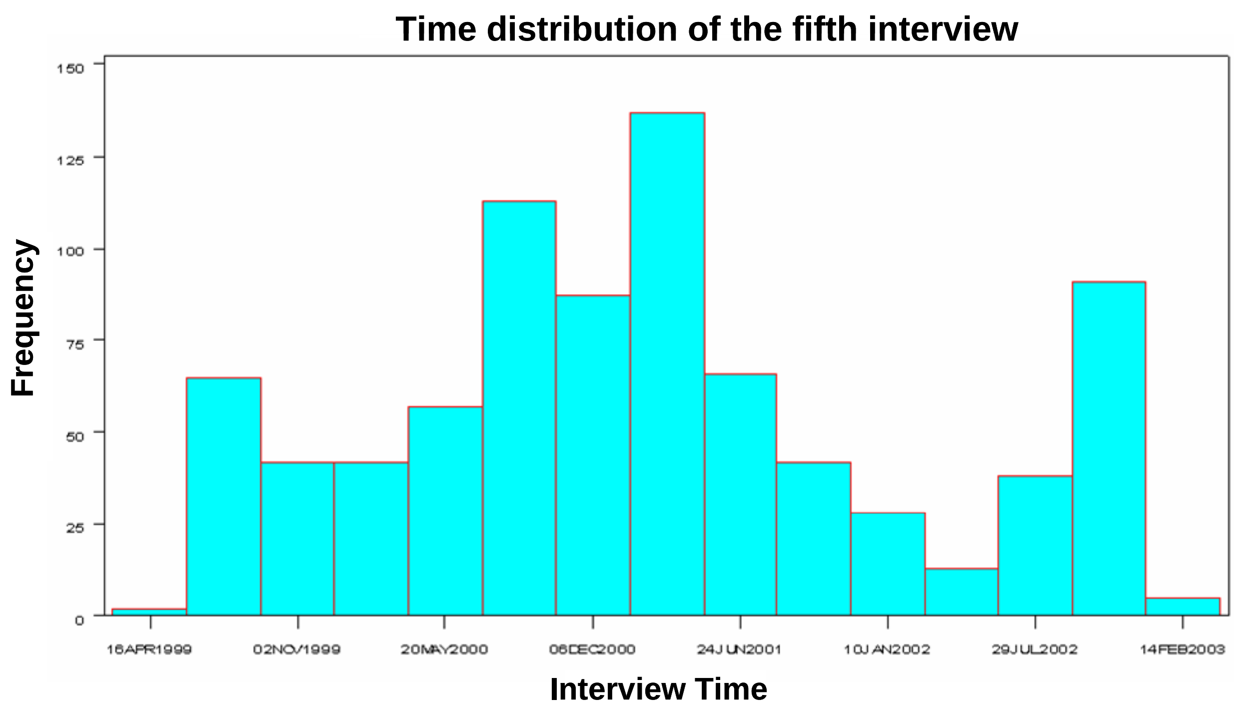
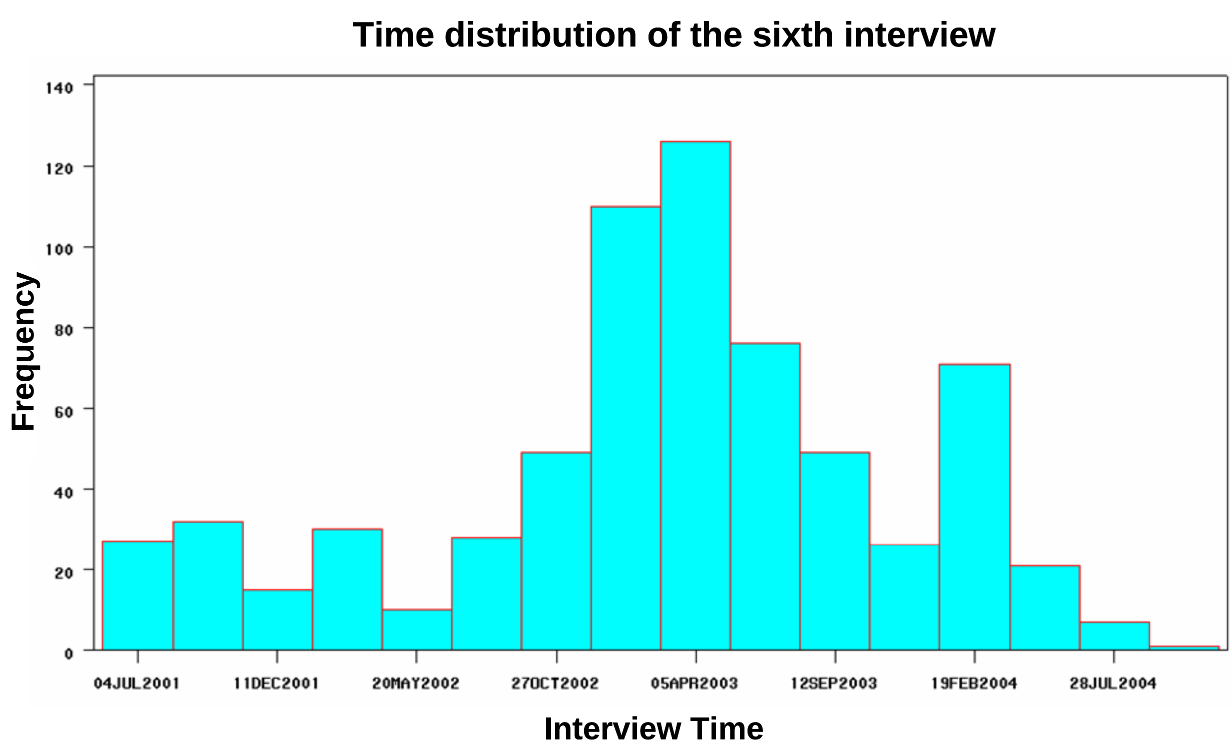


Figure 11. Mean of LHRWK through years



(a) Time distribution of the fifth interview



(b) Time distribution of the six interview

Figure 12. There are overlapping periods during the fifth and sixth interview time

3.6 SUMMARY OF DATA CHARACTERISTICS

The above descriptive statistics shows that the longitudinal data have three major problems, i.e., irregular interview time intervals, non-normality, and missing data. Firstly, these data are measured at irregular observation occasions. Although the observation occasions are regular at the first four interview occasions, there are overlapping periods during the fifth and sixth interview time as shown in Figure 12. Secondly, the observation data such as HRWK is not normally distributed. After excluding zero-value HRWK observations and applying log transformation, LHRWK is approximately normally distributed, as shown in Figure 2 ~ Figure 7. However, this introduces more missing data. Thirdly, because of the large time space between the phase 1 study and phase 2 study, many individuals are lost to follow-up. This leads to many missing data, especially in the fifth interview and sixth interview.

4.0 COMPARISIONS OF LGC AND RANDOM COEFFICIENT MODEL FOR DATA WITHOUT MISSING

As described in Chapter 2, the LGC model and random coefficient model are highly similar. They share the same objectives and have a similar representation. If they are used to represent the same set of good longitudinal data, their models would yield identical estimates of the relevant parameters (Hox, 2000). As an illustration of the equivalence of these two models, we use both LGC model and random coefficient model to analyze the decline of physical activity data described in Chapter 3. Specifically, we apply both models to characterize the trajectory of HRWK collected in the first four observation years (from Year 1990 to Year 1993) in this Chapter. Because these two models are different in processing the missing data, we only consider the observation data set without missing. As depicted in Figure 9, the mean of HRWK during the first four years has nearly linear relationship with the observation time. We hence assume for simplicity that HRWK has a linear change shape. This Chapter consists of three sections: Section 4.1 presents the formulation of LGC model for HRWK. Section 4.2 uses the random coefficient model to analyze the HRWK data. Finally, we discuss and compare the estimate results of these two models in the last section.

4.1 LATENT GROWTH CURVE MODEL FOR HRWK

As described in Chapter 3, HRWK values are different for individuals with different SEX, SES, RACE, and GRADE. We use LGC model to characterize the change of HRWK with these four effects. Specifically, we define y_{it} as HRWK value of individual i ($i = 1,2,..., N$) at measurement occasion t ($t = 0,1,2,3$). That's, $y_{i0} = \text{HRWK90}$, $y_{i1} = \text{HRWK91}$, $y_{i2} = \text{HRWK92}$, and $y_{i3} = \text{HRWK93}$. We also denote four fixed effects SEX, RACE, SES, GRADE as $x_{SEX,i}$, $x_{RACE,i}$, $x_{SES,i}$, and $x_{GRADE,i}$, respectively. Similar to (2.1), we have the linear shape LGC model representation for HRWK as

$$\begin{aligned} y_{it} &= \text{int}_i + \lambda_{it}\text{slp}_\text{i} + \varepsilon_{it} \\ \text{int}_i &= \mu_{\text{int}} + \alpha_1 x_{SEX,i} + \alpha_2 x_{RACE,i} + \alpha_3 x_{GRADE,i} + \alpha_4 x_{SES,i} + e_{\text{int},i} \ , \\ \text{slp}_i &= \mu_{slp} + \alpha_5 x_{SEX,i} + \alpha_6 x_{RACE,i} + \alpha_7 x_{GRADE,i} + \alpha_8 x_{SES,i} + e_{\text{slp,i}} \end{aligned} \tag{4.1}$$

where the initial level and linear shape for each individual subject are presented by the latent factors, int_i and slp_i, with initial value μ_{int} and μ_{slp}, and normal deviations, $\begin{pmatrix} e_{\text{int},i} \\ e_{\text{slp,i}} \end{pmatrix} \sim N(0, \sum_e)$,

and $\sum_e = \begin{pmatrix} \sigma^2_{\text{int}} & \sigma^2{}_{slp,\,\text{int}} \\ \sigma^2{}_{slp,\,\text{int}} & \sigma^2_{slp} \end{pmatrix}$. As the LGC is a multivariate approach, the LGC model is flexible in selecting the covariance structure of the residual. In (4.1), $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$, α_7 and α_8 are the fixed effects parameters of $x_{SEX,i}$, $x_{RACE,i}$, $x_{SES,i}$, and $x_{GRADE,i}$. As given in Appendix A, we have the following definitions: $x_{SEX,\,i} = 1$ if 'the individual is female', $x_{SEX,\,i} = 0$ if 'the individual is male'; $x_{RACE,\,i} = 0$ if 'race is White', $x_{RACE,\,i} = 1$ if 'race is Black'; and $x_{SES,\,i} = 0$ if the individual is from the high SES, $x_{SES,\,i} = 1$ if the individual is from the middle SES,

$x_{SES,i} = 2$ if the individual is from the low SES; and $x_{GRADE,i}$ is the grade of the individual received his or her first interview in Year 1990. The possible values of $x_{GRADE,i}$ are 7, 8, 9 in our study data.

