UNIVERSITI PUTRA MALAYSIA

JOINT MODELLING OF LONGITUDINAL AND SURVIVAL DATA IN PRESENCE OF CURE FRACTION WITH APPLICATION TO CANCER PATIENTS DATA

KHALID ALI SALAH

JOINT MODELLING OF LONGITUDINAL AND SURVIVAL DATA IN PRESENCE OF CURE FRACTION WITH APPLICATION TO CANCER PATIENTS' DATA

KHALID ALI SALAH

DOCTOR OF PHILOSOPHY
UNIVERSITI PUTRA MALAYSIA

2008

JOINT MODELLING OF LONGITUDINAL AND SURVIVAL DATA IN PRESENCE OF CURE FRACTION WITH APPLICATION TO CANCER PATIENTS' DATA

By

KHALID ALI SALAH

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirement for the Degree of Doctor of Philosophy

April 2008

To
My Wife,
Sons and Daughters

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy.

JOINT MODELLING OF LONGITUDINAL AND SURVIVAL DATA IN PRESENCE OF CURE FRACTION WITH APPLICATION TO CANCER PATIENTS' DATA

By

KHALID ALI SALAH

April 2008

Chairman: Associate Professor Mohd Rizam Abu Bakar, PhD

Faculty : Science

Analyses involving longitudinal and time-to-event data are quite common in medical research. The primary goal of such studies to simultaneously study the effect of treatment on both the longitudinal covariate and survival. Often in medical research, there are settings in which it is meaningful to consider the existence of a fraction of individuals who have little to no risk of experiencing the event of interest. In this thesis, we focus on such settings with two different data structures.

In early part of the thesis, we focus on the use of a cured fraction survival models performed in a population-based cancer registries. The limitations of statistical models which embodied the concept of a cured fraction of patients lack flexibility for modelling the survival distribution of the uncured group; lead to a not good fit when the survival drops rapidly soon after diagnosis and also when the survival is too high. In this study, a cure mixture model is enhanced by developing a dynamic semi-parametric exponential function with a smoothing parameter.

The latter (major) part of the thesis focuses on modelling the longitudinal and the survival data in presence of cure fraction jointly. When there are cured patients in the population, the existing methods of joint models would be inappropriate, since they do not account for the plateau in the survival function. We introduce a new class of joint models in presence of cure fraction. In this joint model, the longitudinal submodel is a combination of a random mixed effect model and a stochastic process. A semi-parametric submodel is also proposed to incorporate the true longitudinal trajectories and other baseline time (dependent or independent) covariates. This model accounts for the possibility that a subject is cured, for the unique nature of the longitudinal data, and is capable to accommodating both zero and nonzero cure fractions. We generalize the two submodels to be multidimensional to investigate the relationship between the multivariate longitudinal and survival data.

Bayesian approach was applied to the data using a conjugate and non-conjugate prior families to obtain parameter estimates for the proposed models. Gibbs sampling scheme is modified for fitting the joint model. Metropolis Hasting and Adaptive Rejection Sampling steps are used to update the Markov chain to estimate parameter whose full conditional densities can not be sampled efficiently from the existing methods, leading us to propose efficient proposal densities.

The simulation studies demonstrate that the joint modelling method results in efficient estimates and good coverage for the population parameters. The analysis of cancer patient's data indicates that when ignoring the association between the longitudinal and the survival data would lead to biased estimates for the most important parameters.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia
sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PEMODELAN TERCANTUN DATA LONGITUDINAL DAN MANDIRIAN DENGAN KEHADIRAN PECAHAN SEMBUH DAN APLIKASI KEPADA PESAKIT KANSER

Oleh

KHALID ALI SALAH

April 2008

Pengerusi: Profesor Madya Mohd Rizam Abu Bakar, PhD

Fakulti : Sains

Analisis melibatkan data longitudinal dan masa sehingga suatu peristiwa berlaku merupakan analisis yang biasa dilaksanakan dalam penyelidikan perubatan. Dalam penyelidikan perubatan terutamanya dalam sesetengah keadaan, kesan rawatan ke atas kovariat longitudinal dan mandirian adalah lebih bermakna jika dipertimbangkan kewujudan pecahan individu yang sedikit dan tidak berisiko terhadap peristiwa yang menjadi tumpuan. Dalam tesis ini, fokus adalah kepada keadaan yang sedemikian dengan melibatkan dua struktur data yang berbeza.

