Fall 2011

Modeling glucose-insulin kinetics and development of type 2 diabetes in offspring of diabetic parents

Chih-Wei Lin
University of Iowa

MODELING GLUCOSE-INSULIN KINETICS AND DEVELOPMENT OF TYPE 2 DIABETES IN OFFSPRING OF DIABETIC PARENTS

by

Chih-Wei Lin

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Pharmacy
in the Graduate College of
The University of Iowa

December 2011

Thesis Supervisor: Professor Peter Veng-Pedersen

ABSTRACT

Type 2 diabetes (T2D) has been studied for decades. Many risk factors of T2D have been identified, but few studies were designed to investigate the pharmacokinetics/ pharmacodynamics (PK/PD) risk factors preceding the onset of T2D. Moreover, although the disease progression of T2D has received considerable attention, little is known about the disease development of T2D. It is important to understand the temporal changes of the risk factors of glucose and insulin kinetics during the development of T2D for a better understanding of the etiology of T2D.

The objectives of this work are: 1) to develop a population-based glucose-insulin PK/PD model and identify the PK/PD risk factors preceding the onset of T2D, 2) to develop a methodology to evaluate the development of T2D, 3) model the time-course of the disease development based on the disease development variables (DDVs) derived from repeated intravenous glucose tolerance tests (IVGTT) and oral glucose tolerance tests (OGTT). The central hypothesis is that the development of T2D can be described and characterized by the glucose-insulin kinetics by employing a population-PK/PD based disease development analysis.

To summarize, a glucose-insulin kinetic model was developed and presented in Chapter 2. The pharmacokinetics/pharmacodynamics (PK/PD) risk factors preceding the onset of T2D were investigated using a population-based Bayesian nonlinear hierarchical model. In Chapter 3, a methodology describing the disease development of T2D was developed based on four important DDVs of T2D, namely fasting blood glucose (FBG), fasting serum insulin (FSI), homeostatic model assessment of insulin resistance (HOMA-IR) and body mass index (BMI). These DDVs were investigated for their temporal

patterns and relationships to the time-course of the development of T2D. The proposed model enables a quantitative, time-based evaluation of the development of T2D in this high risk population. In Chapter 4, the DDVs derived from repeated IVGTTs were evaluated. By applying the mixed effect analysis, important DDVs were identified as potential new biomarkers of T2D. Chapter 5 is an extension of application of the disease development analysis based on the DDVs derived from OGTTs. Chapter 6 is the conclusions and future works of this thesis.

The proposed population model of glucose-insulin kinetics has demonstrated that pharmacokinetic differences exists for the high risk population and can be helpful for prediction of T2D. By applying the proposed disease development analysis, the time-dependency and temporal patterns of the DDVs can be identified. An examination of the temporal changes in DDVs for the glucose-insulin system before the diagnosis of the disease provides a quantitative evaluation of the pathophysiological evolution of T2D and is valuable in predicting T2D.

Abstract Approved:_____
Thesis Supervisor

_____
Title and Department

_____
Date

MODELING GLUCOSE-INSULIN KINETICS AND DEVELOPMENT OF TYPE 2 DIABETES IN OFFSPRING OF DIABETIC PARENTS

by

Chih-Wei Lin

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Pharmacy
in the Graduate College of
The University of Iowa

December 2011

Thesis Supervisor: Professor Peter Veng-Pedersen

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Chih-Wei Lin

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Pharmacy at the December 2011 graduation.

Thesis Committee:

Peter Veng-Pedersen, Thesis Supervisor

Lawrence Fleckenstein

Daryl J. Murry

Jennifer Fiegel

Raymond J. Hohl

To my beloved parents, brother and my friends

Far better an approximate answer to the right question, which is often vague, than the exact answer to the wrong question, which can always be made precise.

John Tukey, *The future of data analysis*

ACKNOWLEDGMENTS

First and foremost I would like to express my greatest gratitude to my advisor, Professor Peter Veng-Pedersen, for his guidance during my graduate educations and for giving me the unique opportunity to work on this project. He has been always approachable, and his dedication to research and intelligent thinking style greatly influenced me and made me develop into a research scientist.

Secondly, I thank Dr. James H. Warram for providing the extremely valuable data and in helping me with the exceptional ideas to my work. I gratefully acknowledge Theresa Smith, RN, for performing the fasting blood tests; Anna Karass for measuring glucose concentrations; Marta Grinsbergs for insulin assays; Derek Jones for collecting blood samples; and also Adam Smiles for helping the data collection for the follow-up. I also thank Drs. Mats Karlsson and Nick Holford for the ideas and suggestions in the design and analysis of the data.

I would like to thank all of the faculties, especially the committee members of my comprehensive exam and dissertation defense, Dr. Fleckenstein, Dr Murry, Dr. Fiegel, Dr. Hohl, Dr. Flanagan, and Dr. Salem, for their excellent suggestions on my research. I am appreciative to the Department of Pharmaceutics for offering the teaching and research assistantships. I also want to thank all former and current students in Dr. Veng-Pedersen's lab including Nidal Al-Huniti, Srividya Neelakantan, Neeraj Gupta, Kevin Freise, Lanyi Xie, Mohammad Saleh, Matthew Rosebraugh, Mohammed EL-Komy, and Denison Kuruvilla.

Last but not least, I want to thank my parents, brother, and all of my friends for their support.

ABSTRACT

Type 2 diabetes (T2D) has been studied for decades. Many risk factors of T2D have been identified, but few studies were designed to investigate the pharmacokinetics/ pharmacodynamics (PK/PD) risk factors preceding the onset of T2D. Moreover, although the disease progression of T2D has received considerable attention, little is known about the disease development of T2D. It is important to understand the temporal changes of the risk factors of glucose and insulin kinetics during the development of T2D for a better understanding of the etiology of T2D.

The objectives of this work are: 1) to develop a population-based glucose-insulin PK/PD model and identify the PK/PD risk factors preceding the onset of T2D, 2) to develop a methodology to evaluate the development of T2D, 3) model the time-course of the disease development based on the disease development variables (DDVs) derived from repeated intravenous glucose tolerance tests (IVGTT) and oral glucose tolerance tests (OGTT). The central hypothesis is that the development of T2D can be described and characterized by the glucose-insulin kinetics by employing a population-PK/PD based disease development analysis.

To summarize, a glucose-insulin kinetic model was developed and presented in Chapter 2. The pharmacokinetics/pharmacodynamics (PK/PD) risk factors preceding the onset of T2D were investigated using a population-based Bayesian nonlinear hierarchical model. In Chapter 3, a methodology describing the disease development of T2D was developed based on four important DDVs of T2D, namely fasting blood glucose (FBG), fasting serum insulin (FSI), homeostatic model assessment of insulin resistance (HOMA-IR) and body mass index (BMI). These DDVs were investigated for their temporal

patterns and relationships to the time-course of the development of T2D. The proposed model enables a quantitative, time-based evaluation of the development of T2D in this high risk population. In Chapter 4, the DDVs derived from repeated IVGTTs were evaluated. By applying the mixed effect analysis, important DDVs were identified as potential new biomarkers of T2D. Chapter 5 is an extension of application of the disease development analysis based on the DDVs derived from OGTTs. Chapter 6 is the conclusions and future works of this thesis.

The proposed population model of glucose-insulin kinetics has demonstrated that pharmacokinetic differences exists for the high risk population and can be helpful for prediction of T2D. By applying the proposed disease development analysis, the time-dependency and temporal patterns of the DDVs can be identified. An examination of the temporal changes in DDVs for the glucose-insulin system before the diagnosis of the disease provides a quantitative evaluation of the pathophysiological evolution of T2D and is valuable in predicting T2D.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A1C	glycated hemoglobin test
AIC	Akaike information criterion
ADA	American Diabetes Association
BMI	body mass index
C.I.	confidence interval
C.S.	credible set
CV	coefficient of variation
Dbar	posterior mean of deviance
DIC	deviance information criterion
DDV	disease development variable
FBG	fasting blood glucose
FOCE	first order conditional estimate
FPG	fasting plasma glucose
FSI	fasting serum insulin
HOMA	homeostasis model assessment
HOMA-IR	homeostasis model assessment index of insulin resistance
HOMA-β	homeostasis model assessment index of β-cell function
IGT	impaired glucose tolerance
IVGTT	intravenous glucose tolerance test
MCMC	Markov chain Monte Carlo
MCPEM	Monte Carlo expectation maximization
MVN	multivariate normal distribution

N	total number of subjects
NGT	normal glucose tolerance
OGTT	oral glucose tolerance test
PD	pharmacodynamics
PG	progressor group indicator
PK	pharmacokinetics
REML	restricted maximum likelihood
SAEM	stochastic approximation expectation maximization
S_I	insulin sensitivity index derived by minimal model
S_G	glucose effectiveness index derived by minimal model
T1D	type 1 diabetes
T2D	type 2 diabetes
TIMEtoD	time prior to the diagnosis of T2D, which is calculated as the testing age of the IVGTT minus the age of diagnosis of T2D
TIRFM	total internal reflection fluorescence microscopy
WHO	World Health Organization

CHAPTER 1. INTRODUCTION

1.1 Background

The pathogenesis and nature history of type 2 diabetes (T2D) has been studied for decades. Although various risk factors of T2D have been identified and reported in the literature, including age, ethnicity, obesity, and family history, little is known about the risk factors based on the pharmacokinetics and pharmacodynamics (PK/PD) of glucose-insulin. With the development of population PK/PD, the study of population-based glucose-insulin PK/PD modeling is able to identify the "PK/PD" risk factors of T2D and addresses the unbalanced glucose and insulin kinetics relative to the development of T2D. The abnormalities of glucose-insulin regulation and the biphasic insulin secretion would be helpful for early detection of T2D for individuals with high risk. Furthermore, the relationship between disease development of T2D and the glucose-insulin kinetics can be investigated by disease development analysis. The study of the disease development analysis of T2D would enable a quantitative, time-based evaluation of the development of T2D based on temporal changes of glucose-insulin kinetics during the disease development period. The disease development model may be used to quantify the effect of interventions/prevention strategies such as drug treatment and lifestyle modification, which is important for evaluating regimens that prevent or delay the onset of T2D.

1.2 Epidemiology of diabetes mellitus

1.2.1 Prevalence of diabetes mellitus

The prevalence of diabetes mellitus has been increasing dramatically worldwide in the last few decades due to changes in human behavior and lifestyle (1). In the United states, approximately 25.8 million people have diabetes, accounting for 8.3% of the population in 2011 (2). The epidemic is also rapidly growing in developing nations in Asia, Africa, and South America. The estimated percentage increase of people with diabetes from 2000 to 2010 was 57% in Asia, 50% in Africa, and 44% in South America (1). According to a report from the International Diabetes Federation, an estimated 258 million people had diabetes in 2010, which corresponds to 6.4% of the world's adult population (3). The incidence rate of diabetes is still increasing. It is estimated there will be over 324 million people who will live with diabetes by 2025 and the number is expected to be as high as 438 million by 2030 (3, 4).

Diabetes is a global burden due to its long-term cost and complications. In the United States, the total cost of disease in 2007 was $174 billion including $116 billion in excess medical expenditures, $31 billion on general medical costs, $27 billion directly related to treatment of diabetes and $58 billion spent on treatment of the complications of diabetes including neurological symptoms, peripheral vascular diseases, cardiovascular diseases, renal complications, metabolic complications, and ophthalmic complications (1).

1.2.2 Epidemiology of type 1 diabetes

Type 1 diabetes (T1D) accounts for about 5%-10% of the cases of diabetes (2). Type 1 diabetes is also called juvenile diabetes and typically occurs in children under 15

years old (5). The incidence rate of T1D has high variability worldwide ranging from 36.5 per 100,000 person-years in Finland to 0.1 per 100,000 person-years in China (6). Despite T1D occuring in most ethnic groups worldwide, the white population has the highest incidence rate. In addition, the incidence rate has gradually increased in children in the white population in the United States from 4 per 100,000 person-years in 1900 to 20 per 100,000 person-year in 1976 (7).

1.2.3 Epidemiology of type 2 diabetes

T2D accounts for over 90% of cases of diabetes (2). In contrast to T1D, the incidence rate of T2D significantly increases with age and ranges from 10 per 100,000 person-years for the individuals 20-29 years old to 612 per 100,000 person-years for individuals between 60 and 69 years old (7). The incidence rate of T2D is also associated with ethnicity, obesity, and family history of T2D. It had been reported the prevalence in the United States is 7.1% in non-Hispanic whites, 8.4% in Asian Americans, 12.6% in non-Hispanic blacks, and 11.8% in Hispanics (2). The incidence rate is relatively high in Asian, African, and Hispanic groups compared to Caucasian group. Colditz *et al*. showed the incidence rate of T2D increased exponentially by 27% per increment in BMI for overweight or obese US women (8). The incidence rate of T2D in this population rose from an estimated 93 per 100,000 person-years for the individuals with a BMI of 25 to 1016 per 100,000 person-years for the individuals with a BMI of 35. Another commonly known risk factor of T2D is family history. Family history of T2D has been found as an important risk factor in many studies (9-13). The cumulative risk of T2D is 29.2% and 41.9% respectively for individuals with one diabetic parent and two diabetic parents,

which are significantly higher than the 14.0% for the individuals with no diabetic parent (11).

T2D can be prevented or its onset can be delayed by lifestyle modification and/or pharmacological interventions. Therefore, prevention is one of the most effective ways to reduce the cost, mobility and mortality caused by the disease and complications of the disease.

1.3 Development and prevention of type 2 diabetes

1.3.1 Development of type 2 diabetes

According to the diagnostic criteria of American Diabetes Association (ADA), the development of T2D involves three stages: normal, prediabetes, and diabetes (2). Prediabetes is an intermediate stage before T2D and characterized by impaired glucose tolerance. Three diagnostic tests are currently used to determine the stages of T2D: fasting plasma glucose (FPG) test, oral glucose tolerance test (OGTT), and glycated hemoglobin (A1C) test. The diagnostic criterion used by ADA are shown in Table 1.1 (2).

Imparied fasting glucose refers to the prediabetes condition detected by FPG test, and impaired oral glucose refers to the prediabetes condition detected by OGTT. Approximately 25% of the individuals with prediabetes (both impaired fasting glucose and impaired oral glucose) progress to T2D and 25% of the individuals revert back to normal over a 3 to 5 years period (14-16). This not only indicates that prediabetes is an important predictor of T2D but also shows that the development of T2D is reversible at this early stage. Thus, early intervention of the development of the disease is one of the best and practical solutions for reducing the high cost and incidence rate of T2D. A recent

study by Appleton investigated the reproducibility of OGTT for diagnosis of normal, impaired oral glucose, or diabetes with two OGTTs performed on the same subject within 2 to 6 weeks periods (17). The rate of having the same classifications from the two OGTTs as normal, impaired oral glucose, and diabetes are only 91%, 48%, and 78% respectively. In order to adequately handle the intra-subject variability from an OGTT during the development of the disease at the early stage, periodically repeated measurements or combined tests are helpful to increase the precision of the tests and reduce the false positives and negative results for diagnostic purposes (18-21).

Insulin resistance

Insulin resistance refers to the reduced insulin effect on suppressing hepatic glucose production, on stimulating glucose removal and glycogen synthesis in skeletal muscles tissue, and on inhibiting the lipolysis of adipose tissue (5). Insulin resistance is a metabolic syndrome of T2D and is associated with hypertension, dyslipidemia, atherosclerosis, and cardiovascular diseases (22). Insulin resistance results in increased plasma glucose levels and also higher serum insulin levels to compensate the high glucose concentration. The signs of insulin resistance, including hyperglycemia and hyperinsulinemia, have been reported as early predictors of T2D in longitudinal studies (23-25). The presence of insulin resistance occurs earlier than the presence of prediabetes (impaired glucose tolerance) and persists though the progression of T2D (26, 27).

β-cell dysfunction

Insulin is a peptide hormone produced by β-cells of islets of Langerhans in pancreas (Figure 1.1). β-cell dysfunction refers to the reduced insulin secretion after glucose challenge due to the failure of the β-cells in pancreas to compensate for

hyperglycemia. β-cell dysfunction is present in all cases of T2D and possibly caused by glucotoxicity and β-cell exhaustion due to insulin resistance and impaired glucose tolerance (28-30). β-cell function changes substantially during the development of T2D and also after the diagnosis of T2D. It generally starts with the diminished first phase insulin secretion during the period of prediabetes (impaired glucose tolerance) and deteriorates to the overall reduction of insulin secretion from β-cells in the late phase of T2D (31-33). The diminished first phase insulin secretion has been reported as an early predictor of T2D (25, 31, 34-36).

1.3.2 Risk factors of type 2 diabetes

Based on ADA, the risk factors of T2D include: presence of glucose tolerance/or impaired fasting glucose, age>45, family history of diabetes, overweight, lack of exercise, hypertension, low HDL cholesterol level, high triglyceride level, ethnicity (Non-Hispanic Blacks, Hispanic/Latino Americans, Asian Americans and Pacific Islanders, American Indians and Alaska Natives) and history of gestational diabetes mellitus (37). In addition, other predictors of T2D have been reported in the literature, including the reduced glucose removal rate, hyperinsulinemia, increased body weight, and reduced acute insulin response (the first 10 minutes insulin response after an glucose challenge) (23, 35, 38). These risk factors of T2D can be separated into two types of variables: time-invariant and time-variant variables. Time-invariant risk factors remain the same with time, for example: ethnicity and family history of T2D. This type of risk factors is easy to manipulate when estimating risk of T2D for specific populations, but it provides no information on how the disease progresses over time and the stages of the disease development. The time-variant risk factors contain additional time information

about the pathology the T2D and provide better prediction of risks of T2D based on current conditions of the individuals. However, the time-variant variables are relatively hard to manipulate due to the high inter-occasion variability and intra-individual variability. Therefore, a more in-depth mathematical model and statistical methods are required to perform data analyses on time-variant variables.

1.3.3 Screening for type 2 diabetes

T2D is a chronic disease that develops over years of pre-diabetic stage. In the study of Warram *et al*., 6 selected variables were monitored for 15 years before the diagnosis of T2D (39). Fasting glucose, insulin, and 2-hour glucose post-challenge increased within 5 years before the diagnosis of T2D while rising fasting triglycerides begins at least 10 years before the diagnosis of T2D. It is clear that there are early signs of T2D presented before the diagnosis of the disease and screening is one effective way to identify the individuals proceeding to T2D. In addition, about 27% of the individuals with diabetes remain undiagnosed, which corresponds to 7 million people in the U.S. (2). Without screening, these people could develop severe complications from T2D without detection, such as stroke, blindness, renal failure, and amputation. Therefore, screening of T2D is not only important to prevent or delay the onset of T2D for individuals but also helps to identify the subjects with undiagnosed T2D and further reduce the chances of complications due to T2D.

There are two types of screening methods of T2D. One is questionnaire-based and the other one is laboratory-based. The questionnaire-based screening methods are usually used as initial screening tools to identify the population with high risk of T2D. In Finland, the diabetes risk score was developed based on the presences of the risk factors

of T2D, including age, BMI, waist circumference, use of blood pressure medication and etc. (40). The questionnaire-based screening methods provide fast, simple, convenient low-cost way to identify high risk population of T2D, but the results of these screening methods can not used for diagnostic purposes. On the other hand, the laboratory-based screening methods, such as fasting plasma glucose test, OGTT, and A1C test can be used to diagnose prediabetes and T2D, but they are relatively expensive and require specialists for performing assays on blood samples.

1.3.4 Type 2 diabetes prevention trials

There is evidence that T2D can be prevented or its onset can be delayed with lifestyle modification and pharmacological intervention (41-45). In the lifestyle modification trials, the subjects with high risk of T2D (usually already with impaired glucose tolerance) were enrolled in programs with a restrictive controlled diet and/or exercise. The effects of the lifestyle modifications on preventing T2D were evaluated by the reduction in incidence rate of T2D. It is reported that the lifestyle modification trial can reduce the incidence rate by about 28% to 59% in studies with 2.5 to 6 years follow-up (46). In pharmacological trials, subjects with high risk of T2D were treated by medications, including acarbose, metformin, and orlistat to prevent the onset of T2D (43-45). Pharmacological treatments can also reduce the incidence rate by about 25.5% to 37% (43-45). Although the pharmacological trials significantly reduce the incidence rate of T2D, they are not recommended by ADA (47). Firstly, pharmacological preventions usually were less effective than life-style modifications. For instance, metformin is less effective than exercise/diet to reduce the incidence rate of T2D (43). Secondly, the beneficial effects of lifestyle modification continue after the prevention period (48, 49),

while the pharmacological interventions usually lose their effectness after the treatment period (50, 51). For instance, troglitazone-treated subjects had similar incidence rate of T2D as the subject-treated with placebo after 1 year of discontinuation of treatments (50). In another example, the effect of metformin on T2D prevention did not persist when the drug stopped (51). Thirdly, the lifestyle modifications have other benefits, such as reducing the chance of cardiovascular diseases and dyslipidemia. The pharmacological interventions require monitoring adverse drug effect and their benefits on other diseases are not clear. Finally, the lifestyle modifications are more cost effective than pharmacological interventions. Diet control and exercise cost very little compared to the medications for T2D(47).

1.4 Models of longitudinal data

Longitudinal data analysis arises in many biomedical applications including PK, PD, and disease progression modeling. The longitudinal analysis usually contains "time" as a predictor variable and one or more "biomarkers" of interest as the response variables. The repeatedly measured data of the biomarkers are collected and analyzed in the longitudinal analysis to evaluate the temporal patterns of the biomarkers. For example, PK modeling is important to investigate the mechanisms of drug absorption, disposition, metabolism and elimination. PD analysis is used to evaluate the relationship between drug concentration and drug effect. Disease progression modeling is useful to quantify the development and progression of diseases and determine the best regimen to prevent or treat a disease.

The classic nonlinear longitudinal model (non-population based approach) has been extensively developed and applied in the field of PK/PD. This kind of model is

usually developed based on the physiology of the human body to evaluate or predict drug concentrations and/or responses. With newly developed methodology and software, population-based longitudinal model has received great attention. The main purpose of the population-based PK/PD model is to estimate the population parameters, inter-individual variability, and effects of covariates. This approach is useful to select a dose for a population, to identify the source of the variability, to predict the drug's response for individuals. The structural model of the population PK/PD model is usually less complex compared to the classic PK/PD model. The modeling data of the population-based analysis usually requires more individuals but fewer samples for each individual. The classic PK/PD analysis has fewer individuals but rich-data in each individual to show the detailed longitudinal pattern of the response variables, such as drug concentration and drug effect. In contrast to PK/PD models which have drug concentration or effect as the response variable, the disease progression model has the biomarker of the disease as the response variable. The disease progression model can involve a population-based PK/PD model to evaluate the long term drug effect on the progression of disease and helps to individually optimize the therapy regimen of a disease. In general, the population-based approach involves a complex statistical model structure to address the hierarchy of the data and to evaluate the population parameters as well as the inter-individual variability. The population-based analysis can be done by a frequentist or Bayesian statistical approach, specifically the mixed effect model approach or the Bayesian hierarchical model approach.

1.4.1 Mixed effect model

A mixed effect model is a frequentist statistical model which expresses the response variable as a function of predictor variables with both the fixed effect and random effect parameters. This approach has been widely used in the field of population PK/PD analysis. Some programs and software that are available for this type of analysis include NONMEM, MONOLIX, S-ADAPT, NLMIXED procedures in SAS, nlme package in R and SPLUS. There is an extensive literature on the mixed effect model approach, especially in the field of PK/PD. Ette and Williams provide a comprehensive book review of the linear and non-linear mixed effect models and the applications in NONMEN, R and S-PLUS (52). A review about the statistical aspect of the mixed-effect model and the use of the nlme package is provided by Pinheiro and Bates (53).

The basic mixed effect model, similar to the simple linear regression model, has a response variable y (e.g. drug concentration/effect), and a predictor variable x (e.g. time). The mean of the response variable is characterized by a function $f(\ )$ of the predictor variable, and the intra-individual variability is described by a random variable ε. The model of the data of the jth observation of the ith subject can be written as:

$$y_{ij} = f(x_{ij}, \boldsymbol{\theta}_i) + \varepsilon_{ij} \tag{1.1}$$

The regression function $f(\ )$ can be a linear or nonlinear function of x_{ij}. This function represents for the structural model of PK/PD or disease progression. The mixed effect parameter vector $\boldsymbol{\theta}_i$ of the ith subject is the sum of a fixed effect parameter vector $\boldsymbol{\mu}$ of the population and a random effect parameter vector $\boldsymbol{\eta}_i$ of the ith subject.

$$\boldsymbol{\theta}_i = \boldsymbol{\mu} + \boldsymbol{\eta}_i \tag{1.2}$$

The fixed effect parameter vector $\boldsymbol{\mu}$ represents the population parameter vector and the random effect vector $\boldsymbol{\eta}_i$ represents the inter-individual variability by using a multivariate normal distribution with a mean of zero vector $\mathbf{0}$ and a variance-covariance matrix $\boldsymbol{\Omega}$.

$$\boldsymbol{\eta}_i \sim MVN\left(\mathbf{0}, \boldsymbol{\Omega}\right) \tag{1.3}$$

The variance-covariance matrix $\boldsymbol{\Omega}$ addresses the magnitude and the correlation of the inter-individual variability of the population. The description above only covers the basis of the mixed effect model. Davidian and Giltinan has a detailed review for various mixed effect models and statistical aspect of the model structures (54).

Numerous algorithms have been developed for mixed effect model analyses including, the first order conditional estimate (FOCE) algorithm implemented in NONMEM (55), the stochastic approximation expectation maximization (SAEM) algorithm in MONLIX (56), the Monte Carlo Expectation Maximization (MCPEM) algorithm in S-ADAPT (57), the adaptive Gaussian quadrature method in SAS, and the restricted maximum likelihood (REML) method in nlme package of R. The primary aim of these algorithms is to evaluate the maximum likelihood estimates of the population parameters. Dartois *et al.* provides a review for comparing different algorithms by simulation (58). In general, FOCE used in NONMEM is the most popular algorithm used to estimated population parameters in PK/PD analyses because the early development and wide usage of NONMEM. The algorithms implementing Markov chain Monte Carlo methods, such as SAEM and MCPEM algorithm, have higher successful rate in convergence, but they usually require more calculation time.

1.4.2 Bayesian hierarchical model

A Bayesian hierarchical model is analogous to the mixed effect model in the frequentist statistical approach. An extensive review of Bayesian data analysis is provided n the book by Gelman *et al.* (59). The basic statistical structure of the Bayesian hierarchical model for population PK/PD analysis involves three levels of hierarchy: prior, population, and individual level. In a Bayesian statistical approach, previous experiences and knowledge can be implemented as the prior information. This approach is useful for sequential studies or adaptive design clinical trials. The statistical structure of a Bayesian statistical model is similar to a mixed effect model, but a Bayesian statistical model has an extra layer of hierarchy, the prior distributions. The hierarchy of prior is corresponding to the highest level and needs to be specified before the analysis. WinBUGS is the most extensively used software for Bayesian analysis (60). The WinBUGS differential equation interface written by Lunn extend the flexibility of WinBUGS to various PK/PD or disease progression analysis (61).

In the individual level of the hierarchy, the Bayesian regression model describes the response variable y_{ij} and the predictor variable x_{ij} of the jth observation of the ith subject as:

$$y_{ij} \sim N\left(f\left(x_{ij}, \boldsymbol{\theta}_i\right), \tau\right) \tag{1.4}$$

In general, the observed y_{ij} is assumed to be normally distributed with a mean of $f\left(x_{ij}, \boldsymbol{\theta}_i\right)$ and a precision of τ. The regression function $f\left(x_{ij}, \boldsymbol{\theta}_i\right)$ represents the predicted y_{ij} value at x_{ij} with the individual parameter vector $\boldsymbol{\theta}_i$ of the ith subject. The precision τ is the reciprocal of the variance of the normal distribution. The term "precision" is commonly used in Bayesian literature and software instead of variance for easy

calculation. In the population level of the hierarchy, the individual parameter vector $\boldsymbol{\theta}_i$ is assumed to be multi-normal distributed with a mean vector $\boldsymbol{\mu}$ and an inverse covariance matrix $\boldsymbol{\Omega}^{-1}$.

$$\boldsymbol{\theta}_i \sim MVN\left(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1}\right) \tag{1.5}$$

The vector $\boldsymbol{\mu}$ would be the population parameter vector and $\boldsymbol{\Omega}$ would be the covariance matrix for inter-individual variability. In the highest level of hierarchy, prior distributions are given to τ, $\boldsymbol{\mu}$, and $\boldsymbol{\Omega}^{-1}$. For example, a gamma distribution can be used for τ, a multi-normal distribution can be used to $\boldsymbol{\mu}$, and a Wishart distribution can be used for $\boldsymbol{\Omega}^{-1}$. The book review of Gelman *et al.* describes the prior distribution selection (59). Additional in-depth review about prior selection and the philosophical foundation for Bayesian inference is provided by Kass *et al.* (62).

In a Bayesian analysis, Markov chain Monte Carlo methods, including Gibbs sampling, and Metropolis-Hastings sampling method, are used to estimate the posterior distributions of the parameters (63, 64). The mean and 95% credible set of a parameter can be estimated based on its estimated posterior distribution of the parameter. WinBUGS is the most commonly used software in Bayesian analysis. Although the running time in WinBUGS is longer compared with the algorithms used in mixed effect model analysis, WinBUGS has a higher successful rate of convergence and provides accurate estimates of covariance matrix for inter-individual variability without approximation (58, 65). In addition, WinBUGS provides a flexible platform for various statistical models and a variety of distributions for choosing. WinBUGS can also incorporate other statistical analysis, such as multivariate regression or survival analysis together with the population PK/PD analysis. Although WinBUGS is not the mainstream

software used in population PK/PD analysis, it offers a flexible environment for various statistical analyses and provides additional value in population PK/PD analysis.

1.5 Glucose and insulin kinetics

1.5.1 Glucose kinetics

Diabetes is characterized by a persistent high blood sugar in human body. Information of glucose kinetics is important to understand the pathology of the disease. Plasma glucose can originates exogenously from ingestion of food and also endogenously from the liver by glycogenolysis. The disposition of glucose is speficially evaluated based on radio-labeled glucose tracers in order to prevent the confounding effect from both exogenous absorption and endogenous glucose production. Segal S. *et al.* used C^{14}-labeled glucose to investigate the disposition of glucose (66). A three exponential function has been found best to describe glucose disposition. Later, Ferrannini E. *et al.* proposed a three compartment model to describe body glucose kinetics by a plasma, rapid equilibrating, and slowly equilibrating compartments (67). These types of studies are great contributions to the understanding of the metabolism and disposition of glucose, but the insulin effect on the glucose metabolism and the hepatic glucose production are still two important factors that need to be incorperated in glucose kinetic models.

In 1979, Bergman *et al.* proposed the minimal model (68). The model included the insulin effect on glucose removal in a glucose kinetic model. The model has two compartments to describe the disposition of glucose and an "insulin action" compartment to describe the effect of insulin on glucose removal (Figure 1.2). The amount of "insulin action effect" $X(t)$ is calculated based on the insulin concentration time curve over

baseline. This model became one of the most popular model to describe glucose kinetics for incorporating insulin effects on glucose removal. Later, Caumo A. and Covelli C. evaluated hepatic glucose production using deconvolution technique using labeled IVGTTs (69). A drop in hepatic glucose production has been found in the first 30 minutes after IVGTT, and then hepatic glucose production is back to normal at 90 minutes after IVGTT (Figure 1.3). To overcome the complexity of the physiology of glucose regulation in the human body, a variety of models have already been developed.

1.5.2 Insulin kinetics

Insulin is a peptide hormone with a molecular weight of 5808 Da and is produced by the β-cells of islets of Langerhans in pancreas (Figure 1.1). Insulin is the most important hormone to regulate glucose metabolism by increasing the uptake of glucose from blood in the liver, muscle and fast tissues. Insulin also reduces blood glucose level by inhibiting the release of glucagon. The synthesis of insulin is taking place in β-cells of islets of Langerhans. Proinsulin, a precursor of insulin, converts into insulin and C-peptide in an equalmolar pattern and then is stored in secretory granules in β-cells (70). It has been found that a β-cell contains approximately 13,000 insulin granules (71). The insulin granules can be further divided into different insulin pools according to the different movement patterns and relative locations (72). Readily releasable pools, composed of insulin granules adjacent to the plasma membranes of β-cells, exist in a fully releasable state for the quick response to a sudden glucose increase. In the cell plasma of the β-cells, insulin granules denoted reserve pools secrete insulin granules for maintaining regular insulin level. Gupta *et al.* proposed a detailed insulin kinetic model to analyze the insulin kinetics of the insulin pools in β-cells (73).

The main feature of insulin secretion is its biphasic secretion pattern. After a bolus glucose challenge, a fast insulin secretion (first phase insulin secretion) can be observed in the first 10 minutes and followed by a prolonged insulin secretion (second phase insulin secretion). This pattern was early described by a mathematical model developed by Grodsky (74). Several other kinetic models have been proposed to describe the biphasic secretion pattern from different experimental approaches (75-78). Another important factor affecting insulin kinetics is hepatic extraction of insulin. After insulin is secreted from pancreas, it directly goes into the portal vein before entering the circulatory system. Insulin and C-peptide are secreted in an equalmolar manner, and C-peptide is not extracted by liver. Therefore, data of C-peptide can be used to assess hepatic extraction of insulin. By modeling both insulin and C-peptide kinetics, insulin hepatic extraction can be estimated as the ratio of unextracted insulin and C-peptide. Toffolo G. *et al.* used a minimal model of insulin and C-peptide to assess the insulin hepatic extraction. Their results indicate that the liver can extract 70% insulin in the basal state and 54% during IVGTT (76).

Since insulin is secreted endogenously from pancreas, disposition of insulin can only be evaluated by labeled tracer studies to prevent the confounding effect from endogenous insulin production. Silver A. *et al.* used insulin-^{131}I to evaluate the disposition of insulin (79). It has been found that the kidneys can also greatly affect insulin metabolism. Subjects with renal failure have significantly lower insulin removal rate compared to normal and diabetic subjects. A three compartment model has been found to best describe the disposition of insulin. An insulin kinetic model can not be considered complete if the model does account for the kinetics interaction of glucose and

insulin. This a need for modeling the glucose-insulin kinetics by including a glucose and insulin feedback loops in the model.

1.6 Disease progression model

The disease progression model provides an extensive tool to describe the time-course of the disease progression and quantify the effects of drugs/interventions to the disease by modeling the biomarkers (or surrogates endpoints) in a longitudinal model. The goal of disease progression modeling is not only focusing on the drug effects during the treatment period but also on the post treatment effects on the disease. Disease progression modeling has substantial influences on drugs selection and regimen design of disease treatments. For example, the deterioration of T2D involves insulin resistance and β-cell dysfunction. The disease progression modeling would be useful to identify the most suitable drug and regimen specific to the different stages of insulin resistance or of β-cell dysfunction. An extensive review section of the models of natural disease progression has been done by Mould (52). The mathematical model of disease progression contains two main components: the model of natural progression of disease and the PK/PD model of drug treatments.

The mathematic models of natural disease progression can be divided into three major categories: linear model, asymptotic models, and differential equation models (52). Linear model is a simplest way to describe the nature progression of disease. The linear model has a general form:

$$S(t) = S_0 + \alpha \cdot t \tag{1.6}$$

$S(t)$ represents the disease progression function of a continuous biomarker or rating score which is closely associated with the progression of the disease. S_0 represents the

basal level of the disease status and α is the slope, or the rate parameter of the disease progression over time, t. The effect of drug treatment $E(t)$ can be described as a function of t:

$$S(t) = (S_0 + E(t)) + \alpha \cdot t \tag{1.7}$$

$$S(t) = S_0 + (\alpha + E(t)) \cdot t \tag{1.8}$$

The drug treatment effect can be described as modifying the baseline of the disease (1.7) or modifying the rate of disease progression in (1.8). The linear model has been used to describe the progression of Alzheimer's disease by Holford and Chan (80, 81). The advantage of the linear model is its simplicity and it is best to describe the disease progression for short period of follow-up.

Natural disease progression can also be described by asymptotic models, such as an exponential model, or Emax model.

$$S(t) = S_0 \cdot e^{-\alpha \cdot t} \tag{1.9}$$

$$S(t) = S_0 + \frac{S_{\max} \cdot t}{S_{50} + t} \tag{1.10}$$

In the exponential model (1.9), S_0 represent the initial state of the biomarker of the disease and α represents the rate of the deterioration (first order) of the disease. The exponential function has been used to describe some biomarkers that decrease with time, for instance, the loss of bone density for postmenopausal women (82). The exponential model has the advantage of describing the curvature of the disease progression. In addition, the exponential disease progression model can be easily extended to an accelerated failure time model or proportional hazard model which is commonly used in survival analysis.

The $S_{\max}$ in the Emax model (1.10) represents the maximum level of the biomarkers or score of the disease and S_{50} represents the time progressing to the half level between S_0 and $S_{\max}$. S_0 and $S_{\max}$ are the nature limits of the boundaries of the biomarker. The Emax function has been used to describe the pain resolution by Anderson *et al*. (83). In addition to exponential and Emax models, a combined exponential model was developed to describe the progression of Parkinson's disease by Holford *et al*. as follows (84):

$$S(t) = S_0 e^{-\alpha \cdot t} + S_{ss}\left(1 - e^{-\alpha \cdot t}\right) \tag{1.11}$$

S_0 and S_{ss} represent the basal state and steady state of the disease. α represents the rate constant of the disease progression. This model is an alternative choice of the Emax function that can be used to describe the natural limits of the basal state and the steady state of the disease maturation.

The natural history of the disease progression also can be modeled by differential equations. This type of the disease progression model has the following general form:

$$\frac{d\,S(t)}{dt} = f_{prog}(S) - f_{trt}(S) \tag{1.12}$$

The rate of change in $S(t)$ is modeled by the disease progression function $f_{prog}(S)$ and the treatment effect function $f_{trt}(S)$. Such model has been successfully used to describe the growth of bacteria and effect of antibiotics for a long term therapy (85, 86). For example, Zhi *et al*. proposed the model to describe the growth kinetics of *Pseudomonas aeruginosa* and the effect of piperacillin as follows (86):

$$\frac{d\,S(t)}{dt} = f_{prog}(S) - f_{trt}(S) \tag{1.13}$$

$$f_{prog}(S) = \beta \cdot S \qquad f_{trt}(S) = \frac{k_{\max} \cdot (c - c_{mec})_+}{k_{50} + (c - c_{mec})_+} \cdot S$$

In the function of the disease progression $f_{prog}(S)$, S and β represent the bacterial concentration and apparent growth rate of the organism, respectively. The treatment effect of antibiotics was modeled by the Emax function with the parameters $k_{\max}$ and k_{50} when the concentration of the antibiotics c is larger than the minimum effective antibiotic concentration c_{mec}. In addition to modeling the growth kinetics of bacteria, this type of model can be extended to describe the time course of the tumor growth with the effect of the chemotherapy or the life span of red blood cells with the effect of erythropoietin (87-89).

1.7 Models for type 2 diabetes

T2D is a chronic disease relative to the unbalanced glucose-insulin system. Numerous mathematical models have been developed for various purposes, such as to provide interpretation of the deterioration of the glucose-insulin system processing to T2D, to provide pathophysiology-based diagnosis of the progression of T2D, to quantify and evaluate the drug effects on the glucose-insulin system. Landersdorfer *et al*. discussed that mathematical models can be divided into six categories: models for the intrinsic interaction between glucose and insulin, models for the glucose-insulin system incorporating drug effects, models describing secondary drug effects, models describing effects on ancillary biomarkers, and models of disease progression (90). The purpose of this thesis is to indentify the PK/PD risk factors associated with the development of T2D and to develop a disease development model of T2D. Hence, the following sections

would focus on models for the intrinsic interaction between glucose and insulin used for diagnostic purposes and models for disease progression.

1.7.1 Models for diagnostic purposes

The development of T2D involves the deterioration of insulin resistance and β-cell dysfunction, which are characterized by the reduced insulin effect on glucose removal and the diminished insulin secretion (34, 38). Mathematical models have been developed to quantify the levels of insulin resistance and β-cell dysfunction and to detect abnormalities in the glucose-insulin kinetics for individuals who may have pre-diabetes or diabetes. The most frequently used models for diagnostic purposes of T2D are the Bergman's minimal model and homeostasis model assessment (HOMA) model (68, 91).

Minimal model

The Bergman's minimal model is one the most popular mathematical models used to evaluate insulin sensitivity and glucose effectiveness based on data of intravenous glucose tolerance test (IVGTT). This model was firstly proposed in 1979 by Bergman *et al.* (68). To date, more than 500 papers had been published based on the uses and the extensions of the model (92). The structure and equations of the minimal model is presented in Figure 1.2. The minimal model described the glucose disappearance as the effect from the remote insulin compartment. The plasma insulin $I(t)$ served as a forcing function that the excessive amount of insulin over the baseline produces the "insulin action" $X(t)$ in remote insulin compartment, and the glycogen synthesis in liver and periphery tissues were assumed to be regulated by the amount of "insulin action" $X(t)$ (68). The minimal model describes the glucose-insulin system with a very compact

mathematical structure and provides parameters describing glucose effectiveness (S_G) and insulin sensitivity (S_I), which have been considered in diagnosis of diabetes. In 1980, Toffolo *et al*. extended minimal model to describe the effect of glucose on the insulin production (78). Together with the original minimal model proposed in 1979, the minimal model provides a complete description to the glucose and insulin regulatory system. Some later versions of the minimal model incorporated the changes in hepatic glucose production after an IVGTT (69), and some studies extended the minimal models to estimate insulin sensitivity and glucose effectiveness indexes based on OGTT (93, 94). However, De Gaetano and Arino indicated that the equations of the minimal model are mathematically incorrect and the model cannot be used to fit glucose and insulin data simultaneously (77). Moreover, most versions of the minimal models do not account for the biphasic insulin kinetics which is associated with the deterioration of β-cell dysfunction. Despite these limitations, the minimal model has been used in estimating the insulin sensitivity and providing information for the diagnosis of T2D.

Homeostatic model assessment

The homeostatic model assessment (HOMA) indexes were developed to assess insulin resistance (HOMA-IR) and β-cell function (HOMA- β) based on the fasting glucose and insulin/C-peptide concentrations (91). The HOMA indexes have been widely used for diagnostic purposes of T2D with more than 500 citations (95). The HOMA model contains a series of nonlinear empirical functions describing the glucose and insulin regulations in the organs and tissues at the homeostatic (fasting) state. The numerical solutions for HOMA-IR and HOMA- β can be estimated by either the graph-based method or the equation of the fasting glucose and insulin levels (96). The values of

HOMA-IR and HOMA- β can be directly obtained from the graph based on fasting plasma glucose and serum insulin concentrations in Figure 1.4. The alternative method is to use the simple mathematical approximations shown below:

$$\text{HOMA - IR}: \quad \frac{\text{FPG} \times \text{FSI}}{405}$$

$$\text{HOMA - } \beta: \quad \frac{360 \times \text{FSI}}{\text{FPG - 63}}$$

FPG : Fasting plasma glucose concentration (mg/dL)
FSI : Fasting serum insulin concentration (mU/L)

The advantage of the HOMA indexes is their simplicity compared to the hyperglycemic clamp, euglycemic clamp or IVGTT when estimating insulin sensitivity and β-cell function (68, 97). Thus, they are more practical for use in large scale epidemiological studies. However, the nonlinear empirical functions describing the glucose and insulin regulations in the organs and tissues used in the HOMA model remains unpublished and the model is not validated for patients receiving insulin treatments (90). HOMA indexes are simple measurements of insulin resistance and β-cell function at the homeostasis state (fasting state), but these indexes do not reflect insulin resistance and β-cell dysfunction at the stimulated/dynamic state.

1.7.2 Disease progression models for type 2 diabetes

T2D is a progressive disease. Long-term medical care and treatments are required to maintain normal blood glucose levels. The disease progression model evaluating the long-term anti-diabetic effect of gliclazide on the progression of T2D was developed by Frey *et al*. (98). In the study, the disease progression of T2D was described as a linear

function of fasting plasma glucose concentration. The effect of gliclazide was evaluated by the following equations in the responders and non-responders:

Responders: $FPG(t) = FPG_{base} + \alpha \cdot t - E(t)$ (1.14)

Non-responders: $FPG(t) = FPG_{base} + \alpha \cdot t$ (1.15)

The fasting plasma glucose served as the biomarker of T2D and the time course of the disease progression was described by a function of fasting plasma glucose $FPG(t)$. T2D is assumed to progress from the base glucose level (FPG_{base}) linearly with a rate, α , over time, t . The effect of gliclazide is described by the function $E(t)$. This study has successfully characterized the relationship between the time-course of the progression of T2D and the long term anti-diabetic effect of gliclazide on the progression of T2D. This approach provided a tool to visualize and quantify the effect of the treatment on disease progression.

Later, de Winter *et al.* extended Frey's model and proposed a mechanism-based disease progression model which assesses the progression of T2D based on the physiological homeostasis of glucose and insulin (99). The effects of pioglitazone, metformin and gliclazide were modeled as treatment functions that affect the homeostasis of fasting plasma glucose (FPG), fasting serum insulin (FSI) and glycated hemoglobin A_{1c} (A1C) described by a two linked turn-over model. The model provided systemic-specific parameters which addressed the progression of T2D by a physiological based model and allowed the quantification of the treatment of T2D.

In contrast to the *progression* of the disease after T2D has been diagnosed, the objective of this work is to describe the *development* of the disease. Although prevention of T2D is the best way to prevent/prolong the onset of the disease, the procedures to

identify people with high risk factors and methods to evaluate the status of T2D still need to be improved. The focus of the thesis is on modeling the development of T2D with PK/PD based risk factors derived from IVGTT or OGTT. This work would provide valuable information to accurately identify people with high risk of T2D and adjust the procedures to quantify the development of T2D.

1.8 Objectives and specific aims

The primary objective is to characterize the development of type 2 diabetes with a disease development model based on the glucose-insulin kinetics for better prediction and understanding of the development of the disease.

Specific aim 1: To formulate an integrated population glucose-insulin PK/PD model to characterize the glucose insulin regulation and insulin's biphasic secretion after an intravenous glucose tolerance test and to use the proposed model to evaluate parameters' differences for populations with different level of risks factors preceding type 2 diabetes.

Specific aim 2: To develop a population-based methodology to describe the development of type 2 diabetes based on four important variables namely, fasting blood glucose, fasting serum insulin, homeostatic model assessment of insulin resistance and body mass index and to identify important temporal patterns and time-dependencies of these disease development variables over the disease development period.

Specific aim 3: To identify the important disease development variables derived from an intravenous glucose tolerance test and evaluate the relationship between the temporal changes of these variables and the development of type 2 diabetes.

Specific aim 4: To identify important disease development variables associated with the development of type 2 diabetes based on glucose and insulin measurements from an oral

glucose tolerance test and to evaluate the relationship between the temporal patterns of these variables and the development of the disease.

1.9 Hypotheses

The central hypothesis of this thesis is that the development of type 2 diabetes can be described and characterized by the glucose-insulin kinetics by employing population-PK/PD based disease development analysis.

Hypothesis 1: It is possible to characterize the glucose-insulin regulation and physiology of the insulin biphasic secretion using a population PK/PD model and early pre-diabetic pharmacokinetic differences exist in the population with different level of risk preceding type 2 diabetes.

Hypothesis 2: The specific temporal patterns and time-dependencies of the four important disease development variables are hypothesized to exist and can be identified by applying population-based disease development model.

Hypothesis 3: The important disease development variables derived from an intravenous glucose tolerance test are hypothesized to exist and can be identified by the proposed disease development analysis.

Hypothesis 4: The important disease development variables derived from an oral glucose tolerance test are hypothesized to exist and can be identified by the proposed disease development analysis.

1.10 Outline of the thesis

A glucose-insulin kinetic model is developed and presented in Chapter 2. The pharmacokinetics/pharmacodynamics (PK/PD) risk factors preceding the onset of T2D were investigated in the progressor group of T2D using the population-based Bayesian

nonlinear hierarchical model of glucose-insulin kinetics. Based on the analysis, pharmacokinetic differences exist for the high risk population (progressors) which may be helpful for prediction of T2D.

In Chapter 3, a disease development variable (DDV) based model is developed. Four important DDVs of T2D, namely fasting blood glucose, fasting serum insulin, homeostatic model assessment of insulin resistance and body mass index were investigated for their temporal patterns and relationships to the time-course of the development of T2D by population based approaches. The proposed model enables a quantitative, time-based evaluation of the development of T2D in this higher risk population.