The LGC model (4.1) introduces the time effect by constraining λ_{it} values. As described in Section 3.1, the observation data were acquired uniformly during the first four interviews from Year 1990 to Year 1993. Year 1990 is considered as the baseline time. Thus, we then assign $\lambda_{i0} = 0$, $\lambda_{i1} = 1$, $\lambda_{i2} = 2$, and $\lambda_{i3} = 3$. The LGC model (4.1) for HRWK can then be rewritten as (4.2). The graphic representation of (4.1) is also depicted in Figure 13.

$$\begin{aligned} y_{i0} &= \text{int}_i + \varepsilon_{0i} \\ y_{i1} &= \text{int}_i + \text{slp}_\text{i} + \varepsilon_{1i} \\ y_{i2} &= \text{int}_i + 2\text{slp}_\text{i} + \varepsilon_{2i} \\ y_{i3} &= \text{int}_i + 3\text{slp}_\text{i} + \varepsilon_{3i} \\ \text{int}_i &= \mu_{\text{int}} + \alpha_1 x_{SEX,i} + \alpha_2 x_{RACE,i} + \alpha_3 x_{GRADE,i} + \alpha_4 x_{SES,i} + e_{\text{int},i} \\ \text{slp}_i &= \mu_{slp} + \alpha_5 x_{SEX,i} + \alpha_6 x_{RACE,i} + \alpha_7 x_{GRADE,i} + \alpha_8 x_{SES,i} + e_{\text{slp,i}} \end{aligned} \quad . \tag{4.2}$$

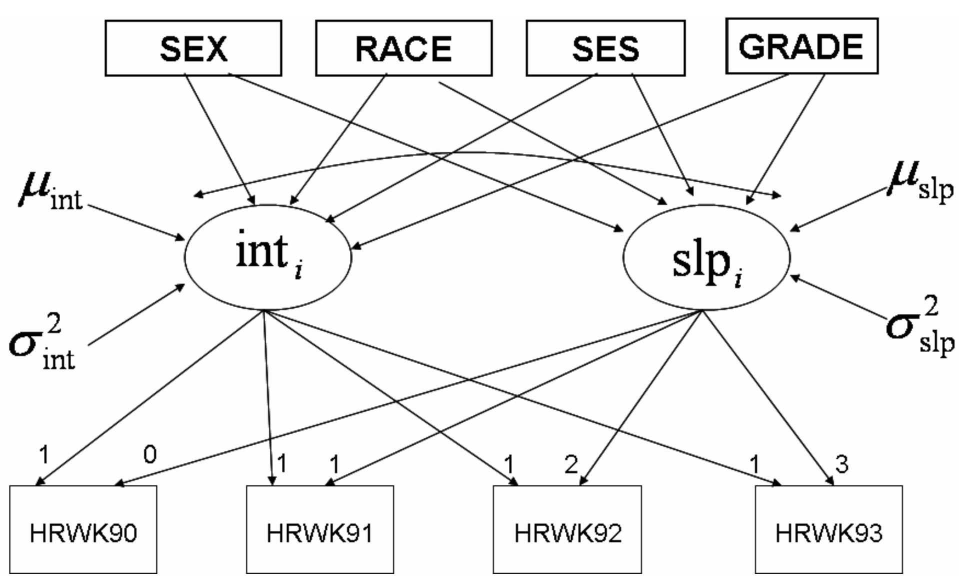


Figure 13. Graphic representation of the LGC model for HRWK with four fixed effects

4.2 RANDOM COEFFICIENT MODEL ANALYSIS

In the random coefficient model analysis, let Y_{it} be the observation HRWK of each individual subject, i ($i = 1,2,\ldots,N$) measured at each of four time occasions $t = 0,1,2,3$ (from Year 1990 to Year 1993). Similar to (2.2), the linear pattern random coefficient model can then be presented as

$$\begin{aligned} Y_{it} &= u_{\text{int}} + e_{\text{int},i} + u_{slp} Time_{it} + e_{slp,i} Time_{it} + a_{SEX} X_{SEX,i} + a_{RACE} X_{RACE,i} + a_{GRADE} X_{GRADE,i} + a_{SES} X_{SES,i} \\ &\quad + a_{SEX,T} X_{SEX,i} Time_{it} + a_{RACE,T} X_{RACE,i} Time_{it} + a_{GRADE,T} X_{GRADE,i} Time_{it} + a_{SES,T} X_{SES,i} Time_{it} + \varepsilon_{it} \end{aligned}, \quad (4.3)$$

where $Time_{it}$ is a variable denoting measurement occasions. For the uniformly acquired HRWK at $t = 0,1,2,3$ during Year 1990 ~ Year 1993, $Time_{i0} = 0$, $Time_{i1} = 1$, $Time_{i2} = 2$ and $Time_{i3} = 3$. The initial level and linear shape for each individual subject have the expectations u_{int} and u_{slp}, and random deviations $e_{\text{int},i}$ and $e_{slp,i}$, which are normally distributed $\begin{pmatrix} e_{\text{int},i} \\ e_{\text{slp,i}} \end{pmatrix} \sim N(0, \sum_e)$, where $\sum_e = \begin{pmatrix} \sigma^2_{\text{int}} & \sigma^2_{slp,\,\text{int}} \\ \sigma^2_{slp,\,\text{int}} & \sigma^2_{slp} \end{pmatrix}$. ε_{it} is a residual at the measurement level and $\varepsilon_{it} \sim N(0, \sigma^2_{\varepsilon})$. Fixed effects $X_{SEX,i}$, $X_{RACE,i}$, $X_{GRADE,i}$, and $X_{SES,i}$ represents the factors of SES, RACE, GRADE and SES for individual i, respectively. They have the same values as those in the LGC model. That's, $X_{SEX,i} = 1$ if 'the individual is female', $X_{SEX,i} = 0$ if 'the individual is Male'; $X_{RACE,i} = 0$ if 'RACE is White', $X_{RACE,i} = 1$ if 'RACE is Black'; and if the individual is from the high SES, $X_{SES,i} = 0$; $X_{SES,i} = 1$ for the individual who is from the middle SES, and $X_{SES,i} = 2$ if the individual is from the low SES; and $X_{GRADE,i}$ is the grade of the individual received his or her first interview in 1990, and has one of the values 7, 8, 9. Finally, a_{SEX}, a_{RACE}, a_{GRADE} and a_{SES}

represent respectively fours fixed effects on the initial level, and $a_{SEX,T}$, $a_{RACE,T}$, $a_{GRADE,T}$ and $a_{SES,T}$ represent respectively fixed effects on the linear shape.

4.3 COMPARISIONS

PROC CALIS routine in SAS 9.0 with the default maximum likelihood (ML) estimation described in Chapter 2 is used to estimate parameters of LGC model (4.2). For random coefficient models (4.3), SAS Proc Mixed procedure is used to estimate model parameters. As the ML method cannot lead to the convergence of model estimation processing, the default Restricted Maximum Likelihood (REML) method is used for random coefficient model parameter estimations. The complete SAS codes for these two model estimations are included in Appendix C, and the corresponding full model fit statistics and model estimation results are presented in Appendix D.

The parameter estimates of LGC model (4.2) and random coefficient model (4.3) are listed in Table 7, where the first two columns present the relevant parameters of (4.2), and the last two columns list the parameters of (4.3). Similar to Figure 13, the estimate LGC model (4.2) is also demonstrated in Figure 14. Due to the different maximum-likelihood methods used, the parameter estimates are close but with different standard errors and random effects. However, the estimated fixed effect parameters of these two models are the same, as listed in Table 7. These two models lead to the same substantive conclusions summarized as follows: after controlling for the effect of the covariates, a mean growth curve of activity time emerges with an initial level of 27.2362 and a growth rate of -2.6418 (the negative sign indicates the activity time is decline). SEX has negative effects on the initial level and positive effect on the growth rate, leading to the conclusion that Male ($X_{SEX} = 0$) had more activity time at the first measurement occasion (Year 1990), and a larger decline rate of physical time than female ($X_{SEX} = 1$). RACE has a negative effect on both the initial level and the growth rate. That is, the White students spent more time in the physical activity at Year 1990, and also had the smaller physical activity decline rate than the Black student. Similarly, GRADE has a negative effect on both the initial level and the change rate, which means the low grade students spent more time in the physical activity at Year 1990, and also had the smaller physical activity decline rate than the high grade students. SES has a positive effect on the initial level and the negative effect on the change rate. This means that the students from the high social status family spent less time in the physical activity at Year 1990, and also had the smaller physical activity decline rate than the students from low social status family.

4.4 SUMMARY

In this Chapter, we apply the LGC model and random coefficient model to analyze the no-missing data in the first four observation occasions. The model estimate results confirm that the two models are identical for the data without missing observations.

Table 7. The estimated parameters of LGC model (4.2) and random coefficient model (4.3)

The LGC model (4.2)		Random coefficient model (4.3)	
Parameter	Estimate	Parameter	Estimate
Fixed effects		Fixed effects	
α_1	-14.4078 (1.1788)	a_{SEX}	-14.4078 (0.9480)
α_2	-0.8937 (0.5252)	a_{RACE}	-0.8937 (1.3684)
α_3	-1.1513 (0.7055)	a_{GRADE}	-1.1513 (0.5673)
α_4	0.6462 (0.9728)	a_{SES}	0.6462 (0.7824)
α_5	1.8742 (0.3796)	$a_{SEX,T}$	1.8742 (0.4371)
α_6	-0.8316 (0.6310)	$a_{RACE,T}$	-0.8316 (0.5480)
α_7	-0.2322 (0.2616)	$a_{GRADE,T}$	-0.2322 (0.2272)
α_8	-0.3266 (0.3608)	$a_{SES,T}$	-0.3266 (0.3133)
μ_{int}	27.2362 (1.8072)	μ_{int}	27.2362 (1.4533)
μ_{slp}	-2.6418 (0.6701)	μ_{slp}	-2.6418 (1.7171)
Random effects		Random effects	
σ_ε^2	107.1895	σ_ε^2	128.18
σ_{int}^2	172.357 (13.382)	σ_{int}^2	6.735e-06
σ_{slp}^2	12.581 (1.970)	σ_{slp}^2	2.6096
$\sigma_{slp,\text{int}}^2$	-41.919 (4.571)	$\sigma_{slp,\text{int}}^2$	0

Note: Standard errors are given in parentheses. The Chi-square test of model fit is $\chi^2(16) = 296.3881$ (p<0.0001); RMSEA=0.1566. GFI=0.9177. For the random coefficient model: -2*Res Log Likelihood=22849.1, AIC=22859.1.