Dalam bahagian awal tesis, fokus adalah terhadap penggunaan model mandirian pecahan sembuh ke atas populasi kanser yang berdaftar. Kekangan bagi model statistik yang mengambilkira konsep pecahan sembuh pesakit adalah ianya tidak begitu anjal untuk memodelkan taburan mandirian bagi kumpulan yang tidak sembuh, justeru mengakibatkan ketidakbagusan penyesuaian apabila mandirian menurun secara mendadak selepas diagnosa dan jika mandirian terlalu tinggi. Dalam

kajian ini, model campuran sembuh diperkasakan dengan membangunkan fungsi dinamik separa parametrik eksponen dengan parametrik pelicin.

Dalam bahagian kemudiannya (utama), tesis ini memfokus kepada pemodelan tercantum data longitudinal dan mandirian dengan kehadiran pecahan sembuh. Apabila terdapatnya pesakit yang sembuh dalam populasi, keadaan model tercantum yang sedia ada tidak bersesuaian disebabkan ianya tidak mengambilkira bahagian mendatar dalam fungsi mandirian. Oleh itu, kami memperkenalkan suatu kelas model tercantum yang baharu dengan kehadiran pecahan sembuh. Dalam model tercantum ini, submodel longitudinal merupakan kombinasi model rawak kesan bercampur dan proses stokastik. Submodel separa parametrik disarankan juga mengambilkira trakjetori longitudinal yang sebenar dan kovariat (bersandar atau merdeka) berdasar masa yang lain. Model ini mengambilkira kemungkinan yang subjek akan sembuh, merupakan keunikan data longitudinal dan berupaya menangani pecahan sembuh sifar dan bukan sifar. Seterusnya kami mengitlak dua submodel ini menjadi multidimensi untuk menyelidik hubungan di antara longitudinal multivariat dan data mandirian.

Pendekatan Bayesian dilaksanakan kepada data menggunakan famili prior konjugat dan bukan konjugat untuk memperolehi anggaran parameter bagi model yang dicadangkan. Skema pensampelan Gibbs diubahsuai untuk penyesuaian model tercantum. Langkah pensampelan penolakan penyesuaian dan Hasting Metropolis digunakan untuk mengemaskini rantaian Markov bagi menganggar parameter yang ketumpatan bersyarat penuh tidak boleh dicerap secara berkesan menggunakan kaedah yang sedia ada. Ini menyebabkan kami usulkan cadangan ketumpatan efisien.

Kajian simulasi menunjukkkan keputusan kaedah pemodelan tercantum memberikan anggaran yang lebih efisien dan libutan yang baik bagi parameter populasi. Analisis data pesakit kanser menunjukkan jika diabaikan hubungan di antara cirian longitudinal dan data mandirian, ianya akan menghasilkan anggaran yang pincang bagi kebanyakan parameter yang penting.

AKNOWLEDGEMENTS

I am extremely indebted to my supervisor, Assoc. Prof. Dr. Hj. Mohd Rizam bin Abu Bakar, for his excellence supervision, invaluable guidance, helpful discussions and understanding throughout the whole period of research. His charisma and enthusiasm while guiding me has made this challenging research period a very useful and enjoyable learning experience. He is the one always leading me to keep faith and be persistent during some difficult times. Special thanks and appreciation goes to my supervisory committee, Assoc. Prof. Dr. Noor Akma Ibrahim and Assoc. Prof. Dr. Kasim bin Haron, for their valuable time, helpful comments, discussions and suggestions.

I am gratefully acknowledge the Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 17 Regs Public-Use, Nov 2005 Sub (1973-2003 varying), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, US, about their valuable helpful suggestions, comments, huge data sets and necessary software.

I am much obliged to all members of Institute for Mathematical Research (INSPEM) and Dept. of Mathematics, University Putra Malaysia, for all their friendly support. This particularly goes to Prof. Dato' Dr. Hj. Kamel Ariffin Mohd Atan, Director of INSPEM. Assoc. Prof. Dr. Habshah Midi, Head of Laboratory of Statistics and Applied Mathematics, INSPEM. Assoc. Prof. Dr. Rozita Rosli , Deputy Dean Faculty of Medicine, UPM. And Dr. Mohd Bakri Adam, INSPEM, UPM.

I wish to acknowledge gratefully Alquds University, Jerusalem, Palestine, for their encouragements and supports during the whole period of my study. This particularly goes to Prof. Dr. Sari Nusseibeh, president of Alquds University and all members of Dept. of Mathematics, Faculty of science, Alquds University.