The important DDVs of T2D derived from the repeatedly intravenous glucose tolerance tests (IVGTT) were identified and presented in Chapter 4. By applying the mixed effect analysis, important DDVs were identified by the mixed effect analysis. Chapter 5 is an extension of application of the disease development analysis on DDVs derived from the repeatedly measured OGTT. The important DDVs were identified and the temporal patterns of these DDVs were evaluated. Chapter 6 is the conclusions and future works of this thesis.

Table 1.1 Diagnostic criterion of type 2 diabetes used by ADA

ADA Criterion	FPG test	OGTT	A1C test
Normal	<100 mg/dL	< 140 mg/dL	< 5.7%
Prediabetes	100 - 125 mg/dL	140 - 199 mg/dL	5.7 - 6.4 %
Diabetes	≥ 126 mg/dL	≥ 200 mg/dL	≥ 6.5 %

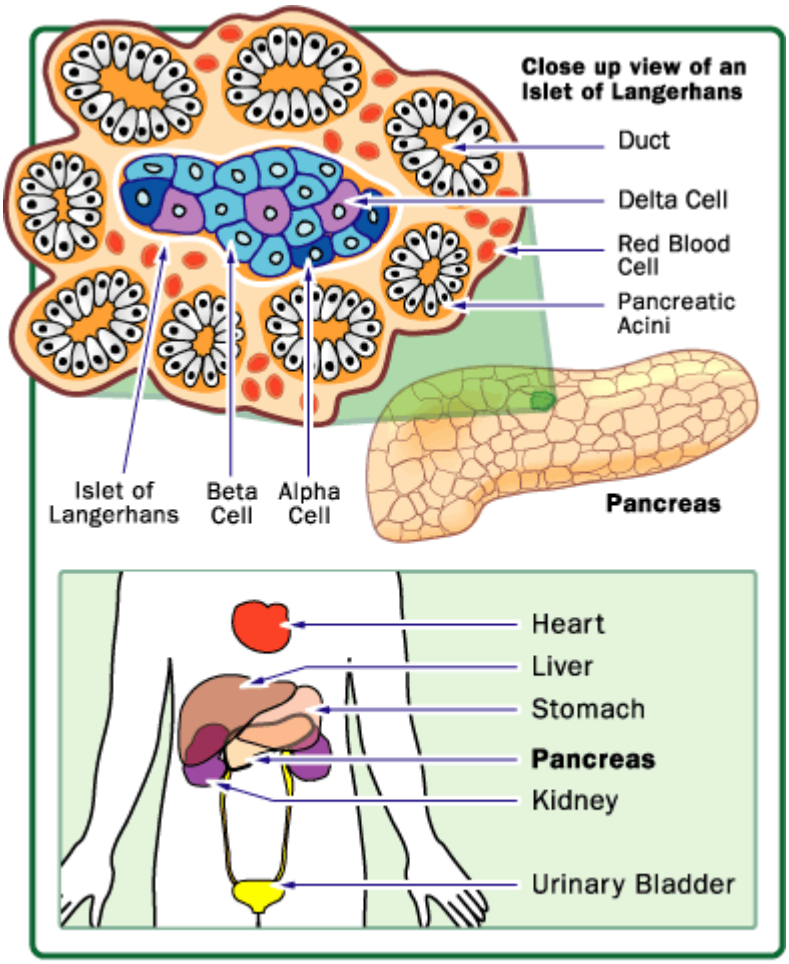
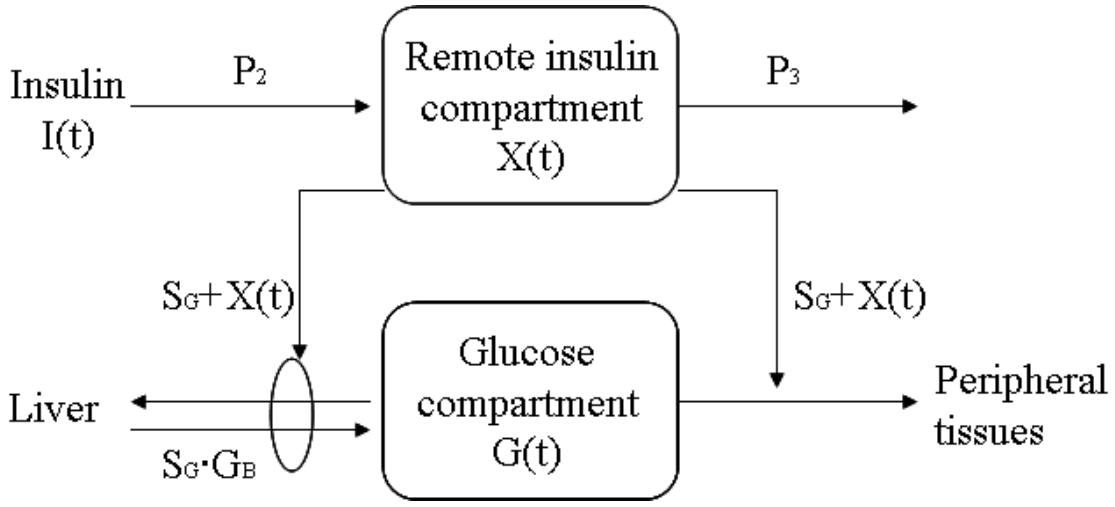


Figure 1.1. β-cells of islets of Langerhans in pancreas(http://trialx.com/curetalk/wp-content/blogs.dir/7/files/2011/05/diseases/Beta_Cell_Function-2.gif)



$$\frac{dX(t)}{dt} = P_2[I(t) - I_b] - P_3 \cdot X(t)$$

$$X(0) = 0 \qquad S_I = {}^{P_2}\!/_{P_3}$$

$$\frac{dG(t)}{dt} = S_G \cdot G_b - \left(S_G + X(t)\right)G(t)$$

$$G(0) = G_b + {}^{Dose}\!/_{V_d}$$

Figure 1.2. The structure and the corresponding differential equations of minimal model (68). The plasma insulin $I(t)$ served as a forcing function that the excessive amount of insulin over the baseline produces the "insulin action" $X(t)$ in remote insulin compartment, and the glycogen synthesis in liver and periphery tissues were assumed to be regulated by the amount of "insulin action" $X(t)$ The minimal model describes the glucose-insulin system with a very compact mathematical structure and provides parameters describing glucose effectiveness (S_G) and insulin sensitivity (S_I).

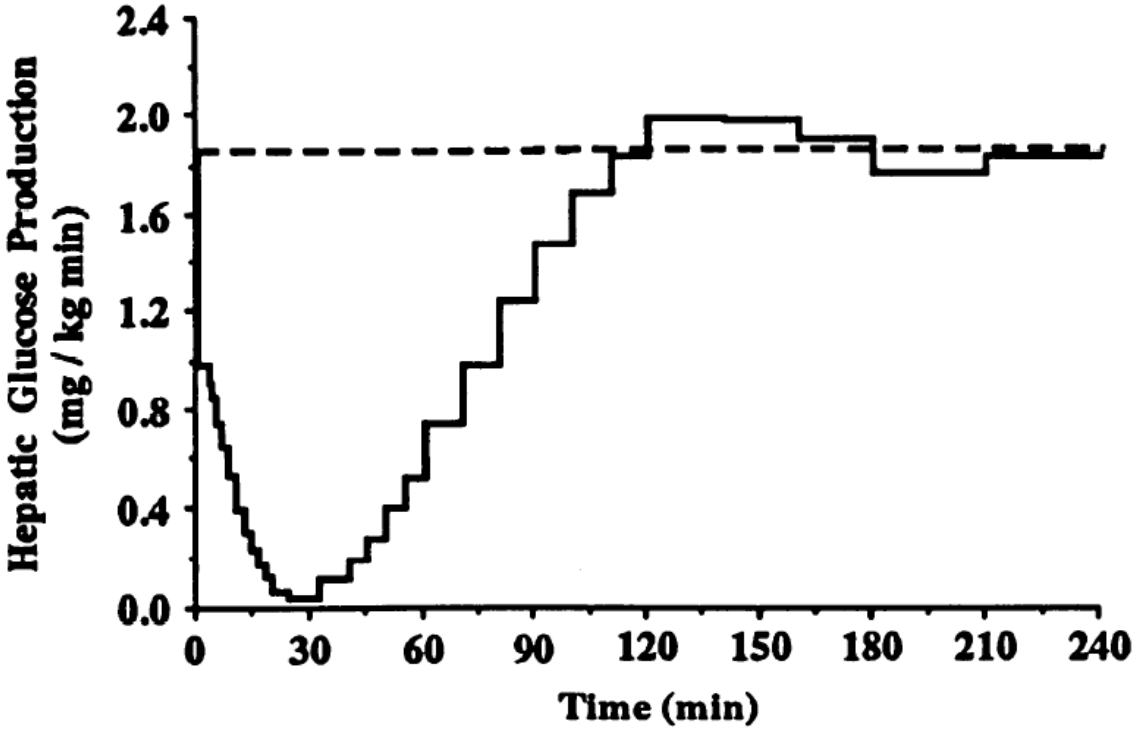


Figure 1.3. Average time course of hepatic glucose production obtained from 2 compartmetn minimal model by deconvolution (69)

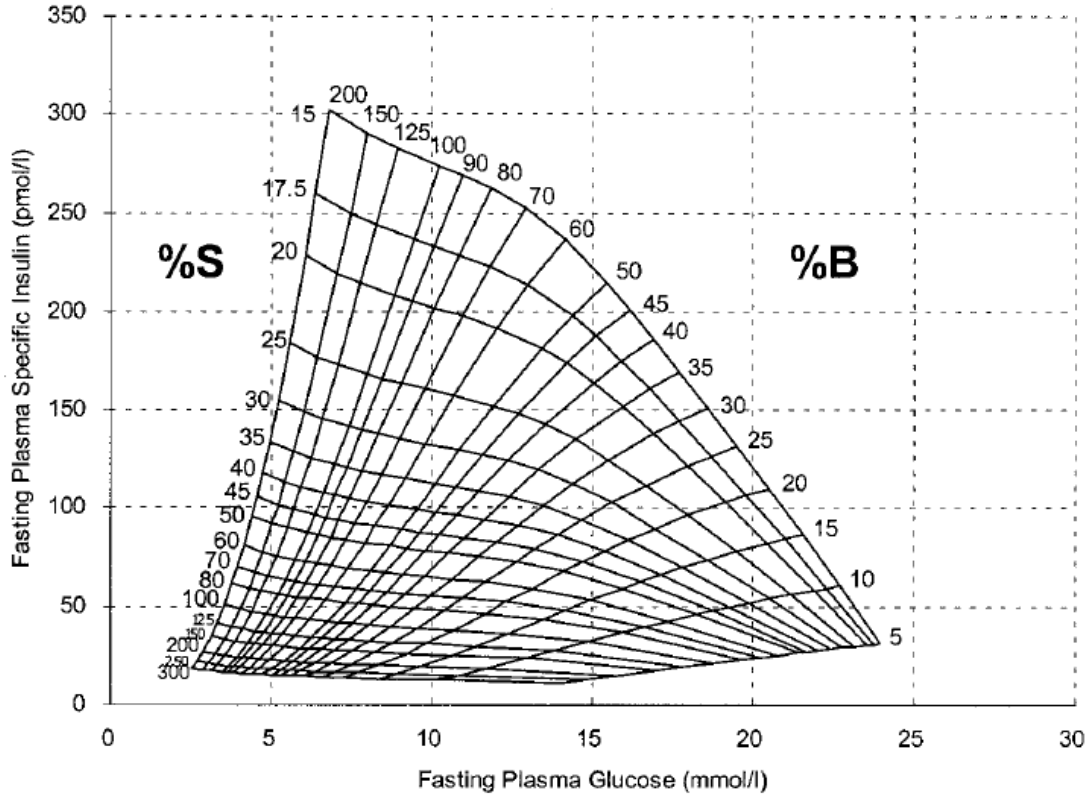


Figure 1.4. The graph-based method for estimating homeostatic model assessment-insulin resistance (HOMA-IR shown as %S) and homeostatic model assessment-β-cell function (HOMA-β shown as %B). Based on fasting plama glucose and insulin concentration, the HOMA-IR and HOMA-β can be evaluated from the graph.

CHAPTER 2. ANALYSIS OF PK/PD RISK FACTORS FOR DEVELOPMENT OF TYPE 2 DIABETES IN HIGH RISK POPULATION USING BAYESIAN ANALYSIS OF GLUCOSE-INSULIN KINETICS

2.1 Abstract

This study was designed to investigate the pharmacokinetics/pharmacodynamics (PK/PD) risk factors preceding the onset of type 2 diabetes using a population-based Bayesian nonlinear hierarchical model to describe the glucose-insulin kinetics. One hundred fifty-two healthy subjects with a family history of type 2 diabetes were recruited. Each subject received an intravenous glucose tolerance test (IVGTT) and glucose and insulin concentrations were collected when entering the study. After the test, subjects were followed for up to 25 years and further divided into the progressor group or the non-progressor group according to the follow-up results. A glucose-insulin kinetic model was developed to account for the physiology and molecular biology of the insulin biphasic secretion and glucose-insulin interactions with a minimal structure. The population PK/PD parameters of the two groups were estimated from the proposed glucose-insulin kinetic model. The relationships between the population PK/PD parameters and the diabetic follow-up results were evaluated. A high insulin baseline concentration, a lower maximum insulin-dependent glucose removal and a lower insulin removal rate constant were found associated with the development of type 2 diabetes in the high risk population. The study shows that very early pre-diabetic pharmacokinetic differences exist and can be helpful for prediction of development of type 2 diabetes.

2.2 Introduction

Type 2 diabetes (T2D) is a disorder of unbalanced glucose and insulin levels. An understanding of glucose-insulin regulation and kinetic relationship is critical to an investigation of the development of T2D. A variety of the mathematical kinetics models has been proposed for describing the glucose and insulin kinetics (73, 75-78, 100-102). Most of these models focus on how insulin affects the glucose metabolism or how glucose affects insulin secretion and do not include the complete mechanism, which includes the mutual effects i.e. glucose's effect on insulin kinetics and *vice versa*.

The main feature of insulin secretion is its biphasic secretion pattern. This pattern was earlier described by a mathematical model developed by Grodsky (74). The model describes insulin as stored packets inside the β-cells, and each insulin packet has a specific threshold-level to glucose concentration. When the glucose concentration increases, a specific number of insulin packets are secreted into the blood. Recently, the movement of insulin granules inside β-cells and the mechanism of exocytosis have been revealed by total internal reflection fluorescence microscopy (TIRFM) (72, 103, 104). The insulin granules can be further divided into different insulin pools according to the different movement patterns and relative locations(104). Readily releasable pools, composed of insulin granules adjacent to the plasma membranes of β-cells, exist in a fully releasable state and are associated with the fast first phase of insulin secretion. These pools provide a quick response of insulin secretion to a sudden glucose increase. In the cell plasma of the β-cells, insulin granules denoted reserve pools secrete insulin granules for maintaining the baseline insulin level and producing the second phase insulin secretion though the nonstandard secretion pathway and provide supplemental insulin

granules to the readily releasable pools(104). Gupta *et al.* proposed a insulin kinetic model to analyze the insulin kinetics of β-cells, the post hepatic insulin delivery and insulin elimination (73). The model has detailed descriptions of the insulin pools inside the β-cells and of how glucose affects the insulin production and secretion. Several other kinetic models have been proposed to describe the biphasic secretion pattern and provided agreement with data from different experimental approaches (75-78).

The Bergman's minimal model, uses remote insulin as a forcing function that regulates the glycogen synthesis in liver and periphery tissues (68, 78). The minimal model describes the glucose-insulin system with a very compact mathematical structure and provides parameters describing glucose effectiveness and insulin sensitivity, which are useful in diagnosis of diabetes. Later versions of the minimal model account for the changes in hepatic glucose production after an intravenous glucose tolerance test (IVGTT) (69, 105). Population-based Bayesian approaches have also been developed (106, 107). However, most versions of the minimal models do not account for the biphasic insulin kinetics and are not designed to fit the glucose and insulin data simultaneously. De Gaetano *et al.* discussed some mathematical issues of the minimal model and proposed a new dynamical model (77). The dynamical model provides simultaneous fits to glucose and insulin data and solves the equilibrium problems of the minimal model by the use of the delayed differential equations.

Recently, an integrated model for glucose and insulin was proposed by Silber *et al.* using a population kinetic approach in NONMEM (102). The model describes the glucose-insulin kinetics in three parts: glucose kinetics, insulin kinetics, and a control mechanism. The control mechanism has three effect compartments that regulate the

glucose-insulin interactions with three power functions. Two glucose effect compartments inhibit the glucose production and stimulate the secretion of insulin. One insulin effect compartment regulates the glucose level by stimulating the glucose clearance. This model provides extensive physiological structures and mechanisms that contain three feedback loops for glucose-insulin regulation, but this model has a complex structure and requires 20 fixed effect parameters in the glucose-insulin kinetic model.

In this study, a simple integrated glucose-insulin kinetic model was proposed. This model was developed though a comparison of the glucose-insulin kinetic models with different dynamic effects functions. With its minimal structure, this model is readily analyzed in a population pharmacokinetic parameter framework. The parameters have intuitive meanings that are associated with the physiology of the glucose removal and insulin production. The parameters' differences were analyzed for the two groups with different follow-up outcomes. The parameter estimates were compared between the groups with the intent to identify PK/PD risk factors for development of T2D.

2.3 Specific aim and hypothesis

The specific aim of this chapter is to formulate an integrated population glucose-insulin PK/PD model to characterize the glucose insulin regulation and insulin's biphasic secretion after an intravenous glucose tolerance test and to use the proposed model to evaluate parameters' differences for populations with different level of risks factors preceding type 2 diabetes.

The specific hypothesis is that it is possible to characterize the glucose-insulin regulation and physiology of the insulin biphasic secretion using a population PK/PD

model and early pre-diabetic pharmacokinetic differences exist in the population with different level of risk preceding type 2 diabetes.

2.4 Methods

2.4.1 Subjects

Between 1963 and 1983, the data of 152 non-diabetic healthy subjects with a family history of type-2 diabetes (diabetic parents) were collected by the Joslin Diabetes Center (Boston MA). No subjects had pre-diabetes or diabetes when entering the study. All participants received an IVGTT at the beginning of the study. After an eight to twelve-hour fasting period the concentrations of blood glucose and plasma insulin were obtained for each subject before a 0.5 g/kg glucose injection. Glucose and insulin blood samples were then collected at 1, 3, 5, 10, 20, 30, 40, 50, 60, 90, 120 and 180 minutes post-injection. The subjects were followed up to 25 years after the IVGTT. At the end of the follow-up period, data of presence or absence of T2D were collected. 25 of 152 participants developed T2D. These subjects were further divided into progressor and non-progressor group with sample sizes of 25 and 127 respectively. The subject demographics at the time entering the study is summarized in Table 2.1.

2.4.2 Assays

Glucose concentrations of the blood samples were measured by the ferricyanide method with a coefficient of variation (CV) of the assay of 1.5% (108). For determination of insulin concentrations, a double-antibody radioimmunoassay technique was used with a 17.6% CV (109). All of the insulin and glucose concentrations were measured by one technician in the Joslin Diabetes Center to control the measurement error.

2.4.3 Glucose-Insulin kinetic model

The goal of the model-building process is to design an integrated population glucose-insulin kinetics model with a simple structure considering the cellular physiologic features of the insulin biphasic secretion and accounting for the mutual kinetic interactions of glucose and insulin after IVGTT. The specific final model is shown in Figure 2.1.

Glucose kinetics model

The disposition function of glucose was described by an one-exponential model and the glucose removal is regulated by the insulin concentration. The blood glucose concentration $C_g(t)$ is given by the following model equations:

$$\frac{d\,C_g(t)}{dt} = \frac{k_{gp}}{V_g} - E_{gr}(C_i(t)) \cdot C_g(t)$$

$$C_{g,ss} = \frac{k_{gp}}{V_g \cdot E_{gr}(C_{i,ss})} \qquad t<0 \quad , \qquad C_g(0) = C_{g,ss} + \frac{D_g}{V_g} \tag{2.1}$$

The rate change of glucose concentration $\frac{d\,C_g(t)}{dt}$ is given as the production of glucose minus the removal of glucose. The glucose concentration is assumed at the steady state $C_{g,ss}$ before the injection of glucose, and is determined as the hepatic glucose production rate k_{gp} divided by glucose volume distribution V_g times insulin effect function of glucose removal $E_{gr}(C_{i,ss})$. After the glucose injection is introduced into the blood stream, the increase in glucose concentration is calculated as the glucose dose D_g divided by the distribution volume V_g. The hepatic glucose production rate, the dose of

glucose (0.5 g/kg) and the distribution volume of glucose are denoted by k_{gp}, D and V_g respectively in (2.1). The volume distribution of glucose is normalized by the body weight. Due to lack of hot glucose tracer data in this study, the changing in the hepatic production rate after IVGTT can not be evaluated. The hepatic glucose production rate was assumed to be unchanged (k_{gp}) after IVGTT. The glucose removal is regulated by an insulin effect function $E_{gr}(C_i(t))$.

Insulin kinetics model

The insulin secretion kinetic model was developed according to the physiology of the insulin granule extocytosis process described in Ohara-Imaizumi's study (104). The first phase of the insulin secretion is created by the fast insulin release from the readily releasable pools docked to the β-cells' membranes in the pancreas. The second phase of the insulin secretion is mainly created by the insulin granules in the reserve pools located in the inner β-cells. These processes are described by the following model equations:

$$\frac{d\,I_R(t)}{dt} = E_{ip}(C_g(t)) - k_{i\sec} I_R(t) \tag{2.2}$$

$$I_{R,ss} = \frac{E_{ip}(C_{g,ss})}{k_{i\sec}} \qquad t \le 0$$

$$\frac{d\,C_i(t)}{dt} = \frac{(1-E)}{V_i} k_{i\sec} I_R(t) - k_{ir} C_i(t) \tag{2.3}$$

$$C_{i,ss} = \frac{(1-E)}{V_i}\frac{k_{i\sec} I_{R,ss}}{k_{ir}} \qquad t<0, \qquad C_i(0) = \frac{(1-E)}{V_i}\frac{k_{i\sec} I_R(0)}{k_{ir}} + \frac{(1-E)}{V_i}\frac{S_p D}{V_g}$$

I_R denotes the amount of insulin in the reserve pools in the β-cells. The rate change of insulin $\frac{d\,I_R(t)}{dt}$ in the reserve pool is given as the production of insulin $E_{ip}(C_g(t))$ minus the secretion rate of insulin from β-cells (2.2). The insulin production rate is assumed to depend on glucose concentration and modeled by an effect function, $E_{ip}(C_g(t))$. The insulin secretion from the reserve pools in the β-cells to the blood stream was modeled by a first order process with an insulin secretion rate constant $k_{i\,sec}$. $C_i(t)$ represents the insulin concentration. In (2.3), the disposition of insulin was described by an one-exponential model with an insulin removal rate constant k_{ir}. The distribution volume of insulin is denoted as V_i and the hepatic extraction ratio is denoted as E. Due to lack of insulin tracer and C-peptide data, the volume distribution of insulin and the hepatic extraction ratio were not identifiable and accordingly fixed at 45 ml/kg body weight and 54%, respectively, according to reported values (76, 110). In the modeling of the first phase insulin secretion, the amount of insulin secreted from the readily releasable pools was approximated as a "bolus injection" to account for its fast secretion pattern. The amount of secreted insulin was assumed to be proportional to the increase in glucose concentration in the IVGTT study. Accordingly, the increase in glucose concentration after glucose injection is $\frac{D}{V_g}$, and the amount of insulin secreted from the readily releasable pools is approximated by $S_p\frac{D}{V_g}$, where S_p represents the proportionality constant that scales the amount of insulin secreted in the first phase. S_p is a variable for the effectiveness of glucose, resulting in a quick insulin secretion.

2.4.6 Glucose and insulin effect function selection

Equations presented in (2.1), (2.2) and (2.3) define the basic structure of the glucose-insulin kinetic model. The dynamics of glucose-insulin interactions were modeled by various effect functions. Four types of effect functions were investigated in the development of the proposed model (2.4-2.7). The effect functions were selected because of their simplicity and providing great stability for mathematical calcaulations. To reduce the complexity of the model of the model and facilitate the statistical analysis the model assumes no delays in the insulin effect on glucose removal and glucose effect on insulin production.

The following effect functions were considered:

A proportional effect function:

$$E_{gr}\left(C_i(t)\right) = S_{gr}\left(\frac{C_i(t)}{C_{i,ss}}\right) \qquad E_{ip}\left(C_g(t)\right) = S_{ip}\left(\frac{C_g(t)}{C_{g,ss}}\right) \tag{2.4}$$

A proportional power effect function:

$$E_{gr}\left(C_i(t)\right) = S_{gr}\left(\frac{C_i(t)}{C_{i,ss}}\right)^{\alpha} \qquad E_{ip}\left(C_g(t)\right) = S_{ip}\left(\frac{C_g(t)}{C_{g,ss}}\right)^{\beta} \tag{2.5}$$

A linear effect function:

$$E_{gr}\left(C_i(t)\right) = k_{gr} + S_{gr}\left[C_i(t) - C_{i,ss}\right]_+ \tag{2.6}$$

$$E_{ip}\left(C_g(t)\right) = k_{ip} + S_{ip}\left[C_g(t) - C_{g,ss}\right]_+$$

$$\left[x\right]_+ = \begin{cases} x, \ if \ \ x > 0 \\ 0, \ x \le 0 \end{cases}$$

An Emax effect function:

$$E_{gr}\left(C_i(t)\right) = k_{gr} + \frac{S_{max\,gr}\left[C_i(t) - C_{i,ss}\right]_+}{kc_{i50} + \left[C_i(t) - C_{i,ss}\right]_+} \tag{2.7}$$

$$E_{ip}\left(C_g(t)\right) = k_{ip} + \frac{S_{max\,ip}\left[C_g(t) - C_{g,ss}\right]_+}{kc_{g50} + \left[C_g(t) - C_{g,ss}\right]_+}$$

$$\left[x\right]_+ = \begin{cases} x, & if \quad x > 0 \\ 0, & x \le 0 \end{cases}$$

The first function (2.4) has the simplest form describing the concentration-effect relationship with only one parameter. S_{gr} and S_{ip} represent the insulin effectiveness on glucose removal and glucose effectiveness on insulin production. The concentrations were normalized by the steady state concentrations to make S_{gr} and S_{ip} equal to the insulin removal rate constant and insulin production rate at steady state. The second function (2.5) has more flexibility with an additional power parameter. The third function (2.6) and the forth function (2.7) describe a linear concentration-effect and an Emax model with steady state corrections. The effects are dependent on the concentrations when the concentrations are higher than the steady state concentration.

Population PK/PD analyses were performed for models with different combinations of effect functions to describe the glucose-insulin interactions. These effect functions were compared based on the posterior mean of deviance (Dbar) and the deviance information criterion (DIC) (111). The Dbar is equal to -2 log-likelihood. The model with a lower Dbar has a better fit to the data according to information theoretical principles. DIC is a model selection criterion developed based on the information theory according to Bayesian principles. It's analogous to the Akaike's information criterion (AIC) in the frequentist's approach (112). DIC is the sum of the Dbar and the effective number of parameters, which are associated with the goodness of fit and the complexity of model.

2.4.4 Bayesian hierarchical model

A nonlinear Bayesian hierarchical model was used to estimate the population glucose-insulin kinetics parameters. Bayesian hierarchical model provides intuitive statistical structure for the population kinetic analysis in terms of individual level of hierarchy, population level of hierarchy and prior distributions. The principles of Bayesian data analysis are well discussed by Gelman *et al.* (59). The model specification follows the directed plot used in WinBUGS (Figure 2.2) (60).The model building process includes three stages.

The first stage is modeling the residual error structure for the observed data in the lowest level of the hierarchy:

$$y_{ij} \sim \text{logNormal}\left(\log\left[f\left(t_{ij}, \boldsymbol{\theta}_i\right)\right],\ \tau\right) \tag{2.8}$$

All estimated concentrations are non-negative ensured by a log-normal distribution used to describe the distribution of the observations. The terms $\log\left[f\left(t_{ij}, \boldsymbol{\theta}_i\right)\right]$ and τ represent the parameters of the log normal distribution. In following the normal parameterization in WinBUGS the precision, τ, is used which is equal to the inverse of the variance. The observation of the ith subject at the jth time point is denoted by y_{ij}. The function $f\left(t_{ij}, \boldsymbol{\theta}_i\right)$ represents the fitted values at t_{ij} with parameter vector $\boldsymbol{\theta}_i$.

In the second stage, the log of individual parameter vectors is modeled by a multivariate normal distribution in the middle level of the hierarchy:

$$log\left(\boldsymbol{\theta}_i\right) \sim MVN\left(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1}\right) \tag{2.9}$$

$\boldsymbol{\mu}$ represents the mean vector of the individual log parameter vectors, and $\boldsymbol{\Omega}^{-1}$ represents the inverse covariance matrix of the individual log parameters. Accordingly,

$\boldsymbol{\Omega}$ is the covariance matrix which accounts for the inter-individual variability and correlations.

In the third stage, conjugate prior distributions are assigned to τ, $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}^{-1}$ in the highest level of the hierarchy. τ has a gamma prior distribution; $\boldsymbol{\mu}$ has a multivariate normal prior distribution, and $\boldsymbol{\Omega}^{-1}$ has a Wishart prior distribution. The structure of this Bayesian hierarchical statistical model is shown in the direct plot in Figure 2.2.

2.4.5 Structure of covariates

The follow-up results of disease status were used as covariates for the population pharmacokinetic parameter.

$$\log(\boldsymbol{\theta}_i) \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1})$$

$$\boldsymbol{\mu} = \boldsymbol{\mu}_{\text{nonDM}} + \mathrm{I}_{\text{DM}}(\mathrm{i}) \cdot \boldsymbol{\mu}_{\text{Diff DM}}$$

$$\mathrm{I}_{\text{DM}}(\mathrm{i}) = \begin{cases} 1 \ , \text{subject i} \in \text{diabetic outcome group} \\ 0 \ , \text{subject i} \in \text{nondiabetic outcome group} \end{cases} \tag{2.10}$$

$$\boldsymbol{\mu}_{\text{DM}} = \boldsymbol{\mu}_{\text{nonDM}} + \boldsymbol{\mu}_{\text{Diff DM}}$$

$$\mathbf{D}_{\text{DM vs nonDM}}(\%) = \frac{\exp(\boldsymbol{\mu}_{\text{DM}}) - \exp(\boldsymbol{\mu}_{\text{nonDM}})}{\exp(\boldsymbol{\mu}_{\text{nonDM}})} \times 100\%$$

The individual log parameter vectors are modeled by a multivariate normal distribution $MVN(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1})$. The population mean vector ($\boldsymbol{\mu}$) are different for subjects in different groups. The reference parameter $\boldsymbol{\mu}_{\text{nonDM}}$ is the population parameter vector of the non-progressor group. The $\boldsymbol{\mu}_{\text{DM}}$ is the population parameter vector of progressor

group. $\mathbf{D}_{\text{DM vs nonDM}}(\%)$ represents the population parameter differences in percentage between the progressor group and non-progressor group.

2.4.9 Prior distributions

No prior information for parameters was available for this newly developed model. Thus, vague prior distributions were used:

$$\tau \sim Gamma(\alpha, \beta) \quad ,\alpha = 0.01 \qquad \beta = 0.01$$

$$\boldsymbol{\mu} \sim MVN(\mathbf{K}, \boldsymbol{\Sigma}^{-1})$$

$$\mathbf{K} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \qquad \boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 0.0001 & 0 & \dots & 0 \\ 0 & 0.0001 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0.0001 \end{bmatrix} \tag{2.11}$$

$$\boldsymbol{\Omega}^{-1} \sim Wishart(\mathbf{R}, \nu) \quad ,\mathbf{R} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad \nu = \text{dimension of } \mathbf{R}$$

The specification of the prior distributions follows the parameterization in WinBUGS. α and β are the shape and rate parameters of the gamma distribution. $\mathbf{K}$ and $\boldsymbol{\Sigma}^{-1}$ are the location parameter vector and inverse covariance matrix of the multivariate normal distribution. $\mathbf{R}$ and ν are the scale matrix and degree of freedom of the Wishart distribution.

2.4.6 Programs and algorithms

The posterior distributions of individual, population parameters and secondary parameters were estimated by Metropolis-Hasting sampling method in WinBUGS version1.4.3 (113). The numerical solutions of the differential equations were calculated by the WinBUGS differential equation interface (WBDiff) (61). The specific codes for

describing the differential equations in the glucose-insulin model were written in Pascal in BlackBox (114). The convergence of the Markov chains is evaluated by the Gelman-Rubin plot (115). The chains are defined as converged when the Gelman-Rubin statistic is smaller then 1.2. Initially, three Markov chains were run simultaneously for 20,000 samples in each chain to access the convergence of log population parameter vector $\boldsymbol{\mu}$. Once the chains of vector $\boldsymbol{\mu}$ were converged to stationary distributions, the samples Dbar, DIC and the all parameters of interest were collected from the extra 5,000 runs of the three chains. The posterior distributions of parameters were estimated according to the 15,000 samples. Medians of the posterior distributions were reported as the Bayesian estimates of the parameters, and the 95% Bayesian credible sets were estimated as the intervals between the 2.5 and 97.5 percentiles of the posterior distributions. R version 2.6.2 was used for graph generation(116).

2.5 Results

2.5.1 Preliminary model independent measurements

The glucose baseline, the insulin baseline and the acute insulin response (AIR) after IVGTT were estimated for subjects in the two groups. The glucose and the insulin baselines were defined as the first measurements of glucose and insulin concentrations before IVGTT. The AIR after IVGTT was calculated by the trapezoidal rule as the insulin concentration vs. time area above the insulin baseline concentration from 0 to 10 minutes. The estimates of means and 95% confidence intervals of glucose baseline, insulin baseline and AIR for the two groups are summarized in Figure 2.3. The means of the two groups were compared by the two sample t-test. The mean of insulin baseline of

the progressor group is significantly higher than the mean of the non-progressor group ($p<0.05$).

2.5.2 Model selection

Dbar and DIC values were estimated for models with different glucose-insulin effect functions. The model for which the glucose removal is described by a steady state corrected Emax model and the insulin production rate is described by a power proportional function provided the overall lowest Dbar and DIC values (27966 and 28936 respectively) among all the models considered (Table 2.2). This final model gave the overall best fit to the data with relatively few parameters (Figure 2.1).

2.5.3 Population parameter estimates

A total of 1,970 glucose and 1,975 insulin samples of 152 subjects were analyzed by the proposed population glucose-insulin kinetic model. The posterior medians and 95% Bayesian credible sets for the population kinetic parameters of glucose and insulin of the progressor group and the non-progressor group are summarized in Table 2.3.

The population PK/PD parameters' differences in percentage between the two groups were estimated. Subjects in the progressor group tended to have a 23.2% lower maximum insulin-dependent glucose removal $S_{\max gr}$ and a 49.7% lower insulin removal rate constant k_{ir} compared to subjects in the non-progressor group (Figure 2.4). The differences were statistically significant (The 95% credible sets of the differences in percentage do not contain zero). The variance-covariance matrix for inter-individual variability is shown in Table 2.4. The highest correlation between any two parameters

was 0.82 which is the correlation between the glucose production rate parameter, k_{gp}, and the glucose removal rate constant at steady state, k_{gr}.

Population model predictions and observed concentrations of glucose and insulin for two groups are shown in semi-log plots (Figure 2.5). The majority of measured glucose blood and insulin plasma concentrations fell within the 95% Bayesian predictive interval. Figure 2.6 shows individual fits to the glucose and insulin concentrations for two representative subjects in each group. For some subjects, the glucose concentrations dropped below the steady state concentration from 60 to 120 minutes post glucose injection due to the effect of insulin (the upper right figure in Figure 2.5). Additionally, the insulin biphasic secretion can be dominated by either the first phase insulin secretion (the lower right figure in Figure 2.5) or the second phase secretion (the lower left figure in Figure 2.5). The proposed Bayesian glucose-insulin kinetics model provided the flexibility to fit these different kinetic scenarios.

2.5.4 Residual error

The exponential proportional error model was used for both glucose and insulin data. The estimate of residual error for glucose was 0.0722 with 95% Bayesian credible set between 0.0695 and 0.0751. This is corresponding to a 7.23% (6.96% ~ 7.52%) of coefficient of variation (CV). The variation includes the assay variability of 1.5% (117). The estimated residual error for insulin was 0.202 with 95% Bayesian credible set between 0.194 and 0.210. The corresponding CV is 20.4% (19.6%~21.2%) which is of similar magnitude to the assay variability of 17.6% (109). The residual errors mainly include the within-subject variability and the variability of the assays.

The goodness of fit was evaluated by the scatter plots of individual predicted values *vs.* the individual observed values and the residual plots. The scatter plot of the individual predicted values *vs.* the individual observed values is shown in Figure 2.7 and the correlation coefficients between the individual values and the observed values are all higher than 0.98. The residual plots for glucose and insulin data are shown in Figure 2.8. The residuals were calculated as the differences between the logarithm of observed data and of the logarithm of estimated concentration. The time scale is transformed as the square root of time in order to better visualize the frequent data of early sampling times. No clear over- and under-estimations were found, and the variability of the residuals seems consistent at different time points.

2.6 Discussion

A higher insulin baseline concentration, a lower maximum insulin-dependent glucose removal $S_{\max gr}$ and a lower insulin removal rate constant k_{ir} were found associated with the development of T2D in the high risk population. These parameters' differences were estimated from the IVGTT data collected at a time when the subjects had no signs of impaired glucose tolerance or diabetes. Considering the follow-up period of up to 25 years (average of 13-14 years), these deviations may be considered as very early signs of the progression towards development of T2D. The pathogenesis of T2D has been investigated in Pima Indians (38). In these previous studies, weight gain, decrease in insulin-stimulate glucose disposal, and decrease in acute insulin response (AIR) were identified longitudinally as factors associated with the progression from normal to impaired glucose tolerance and also from impaired glucose tolerance to diabetes. Based on our proposed model, the low $S_{\max gr}$ is associated with a reduced insulin-dependent

glucose removal effect. This is consistent with the decrease in insulin-stimulated glucose disposal in the study of Pima Indians. But, the estimates of S_p were not significantly different between the progressor and non-progressor group. In addition, the reduced AIR was not found in the progressor group.

AIR is sensitive to the increase in glucose concentration. Although the dose used in IVGTT is adjusted by subject's body weight, the increase in glucose concentration after IVGTT is different in different subjects. In our study, the means of AIR were equal to 825.28 min·mU/L in the progressor group and 577.80 min·mU/L in the non-progressor group. In the study of Xiang *et al.,* the AIR of 86 Hispanic women with prior gestational diabetes was calculated with a mean of 408 min·mU/L (118), which is lower than the mean value reported in our study. This is possibly caused by the different study groups and also the different glucose doses used in IVGTT (0.5mg/kg *vs.* 0.3 mg/kg). S_p in our model is normalized by the increase in glucose concentration after IVGTT, so S_p is an alternative way to quantify the acute insulin response, which is not sensitive to the different glucose doses in IVGTT or different increases in glucose concentration after IVGTT. The estimated insulin removal rate constant k_{ir} is smaller in the progressor group. The first phase of insulin secretion is modeled as a bolus injection, thus k_{ir} is mainly determined by the slope of the curve during the first phase of the insulin secretion. The small insulin removal rate constant may be associated with the slow insulin uptake or utilization.

To investigate the dynamics of glucose-insulin interactions, sixteen combinations of the four different effect functions were evaluated. Although there are many possibilities for describing the concentration-effect relationship, the four functions in this

study were chosen for their easy interpretation and simplicity. The specific parameterizations of the effect functions were designed to get a better stability in the parameter estimation process and get a better estimate of the equilibrium of the glucose-insulin system at steady state. The final model provided the best overall fit to the population data and had relatively few parameters resulting in low Dbar and DIC values.

Figure 2.9a illustrates the glucose removal *vs.* insulin concentration curves of the two groups from simulation based on population parameters of two groups. The glucose removal was considered to be insulin-regulated. According to the model comparison results, the Emax model with a steady state correction was chosen to describe the relationship between insulin level and glucose removal. When insulin concentration is higher than the steady state concentrations, the effect of insulin on glucose removal is determined by the kc_{i50} and $S_{\max gr}$ parameters. However, they are not directly comparable to the various parameters that have been proposed to quantify insulin sensitivity or insulin resistance in other publications (68, 91, 119, 120). The insulin production rate in the reserve pool was modeled by the power proportional function of glucose. The insulin production *vs.* concentration curves of the two groups are shown in Figure 2.9b. Subjects in the progressor group have higher insulin production rate than the subjects in the non-progressor group, but the differences are not statistically significant.

Bayesian based population PK/PD analysis has a similar statistical model structure to the frequentist (non-Bayesian) based analysis. The main difference is the Bayesian approach has prior distributions for the parameters, and the prior distributions are based on experience and belief. Despite the debates between the Bayesian and the frequentist statistical modeling, the concepts of both approaches are accepted and have

been used for decades in the field of statistics. Both types of statistical approaches can be applied to estimate population PK/PD parameter, but the estimation methods are different. WinBUGS is a statistical program widely used in the Bayesian statistical analysis, which applies the Markov Chain Monte Carlo (MCMC) methods to estimate the PK/PD parameters. The algorithm used in WinBUGS offers a very high success rate of convergence which is most helpful in population PK/PD studies (58). In addition the stability of the numeric analysis in the Bayesian approach (WinBUGS) appears considerably better. The Bayesian approach also provides prior information for future studies. In the Bayesian approach, the likelihood function and the Fisher information matrix are calculated without approximation. The population and individuals' parameters can be estimated simultaneously as well as the secondary parameters and predictive intervals. However, running the MCMC methods is computational intensive and typically requires more time than the frequentist approaches.

Study limitations: C-peptide is co-secreted with insulin in an equimolar ratio and has no liver extraction (121). Thus, data of C-peptide is used to estimate the insulin hepatic extraction ratio. Due to lack of C-peptide data, the insulin hepatic extraction ratio is not identifiable and required the use of published values in this study. Thus, to incorporate the hepatic extraction in our model, the mean value of the insulin extraction ratio reported previously was used (76). Similarly, the distribution volume of insulin can not be estimated without an insulin tracer. The volume distribution of insulin was fixed at the mean value of 45 ml/kg reported from a previous study (110). In the propsed model, two scale parameters, S_p and S_{ip}, are sensitive to the fixed distribution volume of insulin and hepatic extraction ratio. Other constant rate parameters and glucose kinetic

parameters are not sensitive to the fixed parameter values. The change in the hepatic glucose production after IVGTT was not addressed in this model due to lack of glucose tracer data. Due to the complexity of the insulin-glucose system some simplifying assumptions are necessary in order to derive a practical, relatively simple model. To not assume a constant endogenous glucose production would significantly confound the analysis.

In the insulin kinetic study of Gupta *et al.*, the insulin disposition was modeled by a simple one-compartmental model (73). The reported estimates of insulin elimination rate constants are 0.124±0.0465 min^{-1} for obese children, 0.113±0.0462 min^{-1} for lean children. The estimate of the insulin removal rate constant was close to the estimate of k_{ir} in the non-progressor group of our study (0.097 min^{-1}). The estimated insulin production rates at steady state were 9.46 and 12 mU/min for the two groups in our study, which is close to the second phase insulin secretion of the healthy volunteers described in Silber's study (9~10 mU/min at the glucose baseline concentration of 88 mg/dL) (102).

In summary, the proposed population-based model has demonstrated the ability to describe the glucose-insulin kinetics by a simple model that address the physiology and molecular biology of insulin's biphasic secretion after IVGTT and the mutual kinetic interactions of glucose and insulin. Although the hepatic extraction of insulin can not be estimated and the change in hepatic glucose production is not addressed in this study, the compact model was able to describe the physiologic features of the glucose-insulin kinetics and insulin's biphasic secretion. The study demonstrates that a Bayesian-based population pharmacokinetic approach can identify early pre-diabetic pharmacokinetic differences and that may be helpful for better prediction of the development of T2D.

Table 2.1. Subject demographics when entering the study

	Progressor group	Non-progressor group	All
Number of subjects	25	127	152
Gender (males in %)	60.0%	42.5%	45.4%
Weight in kg (mean ± sd) range	96.3 ± 29.9 63-193	73.3 ± 17.8 46-136	77.1 ± 21.9 46-193
Height in cm (mean ± sd) range	171.8 ± 9.2 147-191	169.1 ± 11.0 147-206	169.5 ± 10.8 147-206
BMI (mean ± sd) range	32.2 ± 7.6 23-53	25.4 ± 4.3 18-41	26.5 ± 5.6 18-53
Starting age (mean ± sd) range	33.48 ± 8.07 21-50	32.96 ± 10.02 16-59	33.05 ± 9.70 16-59

Table 2.2. Model comparisons.