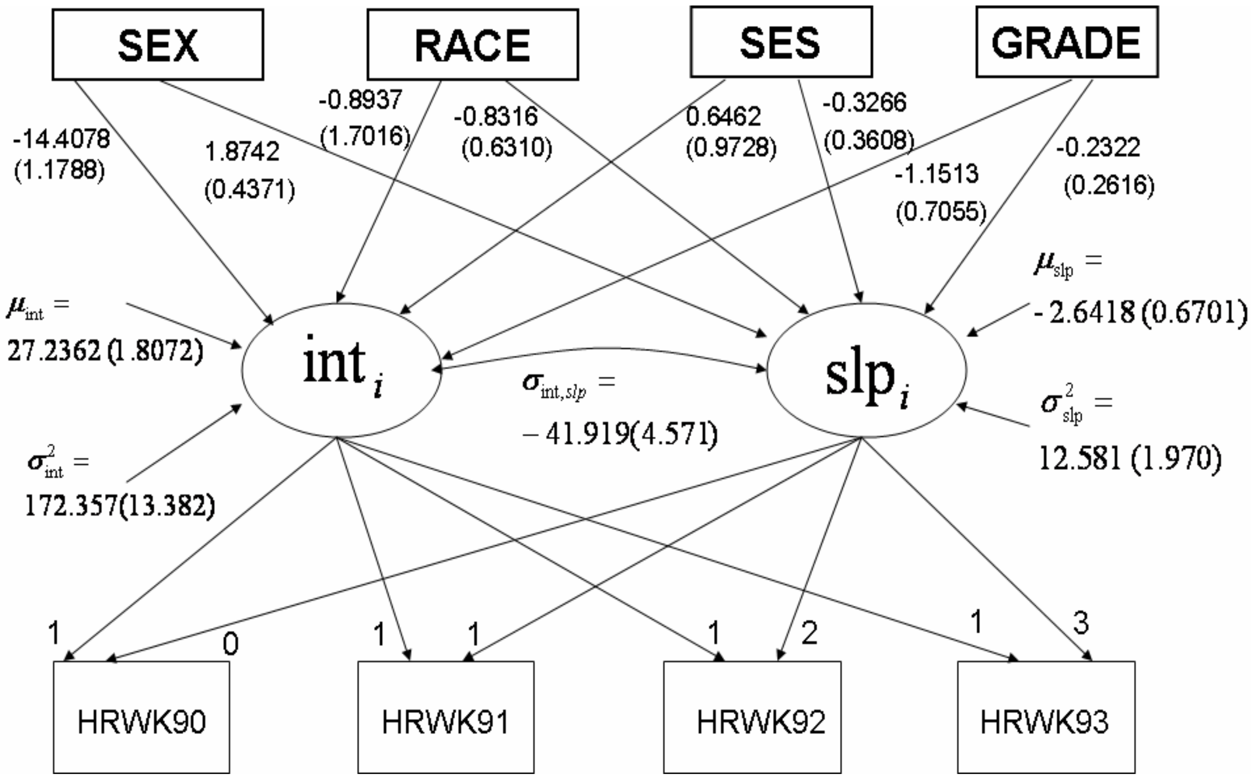


Figure 14. Graphic representation of LGC model (4.2) (standard errors are given in parentheses).

5.0 COMPARISIONS OF LGC AND RANDOM COEFFICIENT MODEL FOR PITTPAS DATA

LGC is a multivariate model, while random coefficient model is a univariate one. They are essentially different, although they have similar results when data sets have no missing observations as demonstrated in Chapter 4. To show their difference, we apply these two models to analyze the decline of activity in all measured PittPAS data described in Chapter 3, and compare their performance. Specifically, we first apply random coefficient model to analyze the decline of physical activity recorded in PittPAS data. Due to the normality and missing observations of HRWK, we conduct a sensitivity analysis and use three random coefficient models to analyze HRWK, LHRWK, and rLHRWK, separately. Then, we use LGC model to analyze the same observations as random coefficient model and compare its performance with random coefficient model. Our study shows that random coefficient model is more appropriate to model the longitudinal data which have a large amount of missing observations and non-uniform observation time.

5.1 RANDOM COEFFICIENT MODEL ANALYSIS

5.1.1 Model Specification

HRWK in PittPAS data is our major observation that reflects the physical activity of each individual. As described in Chapter 3, HRWK observations measured at six occasions are not normally distributed, whereas LHRWK is nearly normal. Therefore, we consider three random coefficient models for observations HRWK, LHRWK, and rLHRWK (recall that *LHRWK=log(HRWK), rLHRWK=log(HRWK+0.0005)*) , respectively. Specifically, three random coefficient models with linear growth pattern can be formulated as

$$\begin{aligned} Y_{i,t,j} &= u_{\text{int},j} + e_{\text{int},i,j} + u_{slp,j} Time_{i,t,j} + e_{slp,i,j} Time_{i,t,j} + a_{SEX,j} X_{SEX,i,j} + a_{RACE,j} X_{RACE,i,j} + a_{GRADE,j} X_{GRADE,i,j} + a_{SES,j} X_{SES,i,j} \\ &+ a_{SEX,T,j} X_{SEX,i,j} Time_{i,t,j} + a_{RACE,T,j} X_{RACE,i,j} Time_{i,t,j} + a_{GRADE,T,j} X_{GRADE,i,j} Time_{i,t,j} + a_{SES,T,j} X_{SES,i,j} Time_{i,t,j} + \varepsilon_{i,t,j} \end{aligned}, \tag{5.1}$$

where observations HRWK, LHRWK, and rHRWK of each individual $i = 1,2,\ldots,N$ measured at six time occasions $t = 0,1,2,3,4,5$ are denoted by $Y_{i,t,j}$ ($j = 0$ is for HRWK, $j = 1$ is for LHRWK, and $j = 2$ is for rHRWK). In (5.1), $\varepsilon_{i,t,j}$ is a residual at the measurement level and $\varepsilon_{i,t,j} \sim N(0, \sigma^2_{\varepsilon,j})$, random effects $e_{\text{int},i,j}$ and $e_{slp,i,j}$ have normal distribution $\begin{pmatrix} e_{\text{int},i,j} \\ e_{\text{slp,i,j}} \end{pmatrix} \sim N(0, \sum_{e,j})$, where $\sum_{e,j} = \begin{pmatrix} \sigma^2_{\text{int},j} & \sigma^2_{slp,\text{int},j} \\ \sigma^2_{slp,\text{int},j} & \sigma^2_{slp,j} \end{pmatrix}$. Fixed effects $X_{SEX,i,j}$, $X_{RACE,i,j}$, $X_{GRADE,i,j}$, and $X_{SES,i,j}$ represent factors SES, RACE, GRADE and SES of individual i at model *j*, respectively. These fixed effects variables are assigned the same values as those presented in Appendix A. That's, $X_{SEX,i,j} = 1$ if 'individual *i* is Female', $X_{SEX,i,j} = 0$ if 'individual *i* is Male'; $X_{RACE,i,j} = 0$ if 'RACE is White', $X_{RACE,i,j} = 1$ if 'RACE is Black'; if individual *i* is from the high SES, $X_{SES,i,j} = 0$; $X_{SES,i,j} = 1$ for the individual who is from the middle SES, and $X_{SES,i,j} = 2$ if the individual is from the low SES. $X_{GRADE,i,j}$ is the grade of the individual received his or her first interview in 1990. The values of $X_{GRADE,i,j}$ are 7, 8, 9.

In (5.1), $Time_{i,t,j}$ is a variable denoting the measurement occasion. For the uniformly acquired observations at first four years, $Time_{i,0,j} = 0$, $Time_{i,1,j} = 1$, $Time_{i,2,j} = 2$ and $Time_{i,3,j} = 3$. As the fifth interview and sixth interview time period are overlapping, we assign the values of time variables $Time_{i,4,j}$ and $Time_{i,5,j}$ as the differences between the interview year and 1990 for each individual *i*. For example, if an individual i received the fifth interview at Year 2001, then $Time_{i,4,j} = 11$.

5.1.2 Sensitivity Analysis

Random coefficient models (5.1) respectively fit into HRWK, LHRWK, rHRWK data using SAS Proc Mixed procedure with the Restricted Maximum Likelihood (REML) estimation method. The corresponding SAS codes are attached in Appendix E. The complete model fit statistics and parameter estimation results of these three random coefficient models are also presented in Appendix F. The final parameter estimates of three models are given in Table 8.

Table 8 shows that random coefficient model for HRWK (*j=0*) has similar performance to that for LHRWK (*j=1*). In both models, factor a_{SES} and its interaction with time $a_{SES,T}$ are non-significant. Moreover, the estimate parameters are quite close in these two models. For example, in the random coefficient model for HRWK (i.e., *j=0*), $u_{\text{int},0} = 27.5203$ means the initial HRWK value in Year 1990 is 27.5203. In the second random coefficient model for LHRWK (i.e., *j=1*), $u_{\text{int},1} = 3.1992$, which means the initial HRWK value in Year 1990 is $e^{u_{\text{int},1}} \approx 24.5129$. However, in the third random coefficient model for rLHRWK (*j=2*), $u_{\text{int},2} = 3.7622$ that indicates the initial HRWK value in Year 1990 is $e^{u_{\text{int},2}} \approx 43.043$. Furthermore, the third random coefficient model (*j=2*) has different significant factors with the first and second models. The major reason is that we manually introduce the small number 0.0005 into observations in random coefficient model for rHRWK. Such operation can prevent data missing when we apply *log* operator. However, it also disturbs the data distribution as shown in Figure 2 ~ Figure 7. Moreover, Table 8 shows that the second random coefficient model for LHRWK has the best data fitness performance, because it has the lowest AIC value compared with other two random coefficient model. This is because LHRWK is more normally distributed which better satisfies the normal assumption of random coefficient model. However, because of *log* operation, there are more missing data of LHRWK observations in the second model estimation. Since the first

Table 8. Comparisons of three random coefficient models (5.1)

Observations		HRWK	LHRWK	rLHRWK
Fixed Effects	$a_{SEX,j}$	-13.3177 (<.0001)	-1.0547 (<.0001)	-1.8179 (<.0001)
	$a_{RACE,j}$	-2.1792 (0.0060)	-0.2307 (0.0003)	-0.4042 (0.0041)
	$a_{GRADE,j}$	-2.1019 (<.0001)	-0.1566 (<.0001)	-0.4467 (<.0001)
	$a_{SES,j}$	**-0.08532 (0.8629)**	**-0.02430 (0.5405)**	-0.2472 (0.0049)
	$a_{SEX,T,j}$	0.9294 (<.0001)	0.04880 (<.0001)	0.09640 (<.0001)
	$a_{RACE,T,j}$	0.2386 (0.0185)	0.02466 (0.0018)	**-0.00490 (0.7933)**
	$a_{GRADE,T,j}$	0.1227 (0.0040)	0.009521 (0.0035)	0.03110 (<.0001)
	$a_{SES,T,j}$	**0.001897 (0.9733)**	**0.000062 (0.9886)**	**0.01581 (0.1311)**
	$u_{\text{int},j}$	27.5203 (<.0001)	3.1992 (0.0209)	3.7622 (<.0001)
	$u_{slp,j}$	**-1.6474 (0.2608)**	**-0.1271 (0.2274)**	**-0.1942 (0.3812)**
Fitness statistics	AIC (the smaller the better)	42924.0	14163.3	24662.9
	-2Resloglikelihood	42914.0	14173.3	24672.9

Note: *p* values are given in the parentheses. *p>0.05* indicates the factor is non-significant.

random coefficient model (*j=0*) for HRWK has the similar performance to the second random coefficient model (*j=1*) for LHRWK, we finally use the first random coefficient model to study the decline of physical activity.

