2010
2011

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Longitudinal analysis of an AB/BA cross-over study in diabetes type 1

Promotor :
Prof. dr. Cristina SOTTO

Promotor :
Mr. JAVIER CASTAÑEDA

Maeregu Woldeyes Arsido

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

Maastricht University

universiteit hasselt
UNIVERSITEIT VAN DE TOEKOMST

2010
2011

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Longitudinal analysis of an AB/BA cross-over study in diabetes type 1

Promotor :
Prof. dr. Cristina SOTTO

Promotor :
Mr. JAVIER CASTAÑEDA

Maeregu Woldeyes Arsido

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

Maastricht University

universiteit hasselt
UNIVERSITEIT VAN DE TOEKOMST

Longitudinal Analysis of an AB/BA Cross-Over Study in Diabetes Type 1

BY

Maeregu Woldeyes Arsido

Internal Supervisor

Prof. dr. SOTTO Cristina

External Supervisor

CASTAÑEDA Javier

Thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Biostatistics

September, 2011

Certification

This is to certify that this report was written by Maeregu Woldeyes Arsido under our supervision.

…… Date ……

Maeregu Woldeyes Arsido Student

…… Date ……

Dr. SOTTO Cristina Internal Supervisor

…… Date ……

CASTAÑEDA Javier External Supervisor

Acknowledgements

First and foremost, I thank the almighty God for being with me in all aspects to realize my dream.

This report has benefited vastly from inspiring communication and joint work with my supervisors. I am grateful to Prof. dr. Sotto Cristina for her valuable guidance, intensified explanations and encouragement. I have learnt much from working with you. Thank you so much. I am heartily thankful to my external supervisor Castaneda Javier whose guidance and support enabled me to understand the subject. I have learnt much from your practical experiences. I never forget your brotherhood approach.

I wish to express my gratitude to VLIR scholarship for the financial support. I am thankful all my professors at center for statistics, Hasselt university. I would also thank my home country friends and relatives in Hasselt. Especially, I thank Birhanu Teshome a final year phd student at Hasselt university for his valuable support. Finally, I would like to thank all my family for their continued encouragement and support.

Abstract

Background*:* Type 1 diabetes is a chronic disease that occurs when the pancreas does not produce enough insulin to properly control blood sugar levels. There are several methods to deliver insulin to a person with type 1 diabetes. Pump therapy is well established as a "gold standard" for insulin delivery, offering improvements over multiple daily injections. In this study, we focus on investigating the effect of real-time continuous glucose monitoring (CGM).

Methods: Subjects were randomized to Sensor ON or Sensor OFF arms for 6 months, after a 1 year run-in period. Following a 4 month washout period, the subjects crossed over the other study arm for 6 months. The primary end point was Glycated hemoglobin (HbA1c). The linear mixed model in the particular case of two-treatment two-period cross-over design with repeated measurements within each period for each subject was employed. Dependence among such measurements, within and between periods was handled by modelling the covariance structure.

Results: An overall significant treatment effect was found in favour of the sensor therapy. The evolution of treatment effect over time is not different between the two periods.

Conclusion: The results established adding real-time continuous glucose monitoring (CGM) to the existing pump therapy can enable better metabolic control.

KEY WORDS: Autocorrelation; Cross-Over design; Diabetes; Linear Mixed Model; period; Pump therapy; Sensor therapy.

Table of Contents

List of Tables

List of Figures

1. Introduction

Type 1 diabetes is a chronic disease that occurs when the pancreas does not produce enough insulin to properly control blood sugar levels. The disease is caused by a complete lack of insulin. Since people with type 1 diabetes can't make their own insulin, the discovery of insulin in 1921 is considered as an event that changed diabetes mellitus from a death sentence to a survivable disease (Elaine and Katja 2010). The disease is difficult to control and patients typically develop long-term vascular and neural problems. According to World Health Organization (WHO), there are 220 million people in the world are affected with diabetes in the year 2011 and this is projected to increase to 366 million by 2030. More than 80 percent of diabetes-related deaths occur in low and middle-income countries. Currently, over 90 percent of the cases are known to be type 2 diabetes, which is caused by inefficient use of insulin, excessive body weight and physical inactivity.

The major problem in type 1 diabetes patients is the maintenance of the amount of blood glucose concentration within physiological limits. Hyperglycemia (high blood glucose) is the culprit behind all the complications related to type 1 diabetes. The closer to the normal blood glucose level are held, the less likely complications are. To maintain blood glucose level, type 1 diabetes patients must take insulin every day. There are several methods to deliver insulin to a person with type 1 diabetes. The most common methods include syringes, insulin pens, insulin pumps and continuous glucose monitoring (CGM). The first two methods essentially do the same thing that is insulin is injected multiple times a day. But, insulin pens gave advantages over syringe method with respect to ease of handling and accuracy.

An alternative to insulin injections is the insulin pump, sometimes called Continuous Subcutaneous Insulin Infusion (CSII). The pump delivers a continuous low (basal) dose of insulin through a cannula (a flexible plastic tube), which attaches to the body through a small needle inserted into the skin. An advantage of the CSII method is greater flexibility with meals, exercise, and daily schedule. Though CSII offers improvements over multiple daily insulin injections, it suffers from severe shortcomings. Amongst them, infections at the site of needle insertions can occur and stoppage of pump for any reason can lead to diabetic ketoacidosis (DKA) in less than half a day. Therefore, blood sugar must be monitored on a regular basis to deliver insulin.

Continuous glucose monitors (CGM) allow people with diabetes to see their glucose levels continuously using a subcutaneous sensor, a transmitter, and a receiving device (Gilliam 2009). Accessibility of glucose values enables recognition of previously undetected glucose levels, direction and rate of change, and glucose trends. Recently, Raccah et al (2009) also demonstrated that in patients previously nave to CSII, are switched to insulin pump, those with a sensor augmented pump system had a better opportunity for haemoglobin A1c (HbA_{1c}) improvements. While glucose sensors can talk directly to insulin pumps, problems with automatically calculating the amount of insulin to dispense mean patients must still make these decisions. In this way such combined devices still fall short of being a true "artificial pancreas." A number of alternative insulin delivery methods exist. Insulin inhalers are now approved for use and insulin patches may be on the horizon.

This study focuses on the comparison of insulin pump (CSII) with real-time continuous glucose monitoring (CGM). The general objective of the study is to compare the efficacy of the two therapeutic methods of treatment on type 1 diabetic patients. Specifically, we investigate a direct comparison of the benefit of Sensor augmented pump therapy compared to pump therapy alone using two measurements on each patient at two different treatment periods. The evolution of treatment effect can be studied using all the repeated measurements within each treatment period. Further, specific time point comparisons may also be interest.

The report is organized as follows. In Section 2, the data set used in the analyses is introduced. Section 3 discusses the methods and design of the study. We provide a perspective on the cross-over design, reviewing the main developments and related issues. Linear mixed model methodology in the context of cross-over design is formulated in Section 3.2, followed by method of sensitivity analysis for missing data which is in Section 3.6. Results of the data are illustrated in Section 4, paying special attention to modelling the covariance structure. Finally a general discussion and concluding remarks are given in Section 6.

2. The Data Set

The data set considered in this study was obtained from a randomized controlled cross-over trial conducted in 2011 on Sensor–augmented insulin pump efficacy in type 1 diabetic patients. The study was conducted in seven European countries (Austria, Denmark, Netherlands, Luxemburg, Italy, Spain and Slovenia). The data set contains information from two patient populations. The first one was based on the *intention-to-treat* (ITT) approach and the second population was based on the *per protocol* (PP) approach. The ITT principle provides unbiased comparisons among the treatment groups, since none of the patients is excluded and the patients are analyzed according to the randomization scheme. The PP approach restricts the comparison of the treatments to the ideal patients, that is, those who adhered perfectly to the clinical trial instructions as stipulated in the protocol. There are 153 randomized patients in total corresponding to the ITT population while 90 patients were found to satisfy the PP principle in the two treatment periods. Patients were visited three times within a period and the primary end point was measured at each visit. The design of the study is presented in Section 3.1. Further, baseline variables and related information are recorded. A more elaborative definition of these variables is given below.

2.1 Primary End Point and Baseline Covariates

The primary end point was Glycated hemoglobin (HbA1c). It is a form of hemoglobin which is measured primarily to identify the average plasma glucose concentration over prolonged period of time. HbA1c serves as a marker for average blood glucose levels over the previous months prior to the measurement. According to American Diabetes Association, HbA_{1c} greater than or equal to 48 mill mol per litter ($\geq$6.5%) serves as criterion for the diagnosis of diabetes. In diabetes mellitus, higher amounts of HbA_{1c} indicate poorer control of blood glucose levels. The available treatments are Sensor augmented pump therapy and pump therapy alone. The two treatments are given for each patient in two different periods (Section 3.1). The baseline covariates include: baseline HbA1c measured before period one, Diabetes duration at randomization (Diabdur) measured in years, Diabetes related complications (DRCO) categorized as no complication or complication, type of insulin at randomization (Instype) categorized as type 1 or type 2, Age group (Agegroup) categorized as pediatric or adult.

3. Study Design and Methods

3.1 AB/BA Cross-Over Design

The study design was based on a cross-over clinical design on subjects suffering from type 1 diabetes disease. In a cross-over design each subject is randomized to a sequence of treatments, with the aim of comparing the effect of individual treatment assignment. Cross-over designs are most suited to investigate treatment differences for ongoing or chronic diseases. Senn (1993) and Pocck (1983) discussed that cross-over designs are highly effective in a disease that show fairly stable condition. The feature that distinguishes the cross-over design from parallel group design is that measurements on different treatments are obtained from each subject. This feature brings with it the advantage that the treatments are compared 'withinsubjects', treatment differences can be obtained without subjective effect.

Although the use of repeated measures on the same subject with different treatment brings with it great advantages, Jones and Kenward (2003) emphasised potential disadvantages. The most known disadvantage for cross-over clinical trials is the presence of a *carry-over* effect. Carry-over is the possibility that the effect of a treatment given in one period might be present at the start of the following period. Though carry-over effect arises in a number of ways, all their effect is to bias the comparative effect of individual treatments. To deal with carry-over, a *wash-out period* is typically used to allow the active effects of a treatment given in one period to be washed out of the body before each subject begins the next period of treatment. In addition to carry-over, *drop-outs* are disadvantages of cross-over trials. For more explanation on these effects we refer to Senn (1993), Jones and Kenward (2003).

The simplest and the most common cross-over design is the *AB/BA* cross-over design. In an AB/BA cross-over design, patients are allocated at random to receive in a first period, either treatment A followed by treatment B in a second period, or treatment B followed by treatment A. Each subject serves as his or her own control, assuming that study conditions are the same from one observation period to the next. We introduce the AB/BA cross-over design with application of therapeutic treatment on type 1 diabetic patients. To give a consistent symbolic representation of the treatments, we use Sensor augmented pump therapy (Sensor ON) as treatment A and insulin pump therapy (Sensor OFF) as treatment B. Based on this representation, sequence AB means 'ONOFF' and sequence BA means 'OFFON'.

The total study duration for the patient was 17 months, including a run-in period, two 6-month treatment periods, and a four month washout period. Subjects who satisfied the initial screening criteria entered a one month run-in period. Eligible patients were randomized to receive either 'Sensor ON' or 'Sensor OFF' for six months of the first period. The subjects then switched over to the alterative therapy for a further six months of the second period. Of the total 185 subjects recruited in the study, 153 were randomized to the treatments: 77 were randomized to sequence ONOFF and 76 were randomized to OFFON sequence. The primary end point HbA1c was measured at three visits at each period.

3.2 Linear Mixed Models

The defining feature of a longitudinal data model is its ability to study changes over time within subjects and changes over time between groups. Cross-over designs in which repeated measurements within treatment periods are an example of such studies. The model can handle these patterns of correlation and variation. First we will describe the standard linear mixed model. Then after, the linear mixed model in the context of cross-over design will be given.