Effect functions		Model selection criterion	
Glucose removal	Insulin production	Dbar	DIC
$S_{gr}\left(\frac{C_i(t)}{C_{i,ss}}\right)$	$S_{ip}\left(\frac{C_g(t)}{C_{g,ss}}\right)$	37898	38541
	$S_{ip}\left(\frac{C_g(t)}{C_{g,ss}}\right)^{\beta}$	29254	30150
	$k_{ip} + S_{ip}\left[C_g(t) - C_{g,ss}\right]_+$	29328	30233
	$k_{ip} + \frac{S_{max\,ip}\left[C_g(t) - C_{g,ss}\right]_+}{kc_{g50} + \left[C_g(t) - C_{g,ss}\right]_+}$	29237	30117
$S_{gr}\left(\frac{C_i(t)}{C_{i,ss}}\right)^{\alpha}$	$S_{ip}\left(\frac{C_g(t)}{C_{g,ss}}\right)$	29944	30712
	$S_{ip}\left(\frac{C_g(t)}{C_{g,ss}}\right)^{\beta}$	29115	30005
	$k_{ip} + S_{ip}\left[C_g(t) - C_{g,ss}\right]_+$	29200	30102
	$k_{ip} + \frac{S_{max\,ip}\left[C_g(t) - C_{g,ss}\right]_+}{kc_{g50} + \left[C_g(t) - C_{g,ss}\right]_+}$	29146	29914
$k_{gr} + S_{gr}\left[C_i(t) - C_{i,ss}\right]_+$	$S_{ip}\left(\frac{C_g(t)}{C_{g,ss}}\right)$	30138	30902
	$S_{ip}\left(\frac{C_g(t)}{C_{g,ss}}\right)^{\beta}$	29057	29995
	$k_{ip} + S_{ip}\left[C_g(t) - C_{g,ss}\right]_+$	29286	30215
	$k_{ip} + \frac{S_{max\,ip}\left[C_g(t) - C_{g,ss}\right]_+}{kc_{g50} + \left[C_g(t) - C_{g,ss}\right]_+}$	29163	30076
$k_{gr} + \frac{S_{max\,gr}\left[C_i(t) - C_{i,ss}\right]_+}{kc_{i50} + \left[C_i(t) - C_{i,ss}\right]_+}$	$S_{ip}\left(\frac{C_g(t)}{C_{g,ss}}\right)$	28612	29469
	$S_{ip}\left(\frac{C_g(t)}{C_{g,ss}}\right)^{\beta}$	27966[(a)]	28936[(a)]
	$k_{ip} + S_{ip}\left[C_g(t) - C_{g,ss}\right]_+$	28312	29282
	$k_{ip} + \frac{S_{max\,ip}\left[C_g(t) - C_{g,ss}\right]_+}{kc_{g50} + \left[C_g(t) - C_{g,ss}\right]_+}$	28248	29131

[(a)]: Model of choice

Table 2.3. Posterior median and 95% Bayesian credible sets (C.S.) of population kinetic parameters

	Progressor group	Non-progressor group
Parameter	Population estimate (95%C.S.)	Population estimate (95%C.S.)
k_{gp} (mg/min)	95.8 (60.1 - 145)	87.9 (71.4 - 104)
V_g (dL/kg)	2.21 (2.05 - 2.38)	2.22 (2.15 - 2.3)
k_{gr} (1/min)	0.00596 (0.00361 - 0.00919)	0.00718 (0.00577 - 0.00855)
$S_{\max gr}$ (1/min)	0.175 (0.0139 - 0.0226)	0.0230 (0.0203 - 0.026)
kc_{i50} (mU/L)	12.4 (5.41 - 25.9)	6.4 (4.25 - 9.23)
S_{ip} (mU/min)	12 (8.11 - 19)	9.46 (8.02 - 11)
β	1.29 (1.13 - 1.47)	1.3 (1.22 - 1.39)
$k_{i\,sec}$ (1/min)	0.0935 (0.0588 - 0.164)	0.0651 (0.0557 - 0.0766)
k_{ir} (1/min)	0.0484 (0.0356 - 0.0716)	0.097 (0.0843 - 0.11)
S_p (mU·dL/mg)	2.65 (1.88 - 3.7)	2.09 (1.79 - 2.42)

Table 2.4. Inter-individual variance covariance matrix (Ω)

	k_{gp}	V_g	k_{gr}	$S_{\max gr}$	kc_{i50}	S_{ip}	β	$k_{i\,sec}$	k_{ir}	S_p
k_{gp}	0.272									
V_g	-0.0202	0.0331								
k_{gr}	0.241	-0.0451	0.316							
$S_{\max gr}$	0.0863	-0.0231	0.0954	0.145						
kc_{i50}	0.378	-0.0584	0.34	0.235	1.44					
S_{ip}	0.0428	-0.0106	-0.0146	0.022	0.231	0.321				
β	-0.00438	0.0068	-0.0221	0.00664	0.0295	-0.0298	0.0792			
$k_{i\,sec}$	-0.00822	0.0059	-0.0343	0.0483	-0.149	0.0281	0.00644	0.418		
k_{ir}	-0.0756	0.00785	-0.0531	-0.00519	-0.141	0.00656	0.00147	-0.0362	0.122	
S_p	0.159	-0.007	0.0891	0.152	0.541	0.194	0.0546	0.215	-0.0434	0.664

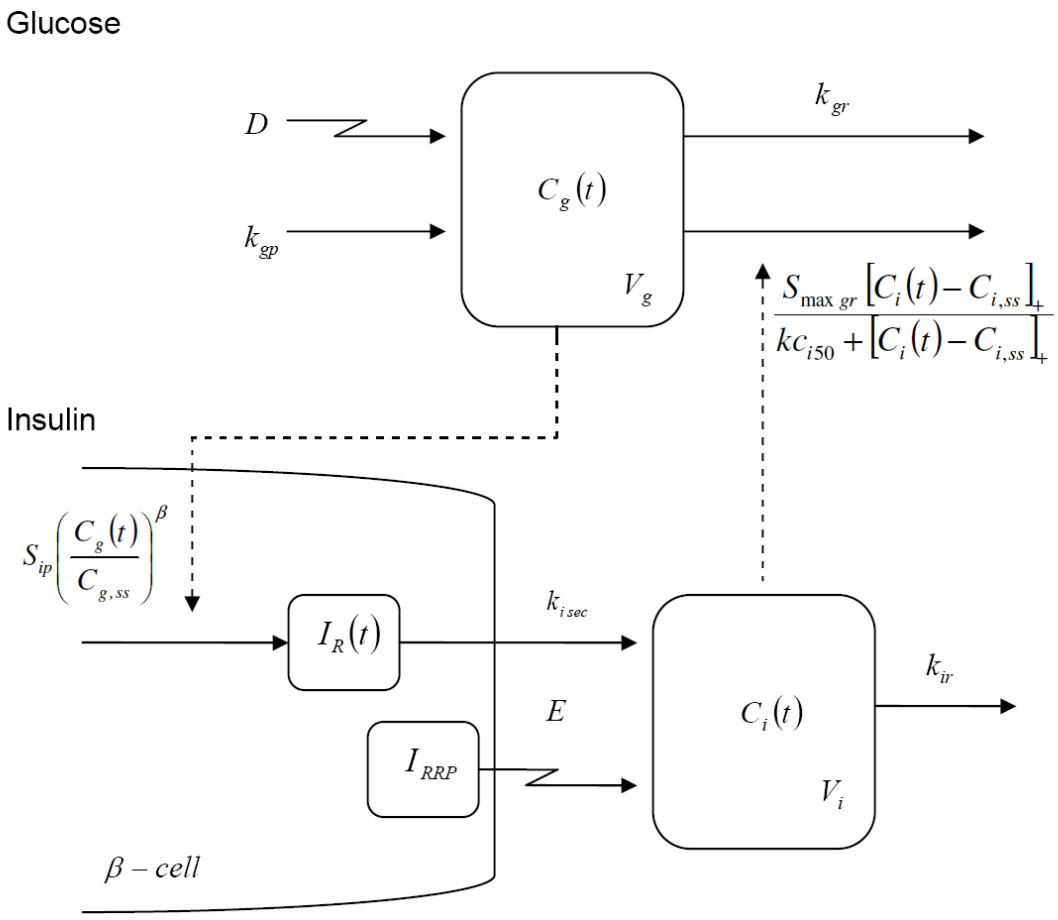


Figure 2.1. The proposed glucose-insulin kinetic model. The dispositions of glucose and insulin were described by one-exponential models. The glucose removal was modeled by an Emax effect function of insulin with a steady state correction, and the insulin production rate was modeled by a proportional power effect function of glucose concentration. The biphasic insulin secretion is controlled by two pools: the readily releasable pool (I_{RRP}) and the reserve pool (I_R).

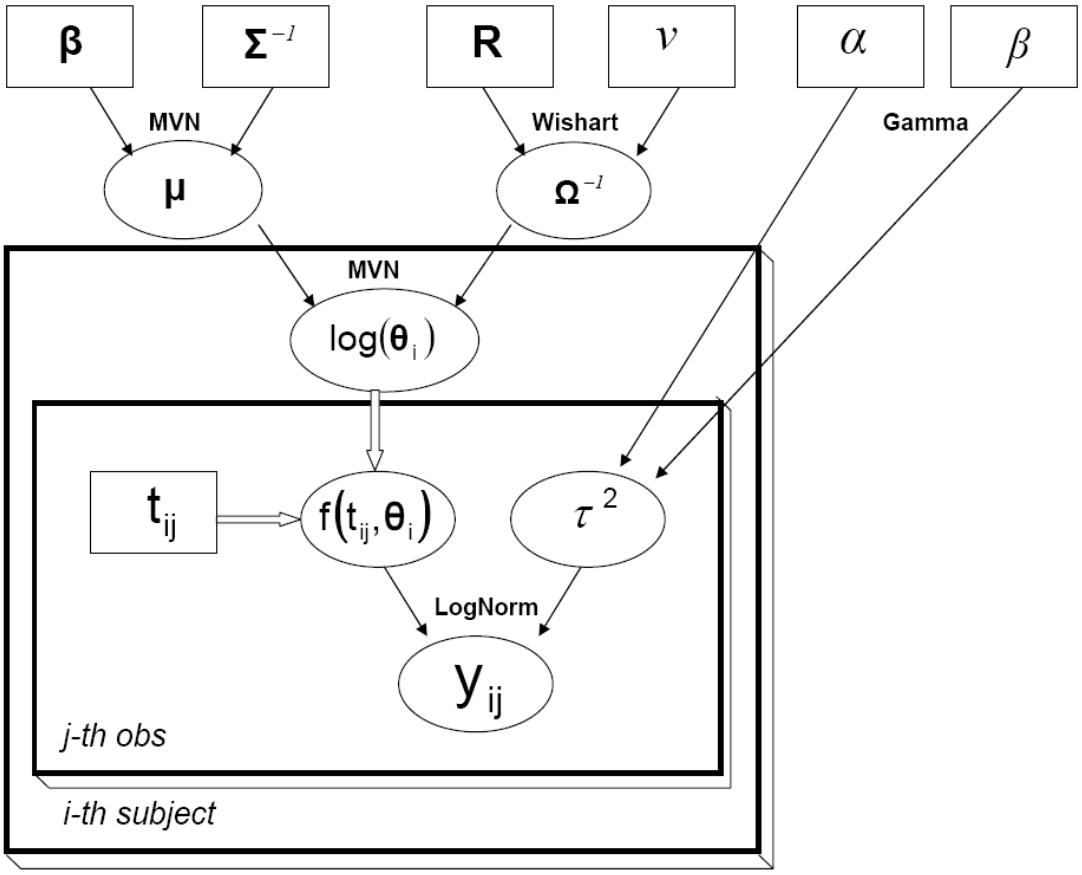


Figure 2.2. Directed plot of Bayesian hierarchical model. The directed plot follows the descriptions in the WinBUGS user manual. The lowest level of hierarchy is shown in the inner plate. The individual's observations have a log-normal distribution with a mean calculated by individual's parameter vector $\boldsymbol{\theta}_i$ and t_{ij}, and a precision parameter τ^2. The second level of the hierarchy is shown in the outer box. Assuming the logarithm of individual parameter vectors are multivariate normal distributed, $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}^{-1}$ represent the population mean and inverse covariance matrix. The Prior distributions in the highest level of the hierarchy are on the pop of the directed plot. The ellipses represent stochastic nodes and the rectangles represent constants.

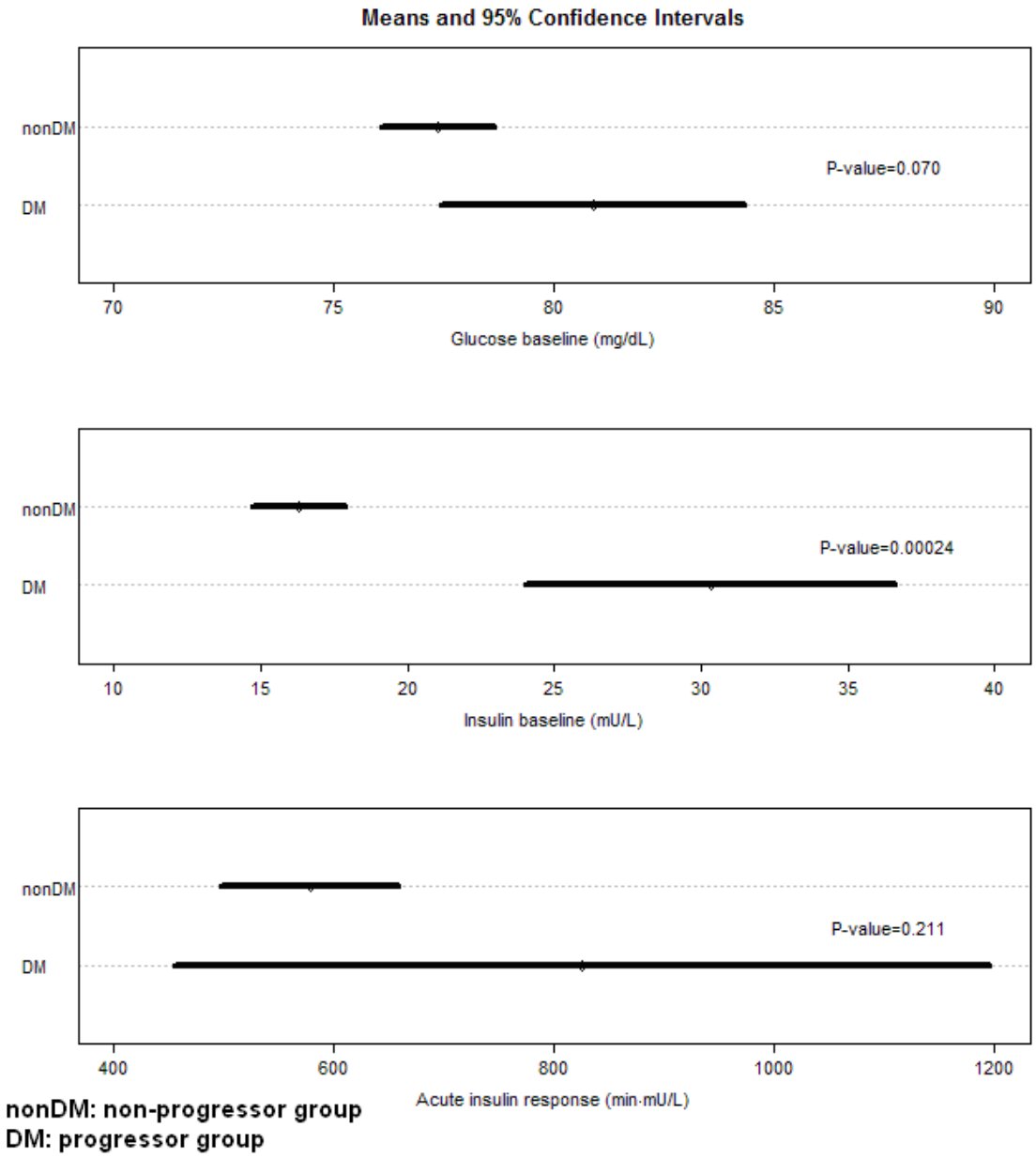


Figure 2.3. Means and 95% confidence intervals for glucose baseline, insulin baseline and acute insulin response of the two groups. The p-values were calculated by the two sample t-tests.

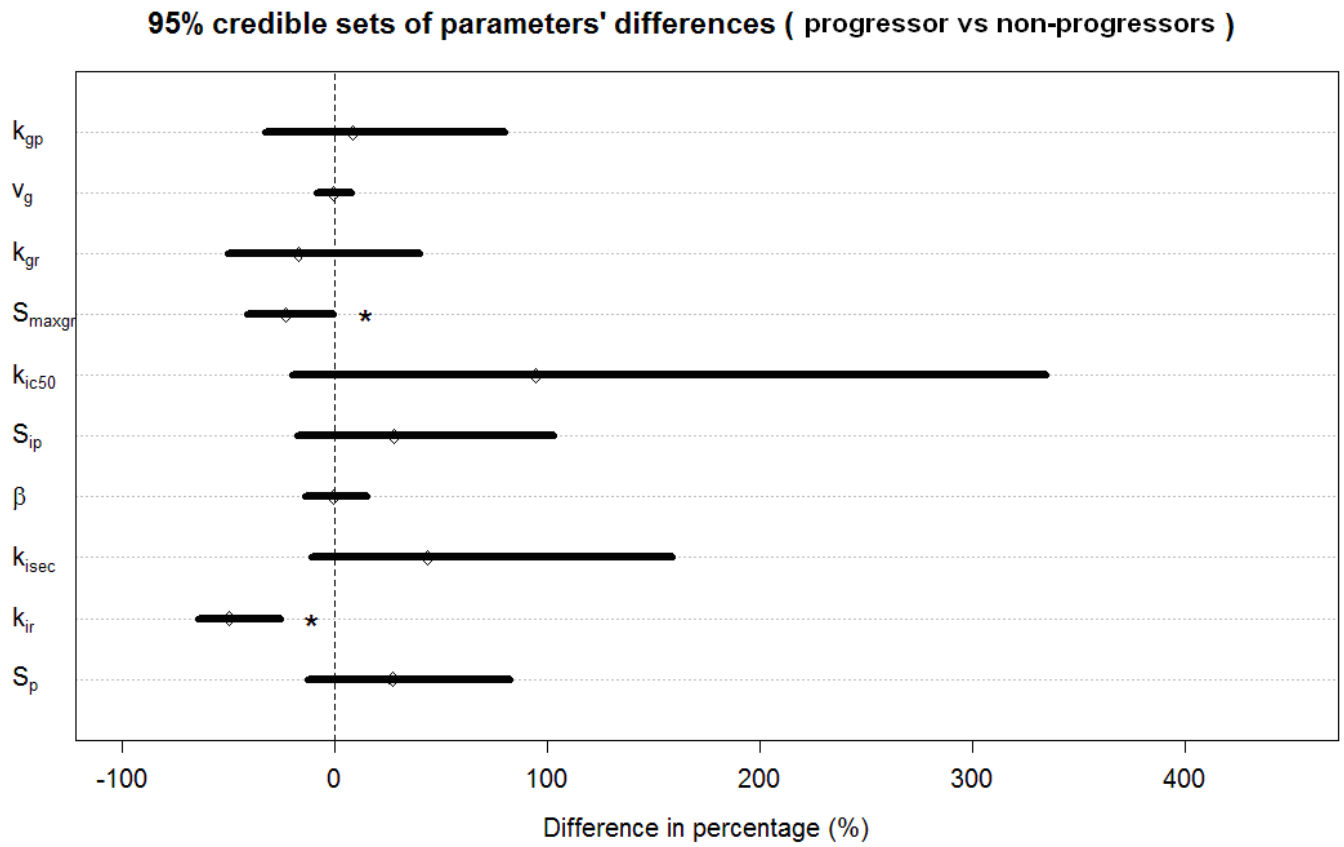


Figure 2.4. 95% Bayesian credible sets of parameters' differences in percentage between the progressor group and the non-progressor group. The diamonds in the figure represent posterior medians of the difference in percentage. The 95% Bayesian credible sets were created according to the posterior distributions. The asterisks indicate the parameters' differences in percentage are statistically significant. (95% Bayesian credible sets do not contain zero)

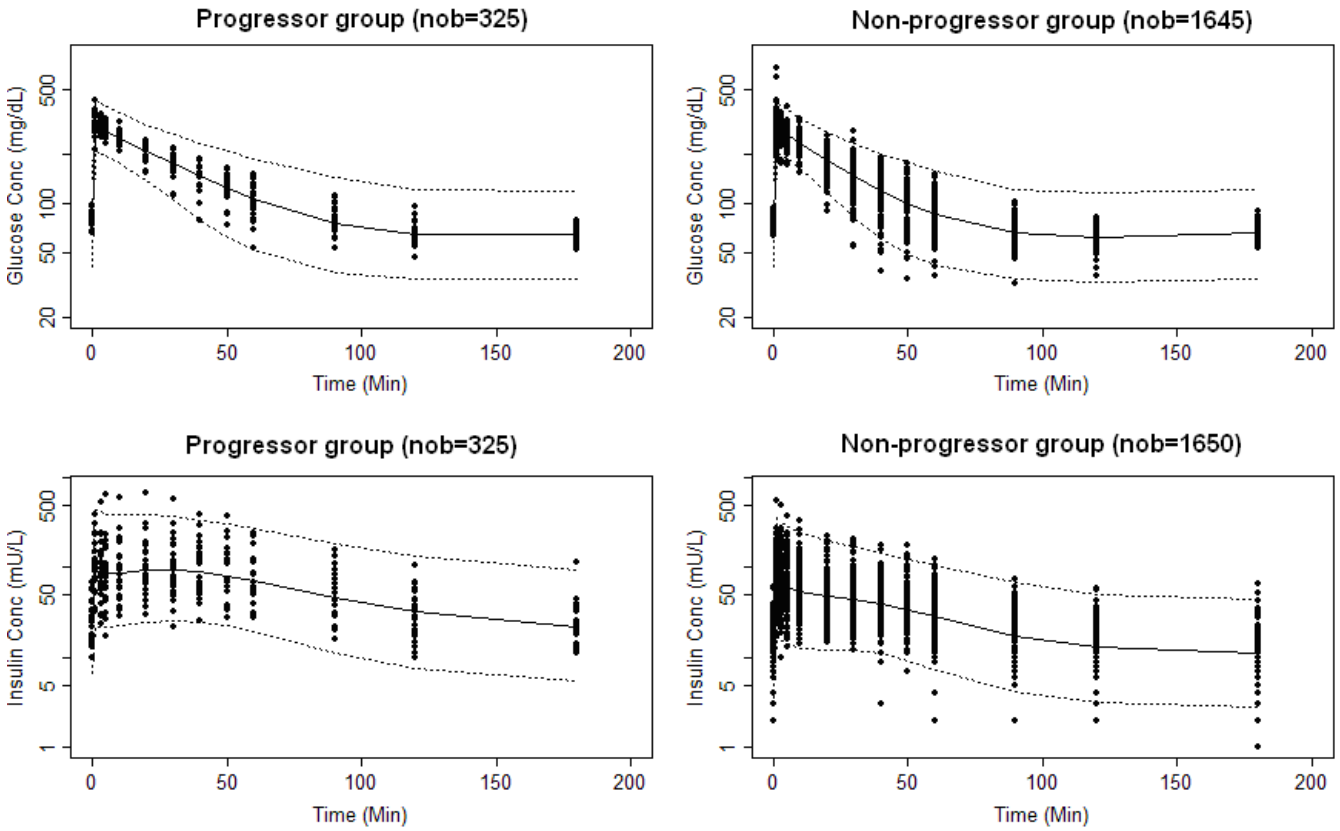


Figure 2.5. Population model predictions and observed concentrations of glucose and insulin for the two groups. The solid lines represent the population prediction curves. The broken lines represent the Bayesian 95% predictive intervals. (nobs: number of observations)

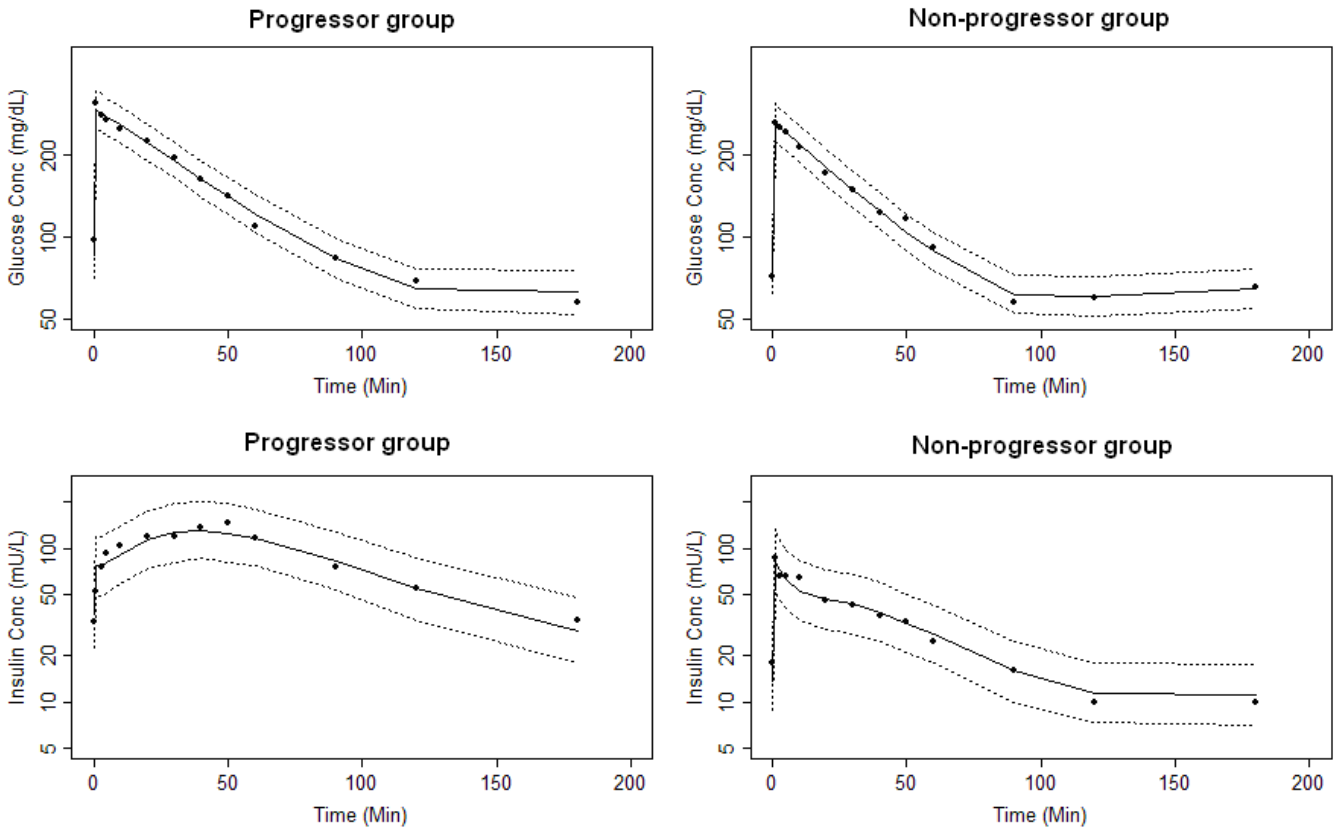


Figure 2.6. Individual predictions and observed concentrations of glucose and insulin for two subjects in the two groups. The solid lines represent the individual fitted curves. The broken lines represent the Bayesian 95% predictive intervals.

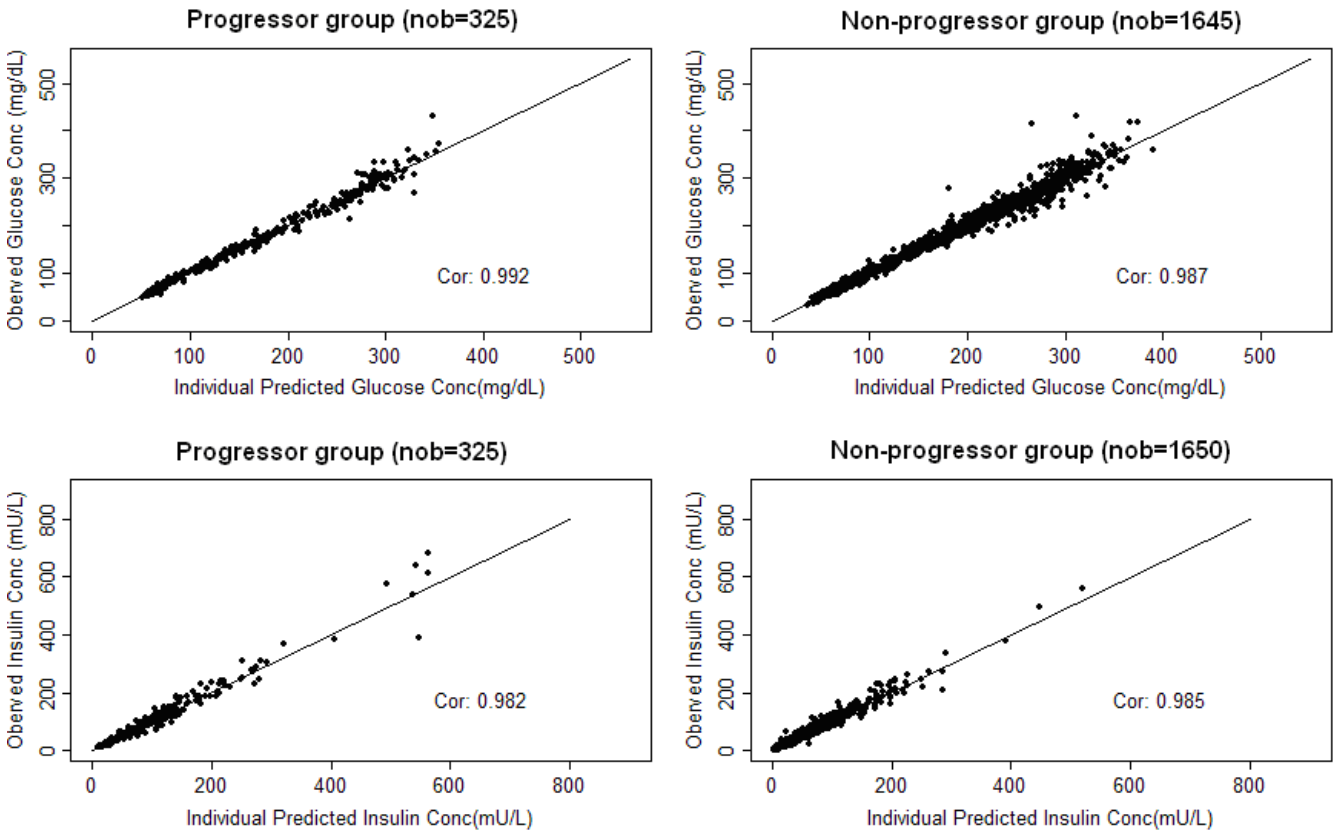


Figure 2.7. Individual predicted *vs.* observed concentrations for glucose and insulin of the two groups. The solid lines represent the identity lines. The correlation coefficients are estimated and all of them are larger than 0.98.

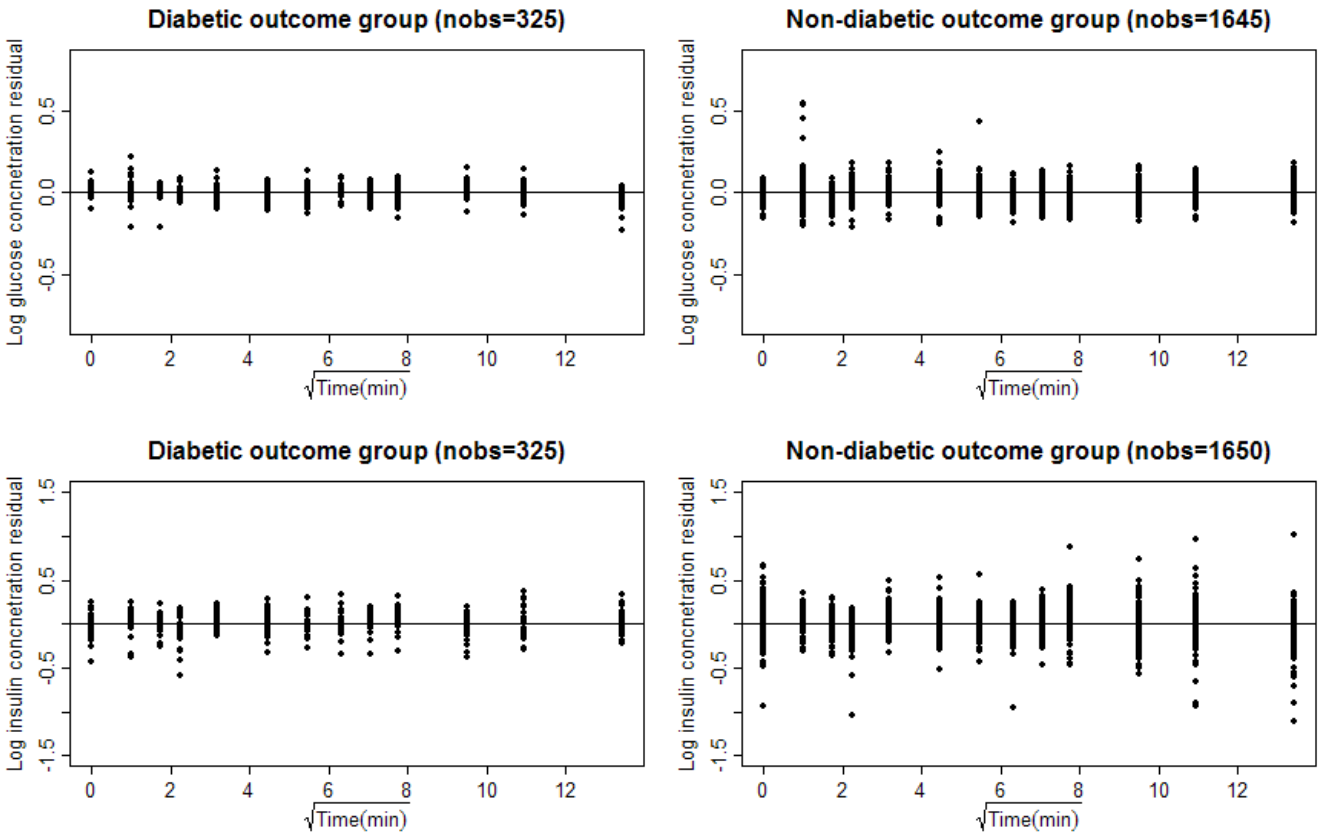


Figure 2.8. Residual plots of the logarithm of glucose and insulin concentrations for progressor and non-progressor groups. The scale of the x-axis is square root of time to better visualize the early sampling time.

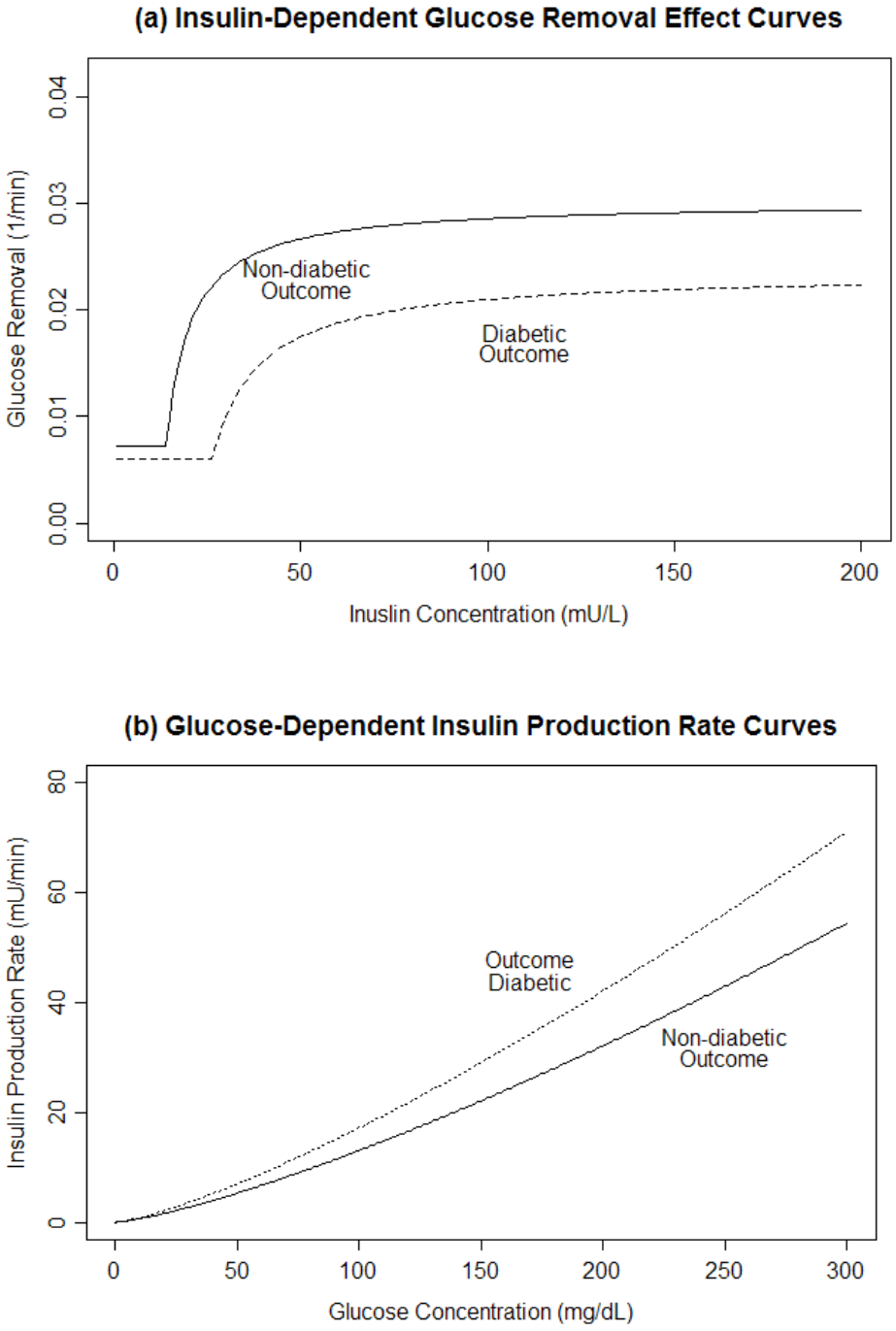


Figure 2.9. Simulations of glucose and insulin regulations based on population parameters from two groups (a)Plot of insulin's glucose removal effect *vs.* insulin concentration of the two groups. (b) Plot of insulin production rate *vs.* glucose concentration of the two groups.

CHAPTER 3. A BAYESIAN POPULATION ANALYSIS OF THE DEVELOPMENT OF TYPE 2 DIABETES IN THE OFFSPRING OF DIABETIC PARENTS

3.1 Abstract

Disease progression of type 2 diabetes (T2D) has received considerable attention, but little is known about the disease development of T2D. The purposes of this study were to identify disease development variables (DDV) for development of T2D and to compare corresponding models for disease development. All subjects included in this study were the offspring of diabetic parents and were followed up to 25 years. Repeated fasting blood samples were collected during the follow-up. Longitudinal data of four DDVs, namely fasting blood glucose (FBG), fasting serum insulin (FSI), homeostatic model assessment of insulin resistance (HOMA-IR) and body mass index (BMI) were recorded and compared. According to the diabetes status at the end of the follow-up, the data analysis involved a progressor group of 25 subjects, and a non-progressor group of 127 subjects. The temporal changes in the four DDVs over the time course of the disease development were evaluated by a single-slope and a two-slope population-based Bayesian model. A two-slope model based on FBG was found to be the best disease development model. For non-progressors, the FBG baseline stayed at 69.2 [66.5, 72.1] mg/dL (Bayes estimate [95% Bayesian credible set]) and increased with age by a rate of 0.227mg/dL [0.149, 0.3] per year. For the progressors, the FBG increase with age the same rate as non-progressors and started to have an additional increase of 2.27 [0.505, 4.52] mg/dL per year, starting 8.73 [-10.8, -6.93] years before the diagnosis of T2D. No significant longitudinal increasing or decreasing temporal pattern was found for FSI,

HOMA-IR and BMI by the population-based Bayesian approach. The proposed model, which enables a quantitative, time-based evaluation of the development of T2D in this higher risk population, may be used to quantify the effect of interventions/prevention strategies such as drug treatment and lifestyle changes.

3.2 Introduction

About 24 million people in the United States (7.8% of the population) were reported to have diabetes in 2007 (122). The total cost of diabetes in 2007 was $174 billion including $116 billion in excess medical expenditures, $31 billion in general medical costs, $27 billion directly for treating diabetes and $58 billion for treating the complications of diabetes (123). The prevalence of diabetes is still increasing. The number of adults with diabetes in the world is predicted to rise from 135 million in 1995 to 300 million in the year 2025 (4). Approximately 90-95% of the whole U.S. diabetes population are affected by type 2 diabetes (T2D) (122). Although T2D has been studied for many decades, currently there is still no cure. Prevention is one of the most effective ways to reduce the incident rate and cost of the disease. Type 2 diabetes is a chronic disease, and the development of the disease involves years of pre-diabetic stage. Many factors had been reported that increase the risk of diabetes by American Diabetes Association (ADA) including family history of diabetes, obesity, race, previously identified impaired glucose tolerance/or impaired fasting glucose, hypertension, low HDL cholesterol level, high triglyceride level, and history of gestational diabetes mellitus (37). However, most of these risks are categorical time-invariant variables and provide little information about the changes of the risk factors over the development process of T2D.

Another important factor associated with the development of T2D is insulin resistance (124). The natural history of insulin resistance in the pathogenesis of T2D was studied by Weyer *et al.* (38). In that study, insulin resistance was evaluated by three variables: rate of glucose disposal, acute insulin response, and endogenous glucose output. These variables were evaluated from normal to impaired glucose tolerance to diabetes over 5.1 ± 1.4 years. The decreases in acute insulin response and in insulin-stimulated glucose disposal were found to be significant during the developing process of T2D. HOMA-IR and BMI have also been reported to be associated with the development of T2D (91, 125-128). HOMA-IR is developed based on the physiology of glucose and insulin regulations in the organs and tissues at homeostasis state and widely used to evaluate insulin resistance in clinical and epidemiological researches (91), and BMI is one of most extensive used index to evaluate the level of obesity (129). The high insulin resistance and BMI have been found to be risk factors in many retrospective studies (38, 130-132). In the studies of Warram *et al.*, 6 selected variables repeatedly measured 15 years before the diagnosis of T2D were compared in five individuals (39). Fasting glucose, insulin, and 2-hour glucose post-challenge increased within 5 years before the diagnosis of T2D while fasting triglycerides began rising at least 10 years before the diagnosis of T2D. As a whole the group was obese from the start of observation, but this did not change during follow-up.

The PK/PD risk factors for development of T2D were investigated previously (133). A high insulin baseline concentration, a low maximum insulin-dependent removal and a low insulin removal rate constant were found to be associated with the development of T2D. Although these variables are known to increase the risk of T2D, the relationships

between the time course of the development of T2D and the magnitude and change in these risk factors were not determined. Recently, a disease progression model was developed by de Winter *et al.* to assess the effects of treatments over the time-course of the progression of diabetes (99). The effects of pioglitazone, metformin and gliclazide were modeled as treatment effect functions of time based on measurements of fasting plasma glucose, fasting serum insulin and glycated hemoglobin A_{1c} (A1C). However, that study only focused on subjects who already had been diagnosed with T2D.

In contrast to the *progression* of the disease after the disease has been diagnosed, the objective of the chapter is to develop a model to describe the *development* of T2D. With the additional longitudinal data of fasting blood glucose, fasting serum insulin, HOMA-IR, and BMI, the current disease development modeling is an extension of previous analysis. Specifically, the disease development model in the present work is aimed at a quantitative longitudinal evaluation of the developing stage of the disease, which hopefully may provide a mechanistic and quantitative basis for evaluating the effect of interventions/prevention strategies of T2D such as drug treatment and lifestyle changes.

3.3 Specific aim and hypothesis

The specific aim of this chapter is to develop a population-based methodology to describe the development of type 2 diabetes based on four important variables namely, fasting blood glucose, fasting serum insulin, homeostatic model assessment of insulin resistance and body mass index and to identify important temporal patterns and time-dependencies of these disease development variables over the disease development period.

The hypothesis is that the specific temporal patterns and time-dependencies of the four important disease development variables are hypothesized to exist and can be identified by applying population-based disease development model.

3.4 Method

3.4.1 Subjects

The study is based on a 25-year follow-up study in Joslin Diabetes Center, Boston, Massachusetts (24). Between 1963 and 1983, 152 healthy off-spring of diabetic parents were recruited in the Joslin Diabetes center. None had diabetes or impaired glucose tolerance. The participants were surveyed to evaluate the status of T2D up to 25 years. During the follow-up, fasting blood glucose concentration (FBG), fasting serum insulin concentration (FSI), body mass index (BMI) and age of the participants were repeatedly collected on all participants in the Joslin Diabetes center. At the end of the follow-up, 25 subjects had developed T2D. All of the subjects were separated into a progressor group and a non-progressor group resulting in group sizes of 25 and 127 respectively. The characteristics of the progressor and the non-progressor groups are summarized in Table 3.1.

In the progressor group, the ages of entering the study ranged from 21 to 50 years and the ages of diagnosis of T2D ranged from 35 to 63 years. Subjects in the progressor group have an average of 6.4 fasting blood tests performed during the follow-up. The follow-up and the ages of diagnosis of T2D for subjects in the progressor group is shown in Figure 3.1. In the non-progressor group, the ages of entering the study ranged from 13 to 59 years. Subjects in the progressor group have an average of 5.8 fasting blood tests

performed. The follow-up and data collection over time in the non-progressor group is summarized in Figure 3.2.

3.4.2 Fasting blood test

All participants were instructed to consume a high-carbohydrate diet (250 to 300 g/day) for 3 days before the fasting blood test. The fasting blood samples of the participants were collected the next morning after over night fasting. Blood glucose and serum insulin concentrations were determined from the fasting blood samples. For the fasting blood tests done before 1983, the blood glucose concentrations were measured by the ferricyanide method with a coefficient of variation (CV) of 1.5% (117). Thereafter, the glucose oxidize method was used with a CV of 1.35% (134). The insulin serum concentrations were measured by a double-antibody radio-immunoassay with a 17.6% CV (109).

3.4.3 Preliminary, naïve pooled data approach

The temporal patterns of the four disease development variables (DDVs), namely fasting blood glucose concentration (FBG), fasting serum insulin concentration (FSI), homeostatic model assessment of insulin resistance (HOMA-IR) and body mass index (BMI) were analyzed by the naïve pooled data approach. The four DDVs were defined as follows:

$$\text{FBG}: \quad \text{fasting blood glucose concentration (mg/dL)}$$

$$\text{FSI}: \quad \text{fasting serum insulin concentration (mU/L)}$$

$$\text{HOMA - IR}: \quad \frac{\text{fasting blood glucose (mg/dL)} \times \text{fasting serum Insulin (mU/L)}}{405}$$

$$\text{BMI}: \quad \frac{\text{weight(kg)}}{\text{height(m)}^2}$$

In the naïve pooled approach, all the data of progressors were considered as collected from one individual and pooled together. Then, the relationship between the DDVs and the time prior to diagnosis of T2D (TIMEtoD) were evaluated by a generalized cross-validation cubic spline approach. TIMEtoD was used to account for the time-dependence of the DDVs in the progressors prior to the diagnosis of the disease and calculated as the age of the fasting blood test minus the age of diagnosis of T2D. The generalized cross-validation cubic spline approach was used to summarize the trend of the DDVs for its advantages of minimum oscillatory behavior and continuously differentiable (135). The smoothness of the cubic spline functions were optimized by generalized cross-validation (135). The fitted curves in Figure 3.3 suggest that the temporal patterns of the DDVs may be described by a single-slope linear model or a two-slope linear model and these two models were tested in the later Bayesian hierarchical analyses. The single-slope model is used to evaluate the direction and rate of the temporal change of the DDV over TIMEtoD, and the two-slope model is the extension of the single-slope model to provide a basis for assessing the start of the deterioration of disease development.

3.4.4 Bayesian hierarchical approach

Both the single-slope and the two-slope model were subsequently applied in a more in-depth analysis employing a population-based Bayesian hierarchical approach. First, the four DDVs (FBG, FSI, HOMA-IR, and BMI) were fitted to a single-slope linear model. The statistical models of the single-slope model of DDV shown as follows for non-progressor and progressor group:

Non - progressor group

$$\mathrm{DDV}_{ij} = k_{base,i} + k_{age,i} \cdot AGE_{ij} + \varepsilon_{ij} \quad (3.1)$$

Progressor group

$$\mathrm{DDV}_{ij} = k_{base,i} + k_{pg,i} \cdot PG_i + k_{age,i} \cdot AGE_{ij} + k_{TIMEtoD,i} \cdot TIMEtoD_{ij} + \varepsilon_{ij} \quad (3.2)$$

The predictor variable, age (AGE), was included in both models to evaluate the natural progression of the DDV over age in the progressor and non-progressor group. The variables, progressor indicator (PG, Progressor:PG=1, Nonprogressor:PG=0) and time to diagnosis (TIMEtoD), were included in the progressor group model so that progressors can be differentiated from the non-progressors after adjusting the age effect. The model (3.2) contains four parameters: a baseline parameter (k_{base}), a progressor effect parameter (k_{pg}), an age effect parameter (k_{age}), and a TIMEtoD effect parameter ($k_{TIMEtoD}$). ε_{ij} represents the residual error from the j^{th} observation of the i^{th} subject. The baseline (k_{base}) and age effect (k_{age}) on DDV are mainly determined by the data of the non-progressor group. The progressor group effect (k_{pg}) and TIMEtoD effect ($k_{TIMEtoD}$) are determined by the extra increasing or decreasing of DDV over TIMEtoD before diagnosis of T2D in the progressor group after adjusting the age effect.

Secondly, the data of DDVs were fitted to a two-slope linear model which is the extension from the single slope model. According to the preliminary naïve pooled data analysis in Figure 3.3, FBG showed a clear increasing pattern about 10 years before diagnosis of T2D. Therefore, FBG was fitted by the following two-slope model:

FBG:

Non - progressor group

$$DDV_{ij} = k_{base,i} + k_{age,i} \cdot AGE_{ij} + \varepsilon_{ij} \quad (3.3)$$

Progressor group

$$DDV_{ij} = k_{base,i} + k_{pg,i} \cdot PG_i + k_{age,i} \cdot AGE_{ij} + k_{TIMEtoD2,i} \cdot \left(TIMEtoD_{ij} - k_{sep}\right)_+ + \varepsilon_{ij}$$

$$\left(a\right)_+ = \begin{cases} a, \text{if } a > 0 \\ 0, \text{if } a \le 0 \end{cases} \quad (3.4)$$

In Figure 3.3, FSI, HOMA-IR, and BMI showed an increasing pattern at the early stage of the development of the disease, and they were fitted by the following two-slope model:

FSI, HOMA-IR, and BMI:

Non - progressor group

$$DDV_{ij} = k_{base,i} + k_{age,i} \cdot AGE_{ij} + \varepsilon_{ij} \quad (3.5)$$

Progressor group

$$DDV_{ij} = k_{base,i} + k_{pg,i} \cdot PG_i + k_{age,i} \cdot Age_{ij} + k_{TIMEtoD1,i} \cdot \left(TIMEtoD_{ij} - k_{sep}\right)_- + \varepsilon_{ij}$$

$$\left(a\right)_- = \begin{cases} a, \text{if } a < 0 \\ 0, \text{if } a \ge 0 \end{cases} \quad (3.6)$$

The two slopes in the two-slope model are separated by a separation-point parameter (k_{sep}). For FBG, the extra increase over TIMEtoD after k_{sep} is described by a slope parameter $k_{TIMEtoD2}$. For FSI, HOMA-IR, and BMI, the early increase before k_{sep} is described by the parameter $k_{TIMEtoD1}$.