5.1.3 Random Coefficient Model for HRWK

Through the above sensitive analysis, the first random coefficient model (*j=0*) is a good trade-off model for our data analysis. Table 8 shows that SES in the model is non-significant factor and can be removed. The random coefficient model for HRWK can then be rewritten as

$$\begin{aligned} Y_{i,t,0} &= u_{\text{int},0} + e_{\text{int},i,0} + u_{slp,0}Time_{i,t,0} + e_{slp,i,0}Time_{i,t,0} + a_{SEX,0}X_{SEX,i,0} + a_{RACE,0}X_{RACE,i,0} + a_{GRADE,0}X_{GRADE,i,0} \\ &+ a_{SEX,T,0}X_{SEX,i,0}Time_{i,t,0} + a_{RACE,T,0}X_{RACE,i,0}Time_{i,t,0} + a_{GRADE,T,0}X_{GRADE,i,0}Time_{i,t,0} + \varepsilon_{i,t,0} \end{aligned} \quad (5.2)$$

Using SAS Proc Mixed procedure with the REML estimation method, we fit model (5.2) to HRWK observations. The model parameter estimates are listed in Table 9. The corresponding estimate of (5.2) is

$$\begin{aligned} \hat{Y}_{i,t,0} &= 27.4589 + \hat{e}_{\text{int},i,0} - 1.6459 Time_{i,t,0} + \hat{e}_{slp,i,0}Time_{i,t,0} - 13.3204 X_{SEX,i,0} - 2.2406 X_{RACE,i,0} - 2.1020 X_{GRADE,i,0} \\ &+ 0.9294 X_{SEX,i,0}Time_{i,t,0} + 0.24 X_{RACE,i,0}Time_{i,t,0} + 0.1227 X_{GRADE,i,0}Time_{i,t,0} + \hat{\varepsilon}_{i,t,0} \end{aligned} \quad , (5.3)$$

where $\hat{\varepsilon}_{it} \sim N(0, 136.25)$, and $\begin{pmatrix} \hat{e}_{\text{int},i} \\ \hat{e}_{\text{slp},i} \end{pmatrix} \sim N(0, \sum_{\hat{e}})$, where $\sum_{\hat{e}} = \begin{pmatrix} 7.156E-6 & 5.291E-9 \\ 5.291E-9 & 6.906E-6 \end{pmatrix}$.

Table 9. Random coefficient model (5.2) for HRWK

Fixed Effects	$a_{SEX,0}$	-13.3204 (<.0001)	$a_{SEX,T,0}$	0.9294 (<.0001)
	$a_{RACE,0}$	-2.2406 (0.0016)	$a_{RACE,T,0}$	0.2400 (0.0104)
	$a_{GRADE,0}$	-2.1020 (<.0001)	$a_{GRADE,T,0}$	0.1227 (0.0040)
	$u_{\text{int},0}$	27.4589 (<.0001)	$u_{slp,0}$	-1.6459 (<.0001)
Fitness measurements	AIC	42920.3	-2Resloglikelihood	42910.3

Note: *p* values are given in the parentheses. And *p>0.05* indicates the corresponding factor is non-significant.

The random coefficient model (5.3) leads to the following conclusions about the changes of HRWK. After controlling for the effect of the covariates, a mean growth curve of activity time emerges with an initial level of 27.4589 and a growth rate of -1.6459 (the negative sign indicates the change of physical activity is decline). The small variation (σ^2_{int} =7.156E-6 and σ^2_{slp} =6.906E-6) between the subjects around these mean values implies that HRWK of different individuals start their growth process at almost the same values with similar rates. The correlation between the initial value and growth rate is almost zero ($\sigma_{\text{int},slp} =$ 5.291E-9), which means the initial level has no predictive value for the growth rate. SEX has negative effects on the initial level and positive effect on the growth rate, leading to the conclusion that Male ($X_{SEX} = 0$) had more activity time at the first measurement occasion, Year 1990, and a larger decline rate of physical time than the female ($X_{SEX} = 1$). RACE has a negative effect on the initial level and positive effect on the growth rate. That means the White students spent more time in the physical activity at Year 1990, and also had the larger physical activity decline rate than the Black student. Similarly, GRADE has a negative effect on the initial level and positive effect on the change rate, which means the low grade students spent more time in the physical activity at Year 1990, and also had the larger physical activity decline rate than the high grade students. As SES is not a significant factor for the change of HRWK, it means the family social status background has little effects on the physical activity time of students. This conclusion is similar to the study by Kimm, who used the questionnaire approach to show the household income is not associated with the decline of activity (Kimm SYK et al, 2002).

5.2 LATENT GROWTH MODEL ANALYSIS

In order to compare LGC model and random coefficient model, we use the LGC model to characterize HRWK observations measured in the six occasions. Similar to (4.1), the linear growth LGC model is adopted and can be formulated as

$$\begin{aligned} y_{it} &= \text{int}_i + \lambda_{it}\text{slp}_\text{i} + \varepsilon_{it} \\ \text{int}_i &= \mu_{\text{int}} + \alpha_1 x_{SEX,i} + \alpha_2 x_{RACE,i} + \alpha_3 x_{GRADE,i} + \alpha_4 x_{SES,i} + e_{\text{int},i} \ , \\ \text{slp}_i &= \mu_{slp} + \alpha_5 x_{SEX,i} + \alpha_6 x_{RACE,i} + \alpha_7 x_{GRADE,i} + \alpha_8 x_{SES,i} + e_{\text{slp,i}} \end{aligned} \qquad (5.4)$$

where y_{it} denotes the HRWK value of individual i ($i = 1,2,..., N$) at measurement occasion t ($t=0,1,2,3,4,5$). That's, $y_{i0} = \text{HRWK90}$, $y_{i1} = \text{HRWK91}$, $y_{i2} = \text{HRWK92}$, $y_{i3} = \text{HRWK93}$, $y_{i4} = \text{HRWK}R0$, and $y_{i5} = \text{HRWK}R1$. Due to the overlapping interview occasions happened during R0 and R1 period, we select the median interview date for measurements at the fifth and sixth occasions. That's, time factor λ_{it} is constrained to be $\lambda_{i0} = 0$, $\lambda_{i1} = 1$, $\lambda_{i2} = 2$, $\lambda_{i3} = 3$, $\lambda_{i4} = 11$ and $\lambda_{i5} = 13$. The other parameters have the same definitions as those in (4.1). Then, the LGC model (5.4) can be rewritten as

$$\begin{aligned} y_{i0} &= \text{int}_i + \varepsilon_{0i} \\ y_{i1} &= \text{int}_i + \text{slp}_\text{i} + \varepsilon_{1i} \\ y_{i2} &= \text{int}_i + 2\text{slp}_\text{i} + \varepsilon_{2i} \\ y_{i3} &= \text{int}_i + 3\text{slp}_\text{i} + \varepsilon_{3i} \\ y_{i4} &= \text{int}_i + 11\text{slp}_\text{i} + \varepsilon_{4i} \\ y_{i5} &= \text{int}_i + 13\text{slp}_\text{i} + \varepsilon_{5i} \\ \text{int}_i &= \mu_{\text{int}} + \alpha_1 x_{SEX,i} + \alpha_2 x_{RACE,i} + \alpha_3 x_{GRADE,i} + \alpha_4 x_{SES,i} + e_{\text{int},i} \\ \text{slp}_i &= \mu_{slp} + \alpha_5 x_{SEX,i} + \alpha_6 x_{RACE,i} + \alpha_7 x_{GRADE,i} + \alpha_8 x_{SES,i} + e_{\text{slp,i}} \end{aligned} \ . \qquad (5.5)$$

PROC CALIS routine in SAS 9.0 with the default maximum likelihood (ML) estimation is used to estimate the parameters of (5.5). The complete SAS codes and model estimation results are presented in Appendix F. The parameter estimates of the LGC model (5.5) are presented in Table 10. Since *t* values of SES and GRADE are less than 2, SES and GRADE are not significant factors in LGC model. Moreover, GFI=0.7827 (<0.90) indicates the fitness of (5.5) is not good. Removing the non-significant factors SES and GRADE from (5.5), we can rewrite the LGC model as (5.6). The corresponding model fitting results are listed in Table 11. It shows that the fitness of LGC model (5.6) (GFI=0.7532) is still not good (GFI<0.9).

$$\begin{aligned}
y_{i0} &= \mathrm{int}_i + \varepsilon_{0i} \\
y_{i1} &= \mathrm{int}_i + \mathrm{slp}_\mathrm{i} + \varepsilon_{1i} \\
y_{i2} &= \mathrm{int}_i + 2\mathrm{slp}_\mathrm{i} + \varepsilon_{2i} \\
y_{i3} &= \mathrm{int}_i + 3\mathrm{slp}_\mathrm{i} + \varepsilon_{3i} \\
y_{i4} &= \mathrm{int}_i + 11\mathrm{slp}_\mathrm{i} + \varepsilon_{4i} \\
y_{i5} &= \mathrm{int}_i + 13\mathrm{slp}_\mathrm{i} + \varepsilon_{5i} \\
\mathrm{int}_i &= \mu_{\mathrm{int}} + \alpha_1 x_{SEX,i} + \alpha_2 x_{RACE,i} + e_{\mathrm{int},i} \\
\mathrm{slp}_i &= \mu_{slp} + \alpha_5 x_{SEX,i} + \alpha_6 x_{RACE,i} + e_{\mathrm{slp,i}}
\end{aligned} \tag{5.6}$$

Table 10. Latent growth curve model (5.5) for HRWK

Fixed Effects	α_1	-12.3410 (1.008)	α_5	0.7800 (0.085)
		t=-12.2415		*t=9.1778*
	α_2	-4.1359 (1.736)	α_6	0.3886 (0.1464)
		t=-2.3825		*t=2.6550*
	α_3	-0.948 (0.6045)	α_7	0.0801(0.0510)
		t=-1.5683		*t=1.5715*
	α_4	-0.1774 (0.815)	α_8	0.0162 (0.0687)
		t=-0.2177		*t=0.2352*
	μ_{int}	24.3325 (1.5574)	μ_{slp}	-1.4464 (0.1313)
		t=15.6235		*t=-11.0162*

Note: Standard errors are given in parentheses, *t* represents t-value. The Chi-square test of model fit is $\chi^2(37) = 793.6933$ (p<0.0001); RMSEA=0.2166. GFI=0.7827.

Table 11. Latent growth curve model (5.6) for HRWK

Fixed Effects	α_1	-12.214 (1.006)	α_5	0.7694 (0.0848)
		T=-12.1418		*T=9.0717*
	α_2	-4.1860 (1.6540)	α_6	0.3936 (0.1394)
		t=-2.5309		*T=2.8226*
	μ_{int}	22.1817 (0.7359)	μ_{slp}	-1.2641 (0.062)
		T=-12.1418		*t=-12.1418*
Random Effects	σ^2_{int}	*82.7743*	σ^2_{slp}	*0.24355*

Note: Standard errors are given in parentheses, *t* represents t-value. The Chi-square test of model fit is $\chi^2(29) = 771.5817$ (p<0.0001); RMSEA=0.2423. GFI=0.7532.

5.3 COMPARISIONS

5.3.1 Random Coefficient Model and LGC Model

From Table 9 and Table 11, one can see that parameter estimates of the LGC model (5.6) and random coefficient model (5.3) are not the same. The major difference is that GRADE and its interaction with time are significant factors in the random coefficient model (5.3), while GRADE has no significant effects on the slope and intercept in the LGC model (5.6).

Except the factor GRADE, fixed effects of these two models are actually quite similar. At least these two models share the same objectives and have similar model representations. Therefore, they can lead to many similar conclusions. Both models show that there are physical activity declining from childhood to adolescent. The initial LHRWK values at Year 1990 and the initial decline rate are quite close. The factors SEX and RACE play the significant roles in both models. Specially, SEX in these two models has significantly negative effects on the intercept, leading to the conclusion that Male participate more physical activity than the Female (since the initial LHRWK value of Male is larger than that of Female). SEX in these two models has significantly positive effects on the growth rate, indicating that the Male has larger physical activity decline rate than the Female. Similarly, both these two models disclose that the White attends more physical activity at Year 1990 (highest initial values of LHRWK) than other races, and have the largest physical decline rate (smallest negative slope values).