Linear mixed models are often appropriate for representing clustered and therefore dependent data or when data are collected hierarchically or when data are gathered over time on the same individuals. The general form of linear mixed models is given by:

$$\mathrm{Y_i} = \mathrm{X_i}\beta + \mathrm{Z_i}\mathrm{b_i} + \varepsilon_\mathrm{i} \ ,$$

where $\mathrm{Y_i}$ is the n_i- dimensional response vector for subject i. X_i and Z_i are the design matrices of the predictor variables and the random effect variables respectively, β is the vector of fixed effects, $\mathrm{b_i}$ is the vector of random effect parameters and follows $b_i \sim N(0, \mathrm{D})$ and ε_i is an n-dimensional vector of residual components that follows: $\varepsilon_i \sim N(0, \Sigma_\mathrm{i})$. D is a general $(q \times q)$ covariance matrix for random effects with (i, j) element corresponding to $d_{ij} = d_{ji}$ and Σ_i corresponds to $(n_i \times n_i)$ covariance matrix for the error terms. b_i and ε_i are independent. For more explanation, we refer to Verbeke and Molenberghs (2000).

The distinguishing features of cross-over designs are time-changing covariates that include individual treatments and other within-individual covariates might change over time. Because of the problems with carry-over effect, there has been disagreement about the appropriate parameterization of AB/BA design.

One possible general model following Jones and Kenward (1989) can be formulated as:

$$Y_{ijk} = \mu + \lambda_g + \pi_k + \tau_{j[k]} + \text{other.fixed} + s_i + \text{other.random} + \varepsilon_{ijk} \quad \ldots (1)$$

where in the fixed effect part μ is the intercept, λ_g is the carry-over effect of sequence g for g=1, 2, π_k is the effect of period k, $\tau_{j[k]}$ is the treatment effect of j in period k, s_i are independent and identically distributed as $N(0, \sigma_s^2)$ denoting random effect of subject i and ε_{ijk} are independent and identically distributed as $N(0, \sigma^2)$ denoting within subject errors. All random effects are independent of each other. A problematic aspect of this parameterization for the AB/BA design is the inclusion of the carry-over effect (see Section 3.1).

3.3 End-of-Period Analysis

Though measurements were taken at three visits in each period, we can analyze the last measurement of period one and the last measurement of period 2 to study the treatment effect. The primary end point, HbA1c is a continuous variable which is taken to be normally distributed (Appendix B Figure 8). A period difference was calculated for each subject. To construct a statistical model, we assume that Y_{ijk} is the response variable that can be represented by a linear mixed model of the form:

$$Y_{ijk} = \mu + \pi_k + \tau_{j[k]} + S_i + \varepsilon_{ijk} ,$$

where each parameter had the same meaning as in model 1 in Section 3.2. The representation of the model did not include carry-over effect, we hope that a four month wash-out period employed in the design possibly avoid the carry-over effect. The interaction between period and treatment was also avoided, since interaction of treatment with period intrinsically aliased with carry-over effect (Jones and Kenward 2003). In addition, such an interaction may emanate from subjects being affected by some factors other than treatment, and/or when the effect of a treatment level might depend on the current state of the subjects (Senn 1993). The subject effect introduced in the model, S_{jk} are declared as random to incorporate the information from subjects on the treatment comparisons. Brown and Kempton (1994) suggested that additional information on treatment comparisons can be recovered from between patients and may be used to increase the precision of the estimate. For estimation purpose, restricted maximum likelihood (REML) can be used.

3.4 Inference: Likelihood Ratio Tests (LRT)

The likelihood ratio test (LRT) is a statistical test used to compare hierarchically nested models. A relatively more complex model is compared to a simpler model to see if it fits a particular data set significantly better. In linear mixed model, a likelihood ratio can be derived for fixed effects and variance components. For variance components, LRT based on restricted maximum likelihood can be derived for comparing nested models with different covariance structures. Valid LRT are also obtained under maximum likelihood estimates. Suppose that the null hypothesis of interest is now given by $H_0: \alpha \in \Theta_{\alpha,0}$, for some subspace $\Theta_{\alpha,0}$ of the parameter space Θ_α of the variance components α. Let L_{REML} denote the restricted maximum likelihood (REML) likelihood function and $-2ln\lambda_N$ be the likelihood ratio test statistic which is defined as:

$$-2ln\lambda_N = -2ln\left[\frac{L_{REML}(\hat{\theta}_{REML,0})}{L_{REML}(\hat{\theta}_{REML})}\right],$$

where $\hat{\theta}_{REML,0}$ and $\hat{\theta}_{REML}$ are the maximum likelihood estimates obtained from maximizing L_{REML} over $\Theta_{\alpha,0}$ and Θ_α, respectively. It then under some regulatory conditions, $-2ln\lambda_N$ follows, asymptotically under H_0, a ch-squared distribution with degrees of freedom equal to the difference between the dimension of Θ_α and the dimension of $\Theta_{\alpha,0}$. One of the regulatory conditions under which the ch-square approximation is valid is that H_0 is not in the boundary of the parameter space Θ_α. In case if H_0 is on the boundary of the parameter space Θ_α, the LRT for testing H_0 is often a mixture of chi-squared distribution rather than the classical single chi-squared distribution (Verbeke and Molenberghs 2000).

LRT for fixed effects based on maximum likelihood can be derived for comparing nested models with different mean structures. The LRT based on REML estimation methods is not recommended for model reduction of mean structures as it needs same error contrasts (Verbeke and Molenbergs 2000). Suppose that the null hypothesis of interest is given by $H_0: \beta \in \Theta_{\beta,0}$, for some subspace of $\Theta_{\beta,0}$ of the parameter space Θ_β of the fixed effects β. Let L_{ML} denote the ML likelihood function and $-2ln\lambda_N$ be the likelihood ratio test statistic which is defined as:

$$-2ln\lambda_N = -2ln\left[\frac{L_{ML}(\hat{\theta}_{ML,0})}{L_{ML}(\hat{\theta}_{ML})}\right],$$

where $\hat{\theta}_{ML,0}$ and $\hat{\theta}_{ML}$ are the maximum likelihood estimates obtained from maximizing L_{ML} over $\Theta_{\beta,0}$ and Θ_β, respectively.

It then, $-2ln\lambda_N$ follows, asymptotically under H_0, a ch-squared distribution with degrees of freedom equal to the difference between the dimension p of Θ_β and the dimension of $\Theta_{\beta,0}$.

3.5 Diagnostic Checks for Outliers

A statistical model represents how one thinks the data were generated. Following model specification and estimation, it is of interest to explore the model-data agreement by raising questions such as to what extent the model assumptions are satisfied, to refine the model components if needed and to identify data points or groups of cases particularly influential on the analysis. We are not interested in a model that is either overly stable or overly sensitive. Changes in the data or model components should produce corresponding changes in the model output.

Unlike in linear models, diagnostics in mixed models is not straight forward. Two kinds of residuals can be considered in a conditional and unconditional sense. The marginal residual reflects how a specific profile deviates from the overall population mean. Alternatively, the subject specific residual (conditional residual) measures how much the observed value deviates from subject's own predicted regression line. The estimated random effect $\hat{b}_i$ can also be used as residuals, since they reflect how much a specific profile deviate from the population averaged profile. Further, linear mixed models involve two kinds of covariates, a matrix of fixed effects and a matrix of random effects. Therefore, it is not clear how leverages should be defined, partially because the matrices are not necessarily of the same dimension (Verbeke and Molenberghs 2000). In linear models for uncorrelated data, changes in the fixed effect estimates, residuals, residual sums of squares, and the variance-covariance matrix of the fixed effects can be computed based on the fit to the full data alone. By contrast, in mixed models data points can affect not only the fixed effects but also the covariance parameter estimates on which the fixed-effects estimates depend.

3.5.1 Residual Diagnostics Analysis

A residual is the difference between an observed quantity and its estimated or predicted value. Following Schabenberger (2004) we can define marginal residuals r_m as the difference between the observed data and the estimated (marginal) mean, $r_m = y_i - x_i'\hat{\beta}$. A conditional residual is the difference between the observed data and the predicted value of the observation, $r_{ci} = y_i - x_i'\hat{\beta} - z_i'b_i$. These residuals are not suited for outliers and potentially influential data points, since the true residuals will exhibit correlations.

To account for the unequal variance of the residuals, various studentization are applied. Once the residuals are studentized, it can be compared to ±2 yardstick which remains useful in mixed models.

3.5.2 Influence Diagnostics

Once potential outliers are detected, the next step is to see how influential they are to our analysis. Quantifying influence in mixed models is done by the following procedure. First, the model is fitted to the data and estimates of all parameters are obtained. Second, one or more data points are removed from the analysis and updated estimates of model parameters are computed. Third, based on full and reduced data estimates, quantities of interest are contrasted to determine how the absence of the observations changes the analysis. The influence statistic used for linear mixed model can be overall influence, change in parameter estimates, change in precision estimates, etc. An overall influence statistic measures the change in the objective function being minimized. In linear mixed models fitted by maximum likelihood (ML) or restricted maximum likelihood (REML), an overall influence measure is the likelihood distance which is also referred as likelihood displacement, given as:

$$LD_{(U)} = 2\{l(\hat{\varphi}) - l(\hat{\varphi}_{(U)})\}$$
$$RLD_{(U)} = 2\{l_R(\hat{\varphi}) - l_R(\hat{\varphi}_{(U)})\},$$

Where $\hat{\varphi}$ and $\hat{\varphi}_{(U)}$ are the parameter estimates based on Full data and reduced data respectively. $l(\hat{\varphi})$ is the loglikelihood function based on the full data set at the full data estimates. $l(\hat{\varphi}_{(U)}$ is the loglikelihood function based on the full data at the reduced data estimates. Once global measures suggest that the points in U are influential, the next is to determine the nature of that influence. Where U denote quantities obtained without the observations in the set U. The points can affect the estimates of fixed effects, the estimates of the precision of the fixed effects, the estimates of the covariance parameters, etc. Cook's distance and covariance ratio were used to capture the change in the entire parameter vector and the effect on the precision of the estimate. Large values of Cook's distance indicate that the change in the parameter estimate is large relative to the variability of the estimate. Data points that have a small Cook's distance can still greatly affect hypothesis tests and confidence intervals, if their influence on the precision of the estimates is large. For more explanation on diagnostics in linear mixed models, we refer to Schabenberger (2004).

3.6 Sensitivity Analysis for Missing Data

As we described in Section 2, ITT means that all patients randomly allocated to treatments in clinical trial should be analyzed together as representing that treatments, whether or not they completed, or indeed received that treatments. Verbeke and Molenberghs (2000) suggested that if one is working within a pragmatic setting, the event of dropout, for example, may well be a legitimate component of the response. There are common approaches to analyze data with dropouts. The common methods used to analyze incomplete data are imputation methods such as last observation carried forward (LOCF), multiple imputation and other non-imputation methods such as likelihood-based approaches. The choice of the method depends entirely on the mechanism (s) generating the missing values. Little and Rubin (1987) make important distinctions between different missing value processes.

A completely random (MCAR) dropout corresponds to independence of the dropout process and the unobserved and observed outcome. If the dropout is independent of unobserved outcome conditional on the observed outcome, then the dropout process is missing at random (MAR). A non-random dropout (MNAR) is one in which dropout is dependent on the unobserved outcome. We will fit a model based on LOCF method and then sensitivity analysis would be carried out to investigate whether the chosen methods were appropriate for the missingness mechanism observed in the data. Assume that the density of the data is given by $f(y_i, d_i|\theta, \varphi)$, where the parameter vectors θ and φ describe the measurement and missingness processes, respectively. The missingness mechanisms are based the factorization of the density of the data;

$$f(y_i, d_i|\theta, \varphi) = f(y_i|\theta)\, f(d_i|y_i, \varphi),$$

where the first factor is the marginal density of the measurement process and the second one is the density of the missingness process conditional on the outcomes. This factorization forms the so called selection modelling approach, on which we want to base our sensitivity analysis. In cases where dropout could be related to the unobserved response (MNAR), dropout is no longer ignorable, implying that treatment effect can no longer be tested or estimated without explicitly taking in to account the dropout model.

We assume that the probability for dropout at occasion j $(j = 2, \dots, 6)$, given the subject was still in the study at the previous occasion, follows a logistic regression model, in line with Diggle and Kenward (1994),

$$logit[P(D_i = j \mid D_i \geq d, y_i)] = \varphi_0 + \varphi_1 y_{ij} + \varphi_2 y_{i,j-1},$$

where D_i is a scalar dropout indicator and it is assumed that $d_i \geq 2$. y_{ij} is the current observation predicted from the assumed model, $y_{i,j-1}$ is the previous observation; and φ_1 is the parameter estimate of y_{ij}, φ_2 is the parameter estimate of $y_{i,j-1}$. Following Verbeke and Molenberghs (2000), the model describes a binary outcome conditional on covariates.