3.4.5 Bayesian statistics

For consistency, the statistical estimation methods used in the analysis of the model and the parameterization were the same for all DDVs. The two-slope model of FBG was used to illustrate the Bayesian statistical analysis. The hierarchy of the two-slope Bayesian model and the distribution assumptions are summarized as follows:

Prior level

$$\begin{aligned}
&\tau \sim gamma(0.01, 0.01) && \sigma = 1/\sqrt{\tau} \\
&\mu_{base} \sim normal(0,\ 1/1000^2) && \sigma_{base} \sim uniform(0,1000) \\
&\mu_{age} \sim normal(0,\ 1/1000^2) && \sigma_{age} \sim uniform(0,1000) \\
&\mu_{pg} \sim normal(0,\ 1/1000^2) && \sigma_{pg} \sim uniform(0,1000) \\
&\mu_{TIMEtoD2} \sim normal(0,\ 1/1000^2) && \sigma_{TIMEtoD2} \sim uniform(0,1000) \\
&\mu_{sep} \sim uniform(-17.97, -2.33) && \sigma_{sep} \sim uniform(0,1000)
\end{aligned} \tag{3.7}$$

Population level

$$\begin{aligned}
&k_{base,i} \sim normal(\mu_{base},\ \tau_{base}) && \sigma_{base} = 1/\sqrt{\tau_{base}} \\
&k_{age,i} \sim normal(\mu_{age},\ \tau_{age}) && \sigma_{age} = 1/\sqrt{\tau_{age}} \\
&k_{pg,i} \sim normal(\mu_{pg},\ \tau_{pg}) && \sigma_{pg} = 1/\sqrt{\tau_{pg}} \\
&k_{TIMEtoD2,i} \sim normal(\mu_{TIMEtoD2},\ \tau_{TIMEtoD2}) && \sigma_{TIMEtoD2} = 1/\sqrt{\tau_{TIMEtoD2}} \\
&k_{sep,i} \sim normal(\mu_{sep},\ \tau_{sep})I(-17.97, -2.33) && \sigma_{sep} = 1/\sqrt{\tau_{sep}}
\end{aligned} \tag{3.8}$$

Individual Level

Non - progressor group : (3.9)

$$\mathrm{FBG}_{ij} = k_{base,i} + k_{age,i} \cdot AGE_{ij} + \varepsilon_{ij}$$

Progressor group

$$\mathrm{FBG}_{ij} = k_{base,i} + k_{pg,i} \cdot PG_i + k_{age,i} \cdot AGE_{ij} + k_{TIMEtoD2,i} \cdot \left(TIMEtoD_{ij} - k_{sep}\right)_+ + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim normal(0, \tau)$$

In the highest level of the hierarchy, prior distributions were given to the population mean, population precision, and the precision (reciprocal of variance) of residuals. Vague normal prior distributions were given to the population mean parameters (μ_{base}, μ_{age}, μ_{pg}, and $\mu_{TIMEtoD2}$). In order to prevent a local convergence at the extremes which results in a false collapse of the two-slope model to a single-slope model, an uniform prior distribution was assigned to the population break-point parameter (μ_{sep}) with a range between -17.97 and -2.33 corresponding to the 10% and 90% percentiles of

TIMEtoD. Vague uniform prior distributions were given to the population precision, σ_{base}, σ_{age}, σ_{pg}, $\sigma_{TIMEtoD2}$, and σ_{sep}. A vague gamma distribution was given to the precision (τ) of the residuals. Uniform and gamma distributions are commonly used as vague or non-informative priors in the literature for Bayesian statistical analysis (136). In the population level of the hierarchy, the individual parameters, $k_{base,i}$, $k_{age,i}$, $k_{pg,i}$, $k_{TIMEtoD2,i}$, and $k_{sep,i}$ were summarized by the normal distributions with means of μ_{base}, μ_{age}, μ_{pg}, $\mu_{TIMEtoD2}$, and μ_{sep}, and precisions of τ_{base}, τ_{age}, τ_{pg}, $\tau_{TIMEtoD2}$, and τ_{sep}, respectively. The precisions were the reciprocals of σ^2_{base}, σ^2_{age}, σ^2_{pg}, $\sigma^2_{TIMEtoD2}$, and σ^2_{sep}, which summarized to the inter-individual variability of $k_{base,i}$, $k_{age,i}$, $k_{pg,i}$, $k_{TIMEtoD2,i}$, and $k_{sep,i}$. In the individual level of the hierarchy, the data was fitted to a two-slope model and the residuals ($\varepsilon_{i,j}$) were assumed to be normal distributed with a mean of zero and a precision of τ, where *j* denoted the sequencing of the observations of the i^{th} individual. Normal distribution has been extensively used in the Bayesian model because of its great simplicity and easy interpretation.

3.4.6 Programs and algorithms

All of the posterior distributions of the parameters were estimated by the Markov chain Monte Carlo method in WinBUGS version 1.4.3 (60). Three Markov chains were run simultaneously for 20,000 samples for each parameter. The convergence of the chains was accessed by the Gelman-Rubin statistic (137). After burn-in with the first 10,000 samples in each chain, a total of 30,000 samples from the three chains were used to estimate the posterior distribution of each parameter. The mean of the parameter's

posterior distribution is defined as the Bayes estimates for the parameter. The 95% Bayesian credible set (95%C.S.) is the interval between the 2.5 and 97.5 percentile of the posterior distribution. The deviance information criteria (DIC) of both single-slope and two-slope models were estimated for model comparison (138). The graphs were generated by R version 2.9.0 (139).

3.5 Results

3.5.1 Bayesian hierarchical approach

First, the data of the four DDVs were fitted by a single-slope model. The summary of Bayesian population parameter estimates and the corresponding 95% Bayesian credible sets (95%C.S.) are shown in Table 3.2. The Bayes estimate of $\mu_{TIMEtoD}$ for FBG was 0.974 [0.551, 1.41] (Bayes estimate [95%C.S.]) mg/dL per year. The significant positive $\mu_{TIMEtoD}$ indicated there was an increasing pattern in FBG over TIMEtoD for the progressors. The estimates of $\mu_{TIMEtoD}$ for FSI, HOMA-IR, and BMI were not significantly different from 0 (95%C.S.s contain 0).

Secondly, the four DDVs were fitted by the two-slope model. The Bayesian estimates of the population means (μ_{base}, μ_{age}, μ_{pg}, $\mu_{TIMEtoD1}$, μ_{sep}, and $\mu_{TIMEtoD2}$) and 95% C.S. for the four DDVs are shown in Table 3.3. The further finding in the two-slope model about FBG was that the estimate of μ_{sep} was -8.73 [-10.8, -6.93] year and $\mu_{TIMEtoD2}$ was significant positive with an estimate of 2.27 [0.505, 4.52] mg/dL per year after the separation point. The population age effect parameter, μ_{age}, was significant with an estimate of 0.227 [0.149, 0.3] mg/dL per year, the estimate of μ_{base} is 69.2 [66.5, 72.1]

mg/dL, and the estimate of μ_{pg} was 1.41 [-1.74, 4.49] mg/dL. The parameter estimates suggest that the temporal pattern of FBG in the progressor group started at 70.6 mg/dL ($\mu_{base} + \mu_{pg}$) and increased by a rate of 0.227 mg/dL (μ_{age}) per year with age, and then at 8.73 years (μ_{sep}) years before the diagnosis of T2D FBG increased by a rate of 2.5 mg/dL ($\mu_{age} + \mu_{TIMEtoD2}$). The estimated DICs of the single-slope and two-slope model of FBG were 5988.6 and 5942.5, respectively, which suggests the two-slope model is more adequate to describe the temporal pattern of FBG than the single-slope model and able to further address the separation time the two slopes, which gives valuable information about the starting time of the deterioration of the disease. For FSI, HOMA-IR, and BMI, the estimates of $\mu_{TIMEtoD1}$ are not significant different from 0.

The two-slope development pattern of FBG in the progressor group and the natural age progression in the non-progressor group are shown in Figure 3.4. For proper reference the TIMEtoD is used as the x-axis of the progressor plot to illustrate the effect of TIMEtoD before diagnosis of T2D. Age is used as the x-axis for the non-progressors to show the natural age progression of T2D. The x-axes of the two plots have the same graphical time scaling to enable a proper visual comparison of the two groups. The overall two-slope development pattern is readily observed in the plot. For the subjects in the non-progressor group, only the baseline of FBG and the natural age progression of FBG were modeled. TIMEtoD is not available in the non-progressor group, so these subjects' data was mainly used as the background comparison between the two groups over age. Representative individual fits of two data-rich progressors are shown in Figure 3.5 to illustrate the two-slope pattern. The inter-individual variability of the individual

parameters ($k_{base,i}$, $k_{pg,i}$, $k_{age,i}$, $k_{sep,i}$, and $k_{TIMEtoD2,i}$) of FBG are summarized by σ_{base}, σ_{pg}, σ_{age}, σ_{sep}, and $\sigma_{TIMEtoD2}$. The parameter estimates of inter-individual variability are shown in Table 3.4. The Bayes estimate of the standard deviation of residual σ for FBG is 6.29 mg/dL [5.97, 6.63] and the residual plot of the individual fit of FBG is shown in Figure 3.6. The variability of the residual is consistent over age and follows a normal distribution.

No significant longitudinal increasing or decreasing was found in FSI, HOMA-IR and BMI over TIMEtoD before the diagnosis of T2D. The single-slope and two-slope models are not applicable. The significant population progressor effect (μ_{pg}) in FSI and BMI (Table 3.3) indicate the progressors have a higher FSI and BMI than the non-progressors. The population age effect parameters of FBG, FSI, and BMI are all significant different from zero. Thus, the age effect on the nature disease progression is needed to be accounted in the model.

3.6 Discussion

Risk factors of T2D have been investigated in many cross-sectional and prospective studies (25, 128, 130). Often, these risk factors are only measured once and treated as time invariant variables. However, some of the risk factors are actually time-variant, such as glucose, insulin concentrations, and HOMA-IR, which are changing along with the development of T2D. Knowing the temporal changes of these time-variant variables is helpful to understand the pathophysiology of T2D and better evaluate the risk of developing diabetes. In the studies by Warram *et al*. and Weyer *et al*., several selected DDVs were analyzed by the average, pooled data approach to investigate the temporal

changes of the DDVs during development of T2D (38, 39). The advantage of this approach is its simplicity. It is easy to visualize the general temporal changes of the averages of the DDVs at different stages. However, due to the effect of averaging, the data means may not represent the individual data, and the results may be greatly biased by the inter-subject variability (52). In this study the temporal patterns of DDVs over the development of T2D were investigated based on additional longitudinal repeatedly measured fasting blood samples, BMI and age.

First, the temporal patterns of the DDVs were summarized by the pooled data analysis to identify possible models for the disease development. These possible models were then further investigated and tested by the Bayesian hierarchical approach. The Bayesian approach enables estimation of individual and population parameters and inter-individual variability and avoids the potential artifacts and bias associated with using averaging based on pooled data.

The results of Bayesian analysis on FBG suggest that there is a significant 2.5 mg/dL increase per year in FBG starting 8.73 years before the diagnosis of T2D. In contrast, the study of Bleich *et al.* showed that the fasting glucose concentration starts rising 1.5 years before the diagnosis of type 1 diabetes (140). This indicates that the fasting glucose follows a two-slope progression pattern in both type 1 and 2 diabetes, but the rise in the fasting glucose concentration starts much earlier in T2D. The development of T2D is characterized by insulin resistance early in the course and then followed by beta-cell dysfunction (5). Those who eventually developed T2D may have insulin resistance for years without having substantial hyperglycemia due to the compensation of the hyper-secreted insulin from beta-cells (141). The failure of beta-cell to compensate

for insulin resistance causes the beta-cell dysfunction and hyperglycemia. The study of Harris *et al.* indicated the onset of T2D was started at least 4-7 years before clinical diagnosis based on linear extrapolation of percentage of retinopathy.(142). Our study shows that the rise in FBG can be used to identify the subjects of high risk of T2D earlier than retinopathy and can assess the starting time of beta-cell dysfunction. According to the current diagnostic criteria of T2D of Would Health Organization (WHO) and ADA, the stages of T2D can be separated into normal glucose tolerance (NGT), impaired glucose tolerance (IGT), and T2D based on the fasting glucose test. Although the standard NGT and IGT classification has been widely used, it provides no information about the temporal changes of the biomarker (FBG) during the development of T2D. The current proposed longitudinal model can identify the important temporal pattern of FBG and can be used to quantify the effect of interventions on FBG, a primary biomarker of T2D, in T2D prevention trials.

Insulin resistance and obesity are both conditions associated with T2D (130, 143). In this study, insulin resistance was evaluated by the indexes of FSI and HOMA-IR, and obesity was evaluated by BMI. Table 3.1 shows the starting BMI in the progressor group is significantly higher than the non-progressor. This suggests high BMI is a risk factor of T2D, which is also reported in many previous studies (128, 130, 131). Despite the increasing temporal pattern in FSI, HOMA-IR, and BMI can be visually observed in the preliminary naïve pooled data approach, the increasing temporal pattern of FSI, HOMA-IR and BMI over TIMEtoD was found not to be significant in the progressor group by the Bayesian hierarchical approach. This conflict can be attributed to the high inter-individual variability and the effect of averaging in the naïve pooled approach (52).

Although the average curves of these DDVs increased over TIMEtoD, the increasing pattern is not clearly shown in the majority of the progressors. None of the $\mu_{TIMEtoD}$ parameters in the single-slope model and the $\mu_{TIMEtoD1}$ parameters in the two-slope model of these DDVs was found to be significant by the Bayesian approach. Nevertheless, the parameter μ_{pg} of BMI is significantly larger than zero, which is consistent with the high BMI in the progressor group shown in Table 3.1. Obesity can be considered as a long-term *static* risk factor, while FBG is a relative short-term, dynamic, time-variant risk factor of T2D.

The subjects in the progressor group were followed up to 25 years. With this long follow-up period, the data provides a broad scope for evaluating the development of T2D. However, the model due to its simple structure and few structural parameters, which are dictated by the sparseness of intra-individual data, enables just a simple assessment of the DDVs over TIMEtoD. The population parameters just show the trend of the majority. Some individuals may not follow the trend of population entirely. The studies of Weyer *et al*. suggest that the change in the acute insulin response may follow a pattern of increase then decrease when approaching the diagnosis of diabetes (38, 144). This type of non-monotonic patterns was not detectable in our study. Some subjects in the non-progressor group might have started developing T2D, but there is no way to know whether these people were going to become diabetic or not due to the limited follow-up period. This may bias the age effect extracted from the non-progressor group. However, given the nature of the data, the non-progressor group is still the only practical choice for providing a correction for the natural age effect. The collinearity between AGE and TIMEtoD in the progressor group may reduce the accuracy and precision when

estimating $k_{TIMEtoD}$ or $k_{TIMEtoD2}$ (145). However, this does not affect the model predictability and the interpretation of the TIMEtoD effect after adjusting the AGE effect.

In this study, the time course of development of T2D was investigated by a population-based Bayesian hierarchical analysis of several potential disease development variables. FBG was found to be the variable best describing the temporal changes during disease development over the long follow-up period. In contrast, no consistent increasing or decreasing patterns of FSI, HOMA-IR and BMI were found. The proposed model provides a quantitative longitudinal evaluation of the development of T2D. Accordingly, the proposed model may be used to quantify the effect of interventions, such as the effect of drugs aimed at preventing or slowing down the development of this disease, or may be used to evaluate the effect of lifestyle changes. In contrast to traditional, long-term, final outcome analysis, a model-based longitudinal evaluation of the disease development provides a much faster assessment of interventions. This is particularly important in dealing with a disease like T2D which has a very long development period before diagnosis

Table 3.1. Summary of characteristics of the progressor and non-progressor groups

	Progressor group	Non-progressor group	p-value[(a)]
Number of subjects	25	127	NA
Gender (males in %)	60%	42.5%	0.119
Starting age (mean ± SD) range	33.6 ± 7.9 21.1-49.8	32.9 ± 10 13.4-59	0.702
Starting BMI (mean ± SD) range	32.3 ± 8.05 23.2-53.1	25.3 ± 4.33 17.2-49.6	0.00015
Years followed (mean ± SD) range	13.5 ± 5.74 3.22-23.9	13 ± 6.55 0.77-25.6	0.677

[(a)]: p-value of a t-test for comparing the means of two groups

Table 3.2. Summary of Bayes estimates and 95% credible sets of population mean parameters in the single-slope model for the four disease development variables (DDV)

DDV	Population mean estimates of the parameters in the single-slope model			
	Baseline μ_{base}	Progressor effect μ_{pg}	Age effect μ_{age}	TIMEtoD effect $\mu_{TIMEtoD}$
FBG (mg/dL)	69.4 [66.5, 72.2]	14.5 [9.43, 19.4]	0.221 [0.145, 0.301]	0.974 [0.551, 1.41]
FSI (mU/L)	20.7 [17.4, 24.6]	69.6 [-15.8, 155]	-0.133 [-0.231, -0.0507]	5.84 [-2.85, 14.6]
HOMA-IR	4.21 [3.48, 5.00]	23.3 [-4.71, 51.1]	-0.0218 [-0.0418, -0.00364]	2.04 [-0.904, 5.03]
BMI (kg/m^2)	23.9 [22.5, 25.2]	8.35 [3.2, 13.5]	0.047 [0.0155, 0.0841]	0.147 [-0.149, 0.446]

Table 3.3. Summary of Bayes estimates and 95% credible sets of population mean parameters in the two-slope model for the four disease development variables (DDV)

DDV	Population mean estimates of the parameters in the two-slope model					
	Baseline μ_{base}	Age effect μ_{age}	Progressor effect μ_{pg}	TIMEtoD effect [(a)] $\mu_{TIMEtoD1}$	Separation point μ_{sep}	TIMEtoD effect [(b)] $\mu_{TIMEtoD2}$
FBG (mg/dL)	69.2 [66.5, 72.1]	0.227 [0.149, 0.3]	1.41 [-1.74, 4.49]	NA	-8.73 [-10.8, -6.93]	2.27 [0.505, 4.52]
FSI (mU/L)	21 [16.8, 24.4]	-0.14 [-0.227, -0.0274]	31.7 [0.891, 63]	5.35 [-16.3, 26.9]	-11.2 [-12.6, -9.6]	NA
HOMA-IR	4.00 [3.23, 4.91]	-0.0161 [-0.0397,0.0027]	9.21 [-0.770, 19.0]	2.06 [-4.48, 8.50]	-11.0 [-12.4, -9.44]	NA
BMI (kg/m^2)	24 [22.7, 25.3]	0.0437 [0.0094, 0.0754]	7.78 [3.53, 12]	0.198 [-0.23, 0.649]	-6.48 [-10.5, -3.83]	NA

[(a)]: the temporal pattern of the DDV over TIMEtoD was modeled before the separation point

[(b)]: the temporal pattern of the DDV over TIMEtoD was modeled after the separation point

Table 3.4. Summary of Bayes estimates and 95% Bayesian credible sets of σ_{base}, σ_{age}, σ_{pg}, σ_{sep}, and $\sigma_{TIMEtoD2}$ in the two-slope model of FBG

DDV	Inter-individual standard deviations of the parameters in the two-slope model				
	Baseline σ_{base}	Age effect σ_{age}	Progressor effect σ_{pg}	Separation point σ_{sep}	TIMEtoD effect $\sigma_{TIMEtoD2}$
FBG (mg/dL)	3.34 [0.791, 5.26]	0.11 [0.0599, 0.154]	2.31 [0.106, 6.21]	1.14 [0.141, 2.43]	3.17 [1.45, 5.96]

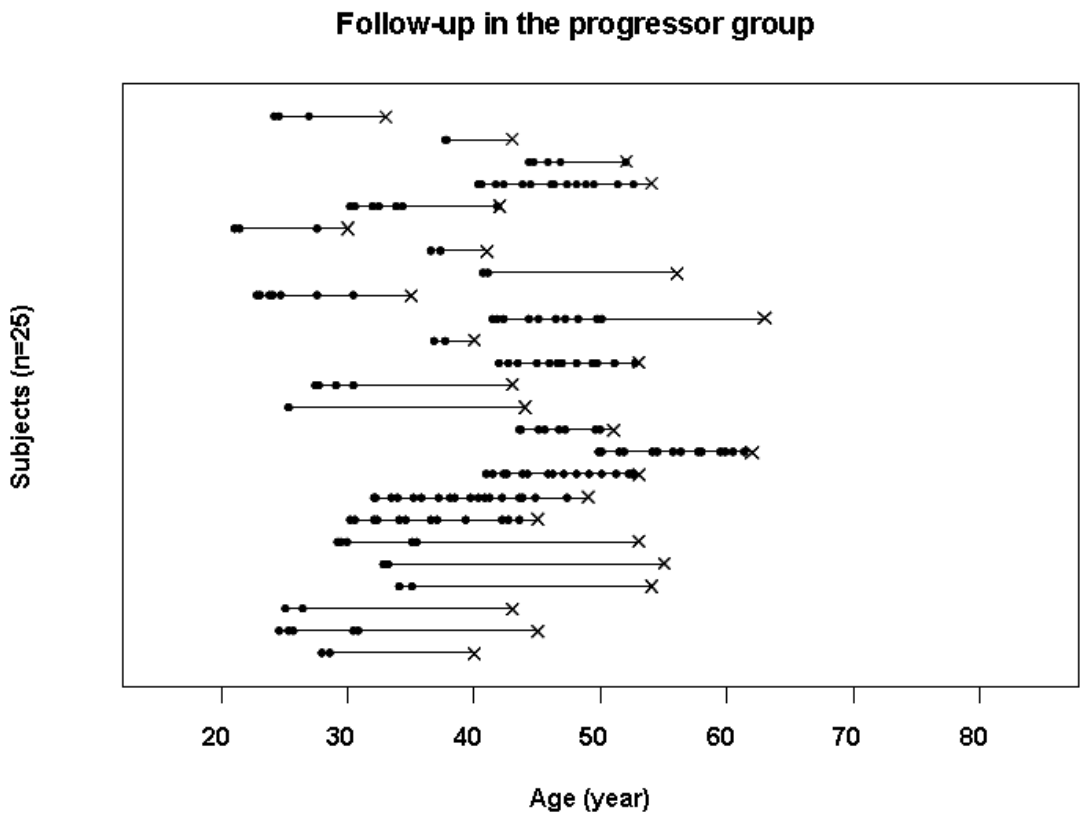


Figure 3.1. Summary of the repeated fasting blood tests and ages of diagnosis of type 2 diabetes in the progressor group. The straight lines show the follow-up periods of the progressors. The dots represent the fasting blood tests and the crosses (x) denote the ages of diagnosis of type 2 diabetes.

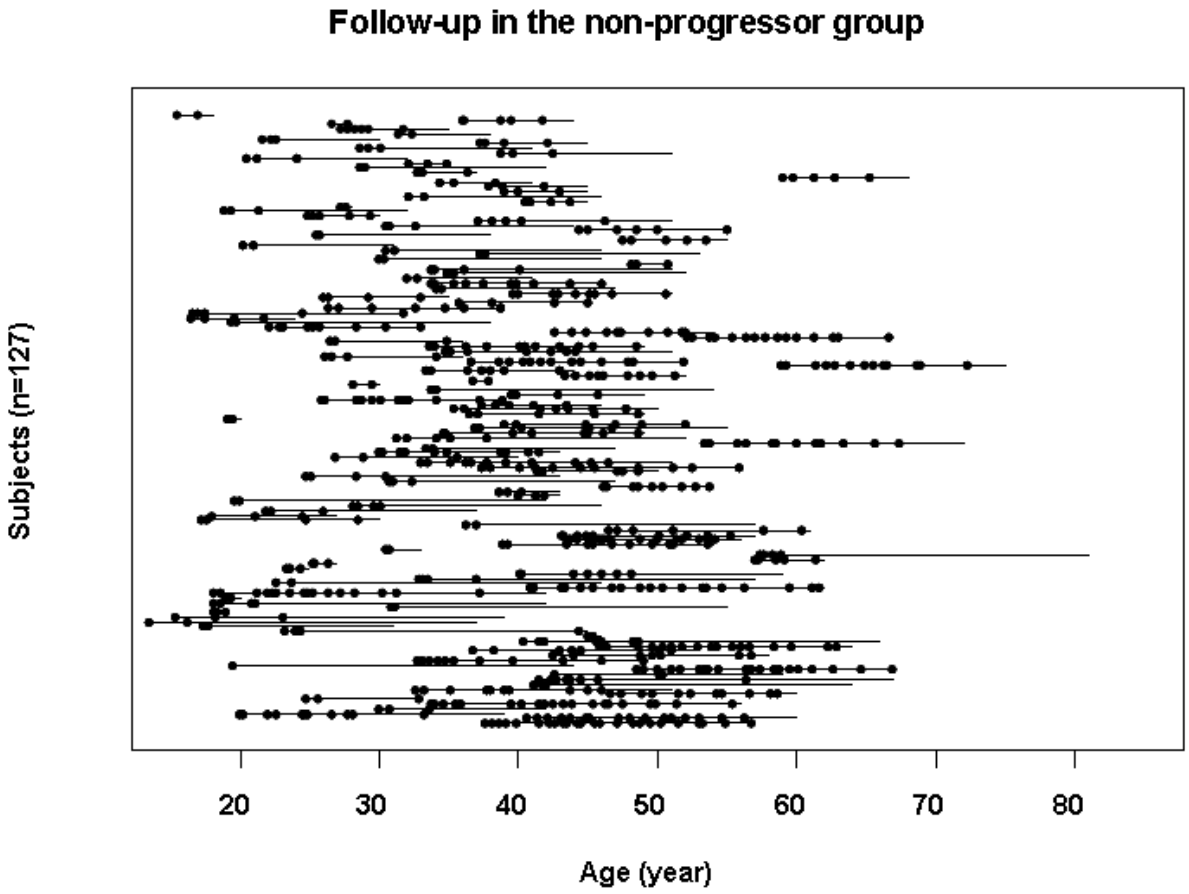


Figure 3.2. Summary of the follow-up and the repeated fasting blood tests in the non-progressor group. The straight lines show the follow-up periods of the non-progressors. The dots represent the ages of the fasting blood tests.

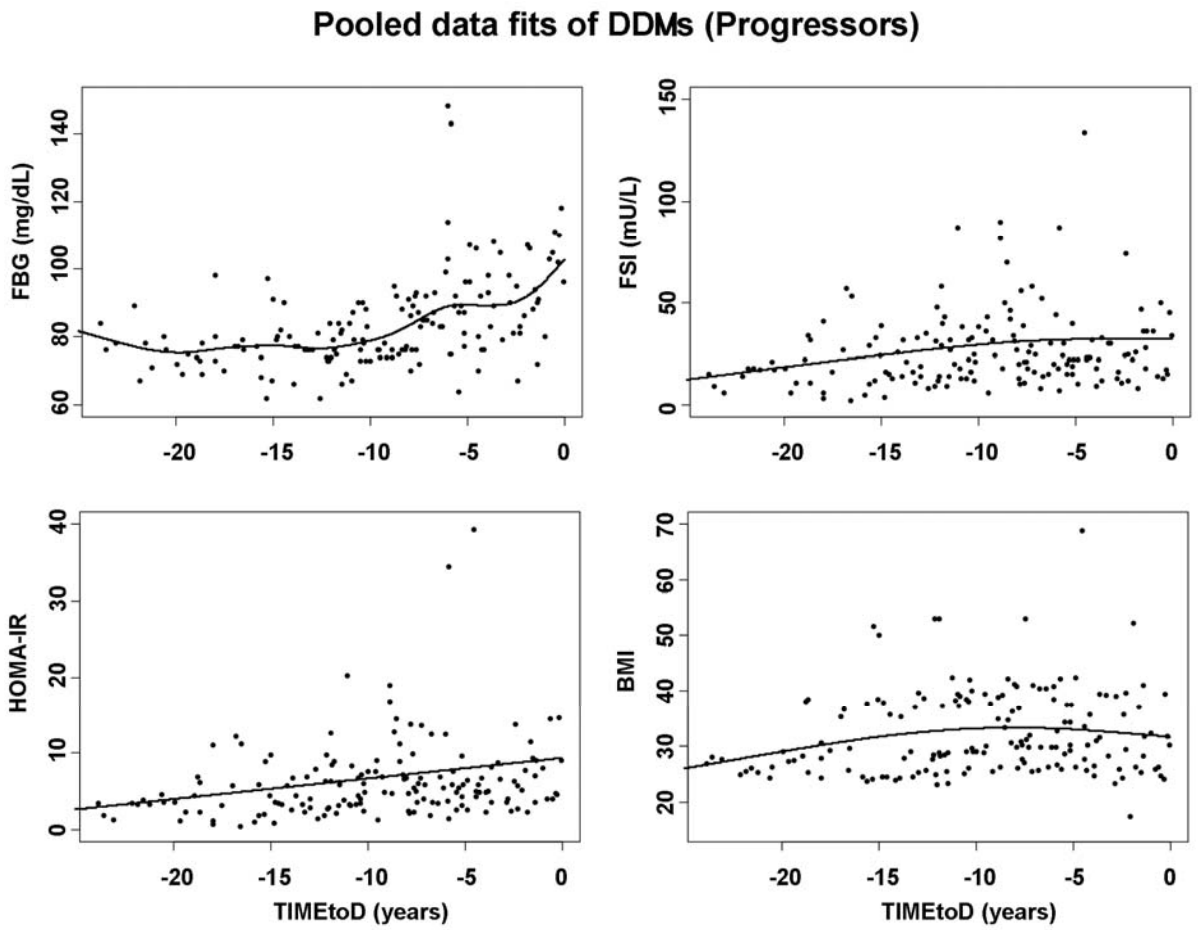


Figure 3.3. Generalized cross-validation cubic spline fits of FBG, FSI, HOMA-IR, and BMI over TIMEtoD for the progressor group by the naïve pooled approach. Repeated samples were collected for each progressor before diagnosis of type 2 diabetes.

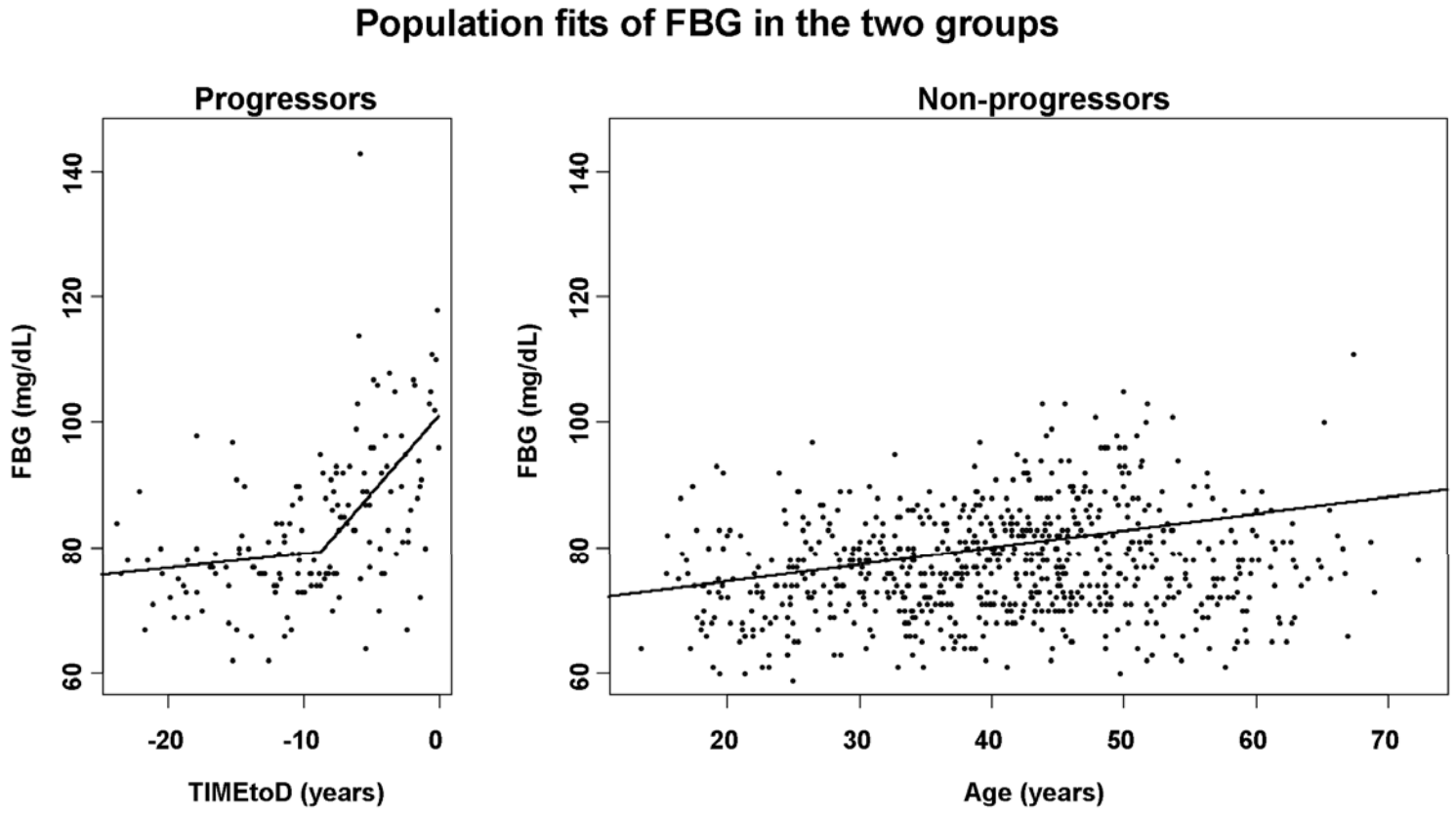


Figure 3.4. The population fits of FBG in the progressor group (n=25) and non-progressor group (n=127) by the Bayesian hierarchical approach. For the progressors, the x-axis represents TIMEtoD to emphasize the two-slope temporal pattern of FBG over TIMEtoD. For the non-progressors, the x-axis represents age to illustrate the nature progression over age in the non-progressor group in contrast to the progressor group.

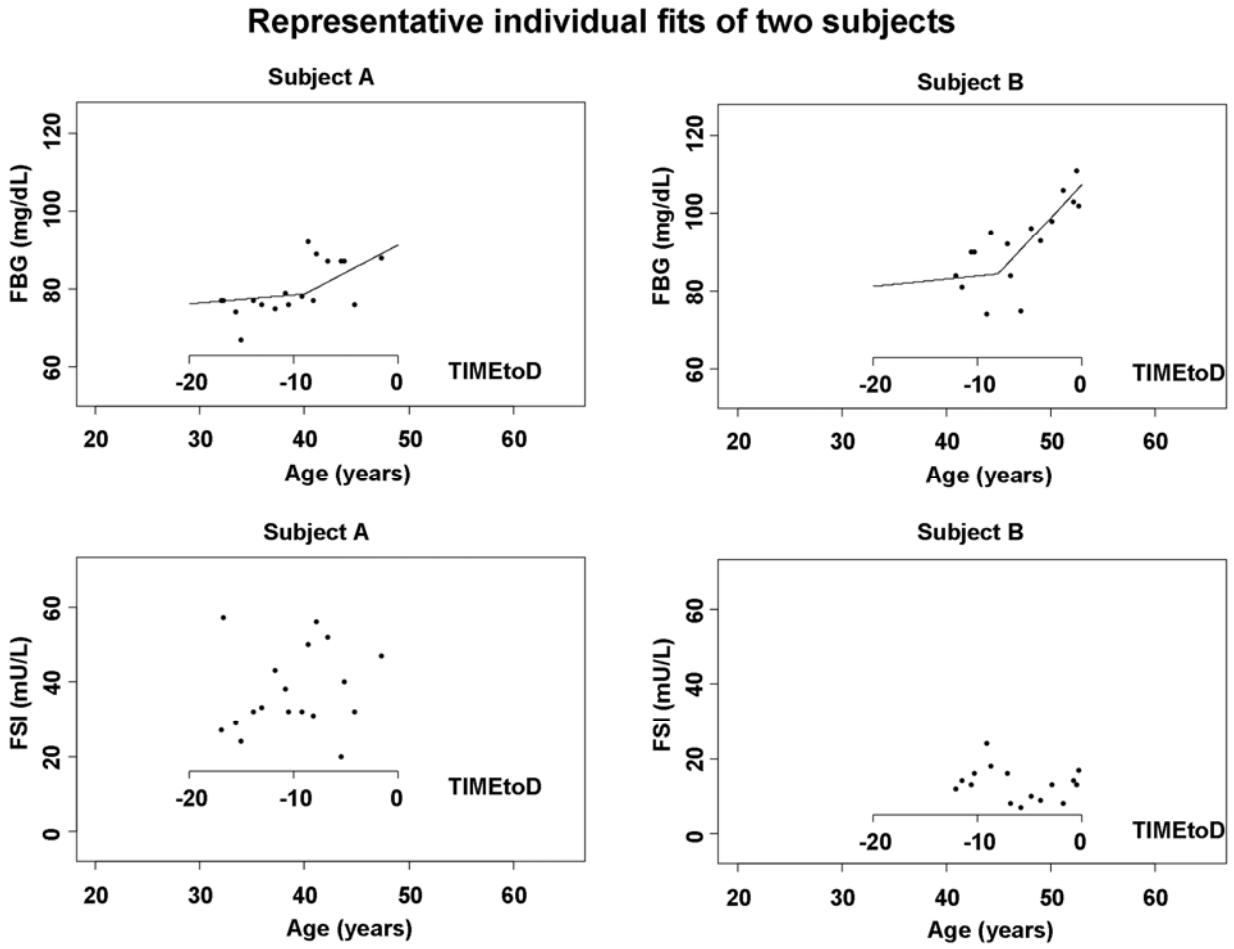


Figure 3.5. Representative individual fits of FBG in two progressors. The repeatedly measured FBG were fitted by the two-slope model while FSI data were included for comparison.

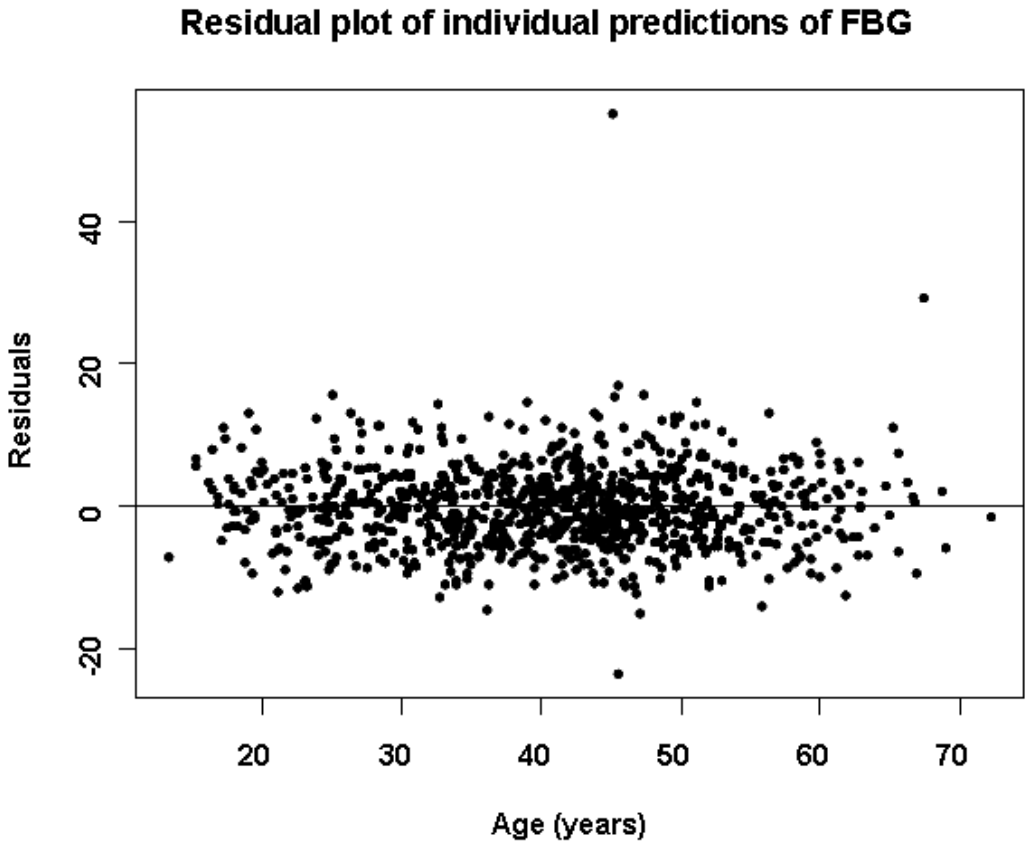


Figure 3.6. The residual plot of FBG over age. The estimate of σ is 6.3 mg/dL with a 95%C.S. [5.98, 6.64] (number of observations=892, number of subjects=152).

CHAPTER 4. DISEASE DEVELOPMENT MODELING OF TYPE 2 DIABETES IN OFFSPRING OF DIABETIC PARENTS BASED ON DISEASE DEVELOPMENT VARIABLES DERIVED FROM IVGTT

4.1 Abstract

The objectives of this study were to find possible disease development variables (DDVs) of type 2 diabetes (T2D) and to evaluate the temporal changes of these DDVs during the development period of T2D. Non-diabetic offspring of diabetic parents were recruited and followed up to 25 years. Repeated intravenous glucose tolerance tests (IVGTTs) were performed during the follow-up. The analysis consisted of a progressor group of 25 subjects who developed T2D and a non-progressor group of 122 subjects who did not develop T2D during the follow-up period. The IVGTT were analyzed by a non-parametric kinetic analysis of the glucose and insulin concentration data. The DDVs from the IVGTTs included single glucose/insulin concentration measurements, differences in glucose/insulin concentrations, and several quantities derived from the non-parametric kinetic analysis of the glucose-insulin data. The DDVs were identified by a mixed effect regression analysis by determining if significant changes in the variables occurred during the disease development period. The results of the analysis indicated that the following DDVs were significantly associated with development of T2D: fasting blood glucose (FBG), the difference in glucose concentrations at 10 and 50 minutes, the difference in insulin concentrations at 3 and 20 minutes, the difference between insulin concentrations at 120 minute and the fasting level, and the non-parametrically estimated insulin concentration at 104 minutes. Taken together these variables indicate that changes in the first and second phase insulin secretion are important indicators of the development

of T2D. This study provided longitudinal information about the pathogenesis of T2D and enables a quantitative, evaluation of the development of T2D in this higher risk population, which may be used to quantify the effect of interventions/prevention strategies such as drug treatments and lifestyle changes.

4.2 Introduction

The pathogenesis and the nature history of type 2 diabetes (T2D) have been studied for decades. However, very few studies were designed to investigate the temporal changes of the glucose and insulin kinetics during the development of T2D. These types of studies are important for understanding the pathology mechanism and pathogenesis of T2D by identifying the specific patterns of the insulin-glucose system associated with the development of the disease. Before the diagnosis of T2D, most people have an intermediary pre-diabetes phase with impaired glucose tolerance characterized by an abnormal fasting blood glucose level. Many subjects with pre-diabetes progress to T2D in just 3 years (146). Insulin resistance and β-cell dysfunction, which are also associated with the development of T2D, result in abnormalities in the glucose-insulin regulations leading to reduced glucose removal, increased insulin baseline level, and diminished first phase insulin secretion (23, 25, 34-36, 38). These early signs are all predictive of T2D, but little is known about how they are changing during the T2D development process.

The intravenous glucose tolerance test (IVGTT) is a standard method to evaluate the glucose-insulin kinetics after a glucose challenge. IVGTT has the distinct advantage that the intravenous administration of glucose, in contrast to an oral administration, is not confounded by the absorption process in gastrointestinal system and generates a better defined first phase insulin secretion (125). The IVGTT enables an accurate evaluation of

the glucose removal and the biphasic insulin secretion pattern, which are critical to assess the risks of T2D by evaluating the insulin sensitivity and the deterioration of the β-cell (38, 147). Many mathematical models of the glucose-insulin kinetics have been developed for the analysis of IVGTT data, including models based on Bergman's empirical minimal model, and several mechanism-based models (68, 73, 77, 102, 133). However, the use of these models to evaluate insulin sensitivity and β-cell function requires mathematical modeling skills and software.

Multiple logistic regression analysis is a very informative statistical methodology useful to identify predictors of T2D (25, 128, 131, 148, 149). Unfortunately, the important variable associated with disease development, namely time, is missing in logistic regression analysis. The differences in the follow-up periods of prospective studies, for example 3 years *v.s.* 20 years follow-up, cannot be accounted for by regular logistic regression analysis. Repeated, longitudinal samples of the predictors of T2D are necessary for better prediction of T2D by properly accounting for the time dependency of the variables associated with the disease development. Evaluation of the temporal changes of the predictors during the development process of T2D is critical for identification of disease development variables (DDV) and the statistical modeling of the disease progression.

Various disease progression models have been developed to evaluate drugs' effects on the progression of diseases (81, 84, 99). These models make use of biomarkers of the disease which are longitudinally measured to evaluate the effect of drugs or other interventions on the bio-markers, as an indirect measure of the drug effect on the disease progression (81, 84). In contrast to the more common modeling of disease *progression*,

this study is focusing on disease *development* i.e. investigates the kinetic changes in the insulin-glucose physiology taking places *before* the diagnosis of T2D. Thus, the major distinction is that this study is relevant to preventive medicine, which essentially is not the case for disease progression studies. In addition, the DDVs derive from IVGTT data were estimated by a model-independent approach for the glucose-insulin kinetics, which provides an easy and fast assessment of the development of T2D. The DDVs were evaluated by a population-based analysis, mixed effect model analysis. An examination of the temporal changes in DDVs for the glucose-insulin system occurring before the diagnosis of the disease provides a quantitative evaluation of the pathophysiological evolution of T2D. This type of preventive-medicine-driven research can lead to a better understanding the etiology of T2D, and can provide a quantitative basis for the evaluation of various interventions strategies such as drugs and lifestyle changes.

4.3 Specific aim and hypothesis

The specific aim of this chapter is to identify the important disease development variables derived from an intravenous glucose tolerance test and evaluate the relationship between the temporal changes of these variables and the development of type 2 diabetes.

The specific hypothesis is that the important disease development variables derived from an intravenous glucose tolerance test are hypothesized to exist and can be identified by the proposed disease development analysis.

4.4 Methods

4.4.1 Subjects

The subjects' data is based on the 25-years follow-up study published previously (24). Between 1963 and 1983, 152 healthy offspring of parents with T2D were recruited in the Joslin Diabetes Center (JDC). The study was approved by the human study review board of JDC and conducted according to the Declaration of Helsinki: "Recommendations guiding physicians in biomedical research involving human subjects", adopted by the 18th World Medical Assembly, Helsinki, Finland, June 1964. Informed consent was obtained from all individuals before participation in the study. All participants had normal glucose tolerance and received an intravenous glucose tolerance test (IVGTT) when entering the study. Thereafter, the participants were followed up to 25 years to ascertain the status of T2D. The detailed methods of follow-up and evaluation of diabetes have been described in previously studies (23, 24). In addition to the follow-up, repeated IVGTTs were performed and the data of the glucose and insulin concentrations after IVGTTs were collected. The subjects were divided into a progressor group and a non-progressor group according to the status of T2D in the end of follow-up.

A total of 339 IVGTTs were performed on the 152 non-progressor and 25 progressors of T2D. Some missing blood samples were found in 38 of the 339 IVGTT. These data were excluded from the analysis. A total of 301 sets of IVGTT data performed on 147 subjects were included in the analysis, which contains 65 and 236 IVGTT data in 25 subjects in the progressor group and 122 subjects in the non-progressor group, respectively. Characteristics of the progressor and the non-progressor groups are summarized in Table 4.1.

4.4.2 Intravenous glucose tolerance tests

All participants of IVGTT received a high-carbohydrate diet over three days (250 to 300 g/day) before an overnight fast prior to the I.V. glucose administration. One blood sample was collected as the fasting sample before the glucose administration. Thereafter 0.5 g/kg glucose was intravenously infused over 3-5 minutes, and twelve blood samples were collected at 1, 3, 5, 10, 20, 30, 40, 50, 60, 90, 120 and 180 minutes post infusion. The blood glucose concentrations were measured by the ferricyanide method for the IVGTTs done before 1983 with a coefficient of variation (CV) of 1.5% (117). After 1983, the glucose oxidize method was used with a 1.35% CV (150). The insulin serum concentrations were measured by a double-antibody radio-immunoassay with a 17.6% CV (109).