To my late wife, who passed away two years ago before I am staring this research, I offer my very special thanks and appreciations for having been a special wife.

Finally, my special thanks and deep gratitude goes to my beloved wife, without her patience, sacrifice and moral support, this thesis would not have been possible. I am also grateful to my children, Haneen, Fatima, Diana, sweet twins (Jihan and Afnan), Ali and the hero Amro. My deep thanks also goes to my special mother (Hjh. Om Khalid), special father (Hj. Abu Khalid), brothers and sisters for all their support.

I certify that an Examination Committee has met on 11th April, 2008 to conduct the final examination of Khalid Ali Salah on his Doctor of Philosophy thesis entitled "Joint Modelling of Longitudinal and Survival Data in Presence of Cure Fraction with Application to Cancer Patients' Data" in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the candidate be awarded the relevant degree.

Members of the Examination Committee are as follows:

Chairman, PhD
Doctor
Mahendran Shitan
Faculty of Sience
Universiti Putra Malaysia
(Chairman)

Habshah Midi, PhD
Associate Professor
Faculty of Sience
Universiti Putra Malaysia
(Internal Examiner)

Isa Daud, PhD
Associate Professor
Faculty of Sience
Universiti Putra Malaysia
(Internal Examiner)

Abdul Aziz Jemain, PhD
Professor
Faculty of Sience and Technology
Universiti Kebangsaan Malaysia
(External Examiner)

HASANAH MOHD GHAZALI, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee are as follows:

Mohd Rizam Abu Bakar, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Noor Akma Ibrahim, PhD
Associate Professor
Institute for Mathematical Research
Universiti Putra Malaysia
(Member)

Kassim Bin Haron, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

AINI IDERIS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 12 – 6 – 2008

DECLARATION

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.

KHALID ALI SALAH

Date: 27 April 2008

TABLE OF CONTENTS

		Page
DEDICATION		ii
ABSTRACT		iii
ABSTRAK		v
ACKNOWLEDGEMENTS		viii
APPROVAL		xi
DECLARATION		xii
LIST OF TABLES		xvi
LIST OF FIGURES		xviii
LIST OF ABBREVIATIONS		xxi

CHAPTER

1	**INTRODUCTION**	
	1.1 Introduction	1
	1.2 Motivation	10
	1.3 Purpose and Objectives of the Thesis	15
	1.4 Scope of the Thesis	17
2	**JOINT MODELS IN LITERATURE**	
	2.1 Introduction	20
	2.2 Standard Joint Models	20
	2.3 Joint Models in Presence of Cure Fraction	26
3	**BACKGROUND AND BASIC CONCEPTS**	
	3.1 Longitudinal Data Analysis	33
	3.1.1 Exploring Longitudinal Data	33
	3.1.2 Approaches to Longitudinal Analysis	34
	3.1.3 General Linear Model for Longitudinal Data	34
	3.1.4 Generalized Linear Model for Longitudinal Data	38
	3.1.5 Multivariate Longitudinal Modelling	40
	3.2 Survival Data Analysis	41
	3.2.1 Parametric and Nonparametric Models	41
	3.2.2 Cure Models	43
	3.3 Joint Modelling for Longitudinal and Survival Data	45
	3.3.1 Two-Stage Approach	45
	3.32 Likelihood Based Approach	46
	3.4 Bayesian-MCMC and Gibbs Sampling	49
	3.4.1 Bayesian Methods	49
	3.4.2 Markov Chain Monte Carlo (MCMC)	50
	3.4.3 The Gibbs Sampling	55
	3.4.4 ARMS Within Gibbs Sampling	59
	3.5 Summary	60

4 CURE FRACTION, MODELLING AND ESTIMATING

4.1 Introduction 62
4.2 Cause-Specific and Relative Survival 63
4.2.1 Estimating Expected Survival 64
4.2.2 Interpretation of Relative Survival 67
4.2.3 Modelling Excess Mortality 69
4.3 Cure Models 71
4.3.1 The Mixture Cure Fraction Model 71
4.3.2 The Non-Mixture Cure Fraction Model 72
4.3.3 The Parametric Distributions and Link Functions 73
4.4 Modified Semiparametric Model and Smoothing Parameter 74
4.4.1 Prior and Posterior Specifications 78
4.4.2 Computational Methods 79
4.5 Application to Breast Cancer Data 82
4.5.1 Aim of the Study 82
4.5.2 Data Description 82
4.5.3 Analysis and Numerical Results 83
4.6 Discussion 97