4. Results

4.1 Exploratory Data Analysis (EDA)

We start with exploratory data analysis to get more insight in to the structure of the data. Any outstanding features were identified by creating and inspecting graphs. Individual and group profiles were assessed to highlight any patterns relevant to the scientific question and identify unusual individuals or observations. The individual profiles help to understand the general trend over time within subjects.

The individual profiles and average evolution shown in Figure 1 plots HbA1c against time points for ITT population. The first three time points corresponds to period 1 and the other three time points are in period 2. The profiles were obtained from a randomly selected 20 subjects from 153 subjects. Variability between subject increases from period 1 to period 2. For most subjects, the variability of measurements at the beginning of the study was smaller than at the end of the study. Variability between subjects was much greater than variability within subjects. There seems to be some outlier measurements in period 2 as compared to period 1. The average evolution indicates, there seems no difference between treatment groups at the beginning of the study. But over time differences start to appear as subjects receiving 'Sensor ON' treatment showed a downward HbA1c level while, the 'Sensor OFF' treatment group had constant HbA1c level throughout period 1. In the 'washout' period in which no treatments were given, HbA1c increases for 'Sensor ON' group. In period 2, the same pattern can be seen as period 1, in which HbA1c decreases for 'Sensor ON' group. The plot over time may indicate significant interaction between time and treatment.

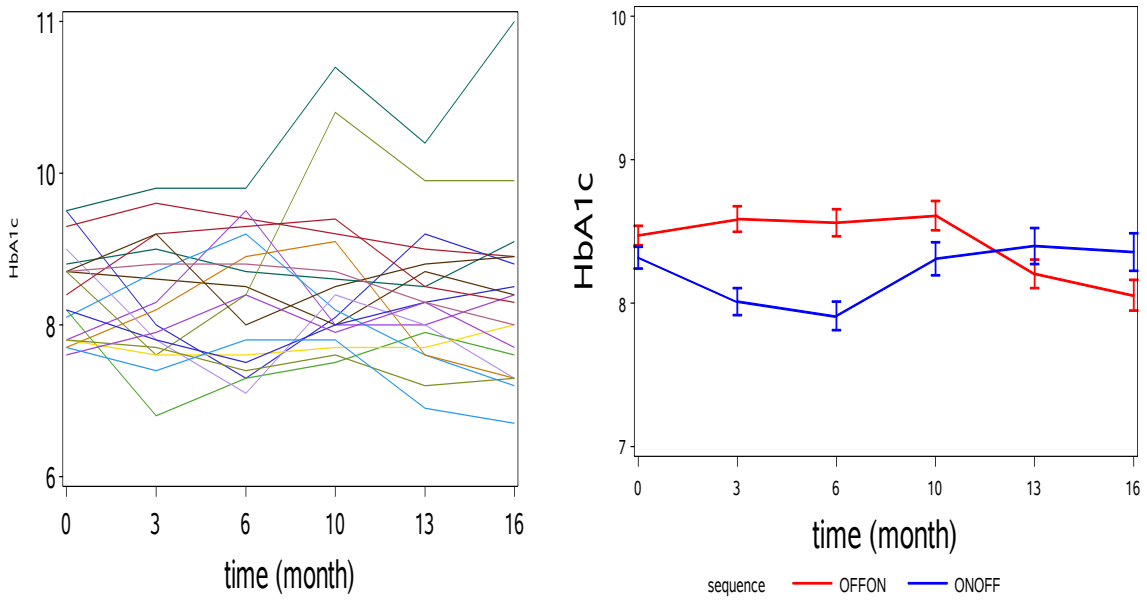


Figure 1: Individual profiles of selected patients (left) and Average evolution (right) for ITT population.

Individual profiles and average evolution by treatment sequence were also plotted for PP population as shown in Figure 2. The general pattern observed was the same as the ITT profile analysis. However, the variability between subjects in PP was more stable than ITT throughout the two periods.

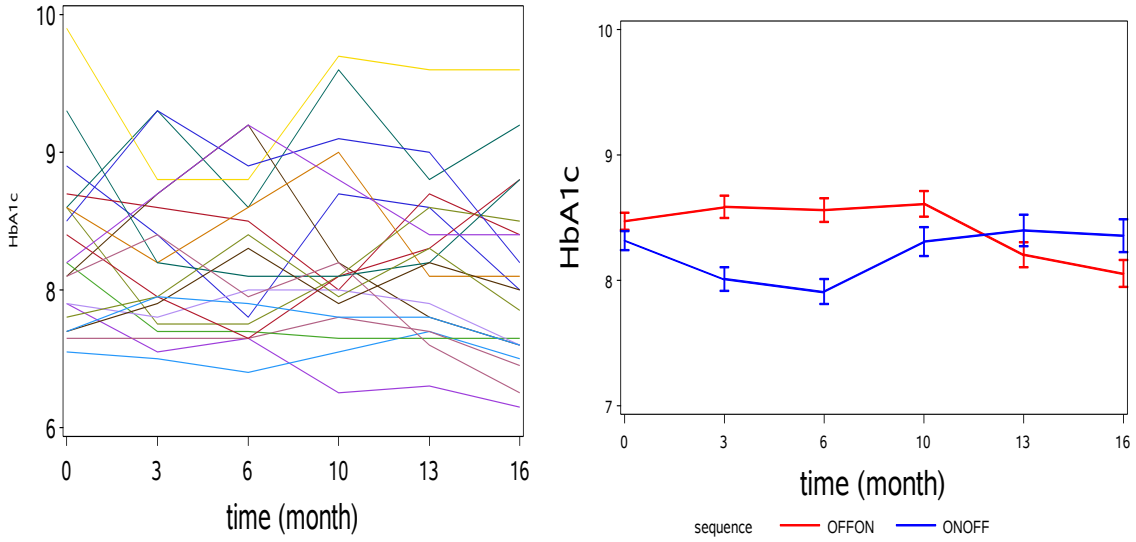


Figure 2: Individual profiles of selected patients (left) and Average evolution (right) for PP.

The summary statistics of the response of interest (HbA1c) at each time point is presented in Appendix A, Table 7. The average HbA1c and the standard deviation in ITT are consistently greater than those of PP population.

Further, the variability increases with time point in both cases. We noted that several measurements are missing for some subjects. We then need to explore the extent of missingness in the data.

4.2 Exploring Missingness Patterns

The problem of dealing with missing values is common in the analysis of longitudinal or repeated measurements data. Indeed, it is common in clinical trials, but the effects are more sever in cross-over design (Senn 1993). As can be seen in Table 1, from the total number of subjects (153) at the time of randomization, 137 subjects or 89.54 percent of the subjects had complete measurements. A total of 16 subjects did not complete the study, from which 14 subjects or 9.15 percent were dropouts. There are six dropouts at early stage of the study in the first period of first time point. We need to account these patterns of missingness to analyze the data under the intent-to-treat (ITT) principle.

Table 1: Overview of missingness patterns and their corresponding frequencies
'O' indicates observed and 'M' indicates missing.

Visit time Point							
Month 1	Month 3	Month 5	Month 6	Month 8	Month 10	Number	Percent
Completers							
O	O	O	O	O	O	137	89.54
Dropouts							
O	O	O	O	O	M	1	0.65
O	O	O	O	M	M	3	1.96
O	O	O	M	M	M	2	1.31
O	O	M	M	M	M	2	1.31
O	M	M	M	M	M	6	3.92
Total dropout						14	9.15
Non-monotone missingness						2	1.31

4.3 End-of-Period Analysis

The results of end-of-period analysis (measurement at the end of each period) are presented in Table 2. The estimated treatment difference for ITT population was -0.4340 with standard error of the treatment difference 0.0593. The P-value < 0.0001 showed that a significant treatment difference was obtained. This signifies that the estimate of HbA1c on Sensor ON group is less than on Sensor OFF group by 0.4340, which indicates that the efficacy of the first treatment (Sensor ON) is better than the second treatment (Sensor OFF). The estimated 95% confidence interval :

$$-0.4340 \pm 1.96(0.0593) = [-0.550, -0.3177].$$

The estimated treatment difference for PP population was -0.5218 with a standard error of 0.0773 and significant p-value =0.0001. Like ITT, the PP analysis also showed that the Sensor augmented therapy improved the biological function. The estimated confidence interval would be:

$-0.5218 \pm 1.96(0.077) = [-0.673, -0.371]$.

As can be seen from the covariance parameters in Table 2, the subject variance (the between) component is larger than the residual variance (the within) component. This is an indication that the cross-over trial design is an appropriate choice for the Sensor augmented therapeutic treatment on type 1 diabetic patients. By declaring patient as a random, the variability of within patient contrast is reduced. For example, for ITT the standard error of the treatment difference for fixed patient effect was 0.1047. The standard error was reduced to 0.0593 by declaring patient as a random; which means the precision of the estimated treatment difference was improved by 48 percent.

Table 2: Parameter Estimates from End-of-period analysis. (S.E = standard Error)

		ITT Population			PP Population		
Effect	Parameter	Estimate	S.E	P-value	Estimate	SD	P-value
Intercept	μ	8.4731	0.08343	0.0001	8.2454	0.0889	0.0001
period	π	-0.0037	0.06395	0.8054	0.0142	0.0773	0.8541
treat	τ	-0.4340	0.06394	0.0001	-0.5218	0.0773	0.0001
Least square Means							
Sensor ON	τ_1	8.0373	0.07393	0.0001	7.7308	0.0780	0.0001
Sensor OFF	τ_2	8.4712	0.07393	0.0001	8.2525	0.0780	0.0001
Variance components							
subject	δ_s^2	0.5601	0.0864		0.2808	0.0651	
residual	σ^2	0.2841	0.0347		0.2655	0.0400	
Overall treatment effect							
	τ_1 - τ_2	-0.4340	0.0593	0.0001	-0.5218	0.0773	0.0001

The variability of the measurements analyzed above can also be verified from Appendix B Figure 7. A randomly selected 20 subjects over the treatment group at the end of each period showed that larger variability between subjects and smaller within variability. It can be seen that subject's HbA1c showed decreasing trend when the subject crossed to the 'Sensor ON' treatment type.

4.4 Linear Mixed Model

In the analysis presented in Section 4.3, measurements from the end of each period were analyzed to compare the treatments. However, the data fall within the realm of continuous longitudinal data and hence can be modelled by use of linear mixed model. Fitting linear mixed models implies that an appropriate mean structure as well as covariance structure needs to be specified. An appropriate covariance model is essential to obtain valid inferences for the parameters in the mean structure.

4.4.1 Modelling the Covariance Structure

In a cross-over design, subjects are administered the treatment in each period in a randomly assigned sequence of treatments. After the subject receives the treatment, a longitudinal series of measurements are taken on each subject on each period. The dependence among such measurements, within and between periods, is important in developing models to account for the structure of the cross-over study. Lindsey (1993) and Byrom et al (1999) argue that where it is possible, modelling stochastic dependence among longitudinal measurements directly by construction of the covariance matrix is more appropriate than doing it indirectly by random coefficients related to time.

In modelling these data we need to take into account two types of covariance patterns among measurements from the same subject: there are dependences (1) among measurements in the same treatment period, and (2) among measurements from different treatment periods. The approach considered is to fit a model with a direct product covariance structure to specify the within subject variance-covariance matrix from two periods, thus to examine the between and within period correlation. Let say the variance of measurement error at first period is σ_1^2 and period two measurement error variance is σ_2^2. The covariance between the two periods is given by σ_{12}^2. The unstructured UN) period covariance (V) can be written of the form:

$$V = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

The within period correlation structure can be handled by first order autoregressive structure AR (1) with parameter ρ.