4.4.3 Analysis of type 2 diabetes development

The disease development analysis assesses the temporal changes of various disease development variables (DDV) using the glucose and insulin measurements data from IVGTTs performed during the disease development period. The longitudinal data of DDVs were analyzed by a linear mixed effect model:

Non - progressor group

$$\mathrm{DDV}_{ij} = k_{base,i} + k_{age,i} \cdot AGE_{ij} + \varepsilon_{ij}$$

Progressor group

$$\mathrm{DDV}_{ij} = k_{base,i} + k_{pg,i} \cdot PG_i + k_{age,i} \cdot AGE_{ij} + k_{TIMEtoD,i} \cdot TIMEtoD_{ij} + \varepsilon_{ij}$$

$$1 \le i \le N \qquad 1 \le j \le n_i \tag{4.1}$$

$$\varepsilon_{ij} \sim Norm\left(0, \sigma^2\right)$$

$$k_{base,i} = \mu_{base} + \eta_{base,i}$$
$$k_{pg,i} = \mu_{pg} + \eta_{pg,i}$$
$$k_{age,i} = \mu_{age} + \eta_{age,i}$$
$$k_{TIMEtoD,i} = \mu_{TIMEtoD} + \eta_{TIMEtoD,i}$$

$$\begin{pmatrix} \eta_{base,i} \\ \eta_{pg,i} \\ \eta_{age,i} \\ \eta_{TIMEtoD,i} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma\right)$$

In the equation above, the time dependency of the DDV is described for both non-progressor and progressor group. In the non-progressor group, the national time-dependency of the DDV is modeled as a linear function of age. While the extra effects from disease development of T2D on the DDV in the progressor group are modeled additively by the predictor variables, TIMEtoD and PG. The predictor variables, PG, AGE and TIMEtoD, represent progressor group indicator, age, and the time prior to the diagnosis of T2D (TIMEtoD). The variable TIMEtoD is the normalized time variable for progressors whose diagnostic times of T2D were known and calculated as the subject's age at IVGTT minus the subject age at the diagnosis of T2D. DDV_{ij} represents the jth DDV of subject i. N denotes the total number of subjects, and n_i denotes the number of IVGTTs performed on subject i. The regression parameters, $k_{base,i}$, $k_{pg,i}$, $k_{age,i}$ and $k_{TIMEtoD,i}$ represent the intercept, progressor effect, age effect and TIMEtoD effect of the i^{th} subject. μ_{base}, μ_{pg}, μ_{age}, and $\mu_{TIMEtoD}$ represent the fixed effects (population) parameters. $\eta_{base,i}$, $\eta_{pg,i}$, $\eta_{age,i}$, and $\eta_{TIMEtoD,i}$ represent the random effects parameters, which are assumed to be multivariate normal distributed. The mixed effect model analyses were performed in R 2.10.1(139). The population parameter, $\mu_{TIMEtoD}$, is the key parameter to address the temporal change of the DDV in the progressor group, and μ_{pg} is

important to distinguish the two groups at TIMEtoD=0. Accordingly, a DDV is considered significant when the p-values of μ_{TIMEtoD} and μ_{pg} are both less than 0.05.

4.4.4 Disease development variables

The DDV candidates include all of the measured single glucose or insulin concentration or difference in glucose/insulin concentrations, or glucose-insulin concentration predicted by a non-parametric analysis at any time point post IVGTT. The non-parametric estimation of the glucose and insulin response is used to provide accurate interpolations of the glucose and insulin concentrations up to 180 minutes after the IVGTT. Accordingly, the glucose and insulin curves were estimated from the IVGTT data by cubic spline functions based on 12 samples collected after the intravenous glucose administration. The smoothness of the cubic spline was determined according to the cross-validation principle and the leave-one-out approach using the R function smooth.spline(). The R package "nlme" was used in the linear mixed effect model analysis. The fixed and random effect parameters were estimated by the function mle() using the restricted maximum likelihood method. The non-parametric analyses and all graphs were done in R version 2.10.1(139).

4.5 Results

4.5.1 DDVs using single measurement

The single measurement DDVs with significant μ_{TIMEtoD} (p-value<0.05) are listed in Table 4.2. Among the single measurement DDV, only FBG gave both significant μ_{TIMEtoD} and μ_{pg} with values 0.474 mg/dL/year (p=0.0247) and 7.99 mg/dL (p<0.01)

respectively. The significant positive value of $\mu_{TIMEtoD}$ indicates an increasing temporal pattern of fasting glucose concentration over TIMEtoD in the progressor group (Figure 4.1). The significance of μ_{pg} suggests the progressor group has higher fasting glucose concentration than the non-progressor group at TIMEtoD=0. The DDV defined by the insulin concentration at 3 minute (I_3) gave the most significant $\mu_{TIMEtoD}$ of -3.43 mU/L/year ($p < 0.01$), but the non-significance of μ_{pg} indicates that this DDV was not different between the groups at the diagnosis time of T2D, TIMEtoD=0, (Figure 4.1). Accordingly, significance of both $\mu_{TIMEtoD}$ and μ_{pg} are needed to qualify a DDV as significant.

4.5.2 DDVs based on two measurements

Glucose data

The DDVs defined as the sum or difference of two glucose or insulin concentrations were also tested in the mixed effect analysis. The three most significant DDVs of glucose are shown in Table 4.3. The DDV defined as the difference between glucose concentrations at 10 and 50 minutes (G_{10-50}) was found to be the most significant DDV based on two glucose measurements. The predicted G_{10-50} for a 45 years progressor at TIMEtoD=0 is 24.8 mg/dL lower than a 45 years non-progressor.

Insulin data

For the DDVs of two measurements of insulin, the three most significant DDVs are shown in Table 4.4. The DDV defined by the difference between insulin concentration at 120 minutes and fasting insulin level (I_{120-0}) was found to be the most

significant combination. This DDV indicates the second phase insulin secretion may be an important indicator of the development of the T2D. The predicted I_{120-0} for a 45 years progressor at TIMEtoD=0 is 15.3 mU/L higher than a 45 years non-progressor. The DDV for the difference between insulin concentrations at 3 and 20 minutes (I_{3-20}) which is almost as significant, is associated with the first phase insulin secretion. The predicted I_{3-20} for a 45 years progressor at TIMEtoD=0 is 62.5 mU/L lower than a 45 years non-progressor. The significance of I_{120-0} and I_{3-20} taken together indicates that changes in the first and second phase insulin responses are important indicators of the development T2D. Three representative population fits are shown in Figure 4.2 to illustrate the mixed effect model estimations of these DDVs. The x-axes of the figures shown are TIMEtoD and age for the progressor group and the non-progressor group, respectively, to visualize the effects of TIMEtoD and age in the two groups. The fitted curves were constructed according to the estimated parameters values of the fixed effect regression model.

4.5.3 DDVs defined from non-parametric analysis

Representative cubic spline fits of glucose and insulin data are shown in Figure 4.3. The glucose and insulin concentrations non-parametrically estimated at each minute from 1 to 180 minutes post glucose administration were used to define the various DDVs candidates considered in the analysis.

Each DDV was examined by the mixed effect model. The estimates and p-values of μ_{TIMEtoD} and μ_{pg} for the 360 kinetic measurements (180 predicted glucose concentrations and 180 predicted insulin concentrations) are shown in Figure 4.4 and Figure 4.5. For the 180 predicted glucose concentrations, μ_{TIMEtoD} is significant (p-value<0.05) at 5-12 minutes and at 137 -180 minutes post intravenous glucose

administration and μ_{pg} is significant at 30-92 minutes (Figure 4.4). For the 180 predicted insulin concentrations, $\mu_{TIMEtoD}$ is significant at 1-8 minutes and at 101-113 minutes and μ_{pg} is significant at 35-173 minutes (Figure 4.5). According to the DDV selection criteria mentioned previously, the DDV of non-parametric kinetic measurement of insulin at 101-113 minutes are all significant and the most significant variable is at 104 minutes (NPI_{104}) with a $\mu_{TIMEtoD}$ of 0.898mU/dL/year ($p=0.0441$) and μ_{pg} of 33.1 mU/dL ($p<0.01$).

4.6 Discussion

Although many predictors of T2D have been identified, very few studies focus on the temporal changes of such predictors during the disease development period. In this study, a linear mixed effect model analysis was used to identify DDVs as important quantitative indicators of the pathogenesis of T2D. In the analyses of the DDVs of single measurements, FBG gave significant $\mu_{TIMEtoD}$ ($p =0.0247$) and μ_{pg} ($p <0.01$) values. FBG is the standard diagnostic test of T2D used by American Diabetes Association and World Health Organization. Thus, FBG was used as the standard DDV reference for comparisons of DDV in this study.

Important DDVs defined by two measurements were also identified by the mixed effect analysis. These DDVs can be divided into three categories: 1. DDVs relative to glucose removal, 2. DDVs relative to the early phase insulin secretion, and 3. DDVs relative to the late phase insulin secretion. The DDV relating to glucose removal is represented by the variable G_{10-50}. This DDV gives a simple model-independent way to measure glucose removal requiring only two samples. The DDV defined by G_{10-50} is more significant than FBG, and this variable has a clear decreasing temporal pattern over

TIMEtoD, and distinguishes the progressors from non-progressors (Figure 4.2). The DDV dealing with the early phase insulin secretion is defined by the insulin difference I_{3-20}. The early phase insulin secretion, or the first phase insulin secretion, is commonly considered as the fast insulin response at the first 10 minutes after glucose administration (151, 152). I_3 may thus be considered a variable for the early phase insulin secretion. Although I_3 is significant decreasing over TIMEtoD, I_3 alone can not distinguish the progressors from non-progressors when they became diabetic. I_{20} may be served as a correction "offset" variable that reduces the between-subject variability so that the variable defined as the insulin difference I_{3-20} results in both μ_{TIMEtoD} and μ_{pg} becoming significant. The fasting serum insulin (FSI) is the most commonly correction variable used in many studies that employ a baseline correction (25, 38, 152). However, I_{20} showed greater significance than FSI in this analysis. The DDV relative to the late phase insulin secretion is represented by the variable I_{120-0}. The late insulin secretion, or second phase insulin secretion, is usually defined as the insulin secretion post 10 minutes after glucose challenge (151), which has a very broad range. The 2 hour insulin concentration after OGTT was studied among Nauruan (148). The high 2 hour insulin response predicted T2D or impaired glucose tolerance for normal subjects, but the low 2 hour insulin response predicted T2D for subjects with impaired glucose tolerance. In our study, the variable of I_{120-0} is closely related to the late phase insulin response. Figure 4.2 shows that this DDV was increasing from normal to diabetic in the progressor group. No clear pattern of decreasing was observed. The controversy may be caused by the baseline correction. In addition, the extrapolated value of I_{120-0} at TIMEtoD=0 in the progressor group is higher than the value of I_{120-0} in the non-progressor group. The significant DDV

defined by the non-parametric estimation of insulin at 104 minute (NPI_{104}) also indicates the late insulin response is increasing during the development of T2D (Figure 4.6). These findings suggest the prolonged or high second phase insulin secretion is associated with development of T2D.

The mixed effect analysis using non-parametrically determined DDVs showed that the glucose curves at 5-12 and 137-180 minutes have a significant decrease and increase respectively during the development period of T2D. Since the amount of glucose infused in IVGTT is normalized by body weight as 0.5 g/kg and no pattern of increasing of insulin concentration was found at 5-12 minutes, the decrease in glucose concentration during this period cannot be caused by an increased body weight or increased insulin level. Thus, the decrease in glucose may be caused by changing in glucose disposition during the disease development period. The insulin curves at 1-8 and 101-113 minutes post IVGTT were significantly decreasing and increasing respectively before the diagnosis of T2D. The decrease in insulin concentration at 1-8 minutes indicates that the first phase insulin secretion is diminishing, but the non-significant μ_{pg} indicates the DDVs can not effectively distinguish the progressors from non-progressor group, which is consistent with the variable I_3 that has a significant $\mu_{TIMEtoD}$ but non-significant μ_{pg}. At 101-113 min both $\mu_{TIMEtoD}$ and μ_{pg} are significant. This suggests that the increase in the late phase insulin secretion is associated with development of T2D.

Study limitations

The non-progressor group was assumed to provide no information of the disease development. However, some subjects in the non-progressor group might have started to develop T2D, but the changes in DDVs for the non-progressor group is evaluated as an

age effect. All DDVs were modeled by the same linear mixed effect model, but the temporal patterns of some DDVs may not be linear. For example, a DDV may have a two stage pattern, e.g. an initial linear change followed by a nonlinearly change. Despite these limitations, the proposed approach provides a "model-independent/non-parametric" analysis that makes simple comparisons and selection of DDVs possible. The simplicity of the evaluations facilitates the interpretation. For example μ_{TIMEtoD} provides basic information about the temporal increase/decrease in the DDV, and μ_{pg} enables a simple determination of how well the DDV distinguishes the progressors and non-progressors groups. The variables defined by $G_{10\text{-}50}$, $I_{3\text{-}20}$, $I_{120\text{-}0}$, and NPI_{104} appear as good or better variables for disease development as FBG. However, these variables require an IV glucose administration and 1-2 blood samples, which is a more involved testing procedure than the simple determination of FBG, but less involved than a standard IVGTT.

This study demonstrates that the proposed disease development analysis is able to identify the important DDVs that show significant temporal changes during the development period of T2D. These DDV may be used to better quantify the risk of development of T2D. Although some of the DDVs are associated with reduced glucose removal and diminished first phase insulin secretion, which have already been found to predict T2D in previous studies (23, 35, 36, 38), our approach gives a new general way to analyze development of T2D. The method determines the temporal pattern of DDV in an intuitive, simple and robust way that due to its non-parametric nature is free of the assumptions associated with model-dependent analysis approaches.

In summary, this study approach provides valuable longitudinal information valuable for a better understanding of the pathogenesis of T2D. The approach also enables a quantitative, time-based evaluation of the development of T2D, which may be used to quantify the effect of interventions/prevention strategies such as drug treatments and lifestyle modifications.

Table 4.1. Summary of characteristics of the progressor and non-progressor groups

	Progressor group	Non-progressor group
Number of subjects	25	122
Number of IVGTTs performed	65	236
Gender (males in %)	60%	41.2%
Starting age (mean ± sd) range	34 ± 7.96 21.5-50.1	33.7 ± 10 16.2-59.7
BMI (mean ± sd) range	32.2 ± 7.62 23.3-53.1	25.5 ± 4.4 17.9-41.3
Years followed (mean ± sd) range	13.2 ± 5.88 2.33-23.9	12.5 ± 6.6 0.31-24.6

Table 4.2. Summary of the estimates and p-values of fixed effect (population) parameters for one measurement-based DDVs which result in a p-value for μ_{TIMEtoD} that is less than 0.05

DDV	μ_{base}		μ_{age}		μ_{pg}		μ_{TIMEtoD}	
	Estimate	(p-value)	Estimate	(p-value)	Estimate	(p-value)	Estimate	(p-value)
FBG[(a)]	70.1	(<0.001)	0.206	(<0.001)	7.99	(<0.001)	0.474	(0.0247)
G_5	262	(<0.001)	0.0826	(0.672)	-12.7	(0.181)	-1.65	(0.0142)
G_{10}	228	(<0.001)	0.26	(0.118)	-8.69	(0.275)	-1.31	(0.0204)
G_{180}	68.9	(<0.001)	-0.065	(0.196)	1.29	(0.588)	0.455	(0.012)
I_1	133	(<0.001)	-1.04	(0.0116)	-15.6	(0.4)	-2.94	(0.0383)
I_3	114	(<0.001)	-0.863	(0.0177)	-19.9	(0.228)	-3.44	(0.0061)
I_{30}	63.7	(<0.001)	-0.292	(0.255)	25.1	(0.0968)	-2.44	(0.0486)

[(a)] μ_{pg} and μ_{TIMEtoD} are both significant

Table 4.3. Summary of estimates and p-values of fixed effect (population) parameters of the 3 most significant DDVs based on two glucose measurements. The DDV is significant when p-values of μ_{TIMEtoD} and μ_{pg} are both smaller than 0.05. When the DDV is significant, a smaller p-value of μ_{pg} indicates a better DDV.

DDV	μ_{base}		μ_{age}		μ_{pg}		μ_{TIMEtoD}	
	Estimate	(p-value)	Estimate	(p-value)	Estimate	(p-value)	Estimate	(p-value)
$G_{10\text{-}40}$	121	(<0.001)	-0.209	(0.198)	-30.3	(<0.001)	-1.32	(0.0283)
$G_{10\text{-}50}$	140	(<0.001)	-0.194	(0.249)	-34.8	(<0.001)	-1.91	(0.00194)
$G_{10\text{-}60}$	152	(<0.001)	-0.148	(0.388)	-33.2	(<0.001)	-2	(0.00522)

Table 4.4. Summary of the estimates and p-values of fixed effect (population) parameters of the 3 most significant DDVs based on two insulin measurements.

DDV	μ_{base}		μ_{age}		μ_{pg}		$\mu_{TIMEtoD}$	
	Estimate	(p-value)	Estimate	(p-value)	Estimate	(p-value)	Estimate	(p-value)
I_{120-0}	0.913	(0.645)	-0.00649	(0.892)	15.3	(<0.001)	0.771	(<0.001)
I_{3-10}	14.4	(0.0381)	0.0166	(0.926)	-51	(<0.001)	-2.5	(<0.001)
I_{3-20}	31.2	(0.00506)	-0.234	(0.390)	-62.5	(<0.001)	-2.46	(0.00794)

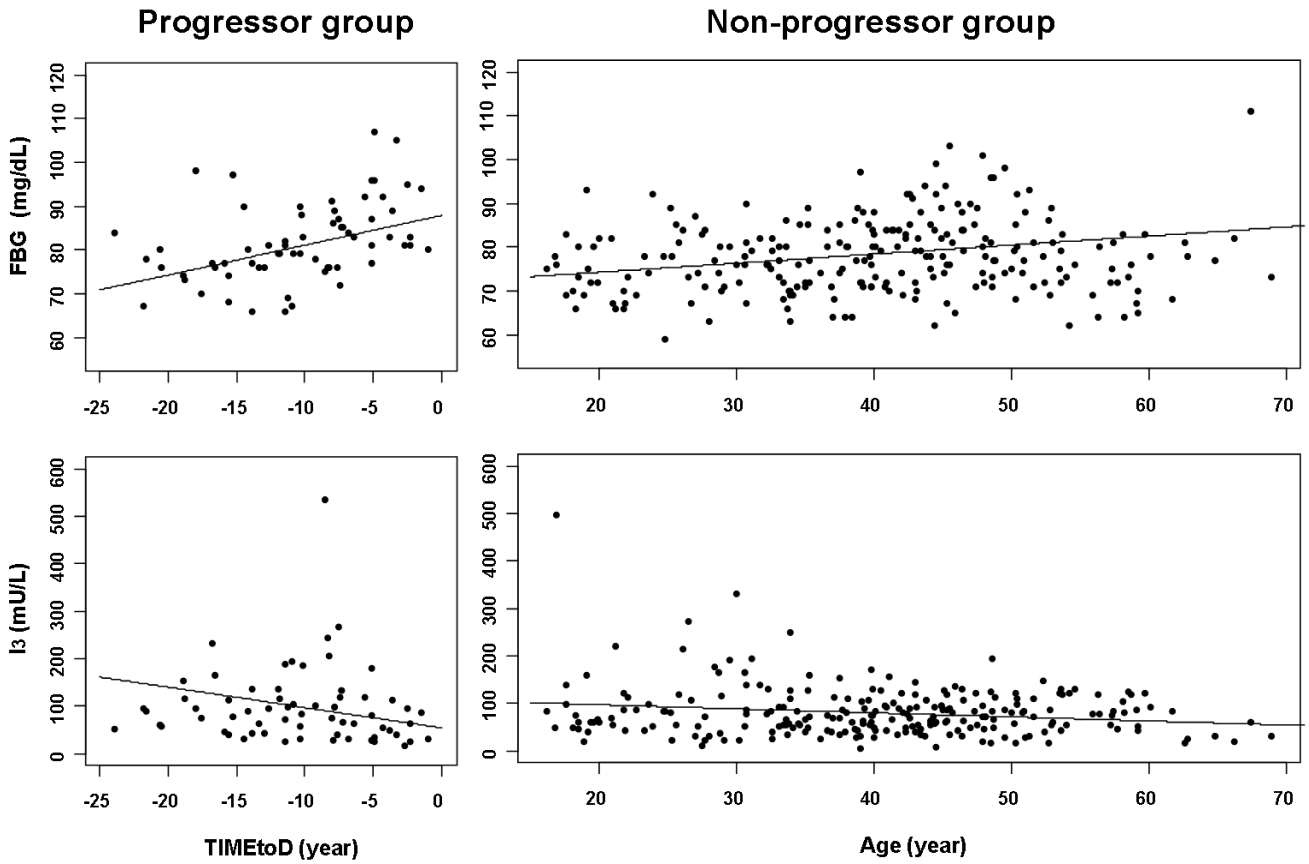


Figure 4.1. Population fits of the FBG and I_3 in the progressor group (n=25) and non-progressor group (n=122). In the population fit for the progressors, the x-axis represents TIMEtoD to emphasize the two-slope pattern over TIMEtoD. In the non-progressors, the x-axis represents subject age to illustrate the natural change over time in the DDV in the non-progressor group.

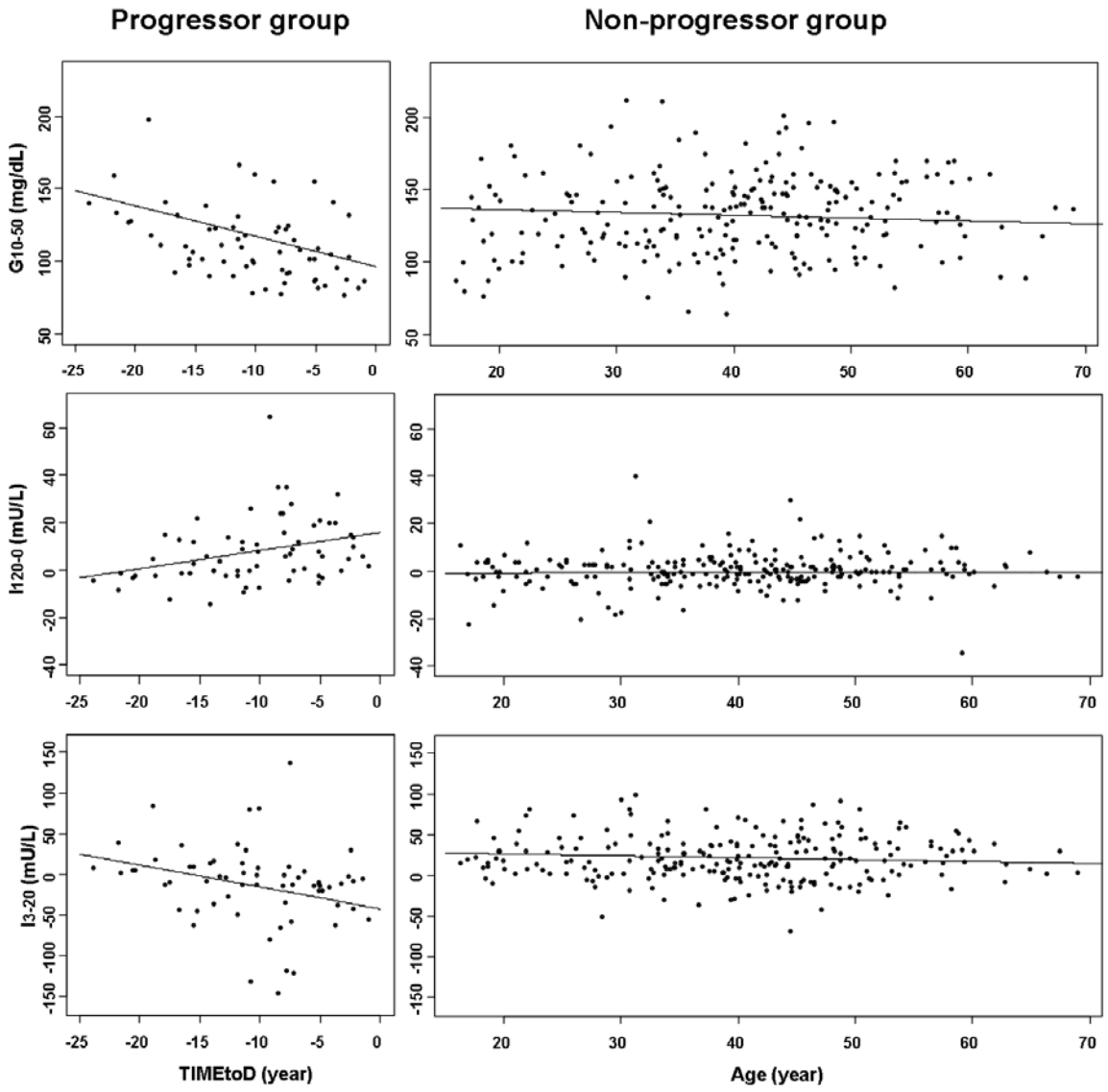


Figure 4.2. Population fits of the DDVs defined by $G_{10\text{-}50}$, $I_{3\text{-}20}$, and $I_{120\text{-}0}$ in the progressor group and non-progressor group.

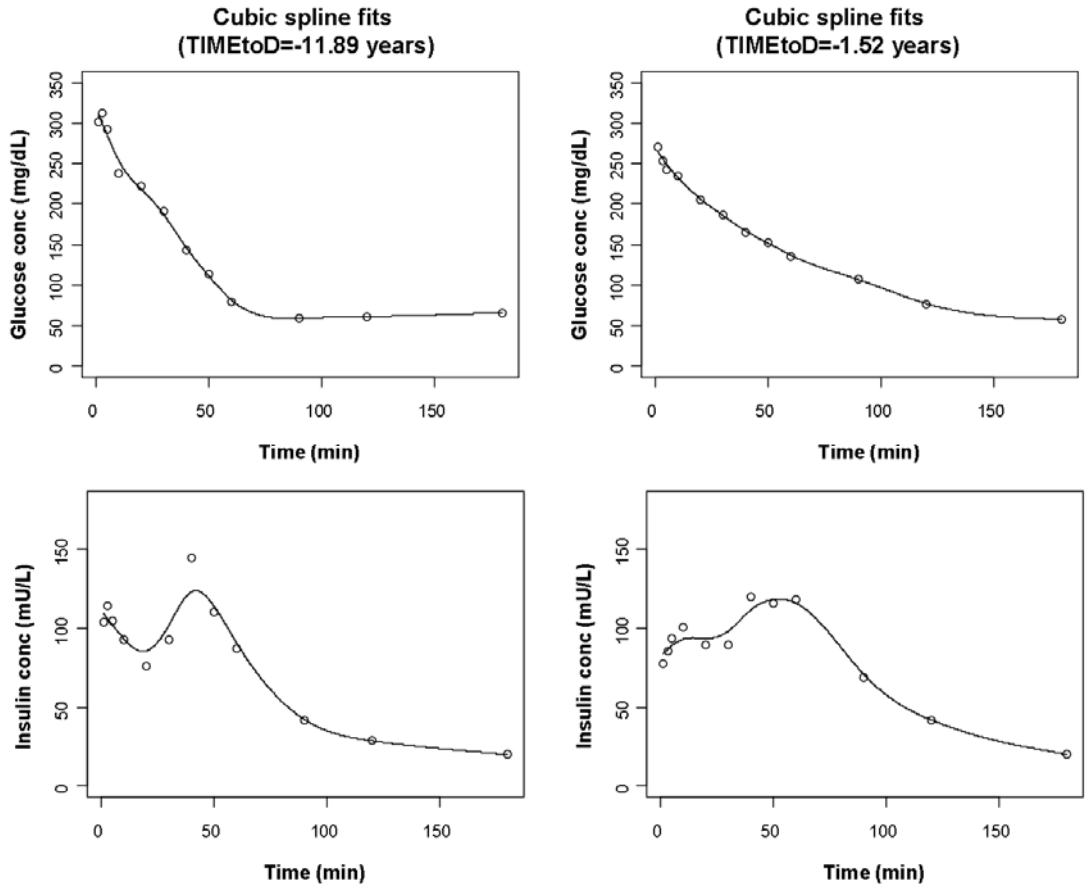


Figure 4.3. Representative cubic spline fits of glucose and insulin data of two IVGTTs from the same subject at TIMEtoD equal to -11.89 and -1.52 years.

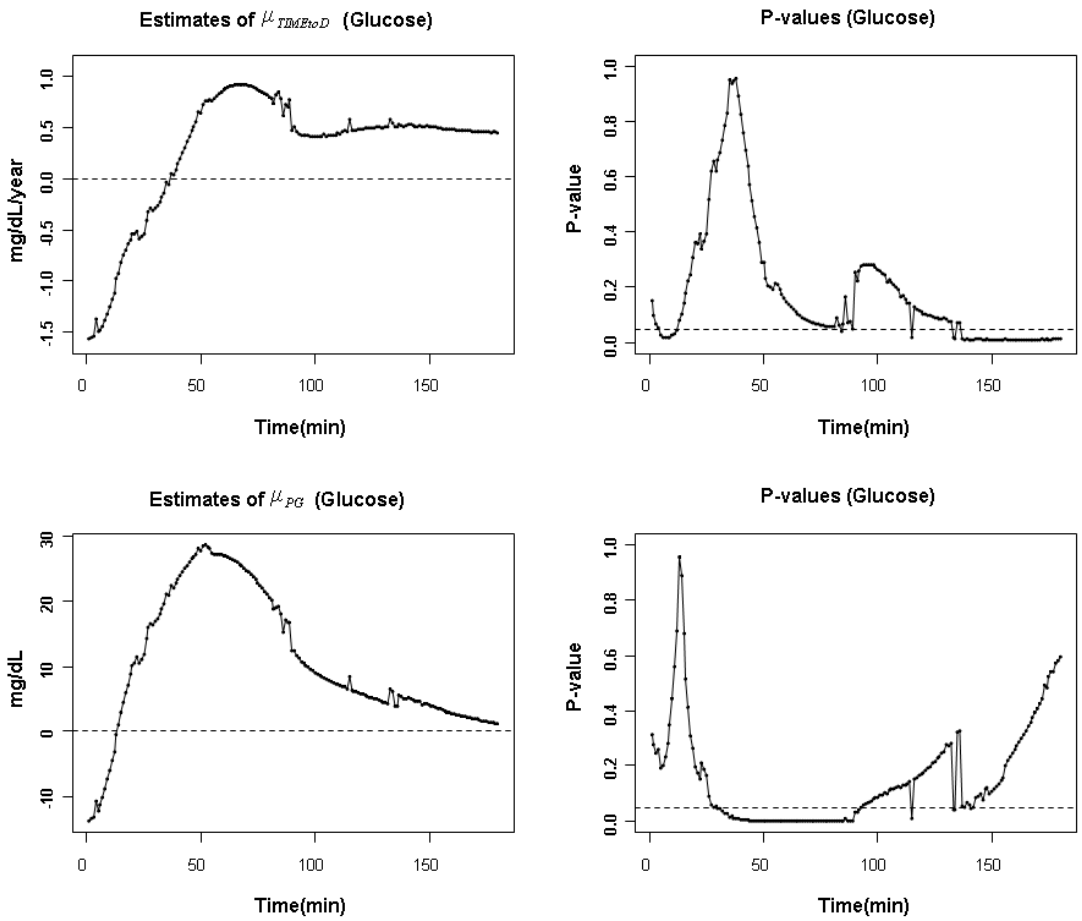


Figure 4.4. Estimates and p-values of μ_{TIMEtoD} and μ_{pg} of DDVs based on non-parametrically estimation of glucose level from 1 to 180 minute. The dashed lines in the p-value plots are drawn at the significance level 0.05.

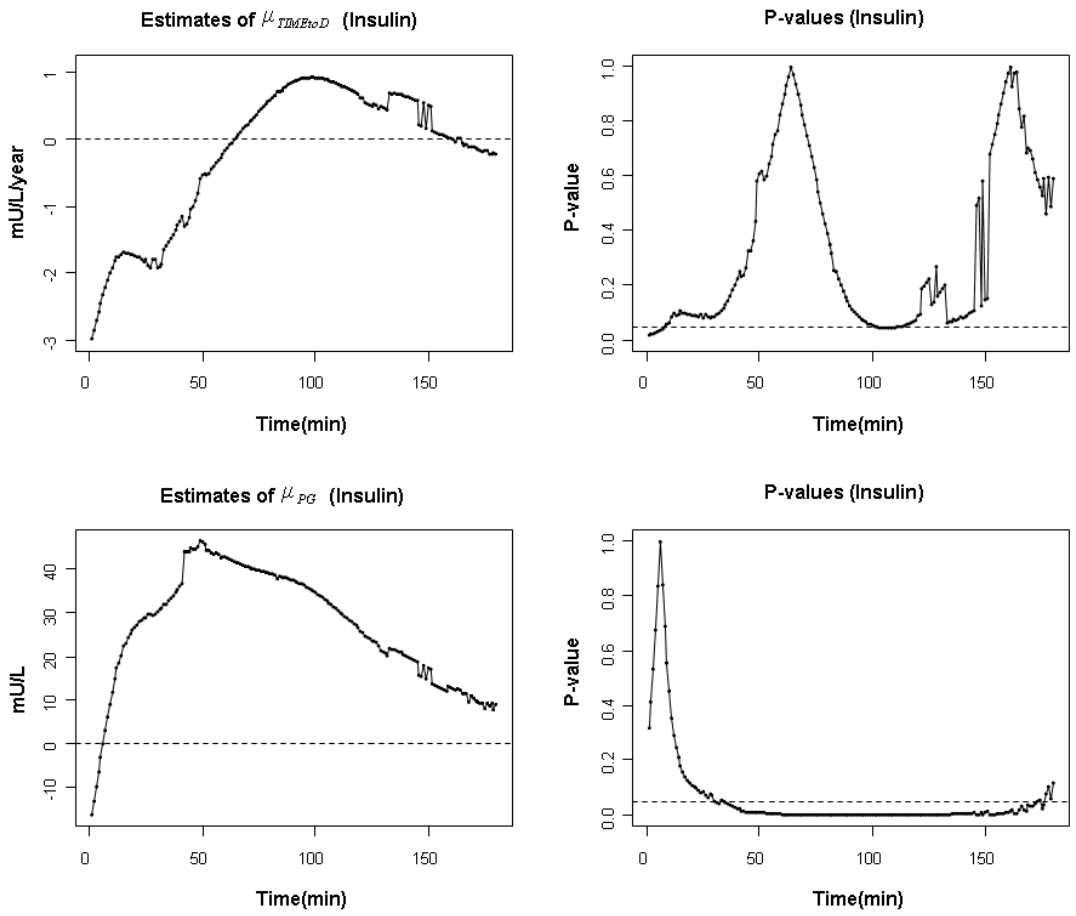


Figure 4.5. Estimates and p-values of μ_{TIMEtoD} and μ_{pg} of the DDVs based on non-parametric estimation of insulin level from 1 to 180 minute. The dashed lines in the p-value plots are drawn at the significance level 0.05

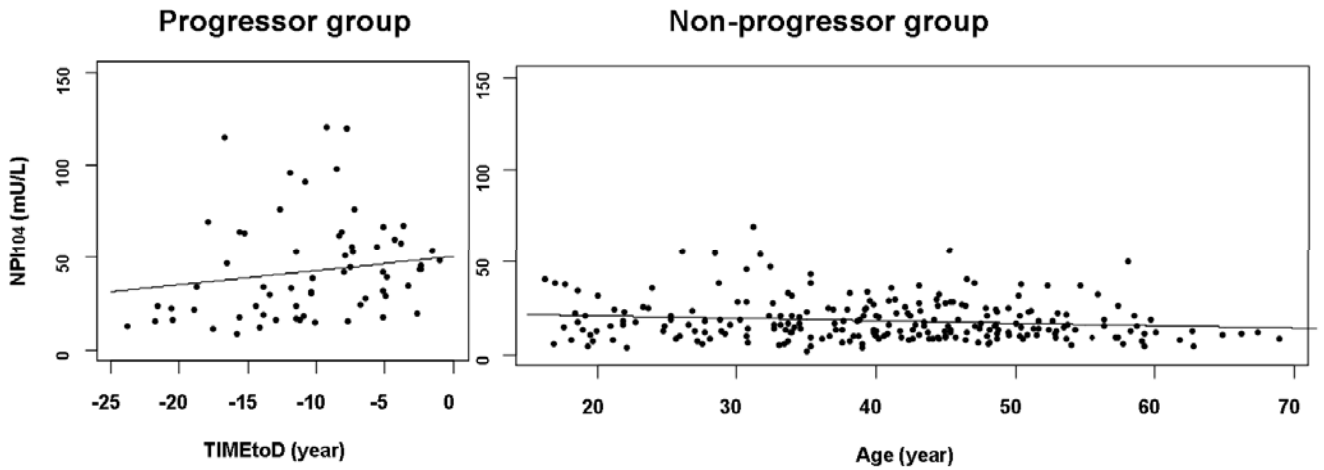


Figure 4.6. The population fits of the DDVs of NPI_{104} in the progressor group and non-progressor group.

CHAPTER 5. DISEASE DEVELOPMENT MODELING OF TYPE 2 DIABETES IN OFFSPRING OF DIABETIC PARENTS BASED ON DISEASE DEVELOPMENT VARIABLES DERIVED FROM OGTT

5.1 Abstract

The objectives of the study were to identify important variables associated with the development of type 2 diabetes (T2D) based on glucose and insulin measurements from an oral glucose tolerance test (OGTT) and to evaluate the temporal patterns of these variables during the development of the disease. Repeated OGTTs were performed for up to 25 years on 149 healthy offspring of two parents with T2D. Types 2 diabetes developed in 24 of the participants and did not develop in the remaining 125 subjects. Disease development variables (DDV) derived from glucose and insulin measurements and the temporal patterns of DDVs were compared by using population-based approaches, including a mixed effect analysis and Bayesian hierarchical analysis. The DDVs defined as the glucose concentration at 90 minute, the difference of glucose concentrations between 60 and 30 minutes, and the area under the glucose curve from 0 to 240 minutes were significant and appeared as good as, or better than the fasting blood glucose or the 2-h glucose OGTT levels commonly used to identify T2D. The study demonstrated that the proposed disease development analysis is able to identify important DDVs which show significant temporal changes during the development period of T2D. The newly identified DDVs are valuable for the understanding of the pathogenesis of T2D.

5.2 Introduction

Oral glucose tolerance test (OGTT) is a standard diagnostic test of type 2 diabetes (T2D) adopted by World Health Organization (WHO) and American Diabetes Association (ADA) (153). The 2 hour OGTT glucose concentration has been found to be one of the most important predictors of "worsening to diabetes" in many longitudinal prospective studies (154-157). OGTT has been used to confirm the diagnosis of T2D and impaired glucose tolerance, identified by the fasting glucose test (158). In addition, several methods have been developed to estimate insulin sensitivity and β-cell function based on data from OGTT (93, 94, 120).

Due to the simplicity and more natural glucose loading of OGTT in comparison to an intravenous glucose tolerance test (IVGTT), OGTT is more frequently used than IVGTT in evaluating the development of T2D. Although OGTT provides valuable information about the metabolic syndrome of T2D, the temporal patterns of the variables derived from OGTT data and the time-dependency of these variables during the development of T2D are not clear. Recently, with the development population-based analysis, various disease progression models have been developed to describe the progression of diseases and evaluate the temporal patterns of related biomarkers (81, 84, 99). In the disease progression model of de Winter *et al.,* (99) the effects of anti-diabetic drugs on the progression of T2D was evaluated based on repeatedly measured fasting glucose, fasting insulin, and glycated hemoglobin A_{1c} (A1C). However, the study only focused on subjects who already had been diagnosed with T2D and did not consider any measurements from OGTT to evaluate the progression of the disease.

In contrast, this study is focusing on the identification of disease *development* variables (DDV) associated with the development of T2D before the diagnosis of the disease. By applying a disease development model analysis of the glucose and insulin data from OGTT, the temporal patterns of DDVs were examined by a population-based analysis to help in the understanding of the pathogenesis of T2D. Such disease development analysis provides a continuous longitudinal evaluation of the developing stage of T2D, which is an important quantitative foundation for evaluating the effect of interventions/prevention strategies such as drug treatment and lifestyle changes.

5.3 Specific aim and hypothesis

The specific aim of this chapter is to identify important disease development variables associated with the development of type 2 diabetes based on glucose and insulin measurements from an oral glucose tolerance test and to evaluate the relationship between the temporal patterns of these variables and the development of the disease.

The specific hypothesis is that the important disease development variables derived from an oral glucose tolerance test are hypothesized to exist and can be identified by the proposed disease development analysis.

5.4 Methods

5.4.1 Subjects

The study is based on a 25-year follow-up study in Joslin Diabetes Center (JDC), Boston, Massachusetts (24). Between 1963 and 1983, 155 healthy offspring of parents with T2D were recruited. The study was approved by the human study review board of JDC and conducted according to the Declaration of Helsinki: “Recommendations guiding

physicians in biomedical research involving human subjects", adopted by the 18th World Medical Assembly, Helsinki, Finland, June 1964. Informed consent was obtained from all individuals before participation in the study. All participants had normal glucose tolerance and received repeated OGTTs during followed-up for up to 25 years for tracking the status of T2D. The detailed methods of follow-up and evaluation of T2D have been described previously (23, 24).

A total of 552 OGTTs were performed during the follow-up for all participants. At the end of follow-up, the subjects were divided into a progressor group and a non-progressor group according to the T2D status. Data of 25 OGTTs with missing blood samples were excluded from the analysis. After data exclusion, a total of 92 and 435 OGTTs from the 24 progressors and 125 non-progressors, respectively, were used in the analysis. Characteristics of the progressor and the non-progressor groups are summarized in Table 5.1.

5.4.2 Oral glucose tolerance tests

All participants were instructed to consume a high-carbohydrate diet over three days (250 to 300 g/day) and have an overnight fast before OGTT. Before the test, a blood sample was collected as the fasting sample and then 100g glucose in solution was taken orally at time zero over a 5 minutes period and blood samples were collected at 15, 30, 45, 60, 90, 120, 180 and 240 minutes. The blood glucose concentrations were measured by the ferricyanide method for the OGTTs done before 1983 with a coefficient of variation (CV) of 1.5% (117). After 1983, the glucose oxidize method was used with a 1.35% CV (150). The insulin serum concentrations were measured by a double-antibody radio-immunoassay with a 17.6% CV (109).

5.4.3 Identifications of disease development variables

A disease development variable (DDV) is a data-derived parameter quantitatively associated with the disease development and can potentially serve as a biomarker for the development of disease. In this study, DDVs of T2D were sought among the glucose and insulin measurements collected from the OGTTs. The potential DDVs were longitudinally repeatedly measured during the follow-up period and then compared by a population-based mixed effect model analysis. The mixed effect model had the following mathematical expectation formulation for non-progressor and progressor group:

$$\begin{aligned}
&\text{Non - progressor group}\\
&\mathrm{DDV}_{ij} = k_{base,i} + k_{age,i} \cdot AGE_{ij} + \varepsilon_{ij}\\
&\text{Progressor group}\\
&\mathrm{DDV}_{ij} = k_{base,i} + k_{pg,i} \cdot PG_i + k_{age,i} \cdot AGE_{ij} + k_{TIMEtoD,i} \cdot TIMEtoD_{ij} + \varepsilon_{ij}\\
&\varepsilon_{ij} \sim Norm\left(0, \sigma^2\right)\\
&k_{base,i} = \mu_{base} + \eta_{base,i}\\
&k_{pg,i} = \mu_{pg} + \eta_{pg,i}\\
&k_{age,i} = \mu_{age} + \eta_{age,i}\\
&k_{TIMEtoD,i} = \mu_{TIMEtoD} + \eta_{TIMEtoD,i}\\
&\begin{pmatrix} \eta_{base,i} \\ \eta_{pg,i} \\ \eta_{age,i} \\ \eta_{TIMEtoD,i} \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma\right)
\end{aligned} \tag{5.1}$$

To summarize, the model was used to evaluate the time-dependency of the DDV over age (AGE) and over the time prior to the diagnosis of T2D (TIMEtoD) in the non-progressor and progressor group. In the non-progressor group, the national time-dependency of the DDV is modeled as a linear function of age. While the extra effects from disease development of T2D on the DDV in the progressor group are modeled additively by the predictor variables, TIMEtoD and progressor group indicator (PG). The variable TIMEtoD is the normalized time variable for progressors whose diagnostic times of T2D

were known and calculated as the subject's age at OGTT minus the subject age at the diagnosis of T2D. PG is the indicator variable for the progressor group. PG equals to one if the subject is progressor and zero if the subject is in the non-progressor group.

The effects of the predictor variables were described as the mixed effect parameters (k), which are the sum of fixed effect parameters (μ) and random effect parameters (η). The fixed effects parameters (population parameters) of the progressor group indicator (μ_{pg}) and the time prior to the diagnosis of T2D ($\mu_{TIMEtoD}$) are the two key parameters used to distinguish the two groups and used to characterize the temporal change of the DDV in the progressor group. Accordingly, a DDV is considered significant when the p-values of the two fixed effect parameters (μ_{pg} and $\mu_{TIMEtoD}$) are both less than 0.05. The mixed effect model analysis was performed in R version 2.10.1 by the lme() function (139).

In the mixed effect analysis, the tested DDV candidates included the direct measurements of glucose and insulin concentrations, the differences of two glucose or insulin concentrations, and non-parametrically estimated glucose-insulin kinetic parameters. The non-parametrically determined glucose-insulin kinetic parameters include the maximum glucose and insulin concentrations (C_{gmax} and C_{imax}), the times to reach C_{gmax} and C_{imax} (t_{gmax} and t_{imax}), and the areas under the curve of glucose and insulin from 0 to 240 minutes ($AUC_{g0\text{-}240}$ and $AUC_{i0\text{-}240}$). The parameters, $C_{gmax,}$ $C_{imax,}$ t_{gmax} and t_{imax} were estimated by generalized cross-validation cubic spline curves fits to glucose and insulin concentrations using the R function smooth.spline(). The areas under the curve ($AUC_{g0\text{-}240}$ and $AUC_{i0\text{-}240}$) were estimated by the trapezoidal method.