The fitness value GFI in Table 11 is only 0.7532, less than the standard value 0.90. It discloses the LGC model (5.6) is not a good fitting model to characterize observation HRWK. As explained in Section 2.2, random coefficient model does not assume time-structure data and has the advantage to handle missing data. Therefore, random coefficient model (5.3) can better disclose the characteristics of physical activity change.

5.3.2 Random Coefficient Model Analysis for First Four Years Data

For the non-uniform acquired PittPAS data, random coefficient model can better disclose the decline trend of physical activity from adolescence to adulthood than LGC model. Here, we further compare the physical activity decline trend characterized by (5.3) with the decline trend happened in the first four years. Fitting random coefficient model (4.3) into the first four years data in Chapter 4, we can find factors a_{SES}, $a_{SES,T}$, $a_{GRADE,T}$, and $a_{RACE,T}$ in (4.3) are all non-significant factors (*p>0.05*) as given in Appendix D. Removing these non-significant factors, we can then rewrite (4.3) as

$$\begin{aligned} Y_{it} = {} & u_{\text{int}} + e_{\text{int},i} + u_{slp} Time_{it} + e_{slp,i} Time_{it} + a_{SEX} X_{SEX,i} + a_{RACE} X_{RACE,i} + a_{GRADE} X_{GRADE,i} \\ & + a_{SEX,T} X_{SEX,i} Time_{it} + \varepsilon_{it} \end{aligned} . \qquad (5.7)$$

The model parameter estimates of (5.7) are listed in Table 12. The complete fit statistic is also included in Appendix G.

Comparing Table 9 and Table 12, we can see the initial HRWK values in (5.3) and (5.7) are quite close to each other. This should be true because μ_{int} in both models represent the mean HRWK values measured in Year 1990. Moreover, the decline rate of (5.3) and (5.7) is also quite close. Both models show that SEX has very distinct effects on the initial HRWK values and the decline trend of HRWK in both long periods and short periods. RACE and GRADE has strong effects on the initial HRWK values (i.e., physical activity in Year 1990), while has little effects on the decline of HRWK in the short time period as given in Table 12. Their effects reflect on the physical activity trend only when the long period data are evaluated (as presented in Table 9). For both models, SES is not significant, which means SES has little effects on the initial HRWK values and the change of HRWK in all interview periods.

Table 12. The parameter estimates of random coefficient model (5.7)

Parameter	Estimate	Parameter	Estimate
Fixed effects			
a_{SEX}	-14.4020 (0.9472)	a_{RACE}	-2.0282 (0.9960)
a_{GRADE}	-1.5009 (0.4533)	$a_{SEX,T}$	1.8739 (0.3794)
μ_{int}	28.6556 (1.1482)	μ_{slp}	-3.5166 (0.2611)
Random effects			
σ^2_{int}	-506E-23 (0.0026)	σ^2_{slp}	3.35E-20 (0.0026)
$\sigma^2_{slp,\text{int}}$	0	σ^2_{ε}	128.49

Note: Standard errors are given in parentheses. -2*Res Log Likelihood=22855.4, AIC=22865.4.

6.0 CONCLUSIONS

Random coefficient model and latent growth model are both popular and useful tools to study longitudinal data. These two models are highly similar. In this thesis, similarities and differences between random coefficient model and latent growth curve regarding growth curve analysis are investigated and illustrated. Both models share the same objectives and have similar model specifications. They have almost the same model estimation results and performance for regular measured ideal data (for example, no missing data). However, they are in essence different methods. As a multivariate approach, the LGC model treats the repeated measurements as different variables thus it has more flexibilities than random coefficient model, which is a univariate method. Such flexibility mainly reflects in the following aspects: (1) estimating the time factors to investigate nonlinear growth curves, and (2) incorporating the growth curve model in a larger and more flexible structural model. The major drawback of LGC model is that the model requires the number of measurement occasions and spacing to be the same for all subjects. As the univariate approach, random coefficient model has no such time-structured data requirements. It allows each subject in the data set can be assessed at a different number of measurement occasions with randomly assigned temporal spacing. For example, random coefficient model allows the explanatory variable ‘time’ to take on different values for each individual. Therefore, random coefficient model is more suitable to analyze the data with a large amount of missing observations and non-uniform observation time.

To illustrate the above similarities and differences between LGC model and random coefficient model, we apply both models to study the natural history of participation in leisure time in physical activity from adolescence to young adulthood. Firstly, we use LGC and random coefficient models to characterize physical activity time spent-per-week (HRWK) at the first four interview years. Using exactly the same complete data, both models yield almost the same model representations, which show that these two models ideally have identical performance. As described in Chapter 2, the normality of HRWK can be improved by applying *log* operator to HRWK. However, direct applying *log* operator makes zero-value HRWK data to be missing. Thus, we evaluate three random coefficient models for HRWK, LHRWK, and rHRWK, separately. Our sensitive analysis shows that the two models for HRWK and LHRWK disclose the similar characteristics of physical activity decline. Therefore, random coefficient model for HRWK is finally used to analyze the change of HRWK. Considering there is a large percentage missing data (71.8%) in the LGC analysis and irregular time intervals, we show that random coefficient model is more appropriate to study PittPAS data than LGC model. Our random coefficient model analysis of HRWK leads to the following conclusions:

1. The leisure time in physical activity declines from adolescence to young adulthood.
2. SES factor is not a significant factor on the decline of physical activity.
3. SEX plays the most significant roles in the decline of physical activity. The Male has larger decline rate than the Female students.
4. White students spent more time in activity per week. But they also have the largest decline rate.
5. The lower grade students spent more time in physical activity per week in Year 1990, and they also have the largest decline rate than other grade students.

There are still some aspects to further improve our analysis results. First, we assume the growth curve of HRWK has the linear growth pattern. Actually, the non-linear pattern property of HRWK has been demonstrated in Figure 8. Quadratic polynomial or other non-linear patterns could be used to further improve the goodness of modeling fitting. Secondly, PROC MIXED and CALIS procedures assume that random variables have an approximately multivariate normal distribution. Therefore, non-normality especially high kurtosis would result in poor estimates and incorrect standard errors, even in large samples. PROC MIXED is stricter on the normal distribution assumption. HRWK in our models is not approximately normally distributed. Therefore, we should further improve our data quality for better estimates. Thirdly, but not the least, the analysis could be more accurate if we could combine some features of random coefficient model and LGC model by using more advanced software package such as Mplus (Muthen B, 2000).

APPENDIX A DATA SET DESCRIPTION

Data set-variable codes

Variable	code	description
ID	7001-9818	integer
Age	12-16	integer
Sex	0	Male
	1	Female
Race	0	White
	1	Black
Ses	0	High SES
	1	Middle SES
	2	Low SES
EASYEX90-EASYEX93		
	1	None
	2	1 to 2 days
	3	3 to 5 days
	4	6 to 8 days
	5	9 or more days
HARDEX90- HARDEX93		
	1	None
	2	1 hour or less
	3	2 to 3 hours
	4	4 to 5 hours
	5	6 or more hours
TV90-TV93	1	None
	2	1 hour or less
	3	2 to 3 hours
	4	4 to 5 hours
	5	6 or more hours