The AR (1) structure assumes that the covariance between two measurements Y_{ij} and Y_{ik} from the same subject i in the same period at time point $j\ and\ k$ is the form $\rho^{|t_{ij}-t_{ik}|}$ and the within period matrix ($\sum$) can be written as:

$$\sum = \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix},$$

from these two set of matrices, we can construct the within subject matrix, a direct product covariance structure (Galecki 1994).

$$V \otimes \sum = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \otimes \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix} =$$

$$\sum_i = \begin{bmatrix} \sigma_1^2 & \sigma_1^2\rho & \sigma_1^2\rho^2 & \sigma_{12} & \sigma_{12}\rho & \sigma_{12}\rho^2 \\ \sigma_1^2\rho & \sigma_1^2 & \sigma_1^2\rho & \sigma_{12}\rho & \sigma_{12} & \sigma_{12}\rho \\ \sigma_1^2\rho^2 & \sigma_1^2\rho & \sigma_1^2 & \sigma_{12}\rho^2 & \sigma_{12}\rho & \sigma_{12} \\ \sigma_{12} & \sigma_{12}\rho & \sigma_{12}\rho^2 & \sigma_2^2 & \sigma_2^2\rho & \sigma_2^2\rho^2 \\ \sigma_{12}\rho & \sigma_{12} & \sigma_{12}\rho & \sigma_2^2\rho & \sigma_2^2 & \sigma_2^2\rho \\ \sigma_{12}\rho^2 & \sigma_{12}\rho & \sigma_{12} & \sigma_2^2\rho^2 & \sigma_2^2\rho & \sigma_2^2 \end{bmatrix}$$

Thus, the above model assumes the two periods share a common within period correlation as measured by AR (1). Some other within period covariance structure can also be compared with AR (1), if they improve the fit of the model. We compared the direct product covariance of unstructured with AR (1) and unstructured with compound symmetry (CS). The CS assumes the same covariance between pairs of measurements within period. Table 3 below presents the covariance parameter estimates with AIC (Akaike information criteria) under UN $\otimes$ AR (1) and UN $\otimes$ CS. Smaller AIC indicates a better fit. It can be seen that the AIC for UN $\otimes$ AR (1) is less than with that of UN $\otimes$ CS. The model with UN $\otimes$ AR (1) better fits the data and we use AR (1) to account within period covariance for the remainder of analysis.

The two periods are independent if the covariance between the periods is zero and the matrix V has the form:

$$V = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix},$$

which can be tested by the null hypothesis $H_0: \sigma_{12} = 0.$ The estimate of these parameters in our data can be seen in Table 3. The estimate of the covariance parameter: $\sigma_{12} = 0.0086$ with $P - value = 0.5749$. We can conclude that there is no significant correlation between periods.

Table 3: Covariance parameter estimates under two direct product structures, S.E=standard error

	UN ⊗ AR (1)			UN ⊗ CS		
Effect	Estimate	S.E	P-value	Estimate	S.E	P-value
σ_1^2	0.2522	0.0224	0.0001	0.2377	0.0203	0.0001
σ_{12}	0.0086	0.0153	0.5749	-0.0049	0.0138	0.7214
σ_2^2	0.2754	0.0281	0.0001	0.2456	0.0232	0.0001
ρ	0.4143	0.0487	0.0001	0.2365	0.0528	0.0001
Fit statistic: AIC						
		1491.7			1529.6	

The within subject matrix can be reduced, since the insignificant σ_{12} was removed. The subjects were declared as a random to account for heterogeneity among subjects. Thus, the correlation of all measurements in the two periods can also be handled by random subject effect (σ_s^2). Further, the within subject matrix, discussed above assumes the two period share the same within period AR (1) structure. However, there is no guarantee that these should take a particular simple form. Jones and Kenward (2003) assumed the stability of a conventional cross-over design to imply that the patterns of variances and correlations that we observed in one treatment period will be similar to these in a second treatment period. This is a very strong assumption and we then assume that the covariance structure in each period is different. Later it will be tested whether the two periods are homogenous or not. Now, the AR (1) structure assumes that the covariance between two measurements Y_{ij} and

Y_{ik} from the same subject i is the form $\rho_1^{|t_{ij}-t_{ik}|}$ if the measurements are in period 1 and $\rho_2^{|t_{ij}-t_{ik}|}$ if the measurements are in period 2.

We had two periods and three measurements within each period. Then, the within subject matrix can be presented of the form:

$$\sum_i = \begin{bmatrix} \sigma_1^2+\sigma_s^2 & \sigma_1^2\rho_1+\sigma_s^2 & \sigma_1^2\rho_1^2+\sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 \\ \sigma_1^2\rho_1+\sigma_s^2 & \sigma_1^2+\sigma_s^2 & \sigma_1^2\rho_1+\sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 \\ \sigma_1^2\rho_1^2+\sigma_s^2 & \sigma_1^2\rho_1+\sigma_s^2 & \sigma_1^2+\sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_2^2+\sigma_s^2 & \sigma_2^2\rho_2+\sigma_s^2 & \sigma_2^2\rho_2^2+\sigma_s^2 \\ \sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_2^2\rho_2+\sigma_s^2 & \sigma_2^2+\sigma_s^2 & \sigma_2^2\rho_2+\sigma_s^2 \\ \sigma_s^2 & \sigma_s^2 & \sigma_s^2 & \sigma_2^2\rho_2^2+\sigma_s^2 & \sigma_2^2\rho_2+\sigma_s^2 & \sigma_2^2+\sigma_s^2 \end{bmatrix}$$

4.4.2 Modelling the Mean Structure

In this Section, it is interest to explore the effects of baseline covariates on the response variable $HbA1c$. In cross-over design covariates might not be used if their inclusion did not improve the precision of the treatment effect. Senn (1993) claimed that the baseline measurement before each period was the only genuine covariate that changes with treatment. However, the covariates might be related to the patient's response and then the possibility exists that some of the between-subject variability can be accounted for by the covariate values. In this way the between subject variance might help to increase the precision of treatment effect. We then investigate the effect of each baseline covariate on the patient's $HbA1c$ level in the presence of treatment and period. Let $HbA1c_{ijkl}$ be the measurement on subject i in period j, for j=1, 2, treatment type k, for k=1, 2 at time in month l, for l=0, 3, 6, 10, 13 and 16. We then have a preliminary model as follows:

$$\begin{aligned} HbA1c_{ijkl} = {} & \beta_0 + \beta_1 Treat_k + \beta_2 Period_j + \beta_3 HbA1cB_i + \beta_4 \text{Diabdur}_i + \beta_5 time_l + \\ & \beta_6 DRC + \beta_7 \text{Instype}_i + \beta_8 Agegroup + \beta_9 Treat_k.time_l + \beta_{10} Period_j.time_l + \\ & \beta_{11} Period_j.Treat_k + \beta_{12} Period_j.Treat_k.time_l + S_i + \varepsilon_{ijkl} \quad \ldots\ldots\ldots\ldots\ldots \text{F} \end{aligned}$$

Where β_g, g = 0, . . , 12 are fixed effect parameters, S_i: the random subjects effect and ε_{ijkl} are random error terms. The assumptions for the random terms are ε_{ijkl} are independent and

normally distributed with zero mean and variance ${\sigma^2}_1$ for period 1 and ${\sigma^2}_2$ for period 2. S_i are independent and normally distributed with zero mean and variance σ_s^2.

4.4.3 Model Reduction

The preliminary model (F) in Section 4.3.2 is bound to yield a large number of parameters, hence the need for parsimony. Further, overparametrization of the covariance structure leads to inefficient estimation and potentially poor assessment of standard errors for estimates of the mean response profile (fixed effects), whereas a too restrictive specification invalidates inference about the mean response profile when the assumed covariance structure does not hold (Altham 1984). Since inference for the mean structure depends on covariance structure, reduction of the covariance structure has been done first.

We can test whether we really need random subject components. The null hypothesis to test the need for random subject effect is given by: $H_0: \sigma_s^2 = 0$. Since testing this hypothesis is on the boundary of the parameter space of the alternative hypothesis as we discussed in Section 3.4, asymptotic chi-squared null distribution for the likelihood ratio test statistic is not valid. Therefore a mixture of chi-squared distribution under 0 (model with no random effect) and 1 (model with random subject effect) degree of freedom ($\chi^2_{0:1}$) can be used. The loglikelihood from model under H_0 is equal to 1550.7 and loglikelihood from model with random subject effects is equal to 1493.6. Likelihood ratio test is give by 1550.7 – 1493.6 = 57.1, with $P - value = 0.0001$, which is extremely significant to reject the null hypothesis and we retain the random subject effect in the model.

As discussed in Section 4.3.1, the within covariance structure of the two periods are assumed to be heterogeneous. We then require test for homogeneity of covariance structure across periods. That is, it tests whether varying the covariance parameters by the period effect provides a significantly better fit compared to a model in which different periods share the same parameter. The null hypothesis to be tested is given by: H_0: $\rho_1 = \rho_2 \; and \; \sigma_1^2 = \sigma_2^2$. Based on the restricted maximum likelihood, the likelihood ratio test statistic between the heterogeneous covariance structure model and model under H_0 is equal to 25.6 with 2 degrees of freedom. The corresponding $P - value = 0.0001$ is significant. Hence, there is insufficient evidence to accept the assumption of equal covariance structure among the two periods. We then conclude that the final covariance structure would be: different first order autoregressive correlations for each period and random subject effects component.

The final covariance structure for the model has been selected; the test discussed above becomes available for the fixed effects in the preliminary mean structure. We employed LRT using ML estimation for reduction of mean structures keeping the same covariance structure. The null hypothesis $H_0: \beta_7 = 0$ tests the significance of the baseline covariate, insulin type (Instype). The LRT from the full (F) model in Section 4.4.2 and the same model under H_0 was found to be 2.4 and $P - value = 0.1213$ with 1 degree of freedom, resulting failure to reject H_0, implying no effect of insulin type. The LRT for the baseline covariates age group (Agegroup) and diabetes related complications (DRC) are 3.2 and 1.8 with their corresponding $P - values$ of 0.0765 and 0.1797 respectively each with 1 degree of freedom. There is no significant effect of these covariates either.

The next step was to test period related covariates, this enables us to understand whether effects are changing across period or the same effects can be obtained in each period. Higher order interaction effects were evaluated first then after the main effects were evaluated and removed if they are insignificant. The null hypothesis $H_0: \beta_{12} = 0$, tests whether the evolution of the treatment effect in each period was different. The calculated LRT= 0.2 and $P - value = 0.6547$ with 1 degree of freedom. We do not reject H_0 and conclude that no effect of period by treatment by time. This indicates the treatment evolution in each period is the same. The other period related factor is period by time interaction which can be tested using the null hypothesis $H_0: \beta_{10} = 0$. The calculated LRT= 1.3 and $P - value = 0.2542$ with 1 degree of freedom. We do not reject H_0 and conclude that no effect of period by time interaction. Again we can conclude that the time trend was the same in each period. The null hypothesis $H_0: \beta_{11} = 0$ tests the effect of treatment in each period. The LRT= 26.9 and $P - value = 0.0001$ with 1 degree of freedom. Since this is a significant test, we reject the null hypothesis and conclude that different treatment effect observed in each period. Period by treatment significant mean, we do not test period as a main effect since the effect is apparent from the interaction term. All the other remaining effects are significant and the so obtained final model can be written of the form:

$$HbA1c_{ijkl} = \beta_0 + \beta_1 Treat_k + \beta_2 Period_j + \beta_3 HbA1cB_i + \beta_4 Diabdur_i + \beta_5 time_l + \beta_6 Treat_k.time_l + \beta_7 Period_j .Treat_k + S_i + \varepsilon_{ijkl} .$$

4.4.4 Results from Final Model

The parameter estimates and standard errors from the final model are shown in Table 4. The same procedures were followed for model building and reduction in PP population data. The same type of covariates and covariance structure were obtained. A significant treatment effect was obtained in both cases. Since treatment by time was found to be significant with p-value = 0.0001, the marginal interpretation of the treatment would depend on the interaction term. The parameter estimate of the interaction for ITT was -0.0851 with standard error 0.0124. This implies HbA1c level will reduce over time in subjects receiving 'Sensor ON' therapeutic method. The reduction was in line with end of period analysis in Section 4.3 and can be observed from the average evolution profile by treatment group. The corresponding 95% confidence interval can be estimated as: $-0.0851 \pm 1.96(0.0124) = [-0.1094, -0.0608]$. The estimated period by treatment interaction was -1.0212 with standard error 0.1959. This parameter estimate also implies that average HbA1c levels decrease when the 'Sensor ON' therapeutic method was applied in the first period compared to 'Sensor ON' was given in the second period.