5 A NEW CLASS OF JOINT MODELS IN PRESENCE OF CURE FRACTION

5.1 Introduction 103
5.2 A New Class of Longitudinal and Survival Joint Model 104
5.2.1 The Longitudinal Process 106
5.2.2 The Time-to-Event Model 107
5.2.3 The Joint Model 112
5.3 Joint Likelihood and Priors 113
5.4 Posterior and Full Conditional Distributions 119
5.5 Specification of Prior Distributions 126
5.6 Bayesian Model Assessment 129
5.6.1 Testing for the Presence of Cure Fraction 133
5.6.2 Testing for Sufficient Follow-Up 138
5.6.3 Testing for Outliers 142
5.7 Summary 144

6 MCMC SAMPLING METHODS AND SIMULATION STUDIES

6.1 Introduction 146
6.2 Simulation Design 146
6.3 Summary Statistics 149
6.4 Posteriors and Sampling Details 150
6.4.1 Sampling Based on Standard Distribution 151
6.4.2 Sampling Based on Standard Distribution Density and A Second Density Term 151

6.4.3 Sampling Based on A non-Standard Distribution Density 155
6.5 MCMC Convergence Diagnostics 158
6.6 Numerical Results Based on MCMC Sampling 160
6.7 Model Selection 164
6.8 Discussion 165

7 MULTIVARIATE JOINT MODEL IN PRESENCE OF CURE FRACTION VIA FRAILTY
7.1 Introduction 170
7.2 The Multivariate Longitudinal Process 170
7.2.1 Multivariate Stochastic Process 171
7.2.2 Parameterizations 172
7.2.3 The Multivariate Model 173
7.2.4 Missing Values and Predictions of Future Observations 175
7.2.5 Approximations and Improvements 176
7.3 A New Multivariate Semi-Parametric Survival Model 178
7.4 The Multivariate Joint Model 184
7.5 The Multivariate Joint Likelihood 185
7.6 Posteriors and Full Conditional Distributions 192
7.7 Priors and Computational Implementation 199
7.8 Discussion 202

8 ANALYSIS OF REAL DATA SET
8.1 Introduction 204
8.2 Data description 205
8.3 Data Analysis 206
8.3.1 Longitudinal Data Analysis 208
8.3.2 Univariate Survival Data Analysis 217
8.3.3 Bivariate Survival Data and Joint Model 224
8.4 Testing of Hypotheses and Model Selection 234
8.4.1 Sufficient Follow-Up Period 234
8.4.2 Presence of Cure Rate 236
8.4.3 Model Selection and Goodness of Fit 237
8.5 Sampling and Computational Issues 239
8.6 Summary and Conclusions 241

9 SUMMARY, CONCLUSION AND IDEAS FOR FURTHER WORKS
9.1 Summary and Conclusion 244
9.2 Ideas for Further Works 255

REFERENCES 258
APPENDICIES 269
BIODATA OF STUDENT 278
LIST OF PUBLICATIONS 279

LIST OF TABLES

Table		Page
4.1	Relative survival. SEER 17. female breast cancer includes cases diagnosed in 1973-2003	84
4.2	Parameter estimates, Standard error and Median survival time of standard Weibull model	86
4.3	Parameter estimates, Standard error and Median survival time of standard Cox model	86
4.4	Parameter estimates, Cure rate (%) and median survival time of mixed cure Weibull model	90
4.5	Estimates and Standard error of the parameters from the Weibull mixture model including all covariates	93
4.6	Posterior estimates of the semi-parametric model parameters for $\alpha = 0.2$, $r = 0.6$ and random J	98
6.1	Scale factor estimation $\sqrt{\hat{R}}$ for different parameters	159
6.2	Monte Carlo Summary statistics of the parameters estimate	160
6.3	The survival's two structure parameters	161
6.4	Posterior estimates from Joint models for J = 1	166
6.5	Posterior estimates from separate models for J = 1	167
6.6	Posterior estimates from Joint models for J = 5	167
6.7	Posterior estimates from separate models for J = 5	168
6.8	Posterior estimates from Joint models for J = 10	168
6.9	Posterior estimates from separate models for J = 10	169
6.10	The LPML and DIC statistics for different models	169
7.2	Kernel functions	177
8.1	Summary statistics for IgG and IgM measurements	207