5.4.4 Investigation of the temporal patterns of DDVs

In the mixed effect analysis, the temporal pattern of each DDV prior to the diagnosis of T2D is assumed to be linear for fast evaluations of all of the DDV candidates in R. However, the DDV may have some distinct different patterns deviating from a straight line relationship. Therefore, the temporal patterns of the important DDVs identified previously were investigated by a non-parametric, pooled data analysis to visualize the possible temporal changes over TIMEtoD. Subsequently, a Bayesian hierarchical analysis was then applied to fit to the corresponding possible pattern of the DDV over TIMEtoD by the function $f(\text{TIMEtoD}_{ij})$ in the following Bayesian hierarchical model (5.2):

$$\begin{aligned}&\text{Non - progressor group}\\ &\text{DDV}_{ij} = k_{base,i} + k_{age,i} \cdot AGE_{ij} + \varepsilon_{ij}\\ &\\ &\text{Progressor group}\\ &\text{DDV}_{ij} = k_{base,i} + k_{pg,i} \cdot PG_i + k_{age,i} \cdot AGE_{ij} + f(TIMEtoD_{ij}) + \varepsilon_{ij}\end{aligned} \tag{5.2}$$

The above model is similar to the model used in the mixed effect analysis. The difference is the function $f(\text{TIMEtoD}_{ij})$ which represents the functional form of the temporal pattern of the DDV over TIMEtoD after adjusting the progressor effect for age effect. Bayesian hierarchical analysis is applied because it provided a better successful rate of convergence for various functions to address the different temporal patterns and flexible statistical models than the mixed effect analysis employed by the lme() or nlme() functions in R. Bayesian hierarchical analysis was performed in WINBUGS 1.4.3(60). The hierarchy of the Bayesian model and the assumptions of parameter distributions are summarized as follows:

Individual level (5.3)

$$\mathrm{DDV}_{ij} = \mathrm{k}_{\mathrm{base,i}} + \mathrm{k}_{\mathrm{pg,\ i}} \cdot \mathrm{PG}_{\mathrm{i}} + \mathrm{k}_{\mathrm{age,i}} \cdot \mathrm{AGE}_{\mathrm{ij}} + \mathrm{k}_{\mathrm{TIMEtoD,\ i}} \cdot \left(\mathrm{TIMEtoD}_{\mathrm{ij}} - \mathrm{k}_{\mathrm{sep,\ i}}\right)_{+} + \varepsilon_{ij}$$

$$(a)_{+} = \begin{cases} a & if\ a > 0 \\ 0 & else \end{cases} \qquad \varepsilon_{ij} \sim normal(0, \tau)$$

Population level (5.4)

$$k_{base,i} \sim normal(\mu_{base}, \tau_{base}) \qquad \sigma_{base} = 1/\sqrt{\tau_{base}}$$
$$k_{age,i} \sim normal(\mu_{age}, \tau_{age}) \qquad \sigma_{age} = 1/\sqrt{\tau_{age}}$$
$$k_{pg,i} \sim normal(\mu_{pg}, \tau_{pg}) \qquad \sigma_{pg} = 1/\sqrt{\tau_{pg}}$$
$$k_{TIMEtoD,i} \sim normal(\mu_{TIMEtoD}, \tau_{TIMEtoD}) \qquad \sigma_{TIMEtoD} = 1/\sqrt{\tau_{TIMEtoD}}$$
$$k_{sep,i} \sim normal(\mu_{sep}, \tau_{sep})I(-22, -4.57) \qquad \sigma_{sep} = 1/\sqrt{\tau_{sep}}$$

Prior (5.5)

$$\tau \sim gamma(0.01, 0.01) \qquad \sigma = 1/\sqrt{\tau}$$
$$\mu_{base} \sim normal(0,\ 1/1000^2) \qquad \sigma_{base} \sim uniform(0,1000)$$
$$\mu_{age} \sim normal(0,\ 1/1000^2) \qquad \sigma_{age} \sim uniform(0,1000)$$
$$\mu_{pg} \sim normal(0,\ 1/1000^2) \qquad \sigma_{pg} \sim uniform(0,1000)$$
$$\mu_{TIMEtoD} \sim normal(0,\ 1/1000^2) \qquad \sigma_{TIMEtoD} \sim uniform(0,1000)$$
$$\mu_{sep} \sim uniform(-22, -4.5) \qquad \sigma_{sep} \sim uniform(0,1000)$$

In the individual level of the hierarchy, the data was fitted to the two-slope model which is close to the temporal patterns shown in Figure 5.1. This two-slope model also provided easy interpretations of parameters and simple function to describe the disease development. The residuals ($\varepsilon_{i,j}$) were assumed to be normal distributed with a mean of zero and a precision of τ, where *j* denoted the sequencing of the observations of the i^{th} subject. In the population level of the hierarchy, the individual parameters, $k_{base,i}$, $k_{age,i}$, $k_{pg,i}$, and $k_{TIMEtoD,i}$ were assumed to be normal distributed with mean (μ) and precision parameters (τ). The individual separation point parameters ($k_{sep,i}$), were summarized by

a truncated normal distribution. In the highest level of the hierarchy, prior distributions were assigned to the population parameters and residuals. Vague normal prior distributions were given to the population parameters (μ_{base}, μ_{age}, μ_{pg}, and $\mu_{TIMEtoD}$). A uniform prior distribution was assigned to the population break-point parameter(μ_{sep}) with a support from -22 to -4.57 corresponding to the 10% and 90% percentiles of TIMEtoD to prevent a local convergence at the extremes which results in a collapse of the two-slope model to a single-slope model. Vague uniform prior distributions were given to σ_{base}, σ_{age}, σ_{pg}, $\sigma_{TIMEtoD}$, and σ_{sep}. A vague gamma distribution was given to the precision (τ) of the residuals.

All of the posterior distributions of the parameters were estimated by the Markov chain Monte Carlo method in WinBUGS version 1.4.3 (159). Three Markov chains were run simultaneously for 20,000 samples for each parameter. The convergence of the chains was accessed by the Gelman-Rubin statistic (137). After burn-in with the first 10,000 samples in each chain, a total of 30,000 samples from the three chains were used to estimate the posterior distribution of each parameter. The means of the parameters' posterior distributions are Bayesian estimates of the parameters. The 95% Bayesian credible sets (95% C.S.) is the intervals between the 2.5 and 97.5 percentiles of the posterior distributions.

5.5 Results

5.5.1 Identifications of disease development variables

DDVs of direct measurements

The repeatedly measured glucose and insulin concentrations in OGTTs were evaluated by the mixed effect model. The DDVs with significant μ_{pg} and $\mu_{TIMEtoD}$ (p-value<0.05) are listed in Table 5.2. All of the significant DDVs are based on the measurements of glucose but not insulin. As expected, the standard diagnosis criteria of T2D, fasting blood glucose (FBG) and 2 hours glucose concentration (OG_{120}), are significant and have temporal increasing patterns during the disease development period (Figure 5.1). In addition to FBG and OG_{120}, the DDV defined as the glucose concentration at 90 minutes (OG_{90}) gave the most significant μ_{pg} and $\mu_{TIMEtoD}$. The DDV, OG_{90}, is potentially at least as good as FBG and OG_{120} for describing the development of T2D.

The DDVs defined as the difference of two glucose or insulin concentrations were also tested by the mixed effect model. The most significant DDVs defined as a difference of two glucose or insulin measurements are shown in Table 5.3. The DDV of the difference between glucose concentrations at 60 and 30 minutes (OG_{60-30}) was found to be the most significant DDV based on two glucose measurements. OG_{60-30} gave more significant $\mu_{TIMEtoD}$ than OG_{120} and more significant μ_{pg} than both FBG and OG_{120}. For the DDVs of the difference of two measurements of insulin, the DDV defined by the difference between insulin concentration at 120 and 45 minutes (OI_{120-45}) was significant

(p-values <0.05). However, the temporal pattern of OI_{120-45} was not as clear as other DDVs determined by glucose measurements mentioned previously (Figure 5.1).

DDVs defined by non-parametric analysis

Representative cubic spline fits of glucose and insulin data of two OGTT data are shown in Figure 5.2. The $C_{gmax,}$ $C_{imax,}$ t_{gmax}, t_{imax}, AUC_{g0-240} and AUC_{i0-240} were estimated non-parametrically and subsequently analyzed by the mixed effect model (5.1). The results of fixed effect parameter estimations of these DDVs are summarized in Table 5.4. The DDV, AUC_{g0-240}, is the most significant DDV determined by the non-parametric approach and gave the smallest p-values for μ_{pg} and $\mu_{TIMEtoD}$.

5.5.2 Temporal patterns of disease development variables

The temporal patterns of six DDVs namely, FBG, OG_{120}, OG_{90}, OG_{60-30}, OI_{120-45}, and AUC_{g0-240}, were investigated by the pooled data approach using cubic spline functions (Figure 5.1). Based on the cubic spline curves, four DDVs including FBG, OG_{120}, OG_{90} and AUC_{g0-240} showed a two-slope pattern, while OG_{60-30} and OI_{120-45} showed a simple linear increase. The DDVs with the two-slope pattern had a stationary stage approximately 10 years prior to the diagnosis of T2D followed by a progression stage for the development of the disease. The two-slope model assumes the DDV changes after the separation point parameter, $k_{sep,\,i}$, which is the time of the separation of the two-slopes. The estimates of the population parameters (μ) are summarized in Table 5.5.

Similar to the mixed effect model, the individual parameters (k_i) were summarized by the population mean parameter (μ) in the Bayesian approach. The estimated separation parameters μ_{sep} of FBG, OG_{120}, OG_{90}, and AUC_{g0-240} ranged from -

10.3 to -6.87 years. This suggests that the development of T2D follows a two-slope pattern starting at 6.87 to 10.23 years before the diagnosis of T2D.

5.6 Discussion

In regular OGTTs, only the blood samples of fasting glucose and the 2 hour glucose concentrations are collected. Although the 2 hour OGTT glucose is one the most sensitive biomarkers for diagnosis of T2D (157, 160), the blood samples collected at other time points are still valuable for evaluating the pathophysiology of T2D. The current study suggests that the blood sample collected at 45, 60, 90 120, 180, 240 post an OGTT are all significantly associated with the development of T2D, and DDVs derived from the additional samples (e.g. OG_{90} and $OG_{60\text{-}30}$) are potential biomarker of T2D that may perform better, or as well as OG_{120}. The OGTT has the potential to provide better prediction and diagnosis of T2D when incorporating such extra glucose measurements.

The temporal pattern of FBG and OG_{120}, which are the two common criteria for diagnosis of impaired glucose tolerance and T2D, were also evaluated in this study. Both FBG and OG_{120} showed the two-slope temporal pattern which suggests there is a stationary stage followed by an exaggeration stage before the diagnosis of T2D. The separation point parameters (μ_{sep}) between the two-slopes were estimated as -9.88 [-12, -8.45] (mean 95% C.S.) years and -6.87 [-8.9, -5.22] years for FBG and OG_{120}, respectively. The FBG and OG_{120} parameters provide different information about the development of T2D. The FBG parameter gives information about the basal line glucose level at homeostasis state, while OG_{120} gives information about glucose tolerance following an oral glucose stimulation. The results suggest that the impaired fasting glucose tolerance may start earlier than the impaired oral glucose tolerance. For the

DDVs defined by the differences of two measurements, OG_{60-30}, showed the most significant and consistent linear increasing pattern over the entire follow-up period in the progressor group. In contrast to OG_{120}, which relates to the late state of an OGTT the OG_{60-30} serves as a DDV of T2D relating to the glucose tolerance in the early phase of OGTT. Although a high fasting insulin and a low 2 hour OGTT insulin concentration has been reported as predictors of T2D (131, 148), these variables did not show an increasing or decreasing pattern in our analysis. The only significant DDV defined by insulin measurements is the difference between insulin concentrations at 120 and 45 minutes (OI_{120-45}), but this DDV just showed a weak increasing pattern (Figure 5.1) and was not as informative as DDVs defined by glucose measurements. The temporal patterns of the FBG, OG_{120}, $OG0_{90}$, OG_{60-30} are shown in Figure 5.3 to visualize the time-course of the development of the disease in the progressor group as compared to the non-progressor group. The DDVs defined by OG_{90} and OG_{60-30} appear as good or better variables for disease development as FBG or OG_{120} and provide added value to an OGTT by only requiring 3 extra samples.

In the non-parametric analysis, AUC_{g0-240} is the most significant DDV associated with the development of T2D. Since AUC_{g0-240} is mainly dependent on all of the measurements of glucose, the increasing pattern of AUC_{g0-240} can be explained by the significant increasing patterns of the individual DDVs derived by glucose levels at 45, 60, 90, 120, 180 and 240 minutes.

The intent of the proposed population-based approach is to deal with the routinely collected OGTT data and addresses the time-dependency of the measurements from OGTT during disease development period of T2D. In comparison to the classical logistic

regression analysis, the proposed approach is focusing on the temporal changes of the time-variant DDVs and provides a time-sensitive quantification of the disease development rather than risks or incidence rate of T2D. Both mixed effect and Bayesian hierarchical statistical analysis were used in the study. For the initial identification of DDVs, the mixed effect analysis profromed in R 2.10.1 gave a fast and general "screening" for all possible DDV candidates. For the later evaluations of the temporal patterns of DDVs, the Bayesian approach was preferred because of its flexibility for different kind linear and non-linear functions and high successful rate of convergence.

Study limitations

In the study, only the off-spring of parents with type 2 diabetes were recruited and the oral glucose dose used in OGTT was 100g which is different from the current standard of 75g suggested by WHO and ADA (161). Hence, the results of the study may not be fully applicable to the current 75g OGTT. Furthermore, the changes in DDVs for the non-progressor group were assumed to be purely an age effect. This assumption may be partly violated because some of the subjects in the non-progressor group may eventually develop T2D following the completion of the follow-up study. Despite these limitations, the proposed approach provides a quantitative way to identify important DDVs and to evaluate the temporal pattern of these DDVs during the development of T2D.

5.7 Conclusion

The proposed disease development analysis is able to identify important DDVs characterized by significant temporal changes during the development period of T2D. Although the few extra blood samples during an OGTT will involve more cost and

extend the testing, this may be offset by the valuable information obtained. The proposed approach offers a new way to analyze development of T2D and provides valuable quantitative longitudinal information about the pathogenesis of T2D. Importantly, the approach is particularly useful for quantifying the effect of interventions/prevention strategies such as drug treatments and lifestyle modifications.

Table 5.1. Summary of characteristics of the progressor and non-progressor groups

	Progressor group	Non-progressor group
Number of subjects	24	125
Number of OGTTs performed	92	435
Gender (males in %)	58%	43.2%
Starting age (mean ± sd) range	34.1 ± 7.89 21.11-49.8	33.4 ± 10 13.4-59
Starting BMI (mean ± sd) range	32.5 ± 7.79 23.2-53.1	25.4 ± 4.42 17.2-40.6
Years followed (mean ± sd) range	13.2 ± 5.72 3.22-23.5	12.5 ± 6.48 0.63-24.3

Table 5.2. Summary of the estimates and p-values of fixed effect (population) parameters for one measurement-based DDVs resulting in p-values for μ_{pg} and $\mu_{TIMEtoD}$ that are both less than 0.05

DDV	μ_{base}		μ_{age}		μ_{pg}		$\mu_{TIMEtoD}$	
	Estimate	(p-value)	Estimate	(p-value)	Estimate	(p-value)	Estimate	(p-value)
FBG	69.9	(<0.001)	0.205	(<0.001)	15.9	(<0.001)	1.27	(<0.001)
OG_{45}	108	(<0.001)	0.409	(0.00480)	37.9	(<0.001)	1.35	(0.00919)
OG_{60}	97.6	(<0.001)	0.535	(0.00109)	47.3	(<0.001)	1.87	(0.00136)
OG_{90}	82.7	(<0.001)	0.552	(<0.001)	58.2	(<0.001)	3.03	(<0.001)
OG_{120}	76.7	(<0.001)	0.407	(<0.001)	49.3	(<0.001)	3.07	(0.00159)
OG_{180}	72.5	(<0.001)	0.0165	(0.879)	35.4	(<0.001)	2.63	(0.00112)
OG_{240}	72.7	(<0.001)	-0.205	(0.00738)	14.6	(0.00921)	1.52	(<0.001)

Table 5.3. Summary of the estimates and p-values of fixed effect (population) parameters of the most significant DDVs defined as the difference of two glucose, or two insulin measurements.

DDV	μ_{base}		μ_{age}		μ_{pg}		$\mu_{TIMEtoD}$	
	Estimate	(p-value)	Estimate	(p-value)	Estimate	(p-value)	Estimate	(p-value)
OG_{60-30}	-19.3	(<0.001)	0.432	(<0.001)	29.5	(<0.001)	1.46	(0.00109)
OI_{120-45}	-40.3	(0.00415)	0.902	(0.0189)	67.2	(0.00127)	2.87	(0.0495)

Table 5.4. Summary of estimates and p-values of fixed effect (population) parameters of non-parametric kinetic parameters, C_{gmax}, t_{gmax}, AUC_{g0-240}, C_{imax}, t_{imax}, and AUC_{i0-240}

DDV	μ_{base}		μ_{age}		μ_{pg}		$\mu_{TIMEtoD}$	
	Estimate	(p-value)	Estimate	(p-value)	Estimate	(p-value)	Estimate	(p-value)
C_{gmax}[(a)]	107	(<0.001)	0.425	(0.00189)	46.0	(<0.001)	2.07	(0.00376)
t_{gmax}	46.1	(<0.001)	0.145	(0.183)	22.4	(0.00528)	1.25	(0.291)
AUC_{g0-240}[(a)]	20100	(<0.001)	49.3	(0.00767)	8620	(<0.001)	513	(<0.001)
C_{imax}	119	(<0.001)	-0.0355	(0.944)	134	(0.0803)	3.07	(0.534)
t_{imax}	76.4	(<0.001)	0.164	(0.223)	24.1	(0.00236)	1.25	(0.266)
AUC_{i0-240}	19200	(<0.001)	-37.0	(0.650)	28900	(0.130)	1480	(0.424)

[(a)]: p-values for μ_{pg} and $\mu_{TIMEtoD}$ are both less than 0.05

Table 5.5. Summary of Bayesian estimates and 95% credible sets (95% C.S.) of population mean parameters of the two-slope model for the four DDVs

DDV	Estimate of μ_{base} (95% C.S.)	Estimate of μ_{age} (95% C.S.)	Estimate of μ_{pg} (95% C.S.)	Estimate of μ_{sep} (95% C.S.)	Estimate of $\mu_{TIMEtoD}$ (95% C.S.)
FBG	69.5	0.214	0.528	-9.88	2.05
	(66.4, 72.8)	(0.128, 0.301)	(-3.16, 4.25)	(-12.0, -8.45)	(0.647, 3.66)
OG_{120}	75.4	0.444	16.1	-6.87	8.72
	(66.7, 84.3)	(0.226, 0.668)	(7.04, 25.0)	(-8.90, -5.22)	(2.49, 16.1)
OG_{90}	85.7	0.469	13.3	-10.3	5.58
	(72.6, 95.6)	(0.194, 0.809)	(2.16, 24.7)	(-12.74, -7.48)	(2.76, 8.87)
$AUC_{g0\text{-}240}$	10650	279	1970	-9.76	757
	(8950, 12300)	(235, 327)	(530, 3400)	(-13.3, -6.96)	(188, 1390)

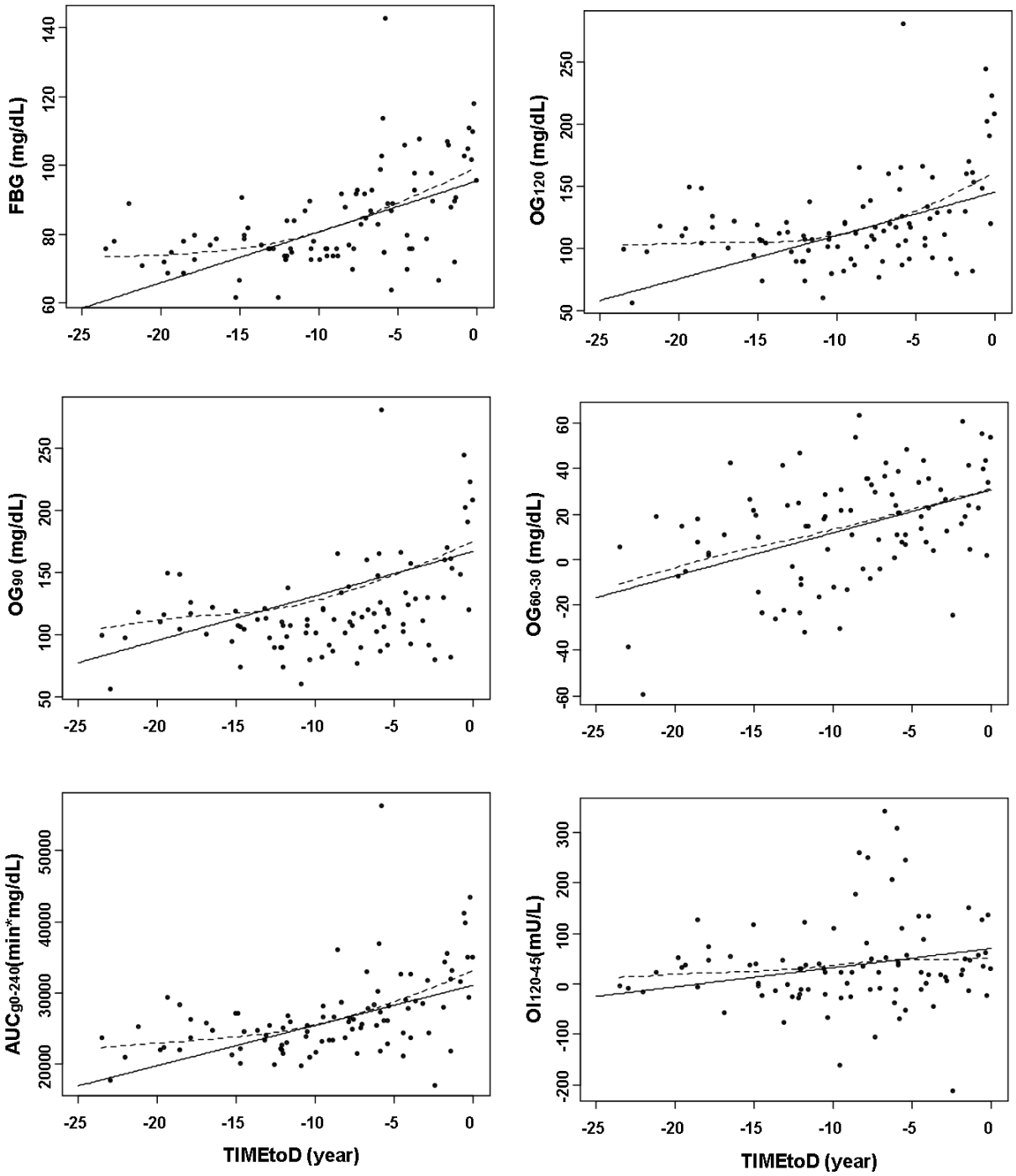


Figure 5.1. Population fits of the disease development variables (DDV) defined by FBG, OG_{120}, OG_{90}, $OG_{60\text{-}30}$, $AUC_{g0\text{-}240}$, and $OI_{120\text{-}45}$ in the progressor group over the time prior to the diagnosis of T2D (TIMEtoD). The solid lines represent the mixed effect model fits and the dashed curves represent the cubic spline fitted curves.

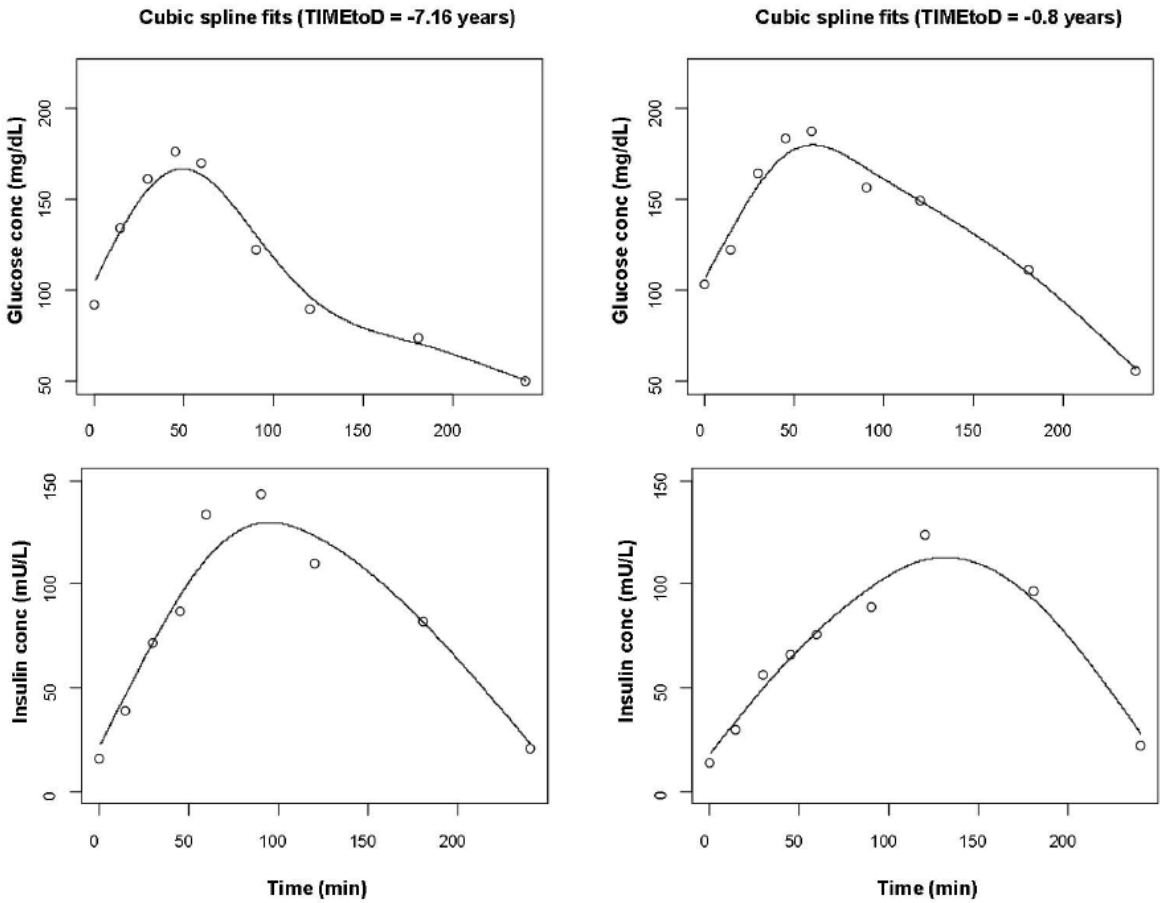


Figure 5.2. Representative cubic spline fits of glucose and insulin data of two OGTTs from the same subject over the time prior to the diagnosis of T2D (TIMEtoD) equal to -7.16 and -0.8 years.

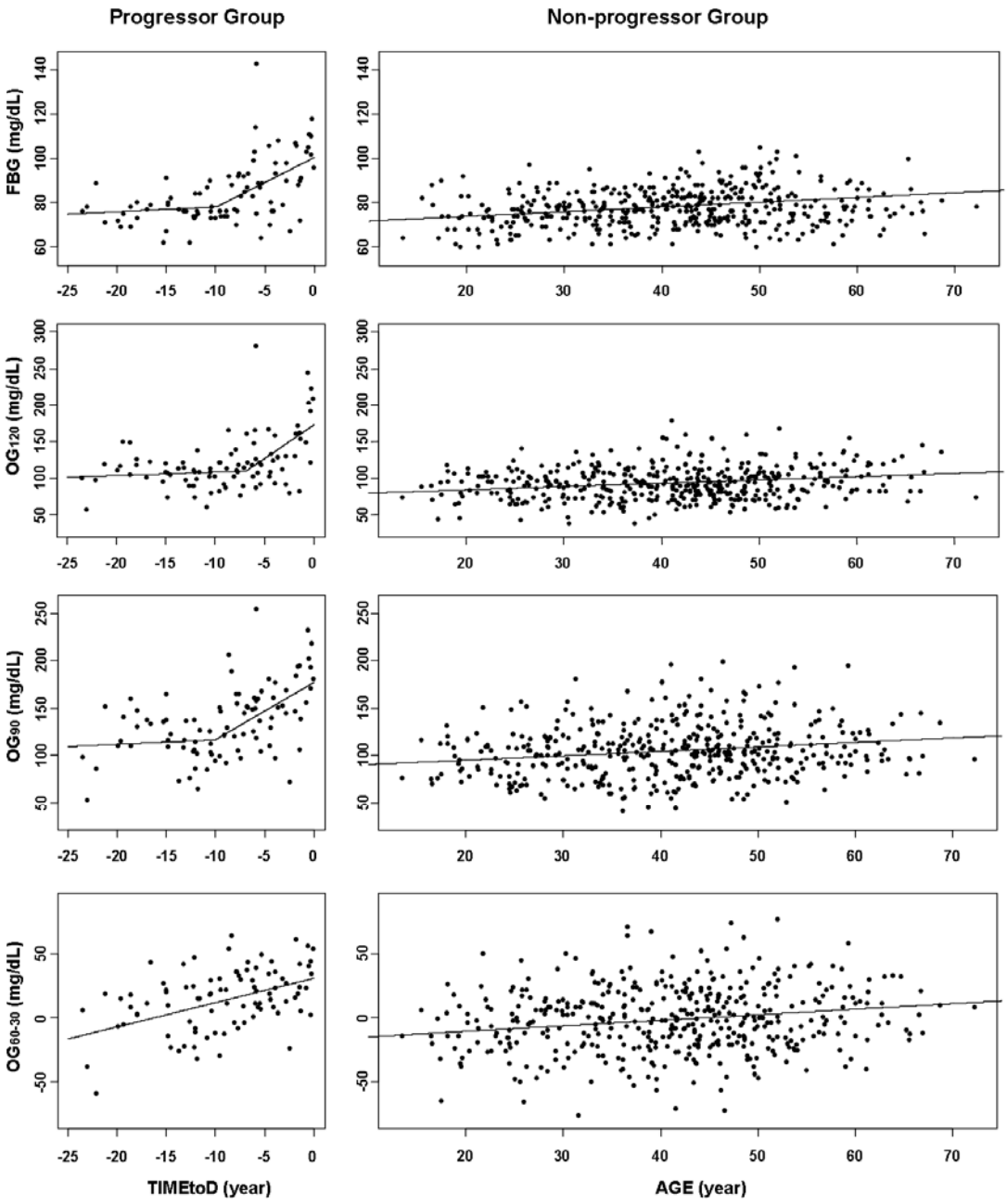


Figure 5.3. Population fits of the DDVs, FBG, OG_{120}, OG_{090}, and $OG_{60\text{-}30}$, in the progressor group (n=24) and non-progressor group (n=125) over the time prior to the diagnosis of T2D (TIMEtoD). In the population fit for the progressors, the x-axis represents TIMEtoD to emphasize the two-slope pattern over TIMEtoD. In the non-progressors, the x-axis represents age to illustrate the natural change over time in the DDV in the non-progressor group.

CHAPTER 6 CONCLUSIONS AND FUTURE WORKS

6.1 Conclusions

In chapter 2, a simple glucose-insulin kinetic model was proposed. This model was developed though a comparison of the glucose-insulin kinetic models with different dynamic effects functions. With its minimal structure, this model is readily analyzed in a population pharmacokinetic framework. The model parameters have intuitive meanings that are associated with the physiology of the glucose removal and insulin production. The proposed model shows it ability to describe the glucose-insulin feedback loops and biphasic insulin secretion. The abnormal changes in glucose-insulin kinetics and biphasic insulin secretion are closely associated with the development of type 2 diabetes. By employing a population Bayesian analysis, glucose-insulin kinetics based differences were found between the progressor and non-progressor groups of type 2 diabetes. Ditermination of early pre-diabetic pharmacokinetic differences may be helpful for better prediction of the development of T2D.

Although many risk factors of type 2 diabetes have been indentified, the temporal patterns and the time-dependency of these risk factors were still not clear. To further understand the pathophysiology of type 2 diabetes, a disease development model of type 2 diabetes was developed (chapter 3). Based on repeated measured fasting blood glucose, fasting serum insulin, HOMA-IR, and BMI, the disease development modeling is an extension of chapter 2. A two-slope model was successfully developed for fasting blood glucose, one of the gold standards for diagnosis of type 2 diabetes. In contrast to the classic logistic regression analysis, this work is aimed at a time-based evaluation of the developing stage of the disease and to provide a mechanistic and quantitative basis for

evaluating status of type 2 diabetes. Accordingly, the proposed model may be used to quantify the effect of interventions aimed at preventing or slowing down the development of this disease. This is particularly important in dealing with diseases like type 2 diabetes which have a very long development period before diagnosis.

By applying the proposed disease development analysis, important DDVs derived from IVGTT were identified in chapter 4. This is the first attempt in the literature to identify risk factors of type 2 diabetes based on their temporal patterns before the diagnosis of the disease using repeated measured IVGTT data. The newly identified DDVs, reduced glucose removal ($G_{10\text{-}50}$), diminished first phase insulin secretion ($I_{3\text{-}20}$), and increased second phase insulin secretion ($I_{120\text{-}0}$), were found to be associated with the pathophysiology of the disease. These new propsed DDVs are easy to calculate and providing a quick assessment of the development of type 2 diabetes.

In chapter 5, important DDVs derived from OGTT were identified by the proposed disease development analysis. Two new identified DDVs, OG_{90} and $OG_{60\text{-}30}$, appear as good or better variables for disease development as FBG or OG_{120} used by American Diabetes Association and World Health Organisation. The study has the potential to complement or replace the classic OGTT and provide more accurate prediction for the development of type 2 diabetes.

The proposed population model of glucose-insulin kinetics has demonstrated that pharmacokinetic differences exists for the high risk population and can be helpful for prediction of T2D. By applying the proposed disease development analysis, the time-dependency and temporal patterns of the DDVs can be identified. An examination of the temporal changes in DDVs for the glucose-insulin system before the diagnosis of the

disease provides a quantitative evaluation of the pathophysiological evolution of T2D and is valuable in predicting T2D.

6.2 Future work

6.2.1 Continue the follow-up

In this study, 25 subjects developed type 2 diabetes out of 152 healthy offspring of diabetic parents during the follow-up. Although this data set appears to be the most comprehensive data set avaliable with such a long follow-up time (up to 25 years), the group of 25 subjects in the progressor group is statistically not considered as large population. Since some non-progressors can develop type 2 diabetes after the last follow-up, continuing the follow-up would be one of the best ways to increase the sample size of the progressor group and provide a better reference group in addition to increasing the data of the repeated IVGTTs and OGTTs. With new follow-up data, it is possible to provide a better estimation for predicting type 2 diabetes and to find important temporal patterns of the DDVs relative to type 2 diabetes.

6.2.2 Incorporate more variables

This thesis is focusing on risk factors of type 2 diabetes based on glucose-insulin kinetics. There are many other risk factors, in addition to the glucose-insulin kinetics, associated with the development of type 2 diabetes, such as reduced HDL and increased triglyceride. It is possible to incorporate these variables in future studies to improve the predictions. Other time-invariant variables could also be included in the model, for example, the diagnosis age of their parents, economic status, and genetic factors may affect the temporal pattern of some DDVs. These variables can also be modeled as

covariates for k_{pg} or $k_{TIMEtoD}$. When incoperating all such variables, the model become more comprehensive in describing the pathophysiology of type 2 diabetes.

6.2.3 Verify the results with other studies

The study is focusing on the Caucasian subjects recruited in Joslin Diabetes Center. It is necessary to exminate the modeling results in different populations. Ethnic status, environment factors, disease status, economic status and genetic status can all affect the reliability of the model predictions. Since IVGTT and OGTT have been performed for many years in numerous diabetes research centers and hospitals, it is possible to conduct model validation studies to consider other group differences.

The proposed disease development analysis provides a novel way to evaluate the development of chronic diseases. By the repeated examinations of the DDVs, it is possible to estimate the time left before the diagnosis of the disease. The time-based information provides additional incentives to encourage people to reduce the risks through lifestyle modifications or preventive medications. The approach could be further extended to quantify the development of other chronic diseases, such as atherosclerosis and osteoporosis. Ideally, the medical examinations in future may will not only provide laboratory values of biomarkers but also give additional time based information of the onset of chronic diseases based on individual's laboratory values hange of such values over time. This would be valuable for early prevention of chronic diseases and reduce the cost and incidence rate of such diseases.

APPENDIX A. WBDIFF CODES FOR GLUCOSE-INSULIN INTERACTIVE MODEL

The following codes written in BlackBox Component Builder are corresponding to the model presented in Figure 2.1.

```
(*1*)   MODULE WBDiffNewhome62;
            IMPORT
                WBDiffODEBlockM,
                Math;
            TYPE
                Equations = POINTER TO RECORD
(WBDiffODEBlockM.Equations) END;
                Factory = POINTER TO RECORD
(WBDiffODEBlockM.Factory) END;
            CONST
                nEq = 2;
            VAR
                fact-: WBDiffODEBlockM.Factory;

(*2*)       PROCEDURE (e: Equations) Derivatives (IN theta, C: ARRAY OF
REAL; n: INTEGER; t: REAL;
(*3*)                                                OUT dCdt: ARRAY OF REAL);
(*4*)       VAR
(*5*)           block: INTEGER;
(*6*)           kgp, vg, kgr, kmaxgr, kic50, kip, beta, kir, vi, gss,iss: REAL;
(*7*)       BEGIN
(*8*)           block := e.Block();
(*22*)      kgp:= theta[0];
(*22*)      vg:= theta[1];
(*22*)      kgr := theta[2];              (*inport data from WinBugs*)
(*22*)      kmaxgr := theta[3];
(*22*)      kic50:=theta[4];
(*22*)      kip := theta[5];
(*22*)      beta:=theta[6];
(*22*)      kir:= theta[7];
(*22*)      vi:= theta[9];

(*13*)      CASE block OF
(*14*)          |0:
(*15*)          |1:
 (*18*)     END;
                                gss:=kgp/(vg*kgr);
                                iss:=kip/(kir*vi);
```

```
(*19*)          dCdt[0] := kgp/vg-(kgr+kmaxgr*MAX(C[1]-
iss,0)/(kic50+MAX(C[1]-iss,0))) *C[0];
(*21*)          dCdt[1] := kip* Math.Power(MAX(C[0],1)/gss,beta)/vi - kir *
C[1];

(*22*)      END Derivatives;

(*23*)      PROCEDURE (e: Equations) Adjust (IN theta: ARRAY OF REAL;
VAR C: ARRAY OF REAL;
(*24*)                                  n: INTEGER; t: REAL);
(*25*)      VAR
(*26*)          block: INTEGER;
                 vi,sp,dose, vg: REAL;
(*27*)      BEGIN
(*28*)          block := e.Block();
(*22*)        vg:= theta[1];
(*22*)         sp:= theta[8];
(*22*)         vi:= theta[9];
(*22*)         dose:= theta[10];
(*29*)          CASE block OF
(*30*)          |0:
(*30*)          |1:
(*31*)                  C[0] := C[0] + dose/vg;
(*31*)                  C[1] := C[1] + sp*(dose/vg)/vi;
(*35*)      END;
(*36*)      END Adjust;

            PROCEDURE (equations: Equations) SecondDerivatives (IN theta,
x: ARRAY OF REAL;
                                        numEq: INTEGER; t: REAL;
                                        OUT d2xdt2: ARRAY OF REAL);
            BEGIN
                HALT(126)
            END SecondDerivatives;
            PROCEDURE (equations: Equations) Jacobian (IN theta, x:
ARRAY OF REAL;
                                        numEq: INTEGER; t: REAL;
                                        OUT jacob: ARRAY OF ARRAY OF
REAL);
            BEGIN
                HALT(126)
            END Jacobian;

            PROCEDURE (f: Factory) New (option: INTEGER):
WBDiffODEBlockM.GraphNode;
            VAR
```

```
                    equations: Equations;
                    node: WBDiffODEBlockM.GraphNode;
                BEGIN
                    NEW(equations);
                    node := WBDiffODEBlockM.New(equations, nEq);
                    RETURN node
                END New;

                PROCEDURE Install*;
                BEGIN
                    WBDiffODEBlockM.Install(fact)
                END Install;

                PROCEDURE Init;
                VAR
                    f: Factory;
                BEGIN
                    NEW(f); fact := f
                END Init;
        BEGIN
                Init
(*1*)   END WBDiffNewhome62.
```

APPENDIX B. WINBUGS CODES FOR BAYESIAN ANALYSIS

The following WinBUGS codes are corresponding to the Bayesian hierarchical model presented in Figure 2.2.

```
model {
        for(i in 1:n.ind)
        {
                for(j in 1:n.grid)
                {
                        data.g[i,j] ~ dlnorm(logmodel.g[i,j],tau.g)
                        preddata.g[i,j] ~ dlnorm(logmodel.g[i,j],tau.g)
                        logmodel.g[i,j]<-log(gluins[i,j,1])
                        data.i[i,j] ~ dlnorm(logmodel.i[i,j],tau.i)
                        preddata.i[i,j] ~ dlnorm(logmodel.i[i,j],tau.i)
                        logmodel.i[i,j]<-log(gluins[i,j,2])
                }
                gluins[i, 1:n.grid, 1:dim]<-newhome62(inits[i,1:dim], grid[1:n.grid],
                theta[i,1:11], origins[i,1:n.block], tol)
                inits[i,1] <- theta[i,1]/(theta[i,2]*theta[i,3])          #kgp/(vg*kgr)
                inits[i,2] <- theta[i,6]/(theta[i,10]*theta[i,8])         #kip/(vi*kir)

                origins[i,1]<- -1.0
                origins[i,2]<- 0.0

                #linear gr & Emax ip

                theta[i,1] <-exp(par[i,1])              #kgp
                theta[i,2] <-wt[i]*exp(par[i,2])          #normalized vg
                theta[i,3] <-exp(par[i,3])              #kgr
                theta[i,4] <-exp(par[i,4])              #kmaxfr
                theta[i,5] <-exp(par[i,5])              #kic50
                theta[i,6] <-exp(par[i,6])              #kip
                theta[i,7] <-exp(par[i,7])              #beta
                theta[i,8] <-exp(par[i,8])              #kir
                theta[i,9] <-exp(par[i,9])              #sp
                theta[i,10] <-wt[i]/10.22222                #vi/(1-E)
                theta[i,11]<-500*wt[i]                          #dose

                par[i, 1:9] ~ dmnorm(theta.mean[i, 1:9], omega.inv[1:9, 1:9])

                for(k in 1:9)
                {
                        theta.mean[i,k]<-mu[k]+diffdm[k]*dm[i]
                }
        }
```

```
        mu[1:9] ~ dmnorm(mu.prior.mean[1:9], mu.prior.precision[1:9, 1:9])
        diffdm[1:9] ~ dmnorm(diffdm.prior.mean[1:9], diffdm.prior.precision[1:9,
1:9])
        omega.inv[1:9, 1:9] ~ dwish(omega.inv.matrix[1:9, 1:9], omega.inv.dof)
        omega[1:9, 1:9] <- inverse(omega.inv[1:9, 1:9])
        for(k in 1:9)
        {
                exp.mu[k]<-exp(mu[k])
                exp.mudm[k]<-exp(mu[k]+diffdm[k])
                diffexpdm.ratio[k]<-(exp.mudm[k]-exp.mu[k])/exp.mu[k]
        }
        tau.g ~ dgamma(0.01, 0.01)
        sigma.g <- 1 / sqrt(tau.g)
        tau.i ~ dgamma(0.01, 0.01)
        sigma.i <- 1 / sqrt(tau.i)
}

Inference data:
list(
dim = 2,
tol = 1.0E-3,
n.ind = 154, n.grid = 13, n.block = 2,
grid = c(-0.01,1,3,5,10,20,30,40,50,60,90,120,180),

#prior
mu.prior.mean = c(
0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0),
mu.prior.precision = structure(
.Data = c(
1.0E-4,         0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,
0.0,    1.0E-4,         0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,
0.0,    0.0,    1.0E-4,         0.0,    0.0,    0.0,    0.0,    0.0,    0.0,
0.0,    0.0,    0.0,    1.0E-4,         0.0,    0.0,    0.0,    0.0,    0.0,
0.0,    0.0,    0.0,    0.0,    1.0E-4,         0.0,    0.0,    0.0,    0.0,
0.0,    0.0,    0.0,    0.0,    0.0,    1.0E-4,         0.0,    0.0,    0.0,
0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    1.0E-4,         0.0,    0.0,
0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    1.0E-4,         0.0,
0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    1.0E-4),
.Dim = c(9, 9)),
diffdm.prior.mean = c(
0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0),
diffdm.prior.precision = structure(
.Data = c(
1.0E-4,         0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,
0.0,    1.0E-4,         0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,
0.0,    0.0,    1.0E-4,         0.0,    0.0,    0.0,    0.0,    0.0,    0.0,
```

```
0.0,    0.0,    0.0,    1.0E-4,         0.0,    0.0,    0.0,    0.0,    0.0,
0.0,    0.0,    0.0,    0.0,    1.0E-4,         0.0,    0.0,    0.0,    0.0,
0.0,    0.0,    0.0,    0.0,    0.0,    1.0E-4,         0.0,    0.0,    0.0,
0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    1.0E-4,         0.0,    0.0,
0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    1.0E-4,         0.0,
0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    1.0E-4),
.Dim = c(9, 9)),
omega.inv.matrix = structure(
.Data = c(
1.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,
0.0,    1.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,
0.0,    0.0,    1.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,
0.0,    0.0,    0.0,    1.0,    0.0,    0.0,    0.0,    0.0,    0.0,
0.0,    0.0,    0.0,    0.0,    1.0,    0.0,    0.0,    0.0,    0.0,
0.0,    0.0,    0.0,    0.0,    0.0,    1.0,    0.0,    0.0,    0.0,
0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    1.0,    0.0,    0.0,
0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    1.0,    0.0,
0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    0.0,    1.0
),
.Dim = c(9,9)),
omega.inv.dof = 9
)
```

The following WinBUGS codes are corresponding to the Bayesian hierarchical model presented in equations (3.7), (3.8), and (3.8).

```
model
{
        SEQ.mn<-mean(SEQ[])
        PTID.mn<-mean(PTID[])
        SEX.mn<-mean(SEX[])
        FSI.mn<-mean(FSI[])
        BMI.mn<-mean(BMI[])
        IBM.mn<-mean(IBM[])
        HOMA.mn<-mean(HOMA[])
        for( i in 1 : N ) {
                for( k in T[i]:(T[i+1]-1) ) {
                        FBG[k] ~ dnorm(Y[k],tau.c)
                        pred[k] ~ dnorm(Y[k],tau.c)
                        Y[k] <- kbase[i] +  kbasediff[i]*DM[k]+ kage[i] * TestAge[k] +
                          ktimetod[i]*step(TIMEtoD[k] -kbp[i])*(TIMEtoD[k] -
kbp[i])*DM[k]
                }
                kbase[i] ~ dnorm(kbase.c, tau.kbase)
                kbasediff[i] ~ dnorm(kbasediff.c, tau.kbasediff)
                kage[i] ~ dnorm(kage.c, tau.kage)
                ktimetod[i] ~ dnorm(ktimetod.c, tau.ktimetod)
                kbp[i] ~ dnorm(kbp.c, tau.kbp)I(-17.97,-2.33)
        }
        tau.c ~ dgamma(0.01,0.01)
        sigma <- 1 / sqrt(tau.c)
        kbase.c ~ dnorm(0.0,1.0E-6)
        kbasediff.c ~ dnorm(0.0,1.0E-6)
        kage.c ~ dnorm(0.0,1.0E-6)
        ktimetod.c ~ dnorm(0.0,1.0E-6)
        kbp.c ~ dunif(-17.97,-2.33)
        sigma.kbase ~ dunif(0,1000)
        sigma.kbasediff ~ dunif(0,1000)
        sigma.kage ~ dunif(0,1000)
        sigma.ktimetod ~ dunif(0,1000)
        sigma.kbp ~ dunif(0,1000)
        tau.kbase<-1/(sigma.kbase*sigma.kbase)
        tau.kbasediff<-1/(sigma.kbasediff*sigma.kbasediff)
        tau.kage<-1/(sigma.kage*sigma.kage)
        tau.ktimetod<-1/(sigma.ktimetod*sigma.ktimetod)
        tau.kbp<-1/(sigma.kbp*sigma.kbp)
}
```

APPENDIX C. R CODES FOR MIXED EFFECT ANALYSIS

The following R codes are corresponding to the mixed effect analysis presented in Figure 4.4 and 4.5.

```
library(nlme)
all_ivgtt<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and IVGTT\\all_ivgtt.txt", header=T)

attach(all_ivgtt)
all_ivgtt<-all_ivgtt[TIMEtoD<=0,]
for(i in 1:339)
{if(all_ivgtt$DM[i]==0) all_ivgtt$TIMEtoD[i]<-0}
#data_ivgtt<-all_ivgtt[is.na(IV_GLU180)==F &
is.na(IV_INS180)==F,][,c(1,7,9,10,12:37)]
#length=317
data_ivgtt<-na.omit(all_ivgtt[,c(1,7,9,10,12:37)])     #length=301  38 sets have NA
detach(all_ivgtt)

attach(data_ivgtt)
id.seq<-unique(PTID)
time<-c(0,1,3,5,10,20,30,40,50,60,90,120,180)
dm<-DM
age<-TESTAGE
ptid<-PTID
testage<-TESTAGE
timetod<-TIMEtoD
GLU<-data_ivgtt[,5:17]
INS<-data_ivgtt[,18:30]

saved.glu<-matrix(,301,180)
saved.ins<-matrix(,301,180)
sub<-1
start<-1

for(i in 1:301)
{
        spl_glu<-
smooth.spline(as.numeric(GLU[i,2:13])~time[2:13],df=8,all.knots=T)
        glu.pred<-predict(spl_glu,seq(1,180,1))
        saved.glu[i,1:180]<-glu.pred$y
}

#lme fit  glu

base.pv.g<-numeric()
```

```
base.mn.g<-numeric()
dm.pv.g<-numeric()
dm.mn.g<-numeric()
age.pv.g<-numeric()
age.mn.g<-numeric()
timetod.pv.g<-numeric()
timetod.mn.g<-numeric()

for(i in 1:180)
{
        for(j in 4:8)   #adjust the msTol from 1e-4 to 1e-8
        {
        testdata<-data.frame(cbind(ptid,dm,age,timetod,saved.glu[,i]))
        names(testdata)<-c("PTID","DM","AGE","TIMEtoD","METRIC")
        testdata.grped<-
groupedData(METRIC~DM+AGE+TIMEtoD|PTID,data=testdata)
        test1.lme<-
try(lme(METRIC~DM+AGE+TIMEtoD,random=~DM+AGE+TIMEtoD|PTID,data=t
estdata.grped,control=list(maxiter=100,opt="optim",msTol=1/(10^j))),TRUE)
if(class(test1.lme)=="try-error") break                    #if test1.lme is a try-error,
then quit loop
        base.pv.g[i]<-summary(test1.lme)$tTable[1,5]
        base.mn.g[i]<-summary(test1.lme)$tTable[1,1]
        dm.pv.g[i]<-summary(test1.lme)$tTable[2,5]
        dm.mn.g[i]<-summary(test1.lme)$tTable[2,1]
        age.pv.g[i]<-summary(test1.lme)$tTable[3,5]
        age.mn.g[i]<-summary(test1.lme)$tTable[3,1]
        timetod.pv.g[i]<-summary(test1.lme)$tTable[4,5]
        timetod.mn.g[i]<-summary(test1.lme)$tTable[4,1]
        }
}

result.glu.ivgtt<-
data.frame(cbind(base.pv.g,base.mn.g,dm.pv.g,dm.mn.g,age.pv.g,age.mn.g,time
tod.pv.g,timetod.mn.g))

write.table(result.glu.ivgtt,"C:\\Documents and Settings\\clin\\Desktop\\Disease
development and IVGTT\\result_glu_ivgtt_final.txt",sep=" ",quote=F)
```

APPENDIX D. R CODES FOR GENERATING FIGURES

The following R codes are used to generate all of the figures presented in the thesis.