Inference can be made not only over time, but also at specific time points might be interest. For example, from Figure 1 exploratory analysis of HbA1c on 'Sensor ON' treatment group starts to decrease at the first month of the study. At month 3, HbA1c continues to decrease further and it starts to increase after month 6. The interest was to test whether the reduction of HbA1c from month 3 to month 6 is statistically significant. The approximate F-statistic of contrast between the two time points is equal to 1.94 with $P - value = 0.164$. The decrement of HbA1c level from month 3 to month 6 using 'Sensor ON' treatment is not significant. However, the mean difference between the two treatment groups at month 3 (F-statistic= 13.73, $P - value = 0.0002$) and at month 6 (F-statistic= 21.20, $P - value = 0.0001$) were significant. As can be seen in Table 4, the positive estimate of the baseline HbA1c measure indicates that patients who have large HbA1c at baseline will also have large subsequent HbA1c score, which in turn imply worse health condition of the patient. While the negative estimate of diabetic duration (in years) may imply that subjects who developed type 1 diabetes earlier, the less controlled their blood sugar, implies the higher the risk of complications.

The estimated treatment evolution in PP population was -0.0999 with standard error 0.0151. The 95% confidence interval could be: $-0.0999 \pm 1.96(0.0151) = [-0.1295, -0.0703]$. It can be seen that the estimate is greater than the ITT in absolute value although the difference is not large. The PP population is more sensitive to show treatment efficacy. This is in line with Garrett (2003) who questioned the conservative nature of the PP. The author illustrated that there can be occasions where the PP analysis produces an unbiased estimate for a particular model chosen. The more widely accepted principle in the literatures suggest that the most liable analysis of a trial is based on ITT, which consists of considering all randomized patients, regardless of any protocol violations. According to the influential ICH E9 guideline issued in 1998, direct comparison of these two populations was more complicated, since the two principles play different roles. Accordingly, for superiority trials the ITT strategy is used in the primary analysis.

Table 4: Parameter estimates from linear mixed model for ITT and PP (S.E=standard error).

		Intention-to-treat (ITT)			**Per-Protocol (PP)**		
Effect	Parameter	Estimate	S.E	P-value	Estimate	S.E	P-value
Intercept	β_0	2.1217	0.5435	0.0001	3.0028	0.6143	0.0001
Treatment	β_1	0.9771	0.1890	0.0001	1.1733	0.2274	0.0001
period	β_2	0.1639	0.1243	0.1880	0.1193	0.1463	0.4152
Baseline	β_3	0.7541	0.0611	0.0001	0.6479	0.0689	0.0001
Diabetic duration	β_4	-0.0120	0.0034	0.0006	-0.0108	0.0035	0.0034
Time	β_5	0.0124	0.0088	0.1576	-0.0024	0.0106	0.8219
Treatment. Time	β_6	-0.0851	0.0124	0.0001	-0.0999	0.0151	0.0001
Treatment. period	β_7	-1.0212	0.1959	0.0001	-1.2690	0.2219	0.0001
Covariance Parameters							
Subject effect	σ_s^2	0.1390	0.0248		0.0771	0.0212	
Autocorrelation 1	ρ_1	0.1570	0.0765		0.1091	0.1006	
Measurement error 1	σ_1^2	0.2014	0.0169		0.1665	0.0178	
Autocorrelation 2	ρ_2	0.5976	0.0467		0.6299	0.0584	
Measurement error 2	σ_2^2	0.3437	0.0358		0.3203	0.0438	

5. Model Adequacy

The adequacy of the final linear mixed model can be checked for both ITT and PP populations. Influential outlier investigation can be done, but removal of outlying subject opposes the principle of ITT; this was done for PP model (Section 5.1). Results for sensitivity analysis of missing data were presented (Section 5.2) for the ITT model.

5.1 Diagnostics for Outliers

To detect unusual observations we use the range of the studentized residuals. From Appendix B Figure 11, the histogram for marginal and conditional residuals can be seen. The marginal residuals are useful to detect the fixed effect components. Conditional residuals are useful to detect the random effect components. In both cases, residuals are observed outside the yardstick of [-2, 2], which needs further investigation for influence. Results from restricted likelihood distance (RLD), given in Figure 3, indicate that subject with id # Eur0309019 influenced the REML solution more than the other subjects.

Subjects with id # Eur0308022, # Eur0307011, # Eur0309029 and # Eur0306021 had relatively high restricted likelihood distance. Since RLD is an overall influence diagnostic, we require Cook's D and COVRATIO to diagnose specific effects. Subjects with the largest effect on the fixed effects estimates are # Eur0309019, # Eur0307011 and # Eur0306021 Appendix B (Figure 10). Except for subject # Eur03060211, the other subjects have COVRATIO values less than one, implying the precision of the parameters was not affected after the subjects were removed from the analysis. The stability of the precision of estimate was important, since hypothesis tests or confidence intervals about β may not be distorted. Influence diagnostic regarding the covariance parameters are shown in Appendix B (Figure 10). Again the influence of subject # Eur0309019 far exceeds that of other subjects in these data. This is expected since its restricted likelihood distance is substantial, while its impact on the fixed effects was rather moderate. The large value Cook's D of covariance parameters shows this subject's impact on the covariance parameter estimates.

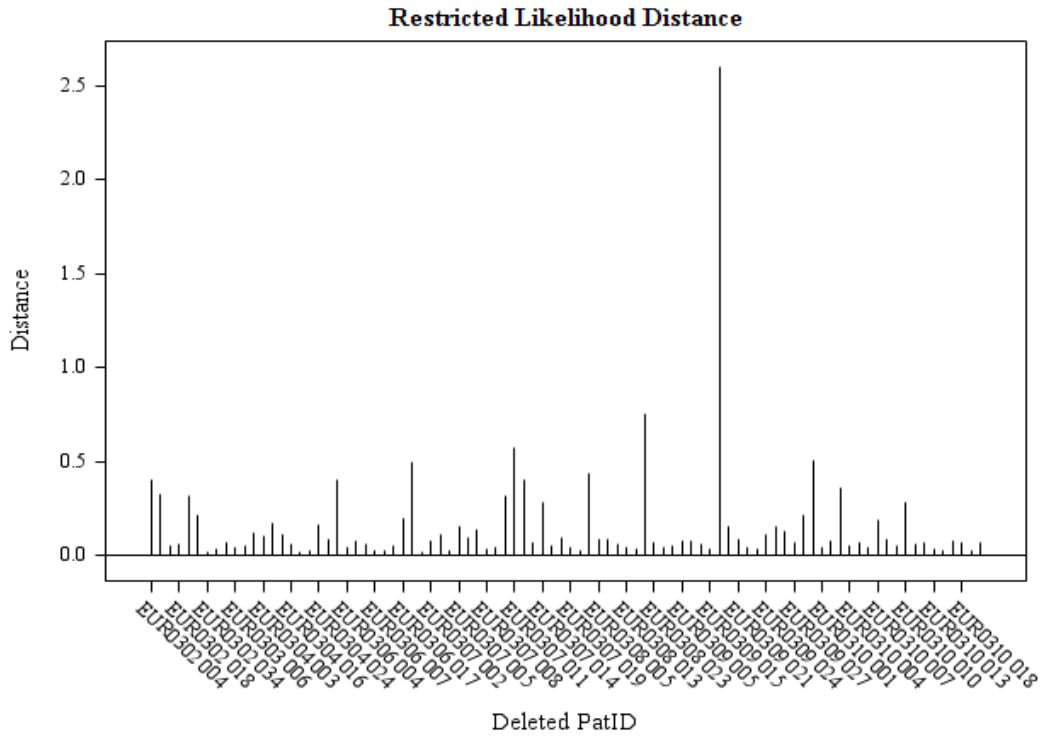
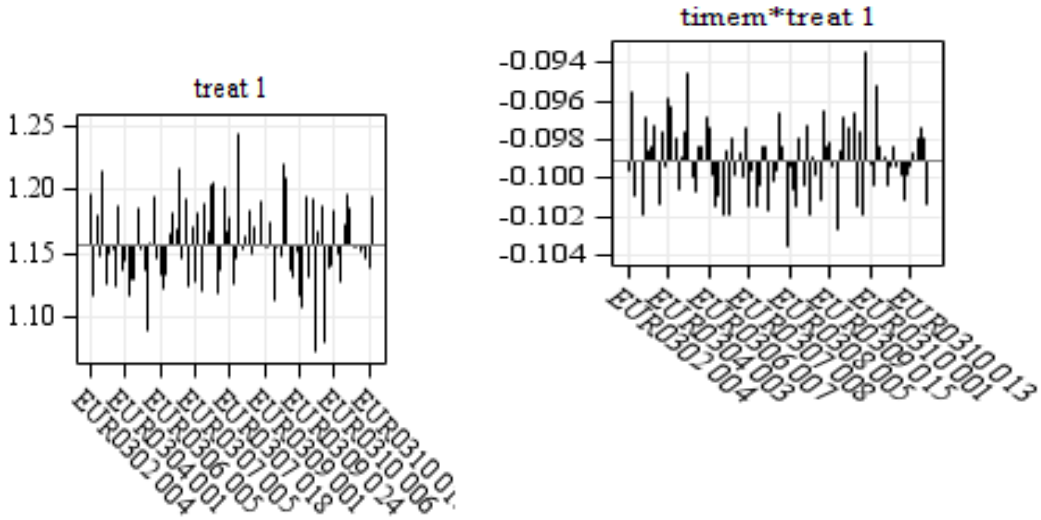


Figure 3: Restricted likelihood distance.

These estimates are not altered extremely by the removal of each subject. The covariance parameter estimates can also be evaluated deleting each subject in turn. The influence plots for the random subject effect and AR (1) for the first period are shown in Figure 4. As expected the suspected subject # Eur0309019 altered the covariance parameter.



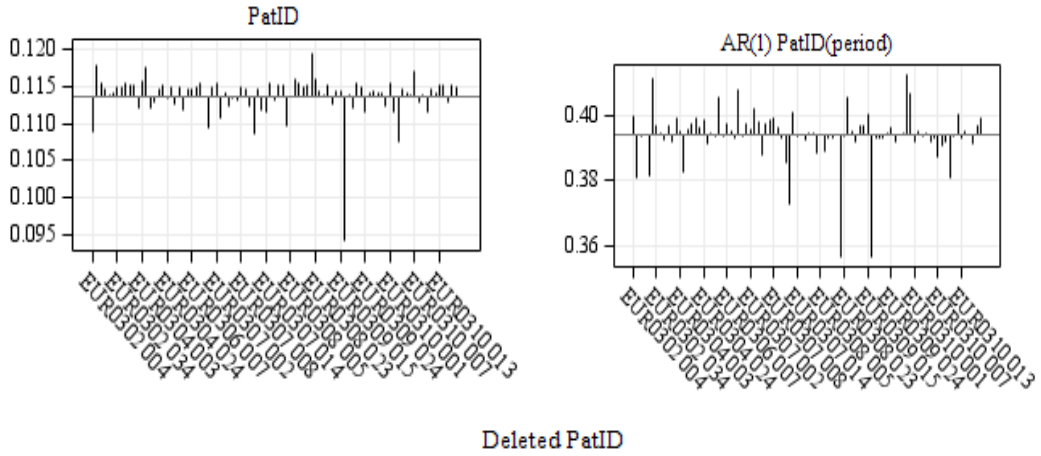
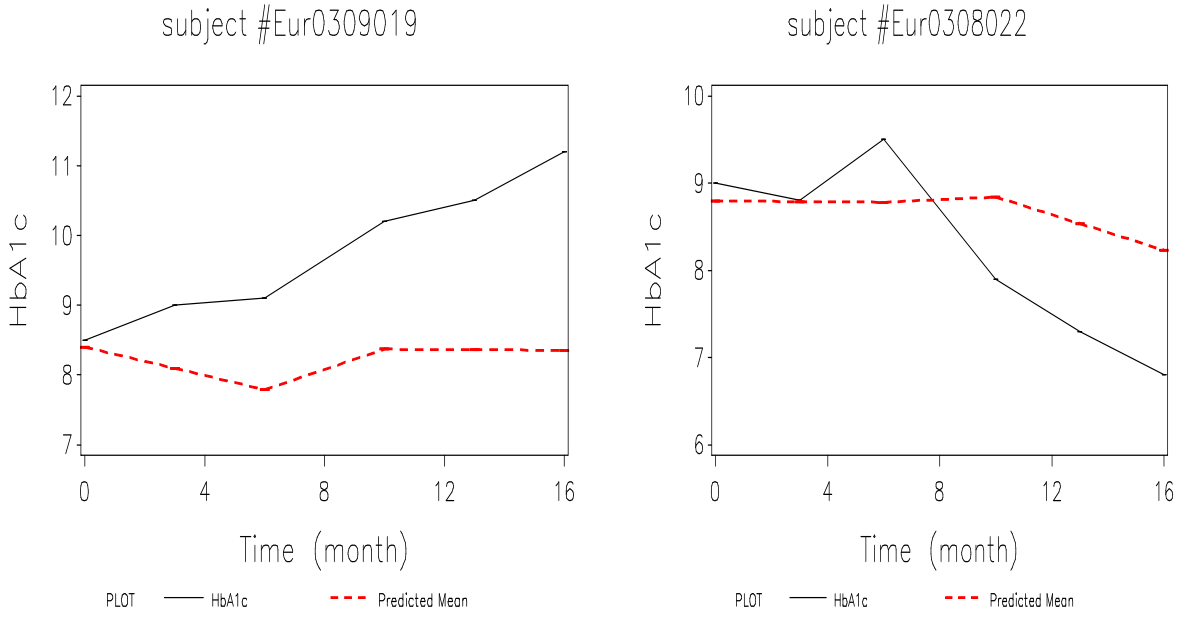


Figure 4: Fixed effect deletion estimates and covariance parameter deletion estimates.