APPENDIX B COMMON SAS CODE

```
title1   "PACS Project" ;

libname   pac      "c:\pittsburgh\pacs";
filename  dfile    "d:\pittsburgh\pacs\baseline_demographics.xls";
run ;


/* data prepration*/
title1   "PACS Project";

PROC IMPORT OUT= WORK.Demo2
            DATAFILE= dfile
            DBMS=EXCEL REPLACE ;
run ;

proc sort data=pac.demo;
by id;
run;


proc sort data=pac.totalhours;
by id;
run;

proc sort data=demo2;
by id;
run;

data pac.newdemo;
merge pac.demo demo2;
by id;
run;


data round1;
set pac.round1;
keep id date1;
run;

proc sort data=round1;
by id;
run;

data round2;
```

```
set pac.round2;
keep id date2;
run;

proc sort data=round2;
by id;
run;


proc format  ;

      value     hsex
             1 = "Female"
             0 = "Male"
             ;
     value     hrace
             0 = "White"
             1 = "Black"
             ;
     value     hses
             0 = "High SES"
             1 = "Middle SES"
             2 = "Low SES"
             ;
 run;

/* add label to the data */
data pac.all;
merge pac.newdemo pac.totalhours round1 round2;
by id;
length grade 4;
*grade=suaiybstr(id,9,1);
if race>=3 then race=.;
time1=0;
time2=1;
time3=2;
time4=3;
time5=year(date1)-1990;
time6=year(date2)-1990;
if sex=1 then sex=0;
if sex=2 then sex=1;
if race=1 then race=0;
if race=2 then race=1;
if ses=1 then ses=0;
if ses=2 then ses=1;
if ses=3 then ses=2;

format sex hsex.;
format race hrace.;
format ses hses.;
run;


data pac.hrwk;
set pac.all;
keep id race sex  ses grade hrwk90 hrwk91 hrwk92 hrwk93 hrwkr1 hrwkr2
 time1 time2 time3 time4 time5 time6;
```

```
run;


data pac.loghrwk ;
set pac.hrwk ;

logHRWK90=log(HRWK90);
logHRWK91=log(HRWK91);
logHRWK92=log(HRWK92);
logHRWK93=log(HRWK93);
logHRWKR1=log(HRWKR1);
logHRWKR2=log(HRWKR2);

label logHRWK90='LHRWK90';
label logHRWK91='LHRWK91';
label logHRWK92='LHRWK92';
label logHRWK93='LHRWK93';
label logHRWKR1='LHRWKR1';
label logHRWKR2='LHRWKR2';
run;


data pac.ologhrwk ;
set pac.hrwk ;

logHRWK90=log(HRWK90+0.0005);
logHRWK91=log(HRWK91+0.0005);
logHRWK92=log(HRWK92+0.0005);
logHRWK93=log(HRWK93+0.0005);
logHRWKR1=log(HRWKR1+0.0005);
logHRWKR2=log(HRWKR2+0.0005);

label logHRWK90='LHRWK90';
label logHRWK91='LHRWK91';
label logHRWK92='LHRWK92';
label logHRWK93='LHRWK93';
label logHRWKR1='LHRWKR1';
label logHRWKR2='LHRWKR2';
 run;
/*******For 6 years data***********************/
/****************** HRWK data***********/
data pac.mixed ;
set pac.hrwk ;

array lognnh(*)  HRWK90    HRWK91    HRWK92    HRWK93    HRWKR1    HRWKR2 ;
array new(*)      time1 time2 time3 time4 time5 time6;
   do i=1 to dim(lognnh) ;
     a = i;
                total_hrs         = lognnh(i)  ;
                 time =new(i);
    output ;
   end ;
label total_hrs         = 'Total hr/wk' ;
label time         = 'Interview time' ;
keep  id   time     total_hrs
      sex  race   ses  grade time ;
run ;
```

```
 data pac.mixed;
 set pac.mixed;
 tsex=time*sex;
tses=time*ses;
trace=time*race;
tgrade=time*grade;
 run;
/***************LHRWK without 0 data*****************/

data pac.logmixed ;
set pac.loghrwk ;
array lognnh(*)  logHRWK90    logHRWK91    logHRWK92    logHRWK93
logHRWKR1    logHRWKR2 ;
array new(*)      time1 time2 time3 time4 time5 time6;
   do i=1 to dim(lognnh) ;
     a = i;
                log_total_hrs         = lognnh(i)  ;
                 time =new(i);
    output ;
   end ;
label log_total_hrs         = 'Log of Total hr/wk' ;
label time         = 'Interview time' ;
keep  id   time     log_total_hrs
      sex  race   ses  grade time;
run ;


data pac.logmixed ;
set pac.logmixed  ;
tsex=time*sex;
tses=time*ses;
trace=time*race;
tgrade=time*grade;
run;





/***************************rLOGHRWK with  0  ********************/

data pac.ologmixed ;
set pac.ologhrwk ;

array lognnh(*)  logHRWK90    logHRWK91    logHRWK92    logHRWK93
logHRWKR1    logHRWKR2 ;
array new(*)      time1 time2 time3 time4 time5 time6;
   do i=1 to dim(lognnh) ;
     a = i;
                log_total_hrs         = lognnh(i)  ;
                 time =new(i);
    output ;
   end ;
label log_total_hrs         = 'Log of Total hr/wk' ;
label time         = 'Interview time' ;
```

```
keep  id   time     log_total_hrs
      sex  race   ses  grade time;
run ;


data pac.ologmixed ;
set pac.ologmixed;
tsex=time*sex;
tses=time*ses;
trace=time*race;
tgrade=time*grade;
run;

data pac.try_LGC1(rename=(HRWK90=y1 HRWK91=y2 HRWK92=y3 HRWK93=y4 hrwkr1=y5
hrwkr2=y6));
set pac.hrwk;
keep id sex  RACE  ses grade  HRWK90 HRWK91 HRWK92 HRWK93 hrwkr1 hrwkr2 ;
run;
/***************For 4 years data ********************/


/***********************************No missing */

data pac.nomiss_hrwk;
set pac.hrwk;
if sex=. or ses=. or race=. or grade=. or hrwk90=. or hrwk91=. or hrwk92=. or
hrwk93=. then delete ;
run;


/* PROC MIXED*/
/*   Restructuring the data for  MIXED procedures   */
data pac.nomiss_try_mixed ;
set pac.nomiss_hrwk ;
array lognnh(*)  HRWK90    HRWK91    HRWK92    HRWK93    ;
array new(*)      time1 time2 time3 time4 ;
   do i=1 to dim(lognnh) ;
     a = i;
                total_hrs         = lognnh(i)  ;
                 time =new(i);
    output ;
   end ;
label total_hrs         = ' Total hr/wk' ;
label time          = 'Interview time' ;
keep  id   time     total_hrs
      sex  race   ses  grade time ;
run ;

data pac.nomiss_try_mixed;
set pac.nomiss_try_mixed;
tsex=time*sex;
tses=time*ses;
trace=time*race;
tgrade=time*grade;
run;

/****************Graph*********************************/
```

```
 /*******************4 year HRWK*********************/
 proc univariate data=pac.hrwk noprint ;
var  HRWK90 HRWK91 HRWK92 HRWK93  ;
output out=make_race mean=mean1990 mean1991 mean1992 mean1993   ;
run;

proc transpose data=make_race out=make2(rename=(col1=mean_HRWK));
run;


proc gplot data=make2;
symbol6 value=dot h=1 interpol=spline  color=red;
axis1  label=('Total mean hour/week of' justify=right 'activities');
 plot mean_HRWK*_NAME_/vaxis=axis1;
run;

/***********6 year HRWK***************************/

proc univariate data=pac.hrwk noprint ;
var  HRWK90 HRWK91 HRWK92 HRWK93 HRWKR1 HRWKR2 ;
output out=make_race mean=mean1990 mean1991 mean1992 mean1993   meanR1 meanR2;
run;

proc transpose data=make_race out=make2(rename=(col1=mean_HRWK));
run;

data make2;
set make2;
if _NAME_="mean1990" THEN time=1;
if _NAME_="mean1991" THEN time=2;
if _NAME_="mean1992" THEN time=3;
if _NAME_="mean1993" THEN time=4;
if _NAME_="meanR1" THEN time=11;
if _NAME_="meanR2" THEN time=13;
run;

proc gplot data=make2;
symbol6 value=dot h=1 interpol=spline  color=red;
axis1  label=('Total mean hour/week of' justify=right 'activities');
title"Six years data";
 plot mean_HRWK*time/vaxis=axis1;
run;

/***********6 year logHRWK***************************/

proc univariate data=pac.loghrwk noprint ;
var  logHRWK90 logHRWK91 logHRWK92 logHRWK93 logHRWKR1 logHRWKR2 ;
output out=make_race mean=mean1990 mean1991 mean1992 mean1993   meanR1 meanR2;
run;

proc transpose data=make_race out=make2(rename=(col1=mean_HRWK));
run;

data make2;
set make2;
if _NAME_="mean1990" THEN time=1;
if _NAME_="mean1991" THEN time=2;
```

```
if _NAME_="mean1992" THEN time=3;
if _NAME_="mean1993" THEN time=4;
if _NAME_="meanR1" THEN time=11;
if _NAME_="meanR2" THEN time=13;
run;

proc gplot data=make2;
symbol6 value=dot h=1 interpol=spline  color=red;
axis1  label=('Total mean hour/week of' justify=right 'activities');
title"Six years data logHRWK";
 plot mean_HRWK*time/vaxis=axis1;
run;

/***********6 year rlogHRWK***************************/
proc univariate data=pac.ologhrwk noprint ;
var  logHRWK90 logHRWK91 logHRWK92 logHRWK93 logHRWKR1 logHRWKR2 ;
output out=make_race mean=mean1990 mean1991 mean1992 mean1993   meanR1 meanR2;
run;

proc transpose data=make_race out=make2(rename=(col1=mean_HRWK));
run;

data make2;
set make2;
if _NAME_="mean1990" THEN time=1;
if _NAME_="mean1991" THEN time=2;
if _NAME_="mean1992" THEN time=3;
if _NAME_="mean1993" THEN time=4;
if _NAME_="meanR1" THEN time=11;
if _NAME_="meanR2" THEN time=13;
run;

proc gplot data=make2;
symbol6 value=dot h=1 interpol=spline  color=red;
axis1  label=('Total mean hour/week of' justify=right 'activities');
title"Six years data rlogHRWK";
 plot mean_HRWK*time/vaxis=axis1;
run;

/****************HRWK distr*************/
proc univariate data=pac.hrwk  noprint;
    var HRWK90 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
                vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;
proc univariate data=pac.hrwk  noprint;
    var HRWK91 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
                vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;
proc univariate data=pac.hrwk  noprint;
    var HRWK92 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
                vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
```

```
run;
proc univariate data=pac.hrwk  noprint;
    var HRWK93 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
               vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;
proc univariate data=pac.hrwk  noprint;
    var HRWKr1 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
               vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;

proc univariate data=pac.hrwk  noprint;
    var HRWKr2 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
               vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;



/*******loghrwk distr***************/
proc univariate data=pac.loghrwk  noprint;
    var logHRWK90 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
               vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;
proc univariate data=pac.loghrwk  noprint;
    var logHRWK91 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
               vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;
proc univariate data=pac.loghrwk  noprint;
    var logHRWK92 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
               vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;
proc univariate data=pac.loghrwk  noprint;
    var logHRWK93 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
               vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;
proc univariate data=pac.loghrwk  noprint;
    var logHRWKr1 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
               vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;

proc univariate data=pac.loghrwk  noprint;
    var logHRWKr2 ;
```

```
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
                vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;



/*************rloghrwk distribution************************/
proc univariate data=pac.ologhrwk  noprint;
    var logHRWK90 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
                vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;
proc univariate data=pac.ologhrwk  noprint;
    var logHRWK91 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
                vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;
proc univariate data=pac.ologhrwk  noprint;
    var logHRWK92 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
                vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;
proc univariate data=pac.ologhrwk  noprint;
    var logHRWK93 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
                vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;
proc univariate data=pac.ologhrwk  noprint;
    var logHRWKr1 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
                vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;

proc univariate data=pac.ologhrwk  noprint;
    var logHRWKr2 ;
       histogram/ normal(noprint) intertile=6 cfill=cyan vscale=count
                vaxislabel='frequency ';
      title 'Distribution of  hours in activity per week';
run;
```

APPENDIX C SAS CODE FOR CHAPTER 4

C.1 PROC CALIS CODE FOR LGC MODEL (4.2)

```
proc calis  data= pac.nomiss_try_lgc1 ucov aug ;
 lineqs
 y1 = F1 + 0F2 + E0,
 y2 = F1 + 1F2 + E1,
 y3 = F1 + 2F2 + E2,
 y4 = F1 + 3F2 + E3,
 F1 = al1 intercept + gamma1 sex +gamma2 RACE + gamma3 ses +gamma4 grade + d1,
 F2 = al2 intercept + gamma5 sex + gamma6 race+ gamma7 ses+gamma8 grade + d2;
 std
 E0-E3= error0 error0 error0 error0  ,
 D1-D2=int slp;
 cov d1-d2=cor1;
 run;
```

C.2 PROC MIXED CODE FOR RANDOM COEFFICIENT MODEL (4.3)

```
PROC MIXED DATA=PAC.NOMISS_TRY_MIXED METHOD=REML;

model total_hrs= sex race  grade ses time tses tsex trace tgrade /s
ddfm=satterth;
random int time/ type=un s G;
repeated /type=cs subject=id ;
run;
```

APPENDIX D ANALYSIS RESULTS OF CHAPTER 4

D.1 ANALYSIS RESULTS OF LGC MODEL (4.2)

```
     The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Fit Function                                            0.4145
Goodness of Fit Index (GFI)                             0.9177
GFI Adjusted for Degrees of Freedom (AGFI)              0.7684
Root Mean Square Residual (RMR)                        14.2004
Parsimonious GFI (Mulaik, 1989)                         0.4078
Chi-Square                                            296.3881
Chi-Square DF                                               16
Pr > Chi-Square                                         <.0001
Independence Model Chi-Square                           5763.8
Independence Model Chi-Square DF                            36
RMSEA Estimate                                          0.1566
RMSEA 90% Lower Confidence Limit                        0.1412
RMSEA 90% Upper Confidence Limit                        0.1724
ECVI Estimate                                           0.4542
ECVI 90% Lower Confidence Limit                         0.4021
ECVI 90% Upper Confidence Limit                         0.5595
Probability of Close Fit                                0.0000
Bentler's Comparative Fit Index                         0.9510
Normal Theory Reweighted LS Chi-Square                288.7251
Akaike's Information Criterion                        264.3881
Bozdogan's (1987) CAIC                                175.2092
Schwarz's Bayesian Criterion                          191.2092
McDonald's (1989) Centrality                            0.8222
Bentler & Bonett's (1980) Non-normed Index              0.8899
Bentler & Bonett's (1980) NFI                           0.9486
James, Mulaik, & Brett (1982) Parsimonious NFI          0.4216
Z-Test of Wilson & Hilferty (1931)                     14.0841
Bollen (1986) Normed Index Rho1                         0.8843
Bollen (1988) Non-normed Index Delta2                   0.9512
Hoelter's (1983) Critical N                                 65
```