8.2	Summary of Survival and Relapse times	207
8.3	Summary statistics for the covariates of E1684 data set	207
8.4	Parameter estimates from MEM of IgG and IgM	211
8.5	The Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Likelihood ratio test of model selection	211
8.6	Covariance parameter estimates of univariate analysis for IgG, IgM	212
8.7	Parameter estimates of the covariates from univariate analysis of IgG and IgM data	213
8.8	Parameter Estimates from Bivariate analysis of IgG and IgM Data	214
8.9	Covariance parameters estimate from bivariate analysis (IgG ,IgM)	214
8.10	Posterior estimates of different parametric models without covariates for RFS and OS (No Cure Presented)	220
8.11	Posterior estimates of different parametric and non parametric models among covariates for RFS (No Cure Presented)	221
8.12	Posterior estimates of different parametric and non parametric models among covariates for OS (No Cure Presented)	222
8.13	Test statistics of different parametric and non parametric models among covariates for RFS and OS (No Cure Presented)	222
8.14	Posterior means, Standard deviations and 95% HPD intervals for parameters in RFS and OS	225
8.15	Posterior means, Standard deviations and 95% HPD intervals for parameters in the bivariate survival model	226
8.16	Posterior means, SE and 95% HPD intervals for regression coefficients in the longitudinal model of IgG and IgM data	229
8.17	Posterior means, SE and 95% HPD intervals for variance covariance parameters in the longitudinal model of (IgG, IgM) data	230
8.18	Posterior means, SE and 95% HPD intervals for all parameters in the bivariate survival model	231
8.19	Summary of Box Plots	235
8.20	LPML statistics for different models	239

LIST OF FIGURES

Figure		Page
3.1	Haematocrit Trajectories for Hip Replacement Patients. The Left Hand Panels are Individual Profiles by Gender; the Right Hand Panels Show a Fitted Lowess Curve for the Mean Superimposed	35
4.1	Hypothetical Cumulative Relative Survival Curve	68
4.2	The Concept of Lead-Time	69
4.3	Comparison of Estimate and Actuarial Survival Function from Standard Cox Models	87
4.4	Comparison of Estimate and Actuarial Survival Function from Standard Weibull Models	88
4.5	Comparison of Estimate and Actuarial Survival Function from Weibull Mixture Cure Models	91
4.6	Observed and Estimated K-year Survival Probability by Diagnosis Year Stratified by Race	95
4.7	Changes in 5-years (RSR) for Different Age Groups	96
4.8	Comparison of Estimate and Actuarial Survival Function from the Cox Mixture Cure Model and the Proposed Semi-Parametric Survival Function for Histic Stage (Distance) of Different Age Groups	100
4.9	Comparison of Estimate and Actuarial Survival Function from the Cox Mixture Cure Model and the Proposed Semi-Parametric Survival Function for Histic Stage (Localized) of Different Age Groups	101
4.10	Comparison of Estimate and Actuarial Survival Function from the Cox Mixture Cure Model and the Proposed Semi-Parametric Survival Function for Histic Stage (Regional) of Different Age Groups	102
5.1	Possible Paths From Initial Intervention to Outcome	104
5.2	Disease progression diagram	108

6.1 Histogram, Time Series and Average Values Plots Respectively for the Parameter Values μ_a at 500, 1000, and 2000 Iterations Respectively, Using Gibbs Sampler 152

6.2 Histogram, Time Series and Average Values Plots Respectively for the Parameter Values β at 500, 1000, and 2000 Iterations Respectively, Using M-H Sampler 154

6.3 Histogram, Time Series and Average Values Plots Respectively for the Parameter Values δ at 500, 1000, and 2000 Iterations Respectively, Using ARMS Sampler 157