We can see that the three subjects that have highest overall influence measure (RLD) appear to have large values of Cook's distance. Further, subject # Eur0309019 consistently altered the fixed effect solution and the covariance parameters. We then study the observed and predicted profile over time for this subject and the other three subjects that influenced the REML solution. As can be seen from Figure 5, the four subjects predicted profiles are not well fitted.



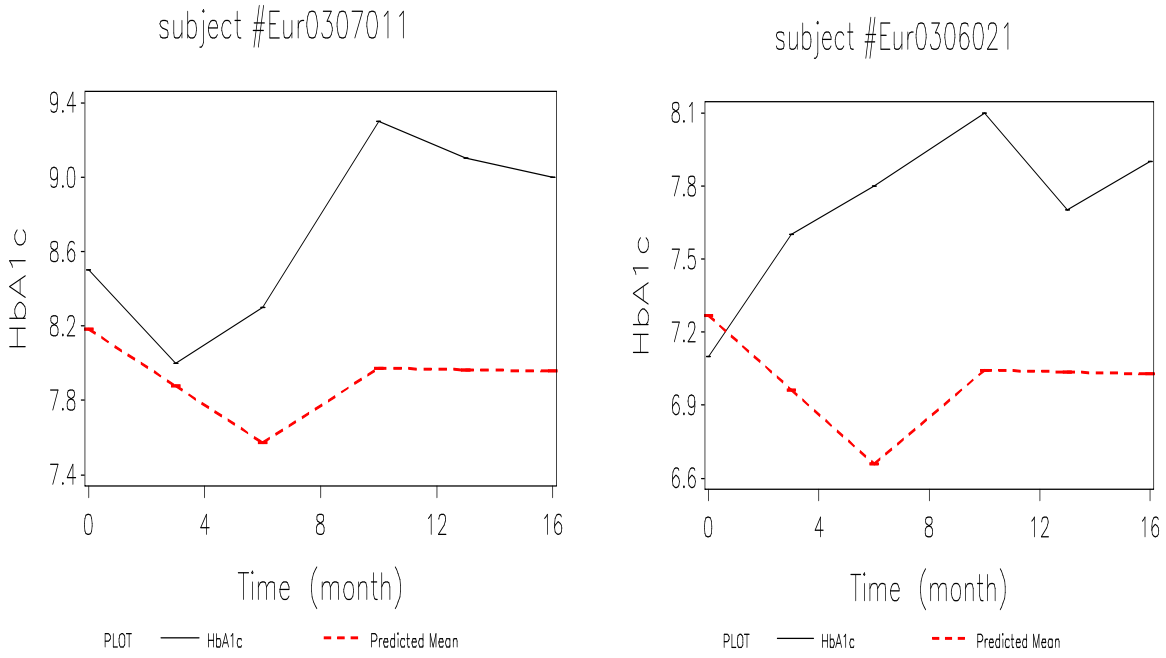


Figure 5: Observed and predicted profile of 4 subjects having highest restricted likelihood distance.

The overall average profile is shown in Appendix B Figure 12 and can be used to observe how the influential subject can influence the evolution. The prediction was fitting pretty well to the data. The influential ability of the subjects discussed above could not be reflected in the overall evolution. This may be reasonable that only few subjects are evolving differently as compared to the crowd of the subjects. The results of model with and without the influential outlying subjects are presented in Table 5. Though there is no dramatic change as a result of excluding the influential subjects, some change occurred in parameter estimates and standard error. The estimated HbA1c was -0.0999 for treatment evolution over time. Exclusion of the said subjects mean the estimate was -0.0936. The precision of the estimate 0.0151 was almost unchanged. As discussed above, the impact on covariance parameter estimates was relatively large as compared to the impact on the fixed effect estimates. For example the estimate of random subject effect 0.0771 was reduced to 0.0704 with exclusion of the subjects. The estimated AR (1) correlation parameters are reduced from 0.1091 to 0.0939 in period 1 and from 0.6299 to 0.5512 in period 2.

Table 5: Parameter estimates with outlying subjects and without outlying subjects ,S.E=standard error

		With outliers			Without outliers		
Effect		Estimate	S.E	P-value	Estimate	S.E	P-value
Intercept	β_0	3.0028	0.6143	0.0001	3.1812	0.5917	0.0001
Treatment	β_1	1.1733	0.2274	0.0001	1.1898	0.2196	0.0001
period	β_2	0.1193	0.1463	0.4152	0.1468	0.1403	0.2961
Baseline	β_3	0.6479	0.0689	0.0001	0.6250	0.0665	0.0001
Diabetic duration	β_4	-0.0108	0.0035	0.0034	-0.0118	0.0035	0.0012
Time	β_5	-0.0024	0.0106	0.8219	-0.0071	0.0105	0.4992
Treatment. Time	β_6	-0.0999	0.0151	0.0001	-0.0936	0.0149	0.0001
Treatment. period	β_7	-1.2690	0.2219	0.0001	-1.3229	0.2137	0.0001
Covariance Parameters							
Subject effect	σ_s^2	0.0771	0.0212		0.0704	0.0196	
Autocorrelation 1	ρ_1	0.1091	0.1006		0.0939	0.1001	
Measurement error 1	σ_1^2	0.1665	0.0178		0.1618	0.0173	
Autocorrelation 2	ρ_2	0.6299	0.0584		0.5512	0.0696	
Measurement error 2	σ_2^2	0.3203	0.0438		0.2476	0.0338	

5.2 Sensitivity Analysis for Missing Data Mechanism

In this Section, the results of the methodology discussed in Section 3.6 are presented. We consider the relation of the dropout to the current and previous observations. The fitted dropout model can be written in form:

$$logit[P(D_i = j \mid D_i \geq d, y_i)] = -10.010 + 0.140y_{ij} + 0.517y_{i,j-1}, \qquad (1)$$

Where y_{ij} is the current observation (unobserved) predicted from the assumed model, $y_{i,j-1}$ is the previous observation; and $\varphi_1 = 0.140$ is the parameter estimate of y_{ij}, $\varphi_2 = 0.517$ is the parameter estimate of $y_{i,j-1}$. By setting the hypothesis $H_0: \varphi_1 = 0$ we claim that dropout is no longer allowed to depend on the current measurement. The likelihood ratio test statistic, comparing the maximized likelihood under model (1) with the maximized likelihood under the same model with $\varphi_1 = 0$, equals $LRT = 1.354$, which is highly insignificant (P-value =0.2446) on 1 degree of freedom. To test the hypothesis corresponding to $H_0: \varphi_1 = \varphi_2 = 0$, this means dropout is independent of outcome (MCAR). The likelihood ratio test statistic,

comparing the maximized likelihood under model (1) with the maximized likelihood under the same model with $\varphi_1 = \varphi_2 = 0$, equals $LRT = 138.68$, which is highly significant (P-value < 0.0001) to reject the null hypothesis on 1 degree of freedom. Since dropout do not depends on unobserved measurement, we then conclude that the sensitivity analysis support MAR dropout mechanism.

The model can be refitted using methods that are valid under MAR assumptions. In line with the ITT principle, imputation by LOCF is prominently used. However, Molenberghs and Kenward (2007) suggested that the likelihood based-ignorable analysis is fully consistent with ITT. It consists in applying likelihood-based models like linear, generalized linear and non-linear mixed models to longitudinal data without modeling the missingness process. The disadvantage in LOCF is that it depends on unrealistic assumptions, such as constant subject profile over time and its validity under MCAR. While the likelihood approach can be seen as a proper way to accommodate information on a subject efficiently, the method is valid only under MAR. An alternative to method which handles missingness under MAR mechanism is the Multiple Imputation (MI). MI was used to acknowledge the uncertainty of missing value by filling 4 times to generate 4 data sets which are analyzed using standard procedures and the results were combined into a single inference. Results from LOCF, likelihood based-ignorable analysis and MI are shown in Table 6.

The application of the methods in our data set indicates that LOCF had more conservative estimates than likelihood based approach. For example, the treatment difference over time using LOCF was -0.0851 with standard error 0.0124; using likelihood based approach, the treatment difference over time was -0.09118 with standard error 0.0132. The estimate using multiple imputations yields -0.0881 with standard error 0.0128. It can be seen that the estimates from multiple imputation and likelihood method are closer. Considering the standard error of the parameters from Table 6, the precision were overestimated in LOCF method which is another disadvantage of the method. To assess the impact of each of the missing data mechanism more clearly, the estimated mean profiles by treatment group are shown in Figure 6. All in all, the graph suggests that there is little difference among the mechanisms with the data at hand. All the mechanism reflects the marginal superiority of treatment 1 (Sensor ON). However caution should be taken with complete case analysis (CC), which is a result of deleting subjects having incomplete measurements.

We have discussed that the aim of ITT is to hold overall balance with respect to the treatment groups. Unless to compare the impact of removing subjects with that of imputing the incomplete measurements, complete case analysis is biased since it undermines any randomization justification.

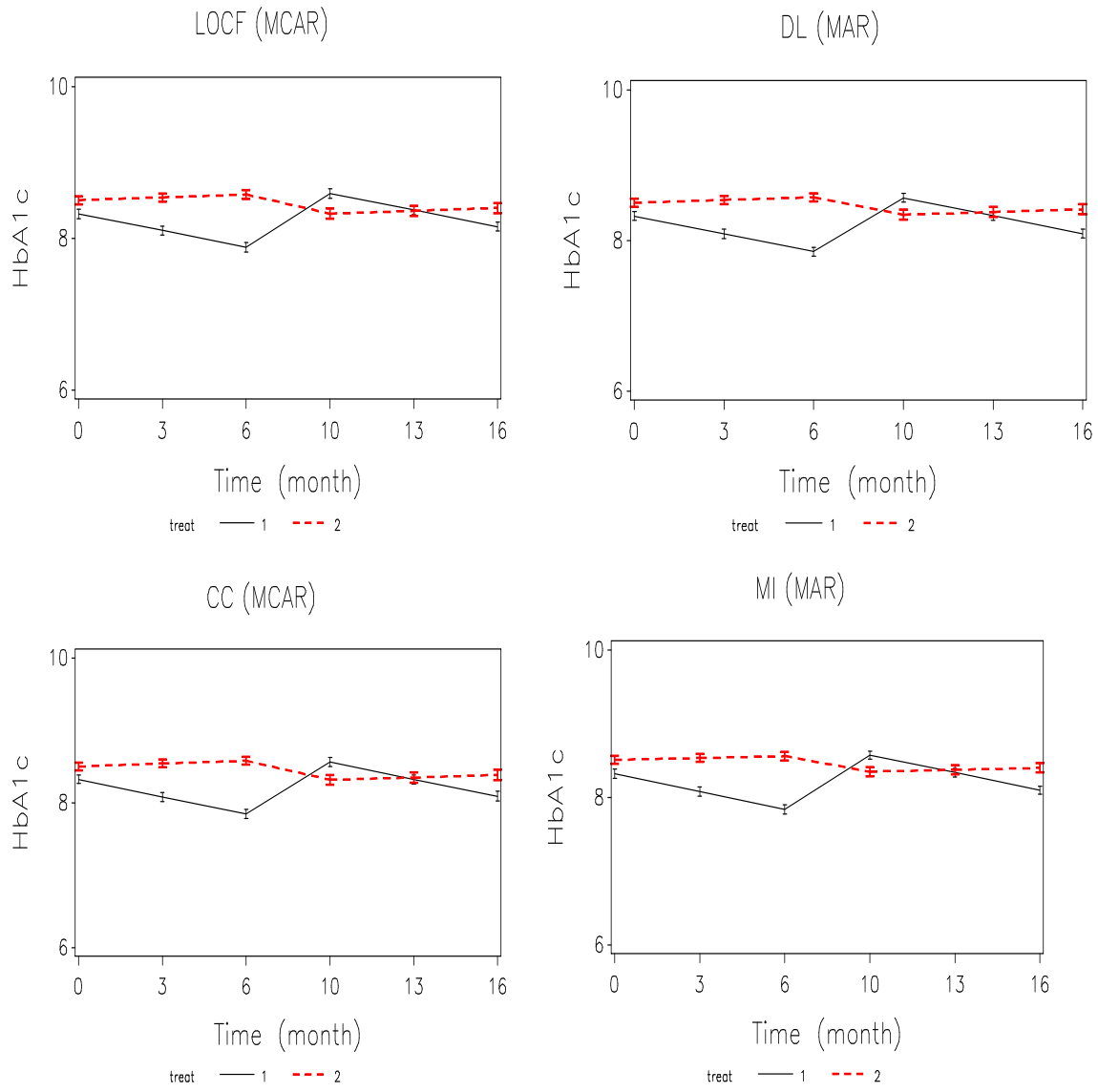


Figure 6: Estimated mean profile by treatment group obtained under missing mechanisms

Table 6: Parameter Estimates and standard error, Estimate (S.E) under missing mechanism: Last observation carried forward (LOCF), direct-likelihood and Multiple Imputation (MI).