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Latent Variable Equations with Estimates

```
F1        = -14.4078*SEX       + -0.8937*RACE      +  0.6462*SES       + -1.1513*GRADE
Std Err       1.1788 gamma1       1.7016 gamma2       0.9728 gamma3       0.7055 gamma4
t Value     -12.2224            -0.5252             0.6643             -1.6319

                     + 27.2362*Intercept +  1.0000 d1
                        1.8072 al1
                       15.0712

F2        =   1.8742*SEX       + -0.8316*RACE      + -0.3266*SES       + -0.2322*GRADE
Std Err       0.4371 gamma5       0.6310 gamma6       0.3608 gamma7       0.2616 gamma8
t Value       4.2875            -1.3179            -0.9052            -0.8877

                     + -2.6418*Intercept +  1.0000 d2
                        0.6701 al2
                       -3.9422
```

Variances of Exogenous Variables

Variable	Parameter	Estimate	Standard Error	t Value
SEX		0.47413		
RACE		0.17483		
SES		1.10070		
GRADE		4.70909		
Intercept		1.00140		
E0	error0	107.18951	4.00866	26.74
E1	error0	107.18951	4.00866	26.74
E2	error0	107.18951	4.00866	26.74
E3	error0	107.18951	4.00866	26.74
d1	int	172.35696	13.38160	12.88
d2	slp	12.58069	1.96974	6.39

D.2 ANALYSIS RESULTS OF RANDOM COEFFICIENT MODEL (4.3)

Estimated G Matrix

Row	Effect	Col1	Col2
1	Intercept	6.735E-6	
2	time		2.6096

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)		6.735E-6
UN(2,1)		0
UN(2,2)		2.6096
CS	ID	70.2069
Residual		128.28

Fit Statistics

-2 Res Log Likelihood	22849.1
AIC (smaller is better)	22859.1
AICC (smaller is better)	22859.1
BIC (smaller is better)	22849.1

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
4	417.68	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	27.2362	1.4533	1	18.74	0.0339
SEX	-14.4078	0.9480	1574	-15.20	<.0001
RACE	-0.8937	1.3684	1574	-0.65	0.5138
GRADE	-1.1513	0.5673	1574	-2.03	0.0426
SES	0.6462	0.7824	1574	0.83	0.4089
time	-2.6418	1.7171	2854	-1.54	0.1240
tses	-0.3266	0.3133	2143	-1.04	0.2974
tsex	1.8742	0.3796	2143	4.94	<.0001
trace	-0.8316	0.5480	2143	-1.52	0.1293
tgrade	-0.2322	0.2272	2143	-1.02	0.3068

The Mixed Procedure

Solution for Random Effects

Effect	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Intercept	-311E-22	0.002595	1	-0.00	1.0000
time	-101E-16	1.6154	2854	-0.00	1.0000

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
SEX	1	1574	230.98	<.0001
RACE	1	1574	0.43	0.5138
GRADE	1	1574	4.12	0.0426
SES	1	1574	0.68	0.4089
time	1	2854	2.37	0.1240
tses	1	2143	1.09	0.2974
tsex	1	2143	24.37	<.0001
trace	1	2143	2.30	0.1293
tgrade	1	2143	1.04	0.3068

APPENDIX E SAS CODE FOR CHAPTER 5

E.1 PROC MIXED CODE FOR RANDOM COEFFICIENT MODEL (5.1)

```
proc mixed data=pac.mixed METHOD=reML;
model total_hrs= sex race  grade ses time tses tsex trace tgrade /s
ddfm=satterth;
random int time/ type=un s G;
repeated /type=CS subject=id ;
run;

proc mixed data=pac.logmixed METHOD=reML;
model log_total_hrs= sex race  grade ses time tses tsex trace tgrade /s
ddfm=satterth;
random int time/ type=un s G;
repeated /type=CS subject=id ;
run;

proc mixed data=pac.ologmixed METHOD=reML;
model log_total_hrs= sex race  grade ses time tses tsex trace tgrade /s
ddfm=satterth;
random int time/ type=un s G;
repeated /type=CS subject=id ;
run;
```

E.2 PROC MIXED CODE FOR RANDOM COEFFICIENT MODEL (5.2)

```
proc mixed data=pac.mixed METHOD=reML;
model total_hrs= sex race  grade  time  tsex trace  tgrade /s ddfm=satterth;
random int time/ type=un s G;
repeated /type=CS subject=id ;
run;
```

E.3 PROC CALIS CODE FOR LATENT GROWTH MODEL (5.5)

```
proc calis  data= pac.try_lgc1 ucov aug ;
 lineqs
 y1 = F1 + 0F2 + E0,
 y2 = F1 + 1F2 + E1,
 y3 = F1 + 2F2 + E2,
 y4 = F1 + 3F2 + E3,
 y5 = F1 + 11F2 +E4,
 y6=  F1 + 13F2+ E5,
 F1 = al1 intercept + gamma1 sex +gamma2 RACE + gamma3 ses +gamma4 grade + d1,
 F2 = al2 intercept + gamma5 sex + gamma6 race+ gamma7 ses+gamma8 grade + d2;
 std
 E0-E5= error0 error0 error0 error0 error0 error0,
 D1-D2=int slp;
 cov d1-d2=cor1;
 run;
```

E.4 PROC CALIS CODE FOR LATENT GROWTH MODEL (5.6)

```
proc calis  data= pac.try_lgc1 ucov aug ;
 lineqs
 y1 = F1 + 0F2 + E0,
 y2 = F1 + 1F2 + E1,
 y3 = F1 + 2F2 + E2,
 y4 = F1 + 3F2 + E3,
 y5 = F1 + 11F2 +E4,
 y6=  F1 + 13F2+ E5,
 F1 = al1 intercept + gamma1 sex +gamma2 RACE + d1,
 F2 = al2 intercept + gamma5 sex + gamma6 race+  d2;
 std
 E0-E5= error0 error0 error0 error0 error0 error0,
 D1-D2=int slp;
 cov d1-d2=cor1;
 run;
```

E.5 PROC MIXED CODE FOR RANDOM COEFFICIENT MODEL (5.7)

```
proc mixed data=pac.nomiss_try_mixed METHOD=ReML;
model total_hrs= sex race  grade time tsex    /s ddfm=satterth;
random int time/ type=un s G; repeated /type=cs subject=id ; run;
```

APPENDIX F ANALYSIS RESULTS OF CHAPTER 5

F.1 ANALYSIS RESULTS OF RANDOM COEFFICIENT MODEL (5.1)

The Mixed Procedure

Estimated G Matrix

Row	Effect	Col1	Col2
1	Intercept	0.4337	0.1157
2	time	0.1157	0.5001

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)		0.4337
UN(2,1)		0.1157
UN(2,2)		0.5001
CS	ID	50.9086
Residual		136.28

Fit Statistics

-2 Res Log Likelihood	42914.0
AIC (smaller is better)	42924.0
AICC (smaller is better)	42924.0
BIC (smaller is better)	42914.0

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
4	449.62	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|

Intercept	27.5203	1.1465	3683	24.00	<.0001
SEX	-13.3177	0.6003	1678	-22.19	<.0001
RACE	-2.1792	0.7915	1761	-2.75	0.0060
GRADE	-2.1019	0.3648	1662	-5.76	<.0001
SES	-0.08532	0.4941	1665	-0.17	0.8629
time	-1.6474	0.7153	1	-2.30	0.2608
tses	0.001897	0.05666	4564	0.03	0.9733
tsex	0.9294	0.07063	4551	13.16	<.0001
trace	0.2386	0.1012	4755	2.36	0.0185
tgrade	0.1227	0.04261	4537	2.88	0.0040

The Mixed Procedure

Solution for Random Effects

Effect	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Intercept	-246E-18	0.6586	5376	-0.00	1.0000
time	-197E-16	0.7072	1	-0.00	1.0000

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
SEX	1	1678	492.25	<.0001
RACE	1	1761	7.58	0.0060
GRADE	1	1662	33.21	<.0001
SES	1	1665	0.03	0.8629
time	1	1	5.30	0.2608
tses	1	4564	0.00	0.9733
tsex	1	4551	173.12	<.0001
trace	1	4755	5.55	0.0185
tgrade	1	4537	8.29	0.0040

j=1

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)		0.005373
UN(2,1)		-0.00006
UN(2,2)		0.01101
CS	ID	0.3581
Residual		0.7459

Fit Statistics

-2 Res Log Likelihood	14163.3

AIC (smaller is better)	14173.3
AICC (smaller is better)	14173.3
BIC (smaller is better)	14163.3

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
4	611.40	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	3.1992	0.1052	1	30.42	0.0209
SEX	-1.0547	0.04837	1655	-21.81	<.0001
RACE	-0.2307	0.06387	1744	-3.61	0.0003
GRADE	-0.1566	0.02940	1637	-5.33	<.0001
SES	-0.02430	0.03969	1625	-0.61	0.5405
time	-0.1271	0.1053	5030	-1.21	0.2274
tses	0.000062	0.004322	4225	0.01	0.9886
tsex	0.04880	0.005398	4233	9.04	<.0001
trace	0.02466	0.007883	4425	3.13	0.0018
tgrade	0				

The Mixed Procedure

Solution for Random Effects

Effect	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Intercept	1.22E-15	0.07330	1	0.00	1.0000
time	9.9E-15	0.1049	5030	0.00	1.0000

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
SEX	1	1655	475.54	<.0001
RACE	1	1744	13.05	0.0003
GRADE	1	1637	28.37	<.0001
SES	1	1625	0.37	0.5405
time	1	5030	1.46	0.2274
tses	1	4225	0.00	0.9886
tsex	1	4233	81.73	<.0001
trace	1	4425	9.78	0.0018
tgrade	1	4227	8.54	0.0035

j=2

Covariance Parameter Estimates

Cov Parm	Subject	Estimate

UN(1,1)		0.01494
UN(2,1)		0.003971
UN(2,2)		0.01718
CS	ID	1.4575
Residual		4.6802

Fit Statistics

-2 Res Log Likelihood	24662.9
AIC (smaller is better)	24672.9
AICC (smaller is better)	24672.9
BIC (smaller is better)	24662.9

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
4	397.57	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	3.7622	0.2067	4350	18.20	<.0001
SEX	-1.8179	0.1066	1867	-17.05	<.0001
RACE	-0.4042	0.1408	1959	-2.87	0.0041
GRADE	-0.4467	0.06478	1849	-6.89	<.0001
SES	-0.2472	0.08776	1853	-2.82	0.0049
time	-0.1942	0.1325	1	-1.47	0.3812
tses	0.01581	0.01047	4666	1.51	0.1311
tsex	0.09640	0.01305	4655	7.39	<.0001
trace	-0.00490	0.01868	4852	-0.26	0.7933
tgrade	0.03110	0.007874	4641	3.95	<.0001

The Mixed Procedure

Solution for Random Effects

Effect	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Intercept	6.85E-16	0.1222	5376	0.00	1.0000
time	6.21E-16	0.1311	1	0.00	1.0000

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
SEX	1	1867	290.65	<.0001
RACE	1	1959	8.24	0.0041
GRADE	1	1849	47.54	<.0001
SES	1	1853	7.93	0.0049
time	1	1	2.15	0.3812
tses	1	4666	2.28	0.1311
tsex	1	4655	54.57	<.0001
trace	1	4852	0.07	0.7933
tgrade	1	4641	15.60	<.0001