Figure 2.3

```
pred.dmnondmg<-read.table("C:\\Documents and
Settings\\clin\\Desktop\\homeostasis project\\data pred BUGS DMnonDM g.txt",
header=T)
data.dmnondmg<-read.table("C:\\Documents and
Settings\\clin\\Desktop\\homeostasis project\\data R DMnonDM g.txt", header=T)
pred.dmnondmi<-read.table("C:\\Documents and
Settings\\clin\\Desktop\\homeostasis project\\data pred BUGS DMnonDM i.txt",
header=T)
data.dmnondmi<-read.table("C:\\Documents and
Settings\\clin\\Desktop\\homeostasis project\\data R DMnonDM i.txt", header=T)
n.dmnondmg<-length(pred.dmnondmg[,1])/13
n.dmnondmi<-length(pred.dmnondmi[,1])/13
conc.v.dmnondmg<-matrix(data.dmnondmg$dv,13,n.dmnondmg)
conc.v.dmnondmi<-matrix(data.dmnondmi$dv,13,n.dmnondmi)
conc.dmg<-matrix(data.dmnondmg$dv[1:(25*13)],,13,byrow=T)
gbase.dmg<-conc.dmg[,1]
conc.nondmg<-matrix(data.dmnondmg$dv[(26*13+1):(153*13)],,13,byrow=T)
gbase.nondmg<-conc.nondmg[,1]
conc.dmi<-matrix(data.dmnondmi$dv[1:(25*13)],,13,byrow=T)
ibase.dmi<-conc.dmi[,1]
conc.nondmi<-matrix(data.dmnondmi$dv[(26*13+1):(153*13)],,13,byrow=T)
ibase.nondmi<-conc.nondmi[,1]
air.dmi<-(conc.dmi[,1]-conc.dmi[,1]+conc.dmi[,2]-conc.dmi[,1])*1/2 +
(conc.dmi[,2]-conc.dmi[,1]+conc.dmi[,3]-conc.dmi[,1])*2/2 +
(conc.dmi[,3]-conc.dmi[,1]+conc.dmi[,4]-conc.dmi[,1])*2/2 + (conc.dmi[,4]-
conc.dmi[,1]+conc.dmi[,5]-conc.dmi[,1])*5/2
air.nondmi<-(conc.nondmi[,1]-conc.nondmi[,1]+conc.nondmi[,2]-
conc.nondmi[,1])*1/2 + (conc.nondmi[,2]-conc.nondmi[,1]+conc.nondmi[,3]-
conc.nondmi[,1])*2/2 +
(conc.nondmi[,3]-conc.nondmi[,1]+conc.nondmi[,4]-conc.nondmi[,1])*2/2 +
(conc.nondmi[,4]-conc.nondmi[,1]+conc.nondmi[,5]-conc.nondmi[,1])*5/2
gbase.dmfam<-c(gbase.dmg,gbase.nondmg)
ibase.dmfam<-c(ibase.dmi,ibase.nondmi)
air.dmfam<-c(air.dmi,air.nondmi)
gbase<-c(mean(gbase.dmg),mean(gbase.nondmg))
ibase<-c(mean(ibase.dmi),mean(ibase.nondmi))
AIR<-c(mean(air.dmi),mean(air.nondmi))
par(mfcol=c(3,1))
par(mex=0.8)
dmlabel=c("DM","nonDM")
```

```
dotchart(gbase,labels=dmlabel,xlim=c(70,90), pch=23,main="Means and 95%
Confidence Intervals",xlab="Glucose baseline (mg/dL)")
lines(c(mean(gbase.dmg)-
qnorm(0.975)*sd(gbase.dmg)/sqrt(25),mean(gbase.dmg)+qnorm(0.975)*sd(gbas
e.dmg)/sqrt(25)),c(1,1),lwd=4)
lines(c(mean(gbase.nondmg)-
qnorm(0.975)*sd(gbase.nondmg)/sqrt(128),mean(gbase.nondmg)+qnorm(0.975)
*sd(gbase.nondmg)/sqrt(128)),c(2,2),lwd=4)
dotchart(ibase,labels=dmlabel,xlim=c(10,40), pch=23,main="",xlab="Insulin
baseline (mU/L)")
lines(c(mean(ibase.dmi)-
qnorm(0.975)*sd(ibase.dmi)/sqrt(25),mean(ibase.dmi)+qnorm(0.975)*sd(ibase.d
mi)/sqrt(25)),c(1,1),lwd=4)
lines(c(mean(ibase.nondmi)-
qnorm(0.975)*sd(ibase.nondmi)/sqrt(128),mean(ibase.nondmi)+qnorm(0.975)*sd
(ibase.nondmi)/sqrt(128)),c(2,2),lwd=4)
dotchart(AIR,labels=dmlabel,xlim=c(400,1200), pch=23,main="",xlab="Acute
insulin response (min mU/L)")
lines(c(mean(air.dmi)-
qnorm(0.975)*sd(air.dmi)/sqrt(25),mean(air.dmi)+qnorm(0.975)*sd(air.dmi)/sqrt(2
5)),c(1,1),lwd=4)
lines(c(mean(air.nondmi)-
qnorm(0.975)*sd(air.nondmi)/sqrt(128),mean(air.nondmi)+qnorm(0.975)*sd(air.n
ondmi)/sqrt(128)),c(2,2),lwd=4)
```

Figure 2.4

```
dmnondmratio<-read.table("C:\\Documents and
Settings\\clin\\Desktop\\homeostasis project\\data pred BUGS
DMnonDMratio.txt", header=T)  #fixed
par(mfrow=c(1,1))
par(mex=0.8)
para<-matrix(dmnondmratio$median,1,10,byrow=T)
kgp<-para[,1]
vg<-para[,2]
kgr<-para[,3]
Smaxgr<-para[,4]
kic50<-para[,5]
Sip<-para[,6]
beta<-para[,7]
kisec<-para[,8]
kir<-para[,9]
Sp<-para[,10]
para.table<-cbind(Sp,kir,kisec,beta,Sip,kic50,Smaxgr,kgr,vg,kgp)
rownames(para.table)<-c("diabetic")
```

```
datlabel<-
c(expression(S[p]),expression(k[ir]),expression(k[isec]),expression(beta),expressi
on(S[ip]),expression(k[ic50]),expression(S[maxgr]),expression(k[gr]),expression(v
[g]),expression(k[gp]))
dotchart(100*para.table[1,],labels=datlabel,xlim=c(-100,450),pch=23,main="95%
credible sets of parameters' differences (progressors vs
nonprogressors)",xlab="Difference in percentage (%)")
for(i in 1:10)
{
        lines(100*c(dmnondmratio$pred2.5[i],dmnondmratio$pred97.5[i]),c(abs(i-
11),abs(i-11)),lwd=5)
}
lines(c(0,0),c(0,34),lty=2)
text(100*dmnondmratio$pred97.5[9]+15,2,cex = 1.5,"*") #kir 2
text(100*dmnondmratio$pred97.5[4]+15,7,cex = 1.5,"*") #Simax 7
```

Figure 2.5

```
j<-13
dm<-c(rep(1,26),rep(0,128))
#Get 95% PI data form WinBugs outpot
pred97.5.v.dmnondmg<-
matrix(pred.dmnondmg$pred97.5[1:length(data.dmnondmg[,1])],j,n.dmnondmg)
pred2.5.v.dmnondmg<-
matrix(pred.dmnondmg$pred2.5[1:length(data.dmnondmg[,1])],j,n.dmnondmg)
median.v.dmnondmg<-
matrix(pred.dmnondmg$median[1:length(data.dmnondmg[,1])],j,n.dmnondmg)
id.v.dmnondmg<-matrix(data.dmnondmg$id,j,n.dmnondmg)
time.v.dmnondmg<-matrix(data.dmnondmg$time,j,n.dmnondmg)
conc.v.dmnondmg<-matrix(data.dmnondmg$dv,j,n.dmnondmg)
dm<-c(rep(2,26),rep(1,128),rep(0,187))
#Get 95% PI data form WinBugs outpot
pred97.5.v.dmnondmi<-
matrix(pred.dmnondmi$pred97.5[1:length(data.dmnondmi[,1])],j,n.dmnondmi)
pred2.5.v.dmnondmi<-
matrix(pred.dmnondmi$pred2.5[1:length(data.dmnondmi[,1])],j,n.dmnondmi)
median.v.dmnondmi<-
matrix(pred.dmnondmi$median[1:length(data.dmnondmi[,1])],j,n.dmnondmi)
id.v.dmnondmi<-matrix(data.dmnondmi$id,j,n.dmnondmi)
time.v.dmnondmi<-matrix(data.dmnondmi$time,j,n.dmnondmi)
conc.v.dmnondmi<-matrix(data.dmnondmi$dv,j,n.dmnondmi)
par(mfrow=c(2,2))
par(mex=0.7)
time.dmg<-data.dmnondmg$time[1:(25*13)]
conc.dmg<-data.dmnondmg$dv[1:(25*13)]
```

```
plot(conc.dmg~time.dmg,pch=20,xlim=c(0,200),ylim=c(20,800), xlab="Time
(Min)",ylab="Glucose Conc (mg/dL)",main="Progressor group (nobs=325)")
lines(pred.dmnondmg$median[(25*13+1):(26*13)]~time.v.dmnondmg[,1],lty=1,lw
d=1)
lines(pred.dmnondmg$pred97.5[(25*13+1):(26*13)]~time.v.dmnondmg[,1],lty=3,l
wd=1)
lines(pred.dmnondmg$pred2.5[(25*13+1):(26*13)]~time.v.dmnondmg[,1],lty=3,lw
d=1)
time.nondmg<-data.dmnondmg$time[(26*13+1):(153*13)]
conc.nondmg<-data.dmnondmg$dv[(26*13+1):(153*13)]
plot(conc.nondmg~time.nondmg,pch=20,xlim=c(0,200),ylim=c(20,800),
xlab="Time (Min)",ylab="Glucose Conc (mg/dL)",main="Non-progressor group
(nobs=1645)")
lines(pred.dmnondmg$median[(153*13+1):(154*13)]~time.v.dmnondmg[,1],lty=1,l
wd=1)
lines(pred.dmnondmg$pred97.5[(153*13+1):(154*13)]~time.v.dmnondmg[,1],lty=
3,lwd=1)
lines(pred.dmnondmg$pred2.5[(153*13+1):(154*13)]~time.v.dmnondmg[,1],lty=3,
lwd=1)
time.dmi<-data.dmnondmi$time[1:(25*13)]
conc.dmi<-data.dmnondmi$dv[1:(25*13)]
plot(conc.dmi~time.dmi,pch=20,xlim=c(0,200),ylim=c(1,800), xlab="Time
(Min)",ylab="Insulin Conc (mU/L)",main="Progressor group (nobs=325)")
lines(pred.dmnondmi$median[(25*13+1):(26*13)]~time.v.dmnondmi[,1],lty=1,lwd
=1)
lines(pred.dmnondmi$pred97.5[(25*13+1):(26*13)]~time.v.dmnondmi[,1],lty=3,lw
d=1)
lines(pred.dmnondmi$pred2.5[(25*13+1):(26*13)]~time.v.dmnondmi[,1],lty=3,lwd
=1)
time.nondmi<-data.dmnondmi$time[(26*13+1):(153*13)]
conc.nondmi<-data.dmnondmi$dv[(26*13+1):(153*13)]
plot(conc.nondmi~time.nondmi,pch=20,xlim=c(0,200),ylim=c(1,800), xlab="Time
(Min)",ylab="Insulin Conc (mU/L)",main="Non-progressor group (nobs=1650)")
lines(pred.dmnondmi$median[(153*13+1):(154*13)]~time.v.dmnondmi[,1],lty=1,l
wd=1)
lines(pred.dmnondmi$pred97.5[(153*13+1):(154*13)]~time.v.dmnondmi[,1],lty=3,l
wd=1)
lines(pred.dmnondmi$pred2.5[(153*13+1):(154*13)]~time.v.dmnondmi[,1],lty=3,l
wd=1)
```

Figure 2.6

```
par(mfrow=c(2,2))
par(mex=0.7)
i<-3
```

```
plot(pred97.5.v.dmnondmg[,i]~time.v.dmnondmg[,i],
type="l",xlim=c(0,200),ylim=c(50,450),log="y"
, xlab="Time (Min)",ylab="Glucose Conc (mg/dL)",main=c("Progressor
Group"),lty=3)
points(conc.v.dmnondmg[,i]~time.v.dmnondmg[,i],pch=20)
lines(median.v.dmnondmg[,i]~time.v.dmnondmg[,i], lty=1)
lines(pred2.5.v.dmnondmg[,i]~time.v.dmnondmg[,i],lty=3)
i<-91
plot(pred97.5.v.dmnondmg[,i]~time.v.dmnondmg[,i],
type="l",xlim=c(0,200),ylim=c(50,450),log="y"
, xlab="Time (Min)",ylab="Glucose Conc (mg/dL)",main=c("Non-progressor
Group"),lty=3)
points(conc.v.dmnondmg[,i]~time.v.dmnondmg[,i],pch=20)
lines(median.v.dmnondmg[,i]~time.v.dmnondmg[,i], lty=1)
lines(pred2.5.v.dmnondmg[,i]~time.v.dmnondmg[,i],lty=3)
i<-3
plot(pred97.5.v.dmnondmi[,i]~time.v.dmnondmi[,i],
type="l",xlim=c(0,200),ylim=c(5,250),log="y"
, xlab="Time (Min)",ylab="Insulin Conc (mU/L)",main=c("Progressor
Group"),lty=3)
points(conc.v.dmnondmi[,i]~time.v.dmnondmi[,i],pch=20)
lines(median.v.dmnondmi[,i]~time.v.dmnondmi[,i], lty=1)
lines(pred2.5.v.dmnondmi[,i]~time.v.dmnondmi[,i],lty=3)
i<-91
plot(pred97.5.v.dmnondmi[,i]~time.v.dmnondmi[,i],
type="l",xlim=c(0,200),ylim=c(5,250),log="y"
, xlab="Time (Min)",ylab="Insulin Conc (mU/L)",main=c("Non-progressor
Group"),lty=3)
points(conc.v.dmnondmi[,i]~time.v.dmnondmi[,i],pch=20)
lines(median.v.dmnondmi[,i]~time.v.dmnondmi[,i], lty=1)
lines(pred2.5.v.dmnondmi[,i]~time.v.dmnondmi[,i],lty=3)
```

Figure 2.7

```
par(mfrow=c(2,2))
par(mex=0.7)
plot(conc.dmg~pred.dmnondmg$median[1:(25*13)],xlim=c(0,550),ylim=c(0,550),
pch=20,xlab="Individual Predicted Glucose Conc(mg/dL)",ylab="Oberved
Glucose Conc (mg/dL)",main="Progressor group (nobs=325)" )
lines(c(0,550),c(0,550))
cor(conc.dmg,pred.dmnondmg$median[1:(25*13)])
text(400,100, c("Cor: 0.992"))
plot(conc.nondmg~pred.dmnondmg$median[(26*13+1):(153*13)],xlim=c(0,550),y
lim=c(0,550),pch=20,xlab="Individual Predicted Glucose
Conc(mg/dL)",ylab="Oberved Glucose Conc (mg/dL)",main="Non-progressor
group (nobs=1645)")
```

```
lines(c(0,550),c(0,550))
cor(conc.nondmg,pred.dmnondmg$median[(26*13+1):(153*13)],use="complete.obs")
text(400,100, c("Cor: 0.987"))
plot(conc.dmi~pred.dmnondmi$median[1:(25*13)],xlim=c(0,900),ylim=c(0,900),pch=20,xlab="Individual Predicted Insulin Conc(mU/L)",ylab="Oberved Insulin Conc (mU/L)",main="Progressor group (nobs=325)")
lines(c(0,800),c(0,800))
cor(conc.dmi,pred.dmnondmi$median[1:(25*13)])
text(650,180, c("Cor: 0.982"))
plot(conc.nondmi~pred.dmnondmi$median[(26*13+1):(153*13)],xlim=c(0,900),ylim=c(0,900),pch=20,xlab="Individual Predicted Insulin
Conc(mU/L)",ylab="Oberved Insulin Conc (mU/L)",main="Non-progressor group (nobs=1650)")
lines(c(0,800),c(0,800))
cor(conc.nondmi,pred.dmnondmi$median[(26*13+1):(153*13)],use="complete.obs")
text(650,180, c("Cor: 0.985"))
```

Figure 2.8

```
par(mfrow=c(2,2))
par(mex=0.7)
residual1<-log(conc.dmg)-log(pred.dmnondmg$median[1:(25*13)])
plot(residual1~sqrt(time.dmg),pch=20,ylim=c(-
0.8,0.8),xlab=expression(paste(sqrt(Time(min)))),ylab="Log glucose
concnetration residual",main="Progressor group (nobs=325)")
abline(0,0)
low1<-qnorm(0.025,0,sd=sd(residual1))
high1<-qnorm(0.975,0,sd=sd(residual1))
#abline(low1,0,lwd=2,lty=3)
#abline(high1,0,lwd=2,lty=3)
residual3<-log(conc.nondmg)-log(pred.dmnondmg$median[(26*13+1):(153*13)])
plot(residual3~sqrt(time.nondmg),pch=20,ylim=c(-
0.8,0.8),xlab=expression(paste(sqrt(Time(min)))),ylab="Log glucose
concnetration residual",main="Non-progressor group (nobs=1645)")
abline(0,0)
low3<-qnorm(0.025,0,sd=sd(residual3,na.rm=T))
high3<-qnorm(0.975,0,sd=sd(residual3,na.rm=T))
#abline(low3,0,lwd=2,lty=3)
#abline(high3,0,lwd=2,lty=3)
residual2<-log(conc.dmi)-log(pred.dmnondmi$median[1:(25*13)])
plot(residual2~sqrt(time.dmi),pch=20,ylim=c(-
1.5,1.5),xlab=expression(paste(sqrt(Time(min)))),ylab="Log insulin concnetration
residual",main="Progressor group (nobs=325)")
abline(0,0)
```

```
low2<-qnorm(0.025,0,sd=sd(residual2))
high2<-qnorm(0.975,0,sd=sd(residual2))
conc.nondmi[181]<-1.0
residual4<-log(conc.nondmi)-log(pred.dmnondmi$median[(26*13+1):(153*13)])
plot(residual4~sqrt(time.nondmi),pch=20,ylim=c(-
1.5,1.5),xlab=expression(paste(sqrt(Time(min)))),ylab="Log insulin concnetration
residual",main="Non-progressor group (nobs=1650)")
abline(0,0)
low4<-qnorm(0.025,0,sd=sd(residual4,na.rm=T))
high4<-qnorm(0.975,0,sd=sd(residual4,na.rm=T))
```

Figure 2.9

```
#(a)Insulin's Glucose Removal Effect vs Insulin Concentration
expmu<-read.table("C:\\Documents and Settings\\clin\\Desktop\\homeostasis
project\\data pred BUGS expmu.txt", header=T)  #fixed
mudm<-read.table("C:\\Documents and Settings\\clin\\Desktop\\homeostasis
project\\data pred BUGS mudm.txt", header=T)  #fixed
par(mfrow=c(1,1))
par(mex=0.8)
kgr<-expmu$median[3]
Simax<-expmu$median[4]
kic50<-expmu$median[5]
I<-c(1:200)
vi<-70.472857/10.2222
Iss<-expmu$median[6]/(vi*expmu$median[9])
gr<-kgr+Simax*pmax(I-Iss,0)/(kic50+pmax(I-Iss,0))
plot(gr~I,type="l",ylim=c(0,0.042),main="(a) Insulin-Dependent Glucose Removal
Effect Curves",ylab="Glucose Removal (1/min)",xlab="Inuslin Concentration
(mU/L)",lty=1)
kgr<-mudm$median[3]
Simax<-mudm$median[4]
kic50<-mudm$median[5]
I<-c(1:200)
vi<-96.272064/10.2222
Iss<-mudm$median[6]/(vi*mudm$median[9])
gr<-kgr+Simax*pmax(I-Iss,0)/(kic50+pmax(I-Iss,0))
lines(gr~I,type="l",lty=2)
text(52,0.024,"Non-progressor")
text(52,0.022,"group")
text(110,0.018,"progressor")
text(110,0.016,"group")
quantile(data.dmnondmi[,3],c(0.025,0.975),na.rm = T)
#(b)  Insulin Production Rate vs Glucose Concentration
par(mfrow=c(1,1))
par(mex=0.8)
```

```
G<-c(0:300)
kip<-expmu$median[6]
beta<-expmu$median[7]
gss<-expmu$median[1]/(expmu$median[2]*70.472857*expmu$median[3])
ip<-kip*(G/gss)^beta
plot(ip~G,type="l",ylim=c(0,80),main="(b) Glucose-Dependent Insulin Production
Rate Curves",ylab="Insulin Production Rate (mU/min)", xlab="Glucose
Concentration (mg/dL)")

kip<-mudm$median[6]
beta<-mudm$median[7]
gss<-mudm$median[1]/(mudm$median[2]*96.272064*mudm$median[3])
ip<-kip*(G/gss)^beta
lines(ip~G,type="l",lty=3)
text(170,44,"Progressor")
text(170,48,"group")
text(250,34,"Non-progressor")
text(250,30,"group")
quantile(data.dmnondmg[,3],c(0.025,0.975),na.rm = T)
```

Figure 3.1

```
WINBUGS_glu_ins<-read.table("C:\\Documents and
Settings\\clin\\Desktop\\Disease development and FBG
FSI\\WINBUGS_glu_ins.txt", header=T)
library(nlme)
attach(WINBUGS_glu_ins)
DM_glu_ins<-na.omit(WINBUGS_glu_ins[DM==1,])
detach(WINBUGS_glu_ins)
attach(DM_glu_ins)
AgeOut<-DM_glu_ins$TestAge-TIMEtoD
#profile of follow-up
testdata.dm.na<-
data.frame(cbind(DM_glu_ins$PTID,DM_glu_ins$SEX,DM_glu_ins$TestAge,
AgeOut,DM_glu_ins$TIMEtoD,DM_glu_ins$FBG,DM_glu_ins$BMI))
names(testdata.dm.na)<-c("PTID","SEX","AGE","AGEOUT",
"TIMEtoD","FBG","BMI")
testdata.dm<-na.omit(testdata.dm.na)
id.seq<-unique(testdata.dm$PTID)
seq.len.dm<-length(id.seq)
plot(NULL,ylim=c(0.5,25.5),yaxt="n",xlim=c(15,85),ylab="Subjects
(n=25)",xlab="Age (year)",
main="Follow-up in the progressor group",font=2, font.lab=2)
SEX.pg<-numeric()
AGE.pg<-numeric()
BMI.pg<-numeric()
```

```
FOLLOW.pg<-numeric()
length.pg<-numeric()
AgeOut.pg<-numeric()
for(i in 1:seq.len.dm)
{
        SEX.pg[i]<-testdata.dm[testdata.dm[["PTID"]]==id.seq[i],]$SEX[1]
        AGE.pg[i]<-testdata.dm[testdata.dm[["PTID"]]==id.seq[i],]$AGE[1]
        BMI.pg[i]<-testdata.dm[testdata.dm[["PTID"]]==id.seq[i],]$BMI[1]
        AgeOut.pg[i]<-testdata.dm[testdata.dm[["PTID"]]==id.seq[i],]$AGEOUT[1]
        FOLLOW.pg[i]<--
testdata.dm[testdata.dm[["PTID"]]==id.seq[i],]$TIMEtoD[1]
        length.pg[i]<-length(testdata.dm[testdata.dm[["PTID"]]==id.seq[i],]$AGE)
        age<-testdata.dm[testdata.dm[["PTID"]]==id.seq[i],]$AGE
        y<-rep(i,length(age))
        points(y~age,pch=20,cex=0.85)
        time.dm<-testdata.dm[testdata.dm[["PTID"]]==id.seq[i],]$AGEOUT[1]
        duration<-c(min(age),time.dm)
        pt<-c(i,i)
        lines(pt~duration)
        points(i~time.dm,pch=4)
}
detach(DM_glu_ins)
```

Figure 3.2

```
attach(WINBUGS_glu_ins)
CON_glu_ins<-na.omit(WINBUGS_glu_ins[DM==0,])
detach(WINBUGS_glu_ins)
attach(CON_glu_ins)
AgeOut<-CON_glu_ins$TestAge-TIMEtoD
#profile of follow-up
testdata.CON.na<-
data.frame(cbind(CON_glu_ins$PTID,CON_glu_ins$SEX,CON_glu_ins$TestAge
, AgeOut,CON_glu_ins$TIMEtoD,CON_glu_ins$FBG,CON_glu_ins$BMI))
names(testdata.CON.na)<-c("PTID","SEX","AGE","AGEOUT",
"TIMEtoD","FBG","BMI")
testdata.CON<-na.omit(testdata.CON.na)
id.seq<-unique(testdata.CON$PTID)
seq.len.CON<-length(id.seq)
plot(NULL,ylim=c(0.5,127.5),yaxt="n",xlim=c(15,85),ylab="Subjects
(n=127)",xlab="Age (year)",
main="Follow-up in the non-progressor group",font=2,font.lab=2)
SEX.nonpg<-numeric()
AGE.nonpg<-numeric()
BMI.nonpg<-numeric()
FOLLOW.nonpg<-numeric()
```

```
length.nonpg<-numeric()
for(i in 1:seq.len.CON)
{
        SEX.nonpg[i]<-testdata.CON[testdata.CON[["PTID"]]==id.seq[i],]$SEX[1]
        AGE.nonpg[i]<-testdata.CON[testdata.CON[["PTID"]]==id.seq[i],]$AGE[1]
        BMI.nonpg[i]<-testdata.CON[testdata.CON[["PTID"]]==id.seq[i],]$BMI[1]
        FOLLOW.nonpg[i]<--
testdata.CON[testdata.CON[["PTID"]]==id.seq[i],]$TIMEtoD[1]
        length.nonpg[i]<-
length(testdata.CON[testdata.CON[["PTID"]]==id.seq[i],]$AGE)

        age<-testdata.CON[testdata.CON[["PTID"]]==id.seq[i],]$AGE
        y<-rep(i,length(age))
        points(y~age,pch=20,cex=0.85)
        time.CON<-testdata.CON[testdata.CON[["PTID"]]==id.seq[i],]$AGEOUT[1]
        duration<-c(min(age),time.CON)
        pt<-c(i,i)
        lines(pt~duration)
        #points(i~time.CON,pch=4)
}
```

Figure 3.3

```
timetod<-seq(-25,0,0.001)
x<-seq(-25,0,0.001)
par(mfrow=c(2,2))
i<-1
plot(DM_glu_ins[,7]~DM_glu_ins$TIMEtoD, pch=19,cex=0.60,
xlab="TIMEtoD (years)", ylab="FBG (mg/dL)", ylim=c(60,150),main="Pooled data
fit (Progressors)",font=2,font.lab=2)
gcv<-smooth.spline(DM_glu_ins[,7]~DM_glu_ins$TIMEtoD)
lines(predict(gcv,x),lwd=2)
i<-2
plot(DM_glu_ins[,8]~DM_glu_ins$TIMEtoD, pch=19,cex=0.60,
xlab="TIMEtoD (years)", ylab="FSI (mU/L)",ylim=c(0,150), main="Pooled data fit
(Progressors)",font=2,font.lab=2)
gcv<-smooth.spline(DM_glu_ins[,8]~DM_glu_ins$TIMEtoD)
lines(predict(gcv,x),lwd=2)
i<-3
plot(DM_glu_ins[,11]~DM_glu_ins$TIMEtoD, pch=19,cex=0.60,
xlab="TIMEtoD (years)", ylab="HOMA-IR", ylim=c(0,40),main="Pooled data fit
(Progressors)",font=2,font.lab=2)
gcv<-smooth.spline(DM_glu_ins[,11]~DM_glu_ins$TIMEtoD)
lines(predict(gcv,x),lwd=2)
i<-4
plot(DM_glu_ins[,9]~DM_glu_ins$TIMEtoD, pch=19,cex=0.60,
```

```
xlab="TIMEtoD (years)", ylab="BMI", ylim=c(15,70),main="Pooled data fit
(Progressors)",font=2,font.lab=2)
gcv<-smooth.spline(DM_glu_ins[,9]~DM_glu_ins$TIMEtoD)
lines(predict(gcv,x),lwd=2)
```

Figure 3.4

```
nf <- layout(matrix(1:2,1,2,byrow=TRUE), c(1,1.97), c(1.5), TRUE)
layout.show(nf)
mar1a<-c(5,4,2,2)+0.1
par(mar=mar1a)
pred.time<-seq(-25,0,0.01)
pred.FBG<-79.33+0.227*as.numeric(I(pred.time< (-8.73)))*(pred.time+8.73)+
(0.227+2.27)*as.numeric(I(pred.time> (-8.73)))*(pred.time+8.73)
plot(testdata.dm$FBG~testdata.dm$TIMEtoD,
pch=20,cex=0.55,ylim=c(60,145),xaxp=c(-20,0,2),
xlab="TIMEtoD (years)", ylab="FBG (mg/dL)",
main="Progressors",font=2,font.lab=2)
lines(pred.FBG~pred.time,lwd=2)
pred.AGE<-seq(0,100,0.01)
pred.FBG<-69.2+pred.AGE*0.27
plot(testdata.CON$FBG~testdata.CON$AGE,pch=20,cex=0.55,ylim=c(60,145),
xlab="Age (years)", ylab="FBG (mg/dL)", main="Non-
progressors",font=2,font.lab=2)
lines(pred.FBG~pred.AGE,lwd=2)
```

Figure 3.5

```
kage<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and FBG FSI\\predkage.txt", header=T)$mean
kbase<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and FBG FSI\\predkbase.txt", header=T)$mean
kbasediff<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and FBG FSI\\predkbasediff.txt", header=T)$mean
kbp<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and FBG FSI\\predkbp.txt", header=T)$mean
ktimetod<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and FBG FSI\\predktimetod.txt", header=T)$mean
testdata<-data.frame(WINBUGS_glu_ins[,c(2,4,5,6,7,8)])
id.seq<-unique(testdata$PTID)
seq.len<-length(id.seq)
seq_DM<-c(20,27)
par(mfrow=c(2,2))
for(i in seq_DM)
{
```

```
        plot(testdata[testdata[["PTID"]]==id.seq[i],]$FBG~testdata[testdata[["PTID"
]]==id.seq[i],]$TestAge,xlim=c(20,65),ylim=c(53,125),
        xlab="Age (years)",ylab="FBG (mg/dL)",main="Individual fits",
pch=19,cex=0.75,font=2, font.lab=2)
        TtoD<-seq(-20,0,0.05)
        DM_status<-testdata[testdata[["PTID"]]==id.seq[i],]$DM[1]
        age_start<-testdata[testdata[["PTID"]]==id.seq[i],]$TestAge[1]-
testdata[testdata[["PTID"]]==id.seq[i],]$TIMEtoD[1]-25
        age<-seq(age_start+5,age_start+25,0.05)
        y<-kbase[i]+ kbasediff[i]*DM_status+ kage[i]*age +
ktimetod[i]*as.numeric(I(TtoD>kbp[i]))*(TtoD -kbp[i])*DM_status
        lines(y~age,lwd=1.5)
        text(age_start+33,61,"TIMEtoD",font=2)
        lines(c(age_start+5,age_start+25),c(63,63))
        text(age_start+5,58,-20,font=2);
lines(c(age_start+5,age_start+5),c(61,63));
        text(age_start+15,58,-
10,font=2);lines(c(age_start+15,age_start+15),c(61,63));
        text(age_start+25,58,0,font=2);lines(c(age_start+25,age_start+25),c(61,63
));
}
i<-20
        plot(testdata[testdata[["PTID"]]==id.seq[i],]$FSI~testdata[testdata[["PTID"]]
==id.seq[i],]$TestAge,xlim=c(20,65),ylim=c(-5,70),
        xlab="Age (years)",ylab="FSI (mU/L)",main="Individual fits",
pch=19,cex=0.75,font=2, font.lab=2)
        text(age_start+33,14,"TIMEtoD",font=2)
        lines(c(age_start+5,age_start+25),c(16,16))
        text(age_start+5,11,-20,font=2);
lines(c(age_start+5,age_start+5),c(14,16));
        text(age_start+15,11,-
10,font=2);lines(c(age_start+15,age_start+15),c(14,16));
        text(age_start+25,11,0,font=2);lines(c(age_start+25,age_start+25),c(14,16
));

i<-27
        plot(testdata[testdata[["PTID"]]==id.seq[i],]$FSI~testdata[testdata[["PTID"]]
==id.seq[i],]$TestAge,xlim=c(20,65),ylim=c(-5,70),
        xlab="Age (years)",ylab="FSI (mU/L)",main="Individual fits",
pch=19,cex=0.75,font=2, font.lab=2)
        text(age_start+33,3,"TIMEtoD",font=2)
        lines(c(age_start+5,age_start+25),c(5,5))
        text(age_start+5,0,-20,font=2); lines(c(age_start+5,age_start+5),c(3,5));
        text(age_start+15,0,-
10,font=2);lines(c(age_start+15,age_start+15),c(3,5));
        text(age_start+25,0,0,font=2);lines(c(age_start+25,age_start+25),c(3,5));
```

Figure 3.6

```
WINBUGS_glu_ins<-read.table("C:\\Documents and
Settings\\clin\\Desktop\\Disease development and FBG
FSI\\WINBUGS_glu_ins.txt", header=T)
pred_FBG<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and FBG FSI\\predFBG.txt", header=T)$mean

residual<-WINBUGS_glu_ins$FBG-pred_FBG
plot(residual~WINBUGS_glu_ins$TestAge,pch=20,cex=0.85,xlab="Age (years)",
ylab="Residuals",
main="Residual plot of individual predictions of FBG",font=2, font.lab=2)
abline(h=0)
qqnorm(residual,pch=20,cex=0.55)
qqline(residual)
```

Figure 4.1

```
library(nlme)
all_ivgtt<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and IVGTT\\all_ivgtt.txt", header=T)
attach(all_ivgtt)
all_ivgtt<-all_ivgtt[TIMEtoD<=0,]     #360-21=339
for(i in 1:339)
{if(all_ivgtt$DM[i]==0) all_ivgtt$TIMEtoD[i]<-0}
data_ivgtt<-na.omit(all_ivgtt[,c(1,7,9,10,12:37)])     #length=301  38 sets have NA
detach(all_ivgtt)
attach(data_ivgtt)
id.seq<-unique(PTID)
time<-c(0,1,3,5,10,20,30,40,50,60,90,120,180)
timetod<-seq(-25,0,0.1)
age<-seq(15,75,0.01)
nf <- layout(matrix(1:4,2,2,byrow=TRUE), c(1,1.81), c(1,1), TRUE)
layout.show(nf)
result<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and IVGTT\\result_ivgtt_new.txt", header=T)
result.new<-na.omit(result)
attach(result.new)
par<-as.numeric(result.new[timetod.pv.g < 0.05 , ][1,1:8])
y<-par[2]+par[4]+par[6]*(47.16+timetod)+par[8]*timetod
plot(data_ivgtt[DM==1,]$IV_GLU000~data_ivgtt[DM==1,]$TIMEtoD,ylim=c(55,12
0),pch=19,cex=0.65,font=2,font.lab=2,
ylab="Fasting blood glucose (mg/dL)",xlab="TIMEtoD (year)",xlim=c(-25,0.1))
lines(y~timetod)
y<-par[2]+par[6]*age
plot(data_ivgtt[DM==0,]$IV_GLU000~data_ivgtt[DM==0,]$TESTAGE,ylim=c(55,1
20),pch=19,cex=0.65,font=2,font.lab=2,
```

```
ylab="Fasting blood glucose (mg/dL)",xlab="Age (year)")
lines(y~age)
par<-as.numeric(result.new[timetod.pv.g < 0.05 , ][6,1:8])
y<-par[2]+par[4]+par[6]*(47.16+timetod)+par[8]*timetod
plot(data_ivgtt[DM==1,]$IV_INS003~data_ivgtt[DM==1,]$TIMEtoD,ylim=c(0,350),
pch=19,cex=0.65,font=2,font.lab=2,
ylab="Insulin at 3 min (mU/L)",xlab="TIMEtoD (year)",xlim=c(-25,0.1))
lines(y~timetod)
y<-par[2]+par[6]*age
plot(data_ivgtt[DM==0,]$IV_INS003~data_ivgtt[DM==0,]$TESTAGE,ylim=c(0,35
0),pch=19,cex=0.65,font=2,font.lab=2,
ylab="Insulin at 3 min (mU/L)",xlab="Age (year)")
lines(y~age)
detach(result.new)
```

Figure 4.2

```
timetod<-seq(-25,0,0.1)
age<-seq(15,75,0.01)
nf <- layout(matrix(1:6,3,2,byrow=TRUE), c(1,1.81), c(1,1,1), TRUE)
layout.show(nf)
result.g<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and IVGTT\\result_gludif_ivgtt_final2.txt", header=T)
attach(result.g)
par<-as.numeric(result.g[timetod.pv.g<0.05 & dm.pv.g < 0.0005,][3,1:8])
#GLU10-50
y<-par[2]+par[4]+par[6]*(47.16+timetod)+par[8]*timetod
plot((data_ivgtt[DM==1,]$IV_GLU010-
data_ivgtt[DM==1,]$IV_GLU050)~data_ivgtt[DM==1,]$TIMEtoD,ylim=c(50,230),
pch=19,cex=0.65,font=2,font.lab=2,
ylab="G10-50 (mg/dL)",xlab="TIMEtoD (year)",xlim=c(-25,0.1))
lines(y~timetod)
y<-par[2]+par[6]*age
plot((data_ivgtt[DM==0,]$IV_GLU010-
data_ivgtt[DM==0,]$IV_GLU050)~data_ivgtt[DM==0,]$TESTAGE,ylim=c(50,230),
pch=19,cex=0.65,font=2,font.lab=2,
ylab="G10-50 (mg/dL)",xlab="Age (year)")
lines(y~age)
detach(result.g)
result.i<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and IVGTT\\result_insdif_ivgtt_final2.txt", header=T)
attach(result.i)
par<-as.numeric(result.i[timetod.pv.i<0.05 & dm.pv.i < 0.000005,][1,1:8])
y<--par[2]-par[4]-par[6]*(47.16+timetod)-par[8]*timetod
```

```
plot((data_ivgtt[DM==1,]$IV_INS120-
data_ivgtt[DM==1,]$IV_INS000)~data_ivgtt[DM==1,]$TIMEtoD,ylim=c(-
40,70),pch=19,cex=0.65,font=2,font.lab=2,
ylab="I120-0 (mU/L)",xlab="TIMEtoD (year)",xlim=c(-25,0.1))
lines(y~timetod)
y<-par[2]+par[6]*age
plot((data_ivgtt[DM==0,]$IV_INS120-
data_ivgtt[DM==0,]$IV_INS000)~data_ivgtt[DM==0,]$TESTAGE,ylim=c(-
40,70),pch=19,cex=0.65,font=2,font.lab=2,
ylab="I120-0 (mU/L)",xlab="Age (year)")
lines(y~age)
#IV_INS003-IV_INS020
par<-as.numeric(result.i[timetod.pv.i<0.05 & dm.pv.i < 0.000005,][3,1:8])
y<-par[2]+par[4]+par[6]*(47.16+timetod)+par[8]*timetod
plot((data_ivgtt[DM==1,]$IV_INS003-
data_ivgtt[DM==1,]$IV_INS020)~data_ivgtt[DM==1,]$TIMEtoD,ylim=c(-
160,160),pch=19,cex=0.65,font=2,font.lab=2,
ylab="I3-20 (mU/L)",xlab="TIMEtoD (year)",xlim=c(-25,0.1))
lines(y~timetod)
y<-par[2]+par[6]*age
plot((data_ivgtt[DM==0,]$IV_INS003-
data_ivgtt[DM==0,]$IV_INS020)~data_ivgtt[DM==0,]$TESTAGE,ylim=c(-
160,160),pch=19,cex=0.65,font=2,font.lab=2,
ylab="I3-20 (mU/L)",xlab="Age (year)")
lines(y~age)
```

Figure 4.3

```
GLU<-data_ivgtt[,5:17]
INS<-data_ivgtt[,18:30]
detach(data_ivgtt)
par(mfcol=c(2,2))
i<-144
        plot(as.numeric(GLU[i,2:13])~time[2:13],pch=19,cex=0.65,font=2,font.lab=
2,
        ylab="Glu conc (mg/dL)",xlab="Time (min)",main="Cubic spline fit of
IVGTT data ",ylim=c(0,700))
        spl_glu<-
smooth.spline(as.numeric(GLU[i,2:13])~time[2:13],df=8,all.knots=T)
        glu.pred<-predict(spl_glu,seq(1,180,0.1))
        lines(glu.pred)
        plot(as.numeric(INS[i,2:13])~time[2:13],pch=19,cex=0.65,font=2,font.lab=2
,
        ylab="Ins conc (mU/L)",xlab="Time (min)",ylim=c(0,300))
        spl_ins<-
smooth.spline(as.numeric(INS[i,2:13])~time[2:13],df=7,all.knots=T)
```

```
        ins.pred<-predict(spl_ins,seq(1,180,0.1))
        lines(ins.pred)
i<-230
        plot(as.numeric(GLU[i,2:13])~time[2:13],pch=19,cex=0.65,font=2,font.lab=
2,
        ylab="Glu conc (mg/dL)",xlab="Time (min)",main="Cubic spline fits of
IVGTT data",ylim=c(0,700))
        spl_glu<-
smooth.spline(as.numeric(GLU[i,2:13])~time[2:13],df=8,all.knots=T)
        glu.pred<-predict(spl_glu,seq(1,180,0.1))
        lines(glu.pred)
        plot(as.numeric(INS[i,2:13])~time[2:13],pch=19,cex=0.65,font=2,font.lab=2
,
        ylab="Ins conc (mU/L)",xlab="Time (min)",ylim=c(0,300))
        spl_ins<-
smooth.spline(as.numeric(INS[i,2:13])~time[2:13],df=7,all.knots=T)
        ins.pred<-predict(spl_ins,seq(1,180,0.1))
        lines(ins.pred)
```

Figure 4.4

```
result.glu.ivgtt<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and IVGTT\\result_glu_ivgtt_final.txt", header=T)
attach(result.glu.ivgtt)
result.ins.ivgtt<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and IVGTT\\result_ins_ivgtt_final.txt", header=T)
attach(result.ins.ivgtt)
par(mfrow=c(2,2))
plot(timetod.mn.g[1:180]~c(1:180),xlab="Time(min)",ylab=expression(mu[TIMEto
D]),main="Fixed effect parameters of TIMEtoD (Glucose)",type="l",ylim=c(-
1.6,1.1))
points(timetod.mn.g[1:180]~c(1:180),pch=19,cex=0.4)
abline(h=0,lty=2)
plot(timetod.pv.g[1:180]~c(1:180),ylim=c(0,1),xlab="Time(min)",ylab="P-
value",main="P-values (Glucose)",type="l")
points(timetod.pv.g[1:180]~c(1:180),pch=19,cex=0.4)
abline(h=0.05,lty=2)
plot(dm.mn.g[1:180]~c(1:180),xlab="Time(min)",ylab=expression(mu[PG]),main=
"Fixed effect parameters of PG (Glucose)",type="l")
points(dm.mn.g[1:180]~c(1:180),pch=19,cex=0.4)
abline(h=0,lty=2)
plot(dm.pv.g[1:180]~c(1:180),ylim=c(0,1),xlab="Time(min)",ylab="P-
value",main="P-values (Glucose)",type="l")
points(dm.pv.g[1:180]~c(1:180),pch=19,cex=0.4)
abline(h=0.05,lty=2)
```

Figure 4.5

```
par(mfrow=c(2,2))
plot(timetod.mn.i[1:180]~c(1:180),xlab="Time(min)",ylab=expression(mu[TIMEto
D]),main="Fixed effect parameters of TIMEtoD (Insulin)",type="l",ylim=c(-3,1.1))
points(timetod.mn.i[1:180]~c(1:180),pch=19,cex=0.4)
abline(h=0,lty=2)
plot(timetod.pv.i[1:180]~c(1:180),ylim=c(0,1),xlab="Time(min)",ylab="P-
value",main="P-values (Insulin)",type="l")
points(timetod.pv.i[1:180]~c(1:180),pch=19,cex=0.4)
abline(h=0.05,lty=2)
plot(dm.mn.i[1:180]~c(1:180),xlab="Time(min)",ylab=expression(mu[PG]),main="
Fixed effect parameters of PG (Insulin)",type="l")
points(dm.mn.i[1:180]~c(1:180),pch=19,cex=0.4)
abline(h=0,lty=2)
plot(dm.pv.i[1:180]~c(1:180),ylim=c(0,1),xlab="Time(min)",ylab="P-
value",main="P-values (Insulin)",type="l")
points(dm.pv.i[1:180]~c(1:180),pch=19,cex=0.4)
abline(h=0.05,lty=2)
```

Figure 4.6

```
saved.glu<-matrix(,301,180)
saved.ins<-matrix(,301,180)
sub<-1
start<-1
for(i in 1:301)
{
        spl_glu<-
smooth.spline(as.numeric(GLU[i,2:13])~time[2:13],df=8,all.knots=T)
        glu.pred<-predict(spl_glu,seq(1,180,1))
        saved.glu[i,1:180]<-glu.pred$y

        spl_ins<-
smooth.spline(as.numeric(INS[i,2:13])~time[2:13],df=7,all.knots=T)
        ins.pred<-predict(spl_ins,seq(1,180,1))
        saved.ins[i,1:180]<-ins.pred$y
}
plot.data<-data.frame(cbind(data_ivgtt[,1:4],saved.ins[,104]))
names(plot.data)<-c("PTID","TESTAGE","DM","TIMEtoD","INS104")
attach(plot.data)
result.ins.ivgtt<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and IVGTT\\result_ins_ivgtt_final.txt", header=T)
attach(result.ins.ivgtt)
parins104<-as.numeric(result.ins.ivgtt[104,])
nf <- layout(matrix(1:2,1,2,byrow=TRUE), c(1,1.81), c(1), TRUE)
layout.show(nf)
```

```
timetod<-seq(-25,0,0.1)
age<-seq(15,75,0.01)
y<-
parins104[2]+parins104[4]+parins104[6]*(47.16+timetod)+parins104[8]*timetod
plot(plot.data[DM==1,]$INS104~plot.data[DM==1,]$TIMEtoD,ylim=c(0,150),pch=
19,cex=0.5,font=2,font.lab=2,
ylab="Insulin at 104 min (mU/L)",xlab="TIMEtoD (year)",xlim=c(-25,0.1))
lines(y~timetod)
y<-parins104[2]+parins104[6]*age
plot(plot.data[DM==0,]$INS104~plot.data[DM==0,]$TESTAGE,ylim=c(0,150),pch
=19,cex=0.5,font=2,font.lab=2,
ylab="Insulin at 104 min (mU/L)",xlab="Age (year)")
lines(y~age)
```