	LOCF	Direct-likelihood	MI
Effect	Estimate(S.E)	Estimate(S.E)	Estimate(S.E)
Intercept	2.1217 (0.5435)	2.1672 (0.5450)	2.1399 (0.5374)
Treatment	0.9771 (0.1890)	1.0010 (0.1992)	0.9682 (0.1957)
period	0.1639 (0.1243)	0.1486 (0.1298)	0.1050 (0.1276)
Baseline	0.7541 (0.0611)	0.7500 (0.0612)	0.7590 (0.0602)
Diabetic duration	-0.0120 (0.0034)	-0.0118 (0.0034)	-0.0116 (0.0034)
Time	0.0124 (0.0088)	0.0122 (0.0093)	0.0083 (0.0091)
Treatment. Time	-0.0851 (0.0124)	-0.09118 (0.0132)	-0.0881 (0.0128)
Treatment. period	-1.0212 (0.1959)	-1.0399 (0.2032)	-1.0194 (0.1986)
Covariance Structure			
Subject effect	0.1390 (0.0248)	0.1343 (0.0246)	0.1274 (0.0243)
Autocorrelation 1	0.1570 (0.0765)	0.1481 (0.0781)	0.1811 (0.0763)
Measurement error 1	0.2014 (0.0169)	0.2072 (0.0177)	0.2097 (0.0181)
Autocorrelation 2	0.5976 (0.0467)	0.5792 (0.0499)	0.6167 (0.0444)
Measurement error 2	0.3437 (0.0358)	0.3587 (0.0382)	0.3971 (0.0410)

6. Discussion and Conclusion

This study was focused on evaluating the impact of real-time continuous glucose monitoring (CGM) on type 1 diabetic patients. A randomized cross-over design was used to evaluate the inherent impact of sensor therapy in children and adults with type 1 diabetes with pump therapy without sensor as a control. The main advantage of this design is that treatments are compared within subjects such that difference between treatment measurements cannot be biased by any subject effect from the comparison, ordinarily greatly increasing precision. In this study, a special case of cross-over design with two-treatment and two-period was considered in which a four-month washout period was used to avoid carry-over effect. The study comprises two type of study population: Intent-to-treat and per-protocol population. The ITT implies all randomized patients are included in the analysis regardless of any protocol violation, whereas PP only includes patients who adhered to the protocol.

We have illustrated a significant treatment difference between the two therapeutic methods under different methodological scenarios. First, two measurements from a subject were taken, one measurement from the end of each period. The results showed that subjects benefited from the sensor therapy technology. The subjects were declared as random to estimate and to recover the between subject variability in the comparison of the treatments. The larger estimated between subject variability as compared to within subject variability might be a witness that cross-over design was appropriate for the data for effective comparison of treatments. Though measurements from the end of each period were analyzed to yield significant treatment effect, the data fall within the realm of continuous longitudinal data and hence can be modelled by use of linear mixed models.

Modelling stochastic dependence among longitudinal measurements directly by construction of the covariance matrix has been done. Two types of covariance pattern among measurements from the same subject were taken in to account. The dependences among measurements in the same treatment period were handled by a first-order autoregressive covariance structure and unstructured covariance matrix was introduced for the period factor to account for the between period dependence. The test for period dependence showed no significant correlation between the two periods. Random subject effects were introduced to handle the correlation of all measurements in a subject. Different autoregressive covariance structures were introduced for each period. A test of homogeneity was carried out to reduce the number of parameters; to test if the same within period covariance structure can be assumed. However, the test for homogeneous covariance structure was rejected.

Following the final covariance structure, several covariates were removed from the model to obtain a parsimonious mean structure. The estimate of the treatment effect over time showed that HbA1c level was significantly reduced for subjects using the sensor therapy as compared to the same subject using therapy without sensor. It was revealed that the evolution of the treatment groups remain the same in the two periods. The trend of time effect was not also different across period. In addition, analysis at specific time points was carried out to test the single treatment effect and the effect of the treatment at specific time points.

The adequacy of the final linear mixed model was checked with respect to outlier diagnostics for the PP model. Sensitivity analysis for missing data was done for The ITT population. Following some suspect outliers based on residual analysis, restricted likelihood distance was

used as an overall measure of influence. Few influential subjects were identified though their influence on the fixed effect and precision were not apparent. A separate individual profile for those influential subject showed that the predicted mean profile deviates from the observed mean profile, even though this could not be reflected in the overall evolution. The model analysis without outlying subjects showed that the impact on covariance parameter estimates was relatively large as compared to the fixed effect estimates. A sensitivity analysis based on logistic regression of dropout model showed that the missingness in the ITT data was missing at random. Selected models were refitted based on this assumption and the subsequent comparison of the models revealed that the parameter estimates and the standard errors were not too different. This might be due to the fact that only ten percent of the subjects had incomplete measurements.

In conclusion, the various analyses showed that the sensor therapy treatment was more effective than the pump therapy without sensor. Patients receiving this modern technology can effectively reduce their HbA1c level to improve their diabetic condition. Therefore the technology can help type 1 diabetic patient to control the HbA1c level continuously and to dispense insulin as needed.

7. References

1. Altham, P.M.E. (1984). Improving the precision of estimation by fitting a model. *Journal of the Royal Statistical Society*, Series B, **46**, 118-119.

2. Ammerman, C. B., Henry, P. R. & Littell, R. C. (1998). Statistical analysis of repeated measures data using SAS procedures. *Journal of Animal Science*, **76**, 1216-1231.

3. Battelino, T., Bolinder, J., Conget, I., Gimenez, M., Gough, H. & Castaneda, J. (2011). The SWITCH (Sensing With Insulin Pump Therapy to Control HbA_{1c}): Design and Methods of a Randomized Controlled Crossover Trial on Sensor-Augmented Insulin Pump Efficacy in Type1 Diabetes Suboptimally Controlled with Pump Therapy. *Diabetes technology & therapeutics,* **13**(1), 49-53.

4. Brown, H. K. & Kempton, R. A. (1994). The application of REMEL in clinical trials. *Statistics in Medicine*, **13**, 1601-1617.

5. Byrom, W.D., Jones, B., Lindsey, J.K. & Wang, J. (1999). Modelling the covariance structure in pharmacokinetic cross-over trials. *Journal of Biopharmaceutical Statistics*, **9**(3), 439 – 450.

6. Diggle, P. J., Liang, K. & Zeger, S. L. (1994). *Analysis of Longitudinal data*. New York: Oxfored Press Inc.

7. Elaine, N. & Katja, H. (2010). *Human Anatomy & Physiology*. 8th edition. San Francisco: Benjamin Cummings.

8. Galecki, A. T. (1994). General class of covariance structures for two or more repeated factors in longitudinal data analysis. *Communication Statistics-Theory and Method*, **23**(11), 3105-3119.

9. Garrett, A. G. (2003). Therapeutic equivalence: fallacies and falsification. *Statistics in Medicine*, **22**, 741-762.

10. Gilliam, L.K. & Hirsch, I.B. (2009). Practical aspects of real-time continuous glucose monitoring. *Diabetes technology & therapeutics,* **11**(1), 75-82.

11. Jones, B. & Kenward, M. G. (1989). *Design and Analysis of Cross-Over Trials.* London: Chapman & Hall.

12. Jones, B. & Kenward, M. G. (2003). *Design and Analysis of Cross-Over Trials*. 2nd edition. Boca Raton: Chapman & Hall/CRC.

13. Lindsey, J. K. (1993): *Models for Repeated Measurements*. Oxford : Oxford University Press.

14. Little, R.J.A. & Rubin, D.B. (1987). Statistical Analysis with Missing Data. New York: John Wiley & Sons.

15. Maringwa, J. T., Helena, G., Ziv, S., Christel, F., Molenberghs, G., Aerts, M., Karel Van, A., Teisman, A., & Bijnens, L. (2008). Tutorial In Biostatistics: Analysis of Cross-Over Designs with Serial Correlation within Periods Using Semi-Parameteric Mixed Models. *Statistics in Medicine*, **27**, 6009-6033.

16. Molenberghs, G. & Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.

17. Molenberghs, G. & Kenward, M. G. (2007). *Missing Data in Clinical Studies*. Chichester: John Wiley & Sons.

18. Patetta, M. (2002). *Longitudinal Data Analysis with Discrete and Continuous Responses Course Notes.* USA: SAS Institute Inc.

19. Pocck, S. J. (1983). *Clinical Trials, a Practical Approach*. Chichester: John Wiley.

20. Raccah, D., Sulmont, V., Reznik, Y., Guerci, B., Renard, E., Hanaire, H., Jeandidier, N.,&Nicolino, M. (2009). Incremental value of continuous glucose monitoring when starting pump therapy in patients with poorly controlled type 1 diabetes: the Real-Trend study. *Diabetes Care*, **32**, 2245-2250.

21. Schabenberger, O. (2004). Mixed Model Influence Diagnostics. Cary, NC: SAS Institute Inc.

22. Senn, S. (1996). The AB/BA Cross-Over: How to perform the two-stage analysis if you can't be persuaded that you shouldn't. Rotterdam: Erasmus University press, 93-100.

23. Senn, S. (1993). *Cross-Over Trials in Clinical Research*. New York: John Wiley & Sons.

24. Verbeke, G. & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: springer.

25. Wallenstein, S. & Fisher, A. C. (1977). The analysis of two-period repeated measures crossover design with application to clinical trials. *Biometrics*, **33**, 261-269.

8. Appendices

8.1 Appendix A: Tables

Table 7: Summary statistics of each visit point within a period

V1HbA1c, V3HbA1c& V5HbA1c are in period 1 and V6HbA1c, V8HbA1c & V10HbA1c are in period 2.