F.2 ANALYSIS RESULTS OF RANDOM COEFFICIENT MODEL (5.2)

Cov Parm	Subject	Estimate
UN(1,1)		7.156E-6
UN(2,1)		5.291E-9
UN(2,2)		6.906E-6
CS	ID	50.8296
Residual		136.25

Fit Statistics

-2 Res Log Likelihood	42910.3
AIC (smaller is better)	42920.3
AICC (smaller is better)	42920.3
BIC (smaller is better)	42910.3

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
4	449.10	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	27.4589	0.8697	1659	31.57	<.0001
SEX	-13.3204	0.5998	1680	-22.21	<.0001
RACE	-2.2406	0.7071	1795	-3.17	0.0016
GRADE	-2.1020	0.3646	1664	-5.77	<.0001
time	-1.6459	0.1017	4557	-16.19	<.0001
tsex	0.9294	0.07050	4548	13.18	<.0001
trace	0.2400	0.09364	4803	2.56	0.0104
tgrade	0.1227	0.04256	4537	2.88	0.0040

Distribution of hours in activity per week 343
16:52 Wednesday, March 21, 2007

The Mixed Procedure

Solution for Random Effects

Effect	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Intercept	-332E-22	0.002675	5378	-0.00	1.0000
time	-959E-22	0.002628	5378	-0.00	1.0000

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
SEX	1	1680	493.25	<.0001
RACE	1	1795	10.04	0.0016
GRADE	1	1664	33.24	<.0001
time	1	4557	261.99	<.0001
tsex	1	4548	173.79	<.0001
trace	1	4803	6.57	0.0104
tgrade	1	4537	8.31	0.004

F.3 ANALYSIS RESULTS OF LGC MODEL (5.5)

The CALIS Procedure

Covariance Structure Analysis: Maximum Likelihood Estimation

Fit Function	1.8204
Goodness of Fit Index (GFI)	0.7827
GFI Adjusted for Degrees of Freedom (AGFI)	0.6125
Root Mean Square Residual (RMR)	30.5674
Parsimonious GFI (Mulaik, 1989)	0.5266
Chi-Square	793.6933
Chi-Square DF	37
Pr > Chi-Square	<.0001
Independence Model Chi-Square	4466.3
Independence Model Chi-Square DF	55
RMSEA Estimate	0.2166
RMSEA 90% Lower Confidence Limit	0.2036
RMSEA 90% Upper Confidence Limit	0.2298
ECVI Estimate	1.8864
ECVI 90% Lower Confidence Limit	1.7176
ECVI 90% Upper Confidence Limit	2.1436
Probability of Close Fit	0.0000
Bentler's Comparative Fit Index	0.8285
Normal Theory Reweighted LS Chi-Square	665.5793
Akaike's Information Criterion	719.6933
Bozdogan's (1987) CAIC	531.7357
Schwarz's Bayesian Criterion	568.7357
McDonald's (1989) Centrality	0.4207
Bentler & Bonett's (1980) Non-normed Index	0.7450
Bentler & Bonett's (1980) NFI	0.8223
James, Mulaik, & Brett (1982) Parsimonious NFI	0.5532
Z-Test of Wilson & Hilferty (1931)	23.0269
Bollen (1986) Normed Index Rho1	0.7358
Bollen (1988) Non-normed Index Delta2	0.8292
Hoelter's (1983) Critical N	30

The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Latent Variable Equations with Estimates

```
F1        = -12.3410*SEX      + -4.1359*RACE     + -0.1774*SES      + -0.9480*GRADE
Std Err       1.0081 gamma1      1.7360 gamma2      0.8150 gamma3      0.6045 gamma4
t Value     -12.2415            -2.3825            -0.2177            -1.5683

                    + 24.3325*Intercept +  1.0000 d1
                       1.5574 al1
                      15.6235

F2        =   0.7800*SEX      +  0.3886*RACE     +  0.0162*SES      +  0.0801*GRADE
Std Err       0.0850 gamma5      0.1464 gamma6      0.0687 gamma7      0.0510 gamma8
t Value       9.1778             2.6550             0.2352             1.5715

                    + -1.4464*Intercept +  1.0000 d2
                       0.1313 al2
                     -11.0162
```

Variances of Exogenous Variables

Variable	Parameter	Estimate	Standard Error	t Value
SEX		0.51835		
RACE		0.10321		
SES		0.89908		
GRADE		5.01147		
Intercept		1.00229		
E0	error0	82.98133	2.81010	29.53
E1	error0	82.98133	2.81010	29.53
E2	error0	82.98133	2.81010	29.53
E3	error0	82.98133	2.81010	29.53
E4	error0	82.98133	2.81010	29.53
E5	error0	82.98133	2.81010	29.53
d1	int	82.13777	7.46958	11.00
d2	slp	0.23900	0.05575	4.29

F.4 ANALYSIS RESULTS OF LGC MODEL (5.6)

```
       The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation

Fit Function                                          1.7697
Goodness of Fit Index (GFI)                           0.7532
GFI Adjusted for Degrees of Freedom (AGFI)            0.6170
Root Mean Square Residual (RMR)                      36.9917
Parsimonious GFI (Mulaik, 1989)                       0.6067
Chi-Square                                          771.5817
Chi-Square DF                                             29
Pr > Chi-Square                                       <.0001
Independence Model Chi-Square                         3197.8
Independence Model Chi-Square DF                          36
RMSEA Estimate                                        0.2423
RMSEA 90% Lower Confidence Limit                      0.2277
RMSEA 90% Upper Confidence Limit                      0.2573
ECVI Estimate                                         1.8166
ECVI 90% Lower Confidence Limit                       1.6294
ECVI 90% Upper Confidence Limit                       2.0495
Probability of Close Fit                              0.0000
Bentler's Comparative Fit Index                       0.7651
Normal Theory Reweighted LS Chi-Square              643.0148
Akaike's Information Criterion                      713.5817
Bozdogan's (1987) CAIC                              566.2636
Schwarz's Bayesian Criterion                        595.2636
McDonald's (1989) Centrality                          0.4276
Bentler & Bonett's (1980) Non-normed Index            0.7085
Bentler & Bonett's (1980) NFI                         0.7587
James, Mulaik, & Brett (1982) Parsimonious NFI        0.6112
Z-Test of Wilson & Hilferty (1931)                   22.7675
Bollen (1986) Normed Index Rho1                       0.7005
Bollen (1988) Non-normed Index Delta2                 0.7657
Covariance Structure Analysis: Maximum Likelihood Estimation

          Latent Variable Equations with Estimates


F1        = -12.2140*SEX      + -4.1860*RACE      + 22.1819*Intercept +  1.0000 d1
Std Err       1.0060 gamma1       1.6540 gamma2       0.7359 al1
t Value     -12.1418             -2.5309             30.1413
F2        =   0.7694*SEX      +  0.3936*RACE      + -1.2641*Intercept +  1.0000 d2
Std Err       0.0848 gamma5       0.1394 gamma6       0.0620 al2
t Value       9.0717              2.8226            -20.3734


              Variances of Exogenous Variables

                                          Standard
Variable  Parameter      Estimate          Error    t Value

SEX                       0.51835
RACE                      0.10321
Intercept                 1.00229
```

E0	error0	82.98133	2.81010	29.53
E1	error0	82.98133	2.81010	29.53
E2	error0	82.98133	2.81010	29.53
E3	error0	82.98133	2.81010	29.53
E4	error0	82.98133	2.81010	29.53
E5	error0	82.98133	2.81010	29.53
d1	int	82.77438	7.51237	11.02
d2	slp	0.24356	0.05604	4.35

F.5 ANALYSIS RESULTS OF RANDOM COEFFICIENT MODEL (5.7)

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)		6.746E-6
UN(2,1)		0
UN(2,2)		6.746E-6
CS	ID	70.0200
Residual		128.49

Fit Statistics

-2 Res Log Likelihood	22855.4
AIC (smaller is better)	22865.4
AICC (smaller is better)	22865.4
BIC (smaller is better)	22855.4

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
4	416.03	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	28.6556	1.1482	907	24.96	<.0001
SEX	-14.4020	0.9472	1577	-15.20	<.0001
RACE	-2.0282	0.9960	712	-2.04	0.0421
GRADE	-1.5009	0.4533	712	-3.31	0.0010
time	-3.5166	0.2611	1	-13.47	0.0472
tsex	1.8739	0.3794	2146	4.94	<.0001

Solution for Random Effects

Effect	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Intercept	-506E-23	0.002597	2858	-0.00	1.0000
time	3.35E-20	0.002597	1	0.00	1.0000

BIBLIOGRAPHY

Bollen KA, Curran PJ. Latent curve models: A structural equation perspective. John Willey & Sons, Inc., New Jersey, 2006.

Diggle P, Heagerty P, Liang K-Y, Zeger S. Analysis of longitudinal data. (Second Edition). Oxford University Press, New York, 2002.

Duncan TE, Duncan SC, Strycker LA. An introduction to latent variable growth curve modeling: Concepts, Issues, and Applications. Lawernce ErlbaumAssociates, New Jersey, London, 2006.

Goldstein H. Multilevel statistical models. Edward Arnold, London, 1995.

Goldstein H. Random coefficient repeated measures models. Encyclopaedia of Biostatistics, Armitage, P., & Colton, T. Edition, 1998.

Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling, 1999; 6(1): 1-55.

Kimm SYK, Glynn NW, Kriska AM, et al. Decline in physical activity in black girls and white girls during adolescence. The New England Journal of Medicine 2002; 347(10): 709-715.

Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics 1982; 38: 963-974.

Li F, Acock AC. Latent growth curve analysis: a manual for research data analysts. Oregon State University, 1999. (http://oregonstate.edu/dept/hdfs/papers/paper.html).

Littell RC, Milliken GA, Stroup WW, et al. SAS system for mixed models. SAS Institute Inc., Cary, NC, 1996.

Llabre MM, Spitzer S, Siegel S, et al. Applying latent growth curve modeling to the investigation of individual differences in Cardiovascular recovery from stress. Psychosomatic Medicine 2004; 66: 29-41.

MacCallum RC, Kim C, Malarkey WB, et al. Studying multivariate change using multilevel models and latent growth curve models. Multivariate Behaviorial Research 1997; 32, 215-253.

McArdle JJ, Epstein D. Latent growth curves within developmental structural equation models. Child Dev 1987; 29: 1110-1133.

Mechelen WV, Twisk JWR, Post GB, et al. Physical activity of young people: the Amsterdam longitudinal growth and health study. Medicine & Science in Sports & Exercise 2000; 32: 1610-1616.

Meredith W, Tisak J. Latent curve analysis. Psychometrika 1990; 55: 107-22.

Muthen B. Methodological issues in random coefficient growth modeling using a latent variable framework: Applications to the development of heavy drinking. Multivariate Applications in Substance use Research 2000: 113-140.

SAS online help. http://v8doc.sas.com/sashtml/. 2006.

Suhr DD. SEM for healthy, business and education. Statistics and Data Analysis, SUGI27, 2004:243-27.

Willett JB, Bub KL. Latent growth curve analysis. Encyclopedia of Behavioral Statistics. Oxford, Wiley, UK. 2004.