Figure 5.1

```
all_ogtt<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and ogtt\\all_ogtt.txt", header=T)
attach(all_ogtt)
all_ogtt<-all_ogtt[TIMEtoD<=0,]
for(i in 1:552)
{if(all_ogtt$DM[i]==0) all_ogtt$TIMEtoD[i]<-0}
data_ogtt<-na.omit(all_ogtt[,c(1,4,7:18,20,22,25:31,33,35,38)])     #length=527
detach(all_ogtt)
attach(data_ogtt)
id.seq<-unique(PTID)
time<-c(0,15,30,45,60,90,120,180,240)
ptid<-PTID
age<-TESTAGE
timetod<-TIMEtoD
dm<-DM
GLU<-data_ogtt[,8:16]
INS<-data_ogtt[,17:25]
pred.time<-c(1:240)
aucg240<-numeric()
for(j in 1:527)
{
        g<-as.numeric(GLU[j,])
        aucg240[j]<-15*(g[1]+g[2])/2 +15*(g[2]+g[3])/2 +15*(g[3]+g[4])/2
+15*(g[4]+g[5])/2 +
                30*(g[5]+g[6])/2 +30*(g[6]+g[7])/2 +60*(g[7]+g[8])/2
+60*(g[8]+g[9])/2
}

timetod.pred<-seq(-25,0,0.1)
par(mfrow=c(3,2))
```

```
result<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and OGTT\\result_ogtt_new.txt", header=T)
result.new<-na.omit(result)
attach(result.new)
par<-as.numeric(result.new[timetod.pv.g < 0.05 & dm.pv.g < 0.0001, ][1,1:8])
y<-par[2]+par[4]+par[6]*(47.16+timetod.pred)+par[8]*timetod.pred
plot(data_ogtt[DM==1,]$O_GLU000~data_ogtt[DM==1,]$TIMEtoD,pch=19,cex=0
.65,font=2,font.lab=2,
ylab="FBG (mg/dL)",xlab="TIMEtoD (year)",xlim=c(-25,0.1))
lines(smooth.spline(data_ogtt[DM==1,]$O_GLU000~data_ogtt[DM==1,]$TIMEto
D,df=3),lty=2)
lo<-lowess(data_ogtt[DM==1,]$O_GLU000~data_ogtt[DM==1,]$TIMEtoD)
lines(y~timetod.pred)
par<-as.numeric(result.new[timetod.pv.g < 0.05 & dm.pv.g < 0.0001, ][5,1:8])
y<-par[2]+par[4]+par[6]*(47.16+timetod.pred)+par[8]*timetod.pred
plot(data_ogtt[DM==1,]$O_GLU120~data_ogtt[DM==1,]$TIMEtoD,pch=19,cex=0
.65,font=2,font.lab=2,
ylab="OG120 (mg/dL)",xlab="TIMEtoD (year)",xlim=c(-25,0.1))
lines(smooth.spline(data_ogtt[DM==1,]$O_GLU120~data_ogtt[DM==1,]$TIMEto
D,df=3),lty=2)
lines(y~timetod.pred)
par<-as.numeric(result.new[timetod.pv.g < 0.05 & dm.pv.g < 0.0001, ][4,1:8])
y<-par[2]+par[4]+par[6]*(47.16+timetod.pred)+par[8]*timetod.pred
plot(data_ogtt[DM==1,]$O_GLU120~data_ogtt[DM==1,]$TIMEtoD,pch=19,cex=0
.65,font=2,font.lab=2,
ylab="OGL90 (mg/dL)",xlab="TIMEtoD (year)",xlim=c(-25,0.1))
lines(smooth.spline(data_ogtt[DM==1,]$O_GLU090~data_ogtt[DM==1,]$TIMEto
D,df=3),lty=2)
lines(y~timetod.pred)
detach(result.new)
result<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and oGTT\\result_gludif_ogtt.txt", header=T)
attach(result)
par<--as.numeric(result[timetod.pv.g<0.05 & dm.pv.g < 0.00000001,1:8])
y<-par[2]+par[4]+par[6]*(47.16+timetod.pred)+par[8]*timetod.pred
plot((data_ogtt[DM==1,]$O_GLU060-
data_ogtt[DM==1,]$O_GLU030)~data_ogtt[DM==1,]$TIMEtoD,pch=19,cex=0.65,
font=2,font.lab=2,
ylab="G60-30 (mg/dL)",xlab="TIMEtoD (year)",xlim=c(-25,0.1))
lines(smooth.spline((data_ogtt[DM==1,]$O_GLU060-
data_ogtt[DM==1,]$O_GLU030)~data_ogtt[DM==1,]$TIMEtoD,df=3),lty=2)
lines(y~timetod.pred)
detach(result)
y<-
20137.26123+8617.13402+49.29864*(47.16+timetod.pred)+513.22996*timetod.p
red
```

```
testdata<-data.frame(cbind(ptid,dm,age,timetod,aucg240))
names(testdata)<-c("PTID","DM","AGE","TIMEtoD","METRIC")
attach(testdata)
plot(testdata[DM==1,]$METRIC~testdata[DM==1,]$TIMEtoD,pch=19,cex=0.65,fo
nt=2,font.lab=2,
ylab="AUCg0-240(min*mg/dL)",xlab="TIMEtoD (year)",xlim=c(-25,0.1))
lines(smooth.spline(testdata[DM==1,]$METRIC~testdata[DM==1,]$TIMEtoD,df=3
),lty=2)
lines(y~timetod.pred)
write.table(testdata,"C:\\Documents and Settings\\clin\\Desktop\\Disease
development and OGTT\\WINBUGS\\AUCg.txt",sep=" ",quote=F)
detach(testdata)
result<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and OGTT\\result_insdif_ogtt.txt", header=T)
attach(result)
par<--as.numeric(result[timetod.pv.i<0.05 & dm.pv.i < 0.05,1:8])
y<-par[2]+par[4]+par[6]*(47.16+timetod.pred)+par[8]*timetod.pred
plot((data_ogtt[DM==1,]$O_INS120-
data_ogtt[DM==1,]$O_INS045)~data_ogtt[DM==1,]$TIMEtoD,pch=19,cex=0.65,f
ont=2,font.lab=2,
ylab="I120-45(mU/L)",xlab="TIMEtoD (year)",xlim=c(-25,0.1))
lines(smooth.spline((data_ogtt[DM==1,]$O_INS120-
data_ogtt[DM==1,]$O_INS045)~data_ogtt[DM==1,]$TIMEtoD,df=3),lty=2)
lines(y~timetod.pred)
detach(result)
```

Figure 5.2

```
par(mfcol=c(2,2))
i<-132
        plot(as.numeric(GLU[i,])~time,ylab="Glucose conc (mg/dL)",xlab="Time
(min)",main="Cubic spline fits (TIMEtoD = -7.16 years)",ylim=c(50,220))
        spl_glu<-smooth.spline(as.numeric(GLU[i,])~time,df=5,all.knots=T)
        glu.pred<-predict(spl_glu,seq(1,240,0.1))
        lines(glu.pred)
        plot(as.numeric(INS[i,])~time,ylab="Insulin conc (mU/L)",xlab="Time
(min)",ylim=c(0,150))
        spl_ins<-smooth.spline(as.numeric(INS[i,])~time,df=4,all.knots=T)
        ins.pred<-predict(spl_ins,seq(1,240,0.1))
        lines(ins.pred)
i<-137
        plot(as.numeric(GLU[i,])~time,ylab="Glucose conc (mg/dL)",xlab="Time
(min)",main="Cubic spline fits (TIMEtoD = -0.8 years)",ylim=c(50,220))
        spl_glu<-smooth.spline(as.numeric(GLU[i,])~time,df=5,all.knots=T)
        glu.pred<-predict(spl_glu,seq(1,240,0.1))
        lines(glu.pred)
```

```
        plot(as.numeric(INS[i,])~time,ylab="Insulin conc (mU/L)",xlab="Time
(min)",ylim=c(0,150))
        spl_ins<-smooth.spline(as.numeric(INS[i,])~time,df=4,all.knots=T)
        ins.pred<-predict(spl_ins,seq(1,240,0.1))
        lines(ins.pred)
```

Figure 5.3

```
timetod.pred<-seq(-25,0,0.01)
age.pred<-seq(10,75,0.01)
nf <- layout(matrix(1:8,4,2,byrow=TRUE), c(1,1.985), c(1,1,1,1), TRUE)
layout.show(nf)
y<-
69.48+0.5281+0.2143*(47.33+timetod.pred)+2.047*as.numeric(I(timetod.pred> (-
9.881)))*(timetod.pred-(-9.881))
plot(data_ogtt[DM==1,]$O_GLU000~data_ogtt[DM==1,]$TIMEtoD,ylim=c(55,145
),pch=19,cex=0.65,font=2,font.lab=2,
ylab="FBG (mg/dL)",xlab="TIMEtoD (year)",xlim=c(-25,0.1))
lines(y~timetod.pred)
y<-69.48+0.2143*age.pred
plot(data_ogtt[DM==0,]$O_GLU000~data_ogtt[DM==0,]$TESTAGE,ylim=c(55,14
5),pch=19,cex=0.65,font=2,font.lab=2,
ylab="FBG (mg/dL)",xlab="Age (year)")
lines(y~age.pred)
y<-75.39+16.1+0.4444*(47.33+timetod.pred)+8.718*as.numeric(I(timetod.pred>
(-6.87)))*(timetod.pred-(-6.87))
plot(data_ogtt[DM==1,]$O_GLU120~data_ogtt[DM==1,]$TIMEtoD,ylim=c(30,300
),pch=19,cex=0.65,font=2,font.lab=2,
ylab="G120 (mg/dL)",xlab="TIMEtoD (year)",xlim=c(-25,0.1))
lines(y~timetod.pred)
y<-75.39+0.4444*age.pred
plot(data_ogtt[DM==0,]$O_GLU120~data_ogtt[DM==0,]$TESTAGE,ylim=c(30,30
0),pch=19,cex=0.65,font=2,font.lab=2,
ylab="G120 (mg/dL)",xlab="Age (year)")
lines(y~age.pred)
y<-85.66+13.32+0.4686*(47.33+timetod.pred)+5.58*as.numeric(I(timetod.pred>
(-10.03)))*(timetod.pred-(-10.03))
plot(data_ogtt[DM==1,]$O_GLU090~data_ogtt[DM==1,]$TIMEtoD,pch=19,cex=0
.65,font=2,font.lab=2,
ylab="G90 (mg/dL)",xlab="TIMEtoD (year)",ylim=c(30,260),xlim=c(-25,0.1))
lines(y~timetod.pred)
y<-85.66+0.4686*age.pred
plot(data_ogtt[DM==0,]$O_GLU090~data_ogtt[DM==0,]$TESTAGE,pch=19,cex=
0.65,font=2,font.lab=2,ylim=c(30,260),
ylab="G90 (mg/dL)",xlab="Age (year)")
lines(y~age.pred)
```

```
result<-read.table("C:\\Documents and Settings\\clin\\Desktop\\Disease
development and oGTT\\result_gludif_ogtt.txt", header=T)
attach(result)
par<--as.numeric(result[timetod.pv.g<0.05 & dm.pv.g < 0.00000001,1:8])
y<-par[2]+par[4]+par[6]*(47.33+timetod.pred)+par[8]*timetod.pred
plot((data_ogtt[DM==1,]$O_GLU060-
data_ogtt[DM==1,]$O_GLU030)~data_ogtt[DM==1,]$TIMEtoD,pch=19,cex=0.65,
font=2,font.lab=2,
ylab="G60-30 (mg/dL)",xlab="TIMEtoD (year)",xlim=c(-25,0.1),ylim=c(-80,90))
lines(y~timetod.pred)
y<-par[2]+par[6]*age.pred
plot((data_ogtt[DM==0,]$O_GLU060-
data_ogtt[DM==0,]$O_GLU030)~data_ogtt[DM==0,]$TESTAGE,pch=19,cex=0.6
5,font=2,font.lab=2,
ylab="G60-30 (mg/dL)",xlab="AGE (year)",ylim=c(-80,90))
lines(y~age.pred)
detach(result)
```

APPENDIX E. PUBLICATIONS AND SUBMITTED MANUSCRIPTS

1. **Lin CW**, Veng-Pedersen P. Analysis of PK/PD risk factors for development of type 2 diabetes in high risk population using Bayesian analysis of glucose-insulin kinetics. *J Pharmacokinet Pharmacodyn.* 2009 Oct; 36 (5): 421-41.

2. **Lin CW**, JH Warram, Veng-Pedersen P. A Bayesian Population Analysis of Development of Type 2 Diabetes in the Offspring of Diabetic Parents. *J Pharmacokinet Pharmacodyn.* 2011 Aug 11.

3. **Lin CW**, JH Warram, Veng-Pedersen P. Disease Development Analysis of Type 2 Diabetes in Offspring of Parents with Type 2 Diabetes based on Glucose and Insulin variables after IVGTT. (submitted)

4. **Lin CW**, JH Warram, Veng-Pedersen P. Disease Development Analysis of Type 2 Diabetes in Offspring of Parents with Type 2 Diabetes based on Glucose and Insulin Variables after OGTT. (submitted)

REFERENCES

1. Zimmet, P.,Alberti, K. G.,Shaw, J., Global and societal implications of the diabetes epidemic. *Nature* **2001,** *414* (6865), 782-787.

2. American Diabetes Association. The 2011 National Diabetes Fact Sheet. 2011. http://www.diabetes.org/diabetes-basics/diabetes-statistics/?utm_source=WWW&utm_medium=DropDownDB&utm_content=Statistics&utm_campaign=CON.

3. International Diabetes Federation.Global Burden: Prevalence and Projections, 2010 and 2030. http://www.diabetesatlas.org/content/diabetes-and-impaired-glucose-tolerance.

4. King, H.,Aubert, R. E.,Herman, W. H., Global burden of diabetes, 1995-2025: prevalence, numerical estimates, and projections. *Diabetes Care* **1998,** *21* (9), 1414-1431.

5. Joslin, E. P.,Kahn, C. R., *Joslin's diabetes mellitus*. 14th ed.; Lippincott Williams & Willkins: Philadelphia, Pa., 2005; p xiv, 1209 p.

6. Maahs, D. M.,West, N. A.,Lawrence, J. M.,Mayer-Davis, E. J., Epidemiology of type 1 diabetes. *Endocrinol Metab Clin North Am 39* (3), 481-497.

7. Krolewski, A. S.,Warram, J. H.,Rand, L. I.,Kahn, C. R., Epidemiologic approach to the etiology of type I diabetes mellitus and its complications. *N Engl J Med* **1987,** *317* (22), 1390-1398.

8. Colditz, G. A.,Willett, W. C.,Rotnitzky, A.,Manson, J. E., Weight gain as a risk factor for clinical diabetes mellitus in women. *Ann Intern Med* **1995,** *122* (7), 481-486.

9. Karter, A. J.,Rowell, S. E.,Ackerson, L. M.,Mitchell, B. D.,Ferrara, A.,Selby, J. V.,Newman, B., Excess maternal transmission of type 2 diabetes. The Northern California Kaiser Permanente Diabetes Registry. *Diabetes Care* **1999,** *22* (6), 938-943.

10. Yki-Jarvinen, H., Pathogenesis of non-insulin-dependent diabetes mellitus. *Lancet* **1994,** *343* (8889), 91-95.

11. Weijnen, C. F.,Rich, S. S.,Meigs, J. B.,Krolewski, A. S.,Warram, J. H., Risk of diabetes in siblings of index cases with Type 2 diabetes: implications for genetic studies. *Diabet Med* **2002,** *19* (1), 41-50.

12. Meigs, J. B.,Cupples, L. A.,Wilson, P. W., Parental transmission of type 2 diabetes: the Framingham Offspring Study. *Diabetes* **2000,** *49* (12), 2201-2207.

13. Klein, B. E.,Klein, R.,Moss, S. E.,Cruickshanks, K. J., Parental history of diabetes in a population-based study. *Diabetes Care* **1996,** *19* (8), 827-830.

14. Gabir, M. M.,Hanson, R. L.,Dabelea, D.,Imperatore, G.,Roumain, J.,Bennett, P. H.,Knowler, W. C., The 1997 American Diabetes Association and 1999 World Health Organization criteria for hyperglycemia in the diagnosis and prediction of diabetes. *Diabetes Care* **2000,** *23* (8), 1108-1112.

15. Shaw, J. E.,Zimmet, P. Z.,de Courten, M.,Dowse, G. K.,Chitson, P.,Gareeboo, H.,Hemraj, F.,Fareed, D.,Tuomilehto, J.,Alberti, K. G., Impaired fasting glucose or impaired glucose tolerance. What best predicts future diabetes in Mauritius? *Diabetes Care* **1999,** *22* (3), 399-402.

16. Nathan, D. M.,Davidson, M. B.,DeFronzo, R. A.,Heine, R. J.,Henry, R. R.,Pratley, R.,Zinman, B., Impaired fasting glucose and impaired glucose tolerance: implications for care. *Diabetes Care* **2007,** *30* (3), 753-759.

17. Appleton, C. A., Problems with new criteria for diagnosis of diabetes mellitus. *Med J Aust* **1999,** *171* (2), 107.

18. Barrett-Connor, E., The oral glucose tolerance test, revisited. *Eur Heart J* **2002,** *23* (16), 1229-1231.

19. Kilpatrick, E. S.,Maylor, P. W.,Keevil, B. G., Biological variation of glycated hemoglobin. Implications for diabetes screening and monitoring. *Diabetes Care* **1998,** *21* (2), 261-264.

20. Lawrence, J.,Robinson, A., Screening for diabetes in general practice. *Prev Cardiol* **2003,** *6* (2), 78-84.

21. Kosaka, K.,Mizuno, Y.,Kuzuya, T., Reproducibility of the oral glucose tolerance test and the rice-meal test in mild diabetics. *Diabetes* **1966,** *15* (12), 901-904.

22. DeFronzo, R. A., Insulin resistance: a multifaceted syndrome responsible for NIDDM, obesity, hypertension, dyslipidaemia and atherosclerosis. *Neth J Med* **1997,** *50* (5), 191-197.

23. Warram, J. H.,Martin, B. C.,Krolewski, A. S.,Soeldner, J. S.,Kahn, C. R., Slow glucose removal rate and hyperinsulinemia precede the development of type II diabetes in the offspring of diabetic parents. *Ann Intern Med* **1990,** *113* (12), 909-915.

24. Martin, B. C.,Warram, J. H.,Krolewski, A. S.,Bergman, R. N.,Soeldner, J. S.,Kahn, C. R., Role of glucose and insulin resistance in development of type 2 diabetes mellitus: results of a 25-year follow-up study. *Lancet* **1992,** *340* (8825), 925-929.

25. Lundgren, H.,Bengtsson, C.,Blohme, G.,Lapidus, L.,Waldenstrom, J., Fasting serum insulin concentration and early insulin response as risk determinants for developing diabetes. *Diabet Med* **1990,** *7* (5), 407-413.

26. Reaven, G. M.,Bernstein, R.,Davis, B.,Olefsky, J. M., Nonketotic diabetes mellitus: insulin deficiency or insulin resistance? *Am J Med* **1976,** *60* (1), 80-88.

27. DeFronzo, R. A., Lilly lecture 1987. The triumvirate: beta-cell, muscle, liver. A collusion responsible for NIDDM. *Diabetes* **1988,** *37* (6), 667-687.

28. Kosaka, K.,Kuzuya, T.,Akanuma, Y.,Hagura, R., Increase in insulin response after treatment of overt maturity-onset diabetes is independent of the mode of treatment. *Diabetologia* **1980,** *18* (1), 23-28.

29. Vague, P.,Moulin, J. P., The defective glucose sensitivity of the B cell in non insulin dependent diabetes. Improvement after twenty hours of normoglycaemia. *Metabolism* **1982,** *31* (2), 139-142.

30. Garvey, W. T.,Olefsky, J. M.,Griffin, J.,Hamman, R. F.,Kolterman, O. G., The effect of insulin treatment on insulin secretion and insulin action in type II diabetes mellitus. *Diabetes* **1985,** *34* (3), 222-234.

31. Calles-Escandon, J.,Robbins, D. C., Loss of early phase of insulin release in humans impairs glucose tolerance and blunts thermic effect of glucose. *Diabetes* **1987,** *36* (10), 1167-1172.

32. Perley, M. J.,Kipnis, D. M., Plasma insulin responses to oral and intravenous glucose: studies in normal and diabetic sujbjects. *J Clin Invest* **1967,** *46* (12), 1954-1962.

33. Cerasi, E.,Luft, R., The plasma insulin response to glucose infusion in healthy subjects and in diabetes mellitus. *Acta Endocrinol (Copenh)* **1967,** *55* (2), 278-304.

34. Leahy, J. L., Natural history of beta-cell dysfunction in NIDDM. *Diabetes Care* **1990,** *13* (9), 992-1010.

35. Warram, J. H.,Martin, B. C.,Soeldner, J. S.,Krolewski, A. S., Study of glucose removal rate and first phase insulin secretion in the offspring of two parents with non-insulin-dependent diabetes. *Adv Exp Med Biol* **1988,** *246*, 175-179.

36. Mitrakou, A.,Kelley, D.,Mokan, M.,Veneman, T.,Pangburn, T.,Reilly, J.,Gerich, J., Role of reduced suppression of glucose production and diminished early insulin release in impaired glucose tolerance. *N Engl J Med* **1992,** *326* (1), 22-29.

37. Screening for type 2 diabetes. *Diabetes Care* **2000,** *23 Suppl 1*, S20-23.

38. Weyer, C.,Bogardus, C.,Mott, D. M.,Pratley, R. E., The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. *J Clin Invest* **1999,** *104* (6), 787-794.

39. Warram, J. H.,Kopczynski, J.,Janka, H. U.,Krolewski, A. S., Epidemiology of non-insulin-dependent diabetes mellitus and its macrovascular complications. A basis for the development of cost-effective programs. *Endocrinol Metab Clin North Am* **1997,** *26* (1), 165-188.

40. Lindstrom, J.,Tuomilehto, J., The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* **2003,** *26* (3), 725-731.

41. Hanefeld, M.,Fischer, S.,Schmechel, H.,Rothe, G.,Schulze, J.,Dude, H.,Schwanebeck, U.,Julius, U., Diabetes Intervention Study. Multi-intervention trial in newly diagnosed NIDDM. *Diabetes Care* **1991,** *14* (4), 308-317.

42. Tuomilehto, J.,Lindstrom, J.,Eriksson, J. G.,Valle, T. T.,Hamalainen, H.,Ilanne-Parikka, P.,Keinanen-Kiukaanniemi, S.,Laakso, M.,Louheranta, A.,Rastas, M.,Salminen, V.,Uusitupa, M., Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med* **2001,** *344* (18), 1343-1350.

43. Knowler, W. C.,Barrett-Connor, E.,Fowler, S. E.,Hamman, R. F.,Lachin, J. M.,Walker, E. A.,Nathan, D. M., Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* **2002,** *346* (6), 393-403.

44. Chiasson, J. L.,Josse, R. G.,Gomis, R.,Hanefeld, M.,Karasik, A.,Laakso, M., Acarbose for prevention of type 2 diabetes mellitus: the STOP-NIDDM randomised trial. *Lancet* **2002,** *359* (9323), 2072-2077.

45. Torgerson, J. S.,Hauptman, J.,Boldrin, M. N.,Sjostrom, L., XENical in the prevention of diabetes in obese subjects (XENDOS) study: a randomized study of orlistat as an adjunct to lifestyle changes for the prevention of type 2 diabetes in obese patients. *Diabetes Care* **2004,** *27* (1), 155-161.

46. Walker, K. Z.,O'Dea, K.,Gomez, M.,Girgis, S.,Colagiuri, R., Diet and exercise in the prevention of diabetes. *J Hum Nutr Diet 23* (4), 344-352.

47. The prevention or delay of type 2 diabetes. *Diabetes Care* **2002,** *25* (4), 742-749.

48. Li, G.,Zhang, P.,Wang, J.,Gregg, E. W.,Yang, W.,Gong, Q.,Li, H.,Jiang, Y.,An, Y.,Shuai, Y.,Zhang, B.,Zhang, J.,Thompson, T. J.,Gerzoff, R. B.,Roglic, G.,Hu, Y.,Bennett, P. H., The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. *Lancet* **2008,** *371* (9626), 1783-1789.

49. Lindstrom, J.,Ilanne-Parikka, P.,Peltonen, M.,Aunola, S.,Eriksson, J. G.,Hemio, K.,Hamalainen, H.,Harkonen, P.,Keinanen-Kiukaanniemi, S.,Laakso, M.,Louheranta, A.,Mannelin, M.,Paturi, M.,Sundvall, J.,Valle, T. T.,Uusitupa, M.,Tuomilehto, J., Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. *Lancet* **2006,** *368* (9548), 1673-1679.

50. Knowler, W. C.,Hamman, R. F.,Edelstein, S. L.,Barrett-Connor, E.,Ehrmann, D. A.,Walker, E. A.,Fowler, S. E.,Nathan, D. M.,Kahn, S. E., Prevention of type 2 diabetes with troglitazone in the Diabetes Prevention Program. *Diabetes* **2005,** *54* (4), 1150-1156.

51. Effects of withdrawal from metformin on the development of diabetes in the diabetes prevention program. *Diabetes Care* **2003,** *26* (4), 977-980.

52. Ette, E. I.,Williams, P. J., *Pharmacometrics : the science of quantitative pharmacology*. John Wiley & Sons: Hoboken, N.J., 2007; p xix, 1205 p.

53. Pinheiro, J. C.,Bates, D. M., *Mixed-effects models in S and S-PLUS*. Springer: New York, 2000; p xvi, 528 p.

54. Davidian, M.,Giltinan, D. M., *Nonlinear models for repeated measurement data*. Chapman & Hall/CRC: Boca Raton, Fla., 1998; p xv, 359 p.

55. Beal, S.,Sheiner, L., The Nonmem System. *American Statistician* **1980,** *34* (2), 118-119.

56. Delyon, B.,Lavielle, V.,Moulines, E., Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics* **1999,** *27* (1), 94-128.

57. Wei, G. C. G.,Tanner, M. A., A Monte-Carlo Implementation of the Em Algorithm and the Poor Mans Data Augmentation Algorithms. *Journal of the American Statistical Association* **1990,** *85* (411), 699-704.

58. Dartois, C.,Lemenuel-Diot, A.,Laveille, C.,Tranchand, B.,Tod, M.,Girard, P., Evaluation of uncertainty parameters estimated by different population PK software and methods. *J Pharmacokinet Pharmacodyn* **2007,** *34* (3), 289-311.

59. Gelman, A., *Bayesian data analysis*. 2nd ed.; Chapman & Hall/CRC: Boca Raton, Fla., 2004; p xxv, 668 p.

60. Spiegelhalter, D.,Thomas, A.,Best, N., WinBUGS User Manual Version 1.4. *Cambridge: Medical Research Council Biostatics Unit* **2003**.

61. Lunn, D., WinBUGS Differential Interface (WBDiff). *Dept. Epidemiology & Public Health, Imperial College School of Medicine* **2004**.

62. Kass, R. E.,Wasserman, L., The selection of prior distributions by formal rules. *Journal of the American Statistical Association* **1996,** *91* (435), 1343-1370.

63. Geman, S.,Geman, D., Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *Ieee Transactions on Pattern Analysis and Machine Intelligence* **1984,** *6* (6), 721-741.

64. Hastings, W. K., Monte-Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **1970,** *57* (1), 97-&.

65. Duffull, S. B.,Kirkpatrick, C. M.,Green, B.,Holford, N. H., Analysis of population pharmacokinetic data using NONMEM and WinBUGS. *J Biopharm Stat* **2005,** *15* (1), 53-73.

66. Segal, S.,Berman, M.,Blair, A., The metabolism of variously C14-labeled glucose in man and an estimation of the extent of glucose metabolism by the hexose monophosphate pathway. *J Clin Invest* **1961,** *40*, 1263-1279.

67. Ferrannini, E.,Smith, J. D.,Cobelli, C.,Toffolo, G.,Pilo, A.,DeFronzo, R. A., Effect of insulin on the distribution and disposition of glucose in man. *J Clin Invest* **1985,** *76* (1), 357-364.

68. Bergman, R. N.,Ider, Y. Z.,Bowden, C. R.,Cobelli, C., Quantitative estimation of insulin sensitivity. *Am J Physiol* **1979,** *236* (6), E667-677.

69. Caumo, A.,Cobelli, C., Hepatic glucose production during the labeled IVGTT: estimation by deconvolution with a new minimal model. *Am J Physiol* **1993,** *264* (5 Pt 1), E829-841.

70. Rorsman, P.,Renstrom, E., Insulin granule dynamics in pancreatic beta cells. *Diabetologia* **2003,** *46* (8), 1029-1045.

71. Dean, P. M., Ultrastructural morphometry of the pancreatic -cell. *Diabetologia* **1973,** *9* (2), 115-119.

72. Ohara-Imaizumi, M.,Nishiwaki, C.,Kikuta, T.,Nagai, S.,Nakamichi, Y.,Nagamatsu, S., TIRF imaging of docking and fusion of single insulin granule motion in primary rat pancreatic beta-cells: different behaviour of granule motion between normal and Goto-Kakizaki diabetic rat beta-cells. *Biochem J* **2004,** *381* (Pt 1), 13-18.

73. Gupta, N.,Hoffman, R. P.,Veng-Pedersen, P., Pharmacokinetic/pharmacodynamic differentiation of pancreatic responsiveness in obese and lean children. *Biopharm Drug Dispos* **2005,** *26* (7), 287-294.

74. Grodsky, G. M., A threshold distribution hypothesis for packet storage of insulin and its mathematical modeling. *J Clin Invest* **1972,** *51* (8), 2047-2059.

75. Overgaard, R. V.,Jelic, K.,Karlsson, M.,Henriksen, J. E.,Madsen, H., Mathematical beta cell model for insulin secretion following IVGTT and OGTT. *Ann Biomed Eng* **2006,** *34* (8), 1343-1354.

76. Toffolo, G.,Campioni, M.,Basu, R.,Rizza, R. A.,Cobelli, C., A minimal model of insulin secretion and kinetics to assess hepatic insulin extraction. *Am J Physiol Endocrinol Metab* **2006,** *290* (1), E169-E176.

77. De Gaetano, A.,Arino, O., Mathematical modelling of the intravenous glucose tolerance test. *J Math Biol* **2000,** *40* (2), 136-168.

78. Toffolo, G.,Bergman, R. N.,Finegood, D. T.,Bowden, C. R.,Cobelli, C., Quantitative estimation of beta cell sensitivity to glucose in the intact organism: a minimal model of insulin kinetics in the dog. *Diabetes* **1980,** *29* (12), 979-990.

79. Silvers, A.,Swenson, R. S.,Farquhar, J. W.,Reaven, G. M., Derivation of a three compartment model describing disappearance of plasma insulin-131-I in man. *J Clin Invest* **1969,** *48* (8), 1461-1469.

80. Holford, N. H.,Peace, K. E., Methodologic aspects of a population pharmacodynamic model for cognitive effects in Alzheimer patients treated with tacrine. *Proc Natl Acad Sci U S A* **1992,** *89* (23), 11466-11470.

81. Chan, P. L.,Holford, N. H., Drug treatment effects on disease progression. *Annu Rev Pharmacol Toxicol* **2001,** *41*, 625-659.

82. Pors Nielsen, S.,Barenholdt, O.,Hermansen, F.,Munk-Jensen, N., Magnitude and pattern of skeletal response to long term continuous and cyclic sequential oestrogen/progestin treatment. *Br J Obstet Gynaecol* **1994,** *101* (4), 319-324.

83. Anderson, B. J.,Holford, N. H.,Woollard, G. A.,Kanagasundaram, S.,Mahadevan, M., Perioperative pharmacodynamics of acetaminophen analgesia in children. *Anesthesiology* **1999,** *90* (2), 411-421.

84. Holford, N. H.,Chan, P. L.,Nutt, J. G.,Kieburtz, K.,Shoulson, I., Disease progression and pharmacodynamics in Parkinson disease - evidence for functional protection with levodopa and other treatments. *J Pharmacokinet Pharmacodyn* **2006,** *33* (3), 281-311.

85. Yano, Y.,Oguma, T.,Nagata, H.,Sasaki, S., Application of logistic growth model to pharmacodynamic analysis of in vitro bactericidal kinetics. *J Pharm Sci* **1998,** *87* (10), 1177-1183.

86. Zhi, J.,Nightingale, C. H.,Quintiliani, R., A pharmacodynamic model for the activity of antibiotics against microorganisms under nonsaturable conditions. *J Pharm Sci* **1986,** *75* (11), 1063-1067.

87. Friberg, L. E.,Freijs, A.,Sandstrom, M.,Karlsson, M. O., Semiphysiological model for the time course of leukocytes after varying schedules of 5-fluorouracil in rats. *J Pharmacol Exp Ther* **2000,** *295* (2), 734-740.

88. Krzyzanski, W.,Jusko, W. J., Multiple-pool cell lifespan model of hematologic effects of anticancer agents. *J Pharmacokinet Pharmacodyn* **2002,** *29* (4), 311-337.

89. Krzyzanski, W.,Jusko, W. J.,Wacholtz, M. C.,Minton, N.,Cheung, W. K., Pharmacokinetic and pharmacodynamic modeling of recombinant human erythropoietin after multiple subcutaneous doses in healthy subjects. *Eur J Pharm Sci* **2005,** *26* (3-4), 295-306.

90. Landersdorfer, C. B.,Jusko, W. J., Pharmacokinetic/pharmacodynamic modelling in diabetes mellitus. *Clin Pharmacokinet* **2008,** *47* (7), 417-448.

91. Matthews, D. R.,Hosker, J. P.,Rudenski, A. S.,Naylor, B. A.,Treacher, D. F.,Turner, R. C., Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **1985,** *28* (7), 412-419.

92. Ferrannini, E.,Mari, A., How to measure insulin sensitivity. *J Hypertens* **1998,** *16* (7), 895-906.

93. Stumvoll, M.,Mitrakou, A.,Pimenta, W.,Jenssen, T.,Yki-Jarvinen, H.,Van Haeften, T.,Renn, W.,Gerich, J., Use of the oral glucose tolerance test to assess insulin release and insulin sensitivity. *Diabetes Care* **2000,** *23* (3), 295-301.

94. Cretti, A.,Lehtovirta, M.,Bonora, E.,Brunato, B.,Zenti, M. G.,Tosi, F.,Caputo, M.,Caruso, B.,Groop, L. C.,Muggeo, M.,Bonadonna, R. C., Assessment of beta-cell function during the oral glucose tolerance test by a minimal model of insulin secretion. *Eur J Clin Invest* **2001,** *31* (5), 405-416.

95. Wallace, T. M.,Levy, J. C.,Matthews, D. R., Use and abuse of HOMA modeling. *Diabetes Care* **2004,** *27* (6), 1487-1495.

96. Levy, J. C.,Matthews, D. R.,Hermans, M. P., Correct homeostasis model assessment (HOMA) evaluation uses the computer program. *Diabetes Care* **1998,** *21* (12), 2191-2192.

97. DeFronzo, R. A.,Tobin, J. D.,Andres, R., Glucose clamp technique: a method for quantifying insulin secretion and resistance. *Am J Physiol* **1979,** *237* (3), E214-223.

98. Frey, N.,Laveille, C.,Paraire, M.,Francillard, M.,Holford, N. H.,Jochemsen, R., Population PKPD modelling of the long-term hypoglycaemic effect of gliclazide given as a once-a-day modified release (MR) formulation. *Br J Clin Pharmacol* **2003,** *55* (2), 147-157.

99. de Winter, W.,DeJongh, J.,Post, T.,Ploeger, B.,Urquhart, R.,Moules, I.,Eckland, D.,Danhof, M., A mechanism-based disease progression model for comparison of long-term effects of pioglitazone, metformin and gliclazide on disease processes underlying Type 2 Diabetes Mellitus. *J Pharmacokinet Pharmacodyn* **2006,** *33* (3), 313-343.

100. Cobelli, C.,Pacini, G.,Toffolo, G.,Sacca, L., Estimation of insulin sensitivity and glucose clearance from minimal model: new insights from labeled IVGTT. *Am J Physiol* **1986,** *250* (5 Pt 1), E591-598.

101. Lin, S.,Chien, Y. W., Pharmacokinetic-pharmacodynamic modelling of insulin: comparison of indirect pharmacodynamic response with effect-compartment link models. *J Pharm Pharmacol* **2002,** *54* (6), 791-800.

102. Silber, H. E.,Jauslin, P. M.,Frey, N.,Gieschke, R.,Simonsson, U. S.,Karlsson, M. O., An integrated model for glucose and insulin regulation in healthy volunteers and type 2 diabetic patients following intravenous glucose provocations. *J Clin Pharmacol* **2007,** *47* (9), 1159-1171.

103. Nagamatsu, S., TIRF microscopy analysis of the mechanism of insulin exocytosis. *Endocr J* **2006,** *53* (4), 433-440.

104. Ohara-Imaizumi, M.,Fujiwara, T.,Nakamichi, Y.,Okamura, T.,Akimoto, Y.,Kawai, J.,Matsushima, S.,Kawakami, H.,Watanabe, T.,Akagawa, K.,Nagamatsu, S., Imaging analysis reveals mechanistic differences between first- and second-phase insulin exocytosis. *J Cell Biol* **2007,** *177* (4), 695-705.

105. Krudys, K. M.,Dodds, M. G.,Nissen, S. M.,Vicini, P., Integrated model of hepatic and peripheral glucose regulation for estimation of endogenous glucose production during the hot IVGTT. *Am J Physiol Endocrinol Metab* **2005,** *288* (5), E1038-1046.

106. Agbaje, O. F.,Luzio, S. D.,Albarrak, A. I.,Lunn, D. J.,Owens, D. R.,Hovorka, R., Bayesian hierarchical approach to estimate insulin sensitivity by minimal model. *Clin Sci (Lond)* **2003,** *105* (5), 551-560.

107. Andersen, K. E.,Hojbjerre, M., A population-based Bayesian approach to the minimal model of glucose and insulin homeostasis. *Stat Med* **2005,** *24* (15), 2381-2400.

108. Hoffman, W., A rapid photoelectric method for determination of glucose in blood and urine. *J Biochem* **1937,** *120*, 51-55.

109. Soeldner, J. S.,Slone, D., Critical variables in the radioimmunoassay of serum insulin using the double antibody technic. *Diabetes* **1965,** *14* (12), 771-779.

110. Sherwin, R. S.,Kramer, K. J.,Tobin, J. D.,Insel, P. A.,Liljenquist, J. E.,Berman, M.,Andres, R., A model of the kinetics of insulin in man. *J Clin Invest* **1974,** *53* (5), 1481-1492.

111. Spiegelhalter D J, B. N. G., Carlin B P and van der Linde A Bayesian measures of model complexity and fit (with discussion). *J. Roy. Statist.* **2002,** (64), 583-640.

112. Burnham, K. P.,Anderson, D. R., *Model selection and multimodel inference : a practical information-theoretic approach*. 2nd ed.; Springer: New York, 2002; p xxvi, 488 p.

113. Hastings, W. K., Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **1970,** *57(1)*, 1970.

114. *BlackBox Oberon microsystems Inc.*, 2001.

115. Andrew, G.,Donald, B. R., Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* **1992,** *7*, 457-472.

116. Team, R. D. C., R: A Language and Environment for Statistical Computing. *R Fondation for Statistical Computing* **2007**.

117. Hoffman, W. S., A rapid photoelectric method for determination of glucose in blood and urine. *J Biochem* **1937,** *120*, 51-55.

118. Xiang, A. H.,Peters, R. K.,Kjos, S. L.,Marroquin, A.,Goico, J.,Ochoa, C.,Kawakubo, M.,Buchanan, T. A., Effect of pioglitazone on pancreatic beta-cell function and diabetes risk in Hispanic women with prior gestational diabetes. *Diabetes* **2006,** *55* (2), 517-522.

119. Cederholm, J.,Wibell, L., Insulin release and peripheral sensitivity at the oral glucose tolerance test. *Diabetes Res Clin Pract* **1990,** *10* (2), 167-175.

120. Matsuda, M.,DeFronzo, R. A., Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. *Diabetes Care* **1999,** *22* (9), 1462-1470.

121. Van Cauter, E.,Mestrez, F.,Sturis, J.,Polonsky, K. S., Estimation of insulin secretion rates from C-peptide levels. Comparison of individual and standard kinetic parameters for C-peptide clearance. *Diabetes* **1992,** *41* (3), 368-377.

122. National Diabetes Fact Sheet 2007, The American Diabetes Association. *http://www.diabetes.org*.

123. Economic costs of diabetes in the U.S. In 2007. *Diabetes Care* **2008,** *31* (3), 596-615.

124. DeFronzo, R. A.,Ferrannini, E., Insulin resistance. A multifaceted syndrome responsible for NIDDM, obesity, hypertension, dyslipidemia, and atherosclerotic cardiovascular disease. *Diabetes Care* **1991,** *14* (3), 173-194.

125. Del Prato, S.,Tiengo, A., The importance of first-phase insulin secretion: implications for the therapy of type 2 diabetes mellitus. *Diabetes Metab Res Rev* **2001,** *17* (3), 164-174.

126. Robert K. Danish, M. a. B. B. W., BSN, RN, CDE Rapid Progression From Pre-diabetes to Severely Ill Diabetes While Under "Expert Care": Suggestions for Improved Screening for Disease Progression. *Diabetes Spectrum* **2005,** *18* (4), 229-239.

127. Katz, A.,Nambi, S. S.,Mather, K.,Baron, A. D.,Follmann, D. A.,Sullivan, G.,Quon, M. J., Quantitative insulin sensitivity check index: a simple, accurate method for assessing insulin sensitivity in humans. *J Clin Endocrinol Metab* **2000,** *85* (7), 2402-2410.

128. Haffner, S. M.,Stern, M. P.,Mitchell, B. D.,Hazuda, H. P.,Patterson, J. K., Incidence of type II diabetes in Mexican Americans predicted by fasting insulin and glucose levels, obesity, and body-fat distribution. *Diabetes* **1990,** *39* (3), 283-288.

129. Eknoyan, G., Adolphe Quetelet (1796-1874)--the average man and indices of obesity. *Nephrol Dial Transplant* **2008,** *23* (1), 47-51.

130. Kadowaki, T.,Miyake, Y.,Hagura, R.,Akanuma, Y.,Kajinuma, H.,Kuzuya, N.,Takaku, F.,Kosaka, K., Risk factors for worsening to diabetes in subjects with impaired glucose tolerance. *Diabetologia* **1984,** *26* (1), 44-49.

131. Charles, M. A.,Fontbonne, A.,Thibult, N.,Warnet, J. M.,Rosselin, G. E.,Eschwege, E., Risk factors for NIDDM in white population. Paris prospective study. *Diabetes* **1991,** *40* (7), 796-799.

132. Weyer, C.,Tataranni, P. A.,Bogardus, C.,Pratley, R. E., Insulin resistance and insulin secretory dysfunction are independent predictors of worsening of glucose tolerance during each stage of type 2 diabetes development. *Diabetes Care* **2001,** *24* (1), 89-94.

133. Lin, C. W.,Veng-Pedersen, P., Analysis of PK/PD risk factors for development of type 2 diabetes in high risk population using Bayesian analysis of glucose-insulin kinetics. *J Pharmacokinet Pharmacodyn* **2009,** *36* (5), 421-441.

134. Kadish, A. H.,Litle, R. L.,Sternberg, J. C., A New and Rapid Method for Determination of Glucose by Measurement of Rate of Oxygen Consumption. *Clinical Chemistry* **1968,** *14* (2), 116-119.

135. Hastie, T.,Tibshirani, R., *Generalized additive models*. Chapman & Hall/CRC: Boca Raton, Fla., 1999; p xv, 335 p.

136. Gelman, A., Prior distributions for variance parameters in hierarchical models(Comment on an Article by Browne and Draper). *Bayesian Analysis* **2006,** *1* (3), 515-533.

137. Gelman, A.,BR., D., Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* **1992,** *7*, 457-472.

138. Spiegelhalter, D. J.,Best, N. G.,Carlin, B. R.,van der Linde, A., Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **2002,** *64*, 583-616.

139. *R: A Language and Environment for Statistical Computing*, Vienna, Austria, 2009.

140. Bleich, D.,Jackson, R. A.,Soeldner, J. S.,Eisenbarth, G. S., Analysis of metabolic progression to type I diabetes in ICA+ relatives of patients with type I diabetes. *Diabetes Care* **1990,** *13* (2), 111-118.

141. Polonsky, K. S.,Sturis, J.,Bell, G. I., Seminars in Medicine of the Beth Israel Hospital, Boston. Non-insulin-dependent diabetes mellitus - a genetically programmed failure of the beta cell to compensate for insulin resistance. *N Engl J Med* **1996,** *334* (12), 777-783.

142. Harris, M. I.,Klein, R.,Welborn, T. A.,Knuiman, M. W., Onset of NIDDM occurs at least 4-7 yr before clinical diagnosis. *Diabetes Care* **1992,** *15* (7), 815-819.

143. Goldstein, B. J., Insulin resistance as the core defect in type 2 diabetes mellitus. *Am J Cardiol* **2002,** *90* (5A), 3G-10G.

144. Pratley, R. E.,Weyer, C., Progression from IGT to type 2 diabetes mellitus: the central role of impaired early insulin secretion. *Curr Diab Rep* **2002,** *2* (3), 242-248.

145. Bonate, P. L., The effect of collinearity on parameter estimates in nonlinear mixed effect models. *Pharm Res* **1999,** *16* (5), 709-717.

146. Nichols, G. A.,Hillier, T. A.,Brown, J. B., Progression from newly acquired impaired fasting glusose to type 2 diabetes. *Diabetes Care* **2007,** *30* (2), 228-233.

147. Lillioja, S.,Mott, D. M.,Spraul, M.,Ferraro, R.,Foley, J. E.,Ravussin, E.,Knowler, W. C.,Bennett, P. H.,Bogardus, C., Insulin resistance and insulin secretory dysfunction as precursors of non-insulin-dependent diabetes mellitus. Prospective studies of Pima Indians. *N Engl J Med* **1993,** *329* (27), 1988-1992.

148. Sicree, R. A.,Zimmet, P. Z.,King, H. O.,Coventry, J. S., Plasma insulin response among Nauruans. Prediction of deterioration in glucose tolerance over 6 yr. *Diabetes* **1987,** *36* (2), 179-186.

149. Gupta, N.,Al-Huniti, N. H.,Veng-Pedersen, P., Individualized pharmacokinetic risk assessment for development of diabetes in high risk population. *Diabetes Res Clin Pract* **2007,** *78* (1), 93-101.

150. A.H. Kadish, R. L. L., and J.C. Sternberg, A New and Rapid Method for the Determination of Glucose by Measurement of Rate of Oxygen Consumption *Clinical Chemistry* **1968,** *14*, 116-131.

151. McCulloch, D. K.,Bingley, P. J.,Colman, P. G.,Jackson, R. A.,Gale, E. A., Comparison of bolus and infusion protocols for determining acute insulin response to intravenous glucose in normal humans. The ICARUS Group. Islet Cell Antibody Register User's Study. *Diabetes Care* **1993,** *16* (6), 911-915.

152. Boston, R. C.,Stefanovski, D.,Moate, P. J.,Sumner, A. E.,Watanabe, R. M.,Bergman, R. N., MINMOD Millennium: a computer program to calculate glucose effectiveness and insulin sensitivity from the frequently sampled intravenous glucose tolerance test. *Diabetes Technol Ther* **2003,** *5* (6), 1003-1015.

153. WHO expert committee. Diabetes Mellitus First Report. Technical Report 310. WHO, Geneva. **1965**.

154. Ten-year follow-up report on Birmingham Diabetes Survey of 1961. Report by the Birmingham Diabetes Survey Working Party. *Br Med J* **1976,** *2* (6026), 35-37.

155. O'Sullivan, J. B.,Mahan, C. M., Blood sugar levels, glycosuria, and body weight related to development of diabetes mellitus. The Oxford epidemiologic study 17 years later. *JAMA* **1965,** *194* (6), 587-592.

156. Five-year follow-up report on the Birmingham diabetes survey of 1962. Report by the Birmingham Diabetes Survey Working Party. *Br Med J* **1970,** *3* (5718), 301-305.

157. Jarrett, R. J.,Keen, H.,Fuller, J. H.,McCartney, M., Worsening to diabetes in men with impaired glucose tolerance ("borderline diabetes"). *Diabetologia* **1979,** *16* (1), 25-30.

158. World Health Organization. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation. WHO, Geneva. **2006**.

159. Spiegelhalter, D. J.,Thomas, A.,Best, N. G. *WinBUGS User Manual Version 1.4.*, 2003.

160. McCance, D. R.,Hanson, R. L.,Charles, M. A.,Jacobsson, L. T.,Pettitt, D. J.,Bennett, P. H.,Knowler, W. C., Comparison of tests for glycated haemoglobin and fasting and two hour plasma glucose concentrations as diagnostic methods for diabetes. *BMJ* **1994,** *308* (6940), 1323-1328.

161. WHO expert committee. Diabetes Mellitus Second Report. Technical Report 646. WHO, Geneva. **1980**.