	ITT			PP		
Time point	#obse.	Mean	Std	#obse.	Mean	Std
V0HbA1c	153	8.394	0.624	90	8.320	0.573
V3HbA1c	147	8.30	0.829	90	8.103	0.711
V6HbA1c	144	8.240	0.888	90	8.028	0.730
V10HbA1c	142	8.462	0.925	90	8.289	0.841
V13HbA1c	139	8.301	0.950	90	8.079	0.739
V16HbA1c	138	8.209	0.999	90	7.956	0.832

8.2 Appendix B: Figures

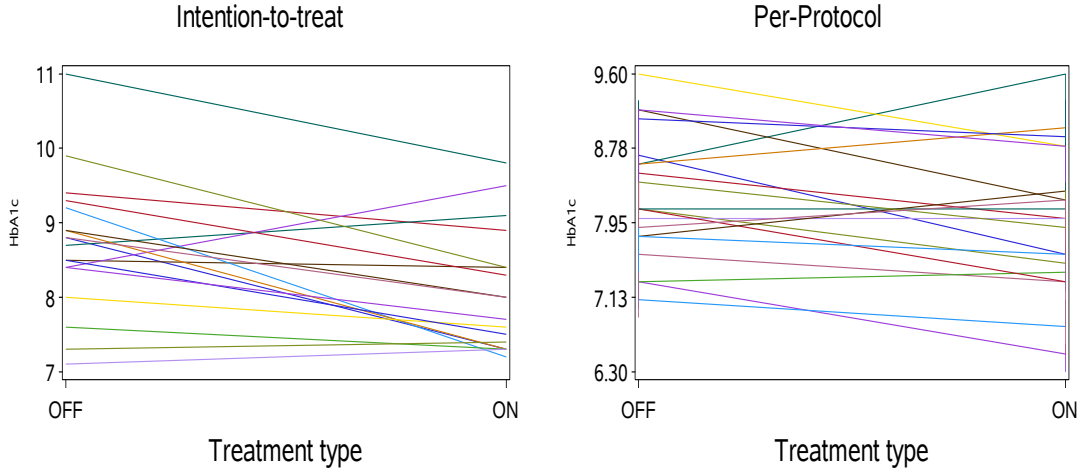
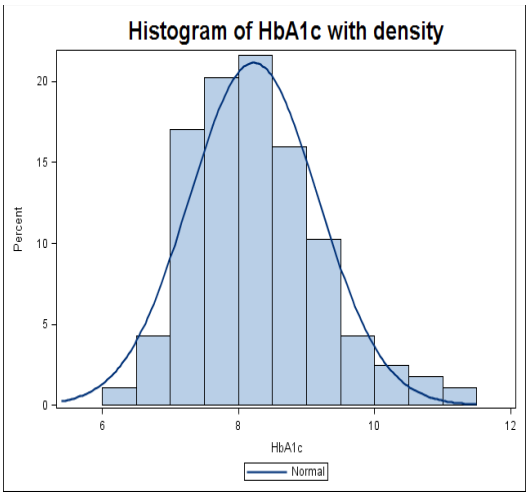
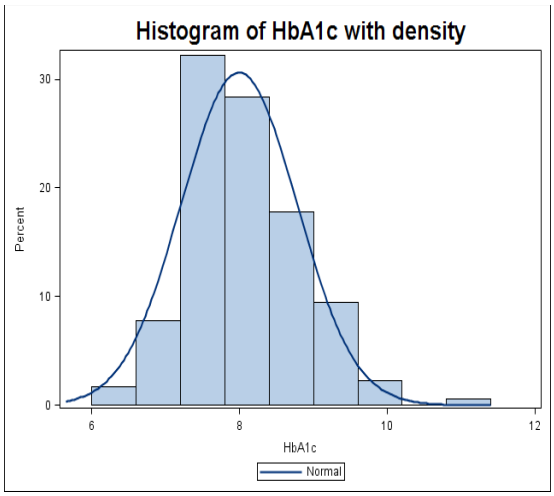


Figure 7: Measurements at the End of each period over treatment group.



For ITT population



For PP population

Figure 8: Normally distributed HbA1c difference between End -of- each period

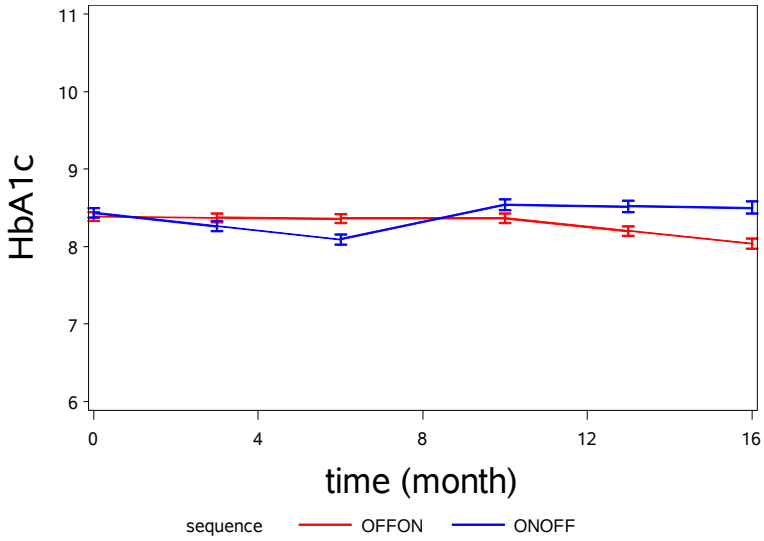


Figure 9: Predicted average evolution by treatment sequence (ITT)

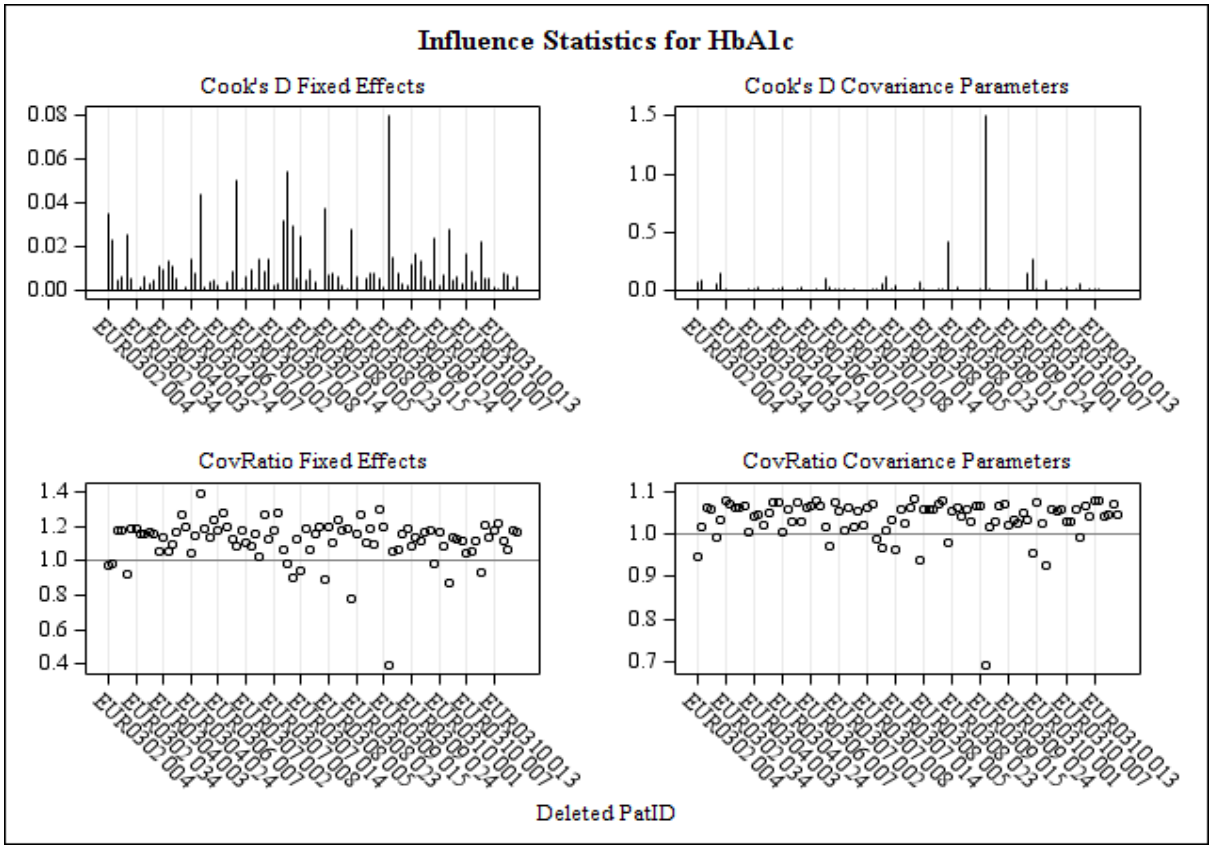


Figure 10: Diagnostic plots for fixed effect and covariance Parameters.

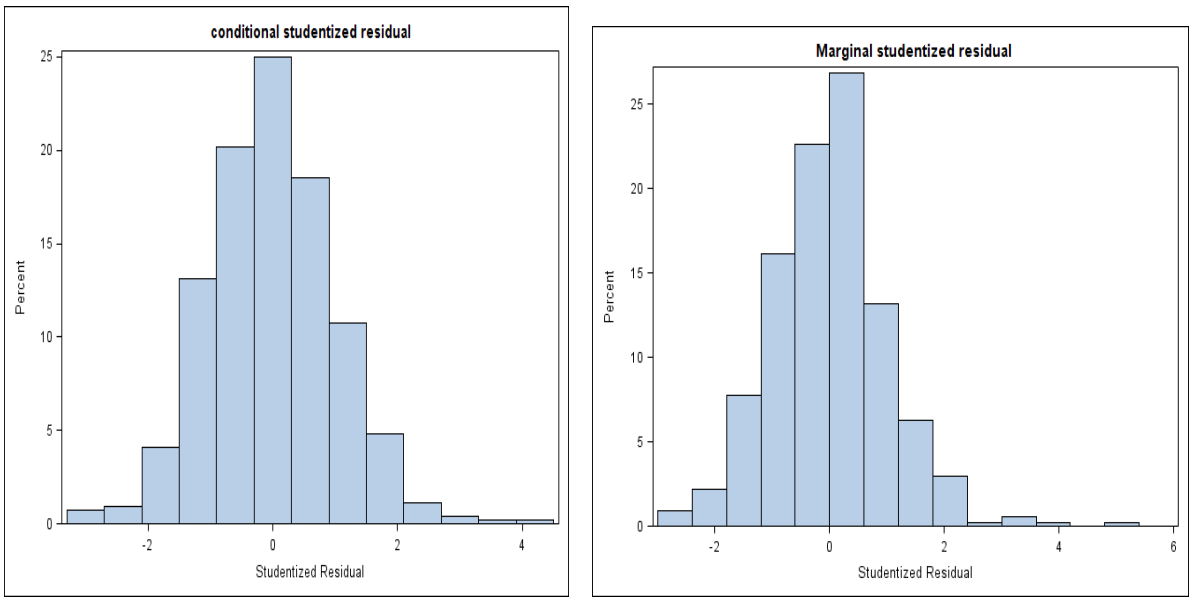


Figure 11: The conditional studentized residual and the marginal studentized residual.

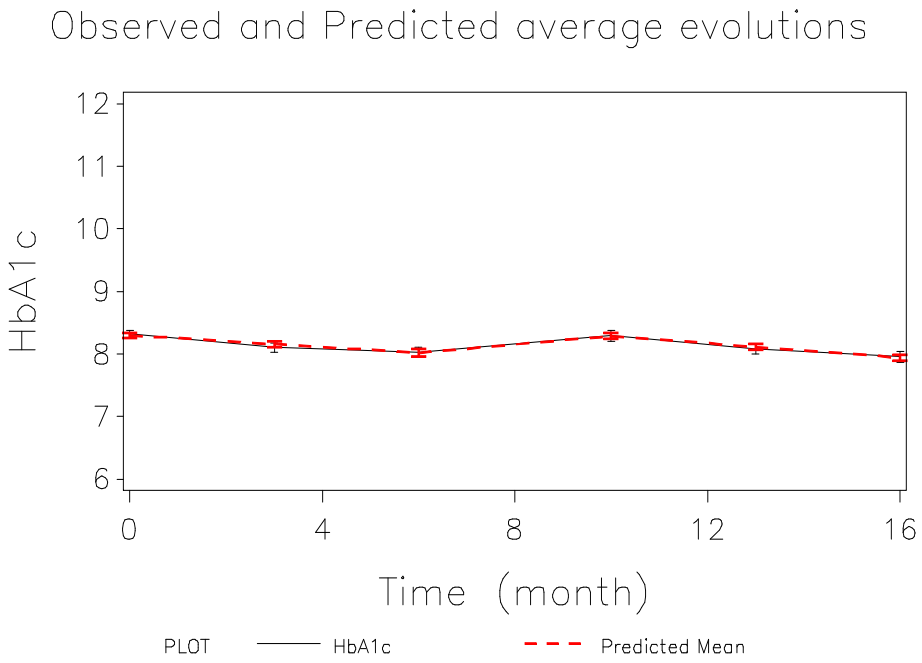


Figure 12: Observed and Predicted average evolution for PP population.

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Longitudinal analysis of an AB/BA cross-over study in diabetes type 1

Richting: **Master of Statistics-Biostatistics**
Jaar: **2011**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Arsido, Maeregu Woldeyes

Datum: **12/09/2011**