6.4 Kaplan Meier Estimator for the Survival Data with Cure Fraction 162

6.5 Kaplan Meier Estimator for the Survival Data with no Cure Fraction 162

6.6 Longitudinal Simulated Data vs. Time 163

8.1 IgG (upper) and IgM (lower) by Time for 30 Randomly Selected Patients 209

8.2 Coefficients for the Within-Subject Regression of IgG and IgM on Time 211

8.3 Prediction of IgG from Random Effects Model 218

8.4 Prediction of IgG from IOU Model 218

8.5 Prediction of IgG from BM Model 219

8.6 Prediction of IgM from IOU Model 219

8.7 Kaplan-Meier Plots for RFS and OS 223

8.8 A log Plot of Cumulative Hazard for RFS and OS 223

8.9 Plot of Marginal Posterior Distributions of α (Different Values of J) 227

8.10 Marginal Posterior Densities for Different Values of γ 233

8.11 Estimate Hazard Rates as a Function of IgG and IgM Taken at Time Point of Peak Measurement 233

8.12 Marginal Posterior Densities of α 234

8.13 Box Plots of the Posterior Means of the Cure Rates for RFS and OS of All Patients 235

8.14 The Bivariate Survival Surface 235

8.15 Plots of Survival Function for Non-Cured Patients of RFS and OS with J=5 236

LIST OF ABBREVIATIONS

AR_1	First Order Autoregressive
ARMS	Adaptive Rejection Metropolis Sampling
ARS	Adaptive Rejection Sampling
BM	Brownian Motion
BP	Bias Percentile
CCR	Confidence Converge Rate
CPO	Conditional Prediction Ordinate
DIC	Deviance Information Criterion
EM	Expectation Maximization
GEE	Generalized Estimated Equation
GLM	General Linear Model
HPD	High Posterior Density
IOU	Integrated Ornstein-Uhlenbeck
KME	Kaplan-Meier Estimator
LLR	Log-Likelihood Ratio
LPML	Logarithm of the Pseudo-Marginal Likelihood
MCMC	Markov Chain Monte Carlo
MCSD	Monte Carlo Standard Deviation
M-H	Metropolis-Hastings
MLE	Maximum Likelihood Estimates
MSE	Mean Square Error
OR	Odd Ratio
OS	Over All Survival
PH	Proportional Hazards
PsBF	Pseudo-Bayesian Factor
REML	Restricted Maximum Likelihood
RFS	Relapse Free Survival
SE	Standard Error

CHAPTER 1

INTRODUCTION

1.1 Introduction

Often in applied statistics, after some empirical data have been collected, the purpose of the analysis is to construct a statistical model. Otherwise; said, we are interested in situations where the aim is to explain how an outcome, or response, variable of particular interest is related to a set of explanatory variables, or covariates.

Longitudinal data is, data in the form of repeated measurements on the same unit over time. Data are routinely collected in this fashion in a broad range of applications, including agriculture and the life sciences, medical and public health research, and physical science and engineering. For example, in a medical study, the antibody immune measures IgG and IgM may be taken at weekly or monthly intervals on patients with cancer vaccine. The main reason and advantage of longitudinal analysis is to study the change over time. That is also how longitudinal analysis differs from repeated measures analysis. In longitudinal analysis, we model both the dependence of the response on the covariates and the associations among responses. Longitudinal study has the ability to distinguish the variation in the outcomes across time for an individual from the ones among the population. To model the random variability in the longitudinal models with continuous outcomes, Diggle *et al.* (1994) distinguish among three components of variability: random effects, serial association and measurement errors.

In many studies, multivariate outcomes are observed and hence multivariate longitudinal models are necessary. Many studies have discrete outcome variables which renders traditional likelihood-based methods that require the multivariate normality assumptions and cumbersome with time-varying covariates inapplicable. Three modern analysis approaches have been developed over the years for the analysis of longitudinal repeated measures study with discrete outcome variables. They are the marginal model, the nonlinear mixed effect model, and the transition model.

The scientific questions of interest often involve not only the usual kinds of questions, such as how the mean response differs across treatments, but also how the change in mean response over time differs and other issues concerning the relationship between response and time. Thus, it is necessary to represent the situation in terms of a statistical model that acknowledges the way in which the data were collected in order to address these questions. Complementing the models, specialized methods of analysis are required. For example, longitudinal data modelling is essential to describe both trend and variation for biological processes, such as growth curves, effects over time of medical intervention on physiological characteristics, monitoring human exposure to carcinogens, and so forth.

A promising approach for longitudinal data analysis is to treat their pathways as realizations of a smooth *stochastic process*, see, e.g., Ramsay and Silverman (2005). This idea originated when researches wanted to describe the effects of certain treatments on a response trajectory and naturally progressed to the modelling of random curves including models for the effects of treatments and covariates on multivariate longitudinal observations.

In this thesis, we describe an approach for capturing the correlation structure between multivariate longitudinal responses, leading to the notion of dynamical correlation to